OXFORD MATHEMATICAL MONOGRAPHS

The Universality of the Radon Transform

LEON EHRENPREIS

OXFORD SCIENCE PUBLICATIONS

OXFORD MATHEMATICAL MONOGRAPHS

Series Editors

J. M. BALL E. M. FRIEDLANDER I. G. MACDONALD
L. NIRENBERG R. PENROSE J. T. STUART
N. J. HITCHIN W. T. GOWERS

OXFORD MATHEMATICAL MONOGRAPHS

L. Ambrosio, N. Fusco and D. Pallara: *Functions of bounded variation and free discontinuity problems*
A. Belleni-Moranti: *Applied semigroups and evolution equations*
A.M. Arthurs: *Complementary variational principles* 2nd edition
M. Rosenblum and J. Rovnyak: *Hardy classes and operator theory*
J.W.P. Hirschfeld: *Finite projective spaces of three dimensions*
A. Pressley and G. Segal: *Loop groups*
D.E. Edmunds and W.D. Evans: *Spectral theory and differential operators*
Wang Jianhua: *The theory of games*
S. Omatu and J.H. Seinfeld: *Distributed parameter systems: theory and applications*
J. Hilgert, K.H. Hofmann, and J.D. Lawson: *Lie groups, convex cones, and semigroups*
S. Dineen: *The Schwarz lemma*
S.K. Donaldson and P.B. Kronheimer: *The geometry of four-manifolds*
D.W. Robinson: *Elliptic operators and Lie groups*
A.G. Werschulz: *The computational complexity of differential and integral equations*
L. Evens: *Cohomology of groups*
G. Effinger and D.R. Hayes: *Additive number theory of polynomials*
J.W.P. Hirschfeld and J.A. Thas: *General Galois geometries*
P.N. Hoffman and J.F. Humphreys: *Projective representations of the symmetric groups*
I. Györi and G. Ladas: *The oscillation theory of delay differential equations*
J. Heinonen, T. Kilpelainen, and O. Martio: *Non-linear potential theory*
B. Amberg, S. Franciosi, and F. de Giovanni: *Products of groups*
M.E. Gurtin: *Thermomechanics of evolving phase boundaries in the plane*
I. Ionescu and M. Sofonea: *Functional and numerical methods in viscoplasticity*
N. Woodhouse: *Geometric quantization* 2nd edition
U. Grenander: *General pattern theory*
J. Faraut and A. Koranyi: *Analysis on symmetric cones*
I.G. Macdonald: *Symmetric functions and Hall polynomials* 2nd edition
B.L.R. Shawyer and B.B. Watson: *Borel's methods of summability*
M. Holschneider: *Wavelets: an analysis tool*
Jacques Thévenaz: *G-algebras and modular representation theory*
Hans-Joachim Baues: *Homotopy type and homology*
P.D.D'Eath: *Black holes: gravitational interactions*
R. Lowen: *Approach spaces: the missing link in the topology–uniformity–metric triad*
Nguyen Dinh Cong: *Topological dynamics of random dynamical systems*
J.W.P. Hirschfeld: *Projective geometries over finite fields* 2nd edition
K. Matsuzaki and M. Taniguchi: *Hyperbolic manifolds and Kleinian groups*
David E. Evans and Yasuyuki Kawahigashi: *Quantum symmetries on operator algebras*
Norbert Klingen: *Arithmetical similarities:prime decomposition and finite group theory*
Isabelle Catto, Claude Le Bris, and Pierre-Louis Lions: *The mathematical theory of thermodynamic limits: Thomas–Fermi type models*
D. McDuff and D. Salamon: *Introduction to symplectic topology* 2nd edition
William M. Goldman: *Complex hyperbolic geometry*
Charles J. Colbourn and Alexander Rosa: *Triple systems*
V. A. Kozlov, V. G. Maz'ya and A. B. Movchan: *Asymptotic analysis of fields in multi-structures*
Gérard A. Maugin: *Nonlinear waves in elastic crystals*
George Dassios and Ralph Kleinman: *Low frequency scattering*
Gerald W Johnson and Michel L Lapidus: *The Feynman Integral and Feynman's Operational Calculus*
W. Lay and S. Y. Slavyanov: *Special Functions: A Unified theory based on singularities*
D Joyce: *Compact Manifolds with Special Holonomy*
A. Carbone and S. Semmes: *A graphic apology for symmetry and implicitness*
Johann Boos: *Classical and modern methods in summability*
Nigel Higson and John Roe: *Analytic K-Homology*
S. Semmes: *Some novel types of fractal geometry*
Tadeusz Iwaniec and Gaven Martin: *Geometric Function Theory and Nonlinear Analysis*
Terry Lyons and Zhongmin Qian: *System Control and Rough Paths*
Andrew Ranicki: *Algebraic and Geometric Surgery*
Leon Ehrenpreis: *The Universality of the Radon Transform*

The Universality of the Radon Transform

LEON EHRENPREIS

Professor of Mathematics
Temple University

CLARENDON PRESS • OXFORD
2003

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Taipei Tokyo Toronto

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2003

A catalogue record for this title is available from the British Library
Library of Congress Cataloging in Publication Data
(Data available)
ISBN 0–19–850978–2

10 9 8 7 6 5 4 3 2 1

Typeset by Newgen Imaging Systems (P) Ltd., Chennai, India
Printed in Great Britain
on acid-free paper by T.J. International Ltd, Padstow, Cornwall

ב ס ״ ד

To Ahava
Many are the
Inspirations of the heart
But that born by love
Surpasses all the rest.

This page intentionally left blank

PREFACE

Functions represent one of the principal objects of study in mathematics. Sometimes we study individual functions by performing various operations on them. At other times we study spaces of functions, in which case functions are individualized by parametrization data; that is, data which picks out the individual functions from the space.

More generally we might start with a large space, then decompose it into subspaces $\{\mathbf{g}\}$ which we can think of as "coarse grains," and then decompose the coarse grains into "fine grains," which are the individual functions. In the case of the Radon transform the coarse grains consist of "spread functions." These are functions which are constant in certain directions or, more generally, which satisfy partial differential or more complicated equations. The parametrization data is data for Cauchy- or Dirichlet-like problems for these equations.

In the theory of Radon the passage from functions to coarse grains is accomplished by integration over geometric objects called leaves of the spread. The leaves are equipped with measures in a consistent fashion. From the averages of f we then form a spread function, $\mathbf{R}^*\mathbf{R}f(\mathbf{g})$, which represents the best approximation to f within this coarse grain.

Instead of merely integrating we can multiply by fixed interesting functions called attenuations before integration. When the leaves are homogeneous spaces of a group a reasonable class of attenuations is defined by representation functions for the group.

Multiplication by a fixed function represents a linear transform of the function spaces. For certain problems, e.g. in number theory, nonlinear transformations such as $f \to \chi(f)$ where χ is a character on the range of f are important.

A natural question is the reconstruction of a function f from its Radon transform $\mathbf{R}f$, meaning the set of its integrals over all leaves of all spreads. How is a function constructed from its averages?

The simplest example is the reconstruction of a function f of a single variable from its indefinite integral

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

which represents the averages of f over the leaves $(-\infty, x]$ (assuming f vanishes for large negative values). The reason that we can find an easy inversion formula is that there is an order on the summation index t. An analog when the order is the natural order on divisors of a number is the Möbius inversion formula: if

$$g(q) = \sum_{d|q} f(d)$$

then

$$f(q) = \sum_{d|q} \mu(q/d) g(d)$$

where μ is the Möbius function.

One way of thinking of the inversion formula for the indefinite integral (the fundamental theorem of calculus) is that we can write

$$F(x^0) = \chi_{(-\infty, x^0)} \cdot f$$

where χ represents the characteristic function and $\cdot$ represents integration of the product. $\chi_{(-\infty,x^0)}$ has a jump singularity at the point x^0 which can be converted into the δ function δ_{x^0} at x^0 by applying the operator d/dx. (The Möbius inversion formula can be given a somewhat more complicated but analogous interpretation.)

For the original Radon transform the leaves are hyperplanes. If we are in dimension >1 there is no natural ordering. Nevertheless if we form

$$\int \chi_{L(x^0)} = \Phi(x^0)$$

over all hyperplanes $L(x^0)$ passing through x^0, integrating with a natural measure, then we would expect that the distribution $\Phi(x^0)$ has a higher order singularity at x^0 than at other points in analogy to the singularity of $\chi_{(-\infty,x^0)}$ at x^0. Thus we might hope to find a local operator ∂ which converts $\Phi(x^0)$ into δ_{x^0}. Such local operators ∂ do not always exist; sometimes more complicated operators are needed.

The above discussion centered around the reconstruction of f from its averages over various subsets such as hyperplanes. A more subtle form of this process can be formulated as

Radon ansatz. *Study properties of f in terms of its restriction to lower dimensional sets.*

Coarse grains represent decompositions of function spaces. When the coarse grains are isomorphic we can sometimes realize the space as a tensor product of the coarse grains with a "Grassmannian" which represents the set of coarse grains.

There is another process of studying spaces and other mathematical objects which we call hierarchy. In contradistinction to decomposition, the hierarchy is a larger object for which the given object is one component. In Radon transform theory such hierarchies arise when the (isomorphic) leaves are given a parametric representation, meaning that they are represented by maps of a given manifold into the ambient space. Usually these maps involve redundant parameters; the redundant parameter space defines the heirarchy. The integrals over the leaves

represent a function h as a function $\mathbf{R_p}h$ of these parameters. Only those functions G of the parameters which satisfy equations determined by the redundancy can be of the form $\mathbf{R_p}h$. Under favorable conditions these equations characterize $\{\mathbf{R_p}h\}$. In this case $\{\mathbf{R_p}h\}$ is a component of the "hierarchy" space of all functions on the parameter space.

One of the principal aims of this book is to show how the Radon transform impinges on such varied branches of pure mathematics as integral geometry, partial differential equations, Lie groups, holomorphic functions of several complex variables, asymptotic analysis, and number theory. Moreover, it has multifold applications to problems in medicine, aerodynamics, etc. (in which case it is generally referred to as "tomography").

One might wonder as to why this Radon "averaging process" has such universality. Averages involve the global structure of the function. We shall see that the Radon transform is most useful when there is a regularity to the structure under consideration; the regularity means that local is determined by global.

This book could never have come into existence without the aid of Tong Banh, Cristian Guriţă, and Paul Nekoranik. In particular Cristian and Paul worked incessantly for several years typing the manuscript and correcting the multifold errors. Banh's contributions to Chapters 5 and 9 were extremely significant. The final (and correct) form of several of the theorems and proofs is due to him. He has also made profound refinements of some of the results.

I also want to thank Marvin Knopp, Karen Taylor, Wladimir Pribitkin, Pavel Gurzhoy, Hershel Farkas, and Arnold Dikanski for the significant outlay of energy that they expended in proofreading. Finally I want to express my appreciation to Peter Kuchment and Eric Todd Quinto for contributing the appendix on tomography; this brings our abstract theory in contact with the practical world.

August 2002 L.E.

CONTENTS

This page intentionally left blank

1

INTRODUCTION

Chapter 1 gives a heuristic treatment of much of the material presented in this book. In Section 1.1 we introduce the notion of the Radon transform $\mathbf{R}f$ of the function f as a set of integrals of f over various sets; the crucial ingredient in our study is the organization of these sets. This leads to the concept of spreads and the relation of the Radon transform to various areas of mathematics. These are outlined briefly in Section 1.1 and in more detail in the remainder of this chapter.

The parametric Radon transform, which was introduced by F. John, is the subject of Section 1.2. The points of the surface of integration are "individualized" by the parameter. The parametric Radon transform leads to interesting partial differential equations.

In Section 1.3 the concept of *spread* is made precise. Roughly speaking, a spread is a family of disjoint sets called *leaves* whose union is the whole space. The families of affine planes of fixed dimension and certain families of algebraic varieties are decomposed into spreads.

Section 1.4 studies ways of parametrizing solutions of partial differential equations by data on various subspaces. When the subspace on which the data is given is of lower dimension than the "natural dimension" of the equations, then an infinite number of data must be prescribed. In this case we call the parametrization problem "exotic"; it is also termed the "Watergate problem." The concept of a well-posed parametrization problem is introduced and is related to H. Weyl's method of orthogonal projection.

The decomposition of solutions of partial differential equations in analogy with the decomposition of harmonic functions in the plane into holomorphic and antiholomorphic functions is one of the main themes of Section 1.5. The John equations for the parametric Radon transform are studied from this viewpoint. The decomposition of harmonic functions in higher dimensions is introduced; this is related to the Penrose transform. The inverse problem, namely the construction of solutions of an "enveloping equation" from decompositions, is studied.

In Section 1.6 the Radon transform is studied in the framework of Lie groups. Given a Lie group G and the closed subgroups H, K we can "intertwine" G/H and G/K by integration; this is the double fibration. The relation of double fibration to spreads is analyzed as is the relation to the parametric Radon transform. In particular the geodesic and horocyclic Radon transforms are introduced.

The projection–slice theorem is studied in Section 1.7. Given a spread $\mathbf{g}$ the integrals over the leaves of $\mathbf{g}$ define a function (projection) on a cross-section S of the leaves. The projection–slice theorem relates the Fourier transform of a function f on the whole space to the Fourier transform on S of its projection, i.e. its Radon transform on $\mathbf{g}$. Generalizations to partial differential equations are studied. Nonabelian analogs related to the Frobenius reciprocity theorem are introduced. The projection–slice theorem is closely related to the Poisson summation formula (PSF). "Cut-offs" of the PSF which are the Euler–Maclaurin formula and some sharpened forms are studied. Selberg's trace formula fits into the same framework.

The space of spread functions for a given spread can be considered as a "coarse grain," meaning a component of a tensor product or "direct integral" decomposition of a space of functions on the whole space. This is one of the types of tensor product studied in Section 1.8. Actually we do not obtain true tensor products but "sub" or "supra" tensor products. The tensor product decomposition leads to insights into the inversion formula for the Radon transform and its relation to classical potential theory.

1.1 Functions, geometry, and spaces

Suppose we want to study the density distribution in a person's brain. Since we do not wish to access it directly, we input some signal such as an X-ray and we examine the outcome. We then make the hypothesis that the relation between the input and output is some sort of average of the effects on the X-ray over the points in the brain traversed by that X-ray.

This "average hypothesis" leads to the mathematical idea that functions should be studied via their averages. Which averages?

From a pure mathematics point of view the idea of describing a function in terms of its averages came to the fore in the calculus of variations and culminated in the theory of distributions of Schwartz [135]. In this setting a function $f(x)$ is determined by its averages

$$\phi \cdot f = \int f(x)\phi(x)\,dx \tag{1.1}$$

for suitable "test functions" ϕ. The functions ϕ are arbitrary except for regularity and growth conditions at infinity.

In 1917, Radon [133] introduced the idea of studying functions f on $\mathbb{R}^n$ in terms of their averages over affine hyperplanes. For this to make sense, f must be suitably small at infinity. Radon showed that such an f is uniquely determined by its averages over all (unoriented) affine hyperplanes. In fact, he produced an inversion formula which reconstructs f in terms of these averages.

Of course, in both the Schwartz and Radon cases we need a continuum of averages to reconstruct f exactly. Various techniques (called *tomography* in the

Radon case) have been developed for approximate reconstruction formulas in case we know only a finite number of averages (see appendix).

The (unoriented) hyperplanes through the origin form the manifold $\mathbf{P}^{n-1}$ which is the $n-1$ dimensional projective space; affine hyperplanes L are obtained by translation of such hyperplanes. The "delta function" δ_L of L is used to denote the averaging process over L, meaning that $\delta_L \cdot f$ is the integral of f over L (with the usual measure). More generally for any set Ω with a given measure we define $\delta_\Omega \cdot f$ by

$$\delta_\Omega \cdot f = \int_\Omega f. \tag{1.2}$$

The Radon transform shows that, in fact, any ϕ average (as in (1.1)) of a suitable type is an integral of averages over hyperplanes. We have thus introduced a new structure in the set of all ϕ averages by producing a "basis" $\{\delta_L\}$ corresponding to the n parameter family of affine hyperplanes. (Of course, the set $\{\delta_x\}_{x\in\mathbb{R}^n}$ is also a basis.)[1]

There is a major difference between the Radon and Schwartz viewpoints. In (1.1) the averages of f depend on ϕ which belongs to an infinite dimensional space. On the other hand the hyperplanes form an n dimensional family. In its most elementary form a function $f(x)$ is determined by x which is a parameter in $\mathbb{R}^n$. Thus both $\{x\}$ and the set $\{L\}$ of hyperplanes are n parameter sets.

Instead of using 0 dimensional planes $\{x\}$ or hyperplanes $\{L\}$ one could form intermediate families. For example, we could search for an n parameter family $\{L^l\}$ of l planes or l dimensional subvarieties of $\mathbb{R}^n$ $(0 < l < n)$, such that the integrals of f over $\{L^l\}$ parametrize $\{f\}$.

Besides parametrization we can search for $\{\tilde{L}^l\}$ such that certain properties of the restriction of f to the $\tilde{L}^l$ imply properties of f on $\mathbb{R}^n$. For example, if ∂ is a partial differential operator (or a system of operators) then $\partial f = 0$ is an n parameter condition, namely $\partial f(x) = 0$ for all x. It sometimes happens (see e.g. Chapters 5 and 9) that there is an $n - l$ parameter family $\{\tilde{L}^l\}$ and other differential operators $\partial_{\tilde{L}^l}$ on $\tilde{L}^l$ such that the conditions $\partial_{\tilde{L}^l} f = 0$ on each $\tilde{L}^l$ imply $\partial f = 0$ on $\mathbb{R}^n$.

We put these ideas under the umbrella of

Radon ansatz. *To determine a function f on $\mathbb{R}^n$ or a property $\mathcal{P}$ of f find suitable subvarieties $\{\tilde{L}\}$ of $\mathbb{R}^n$ so that the restrictions of f to the $\tilde{L}$ determine f, or suitable properties $\mathcal{P}_{\tilde{L}}$ of f on the $\tilde{L}$ imply property $\mathcal{P}$ for all of $\mathbb{R}^n$.*

The basis $\{\delta_L\}$ formed from affine hyperplanes has an additional structure. Each L has two natural coordinates $(s, \mathbf{g})$ which serve to "organize" the set of planes, as follows.

[1]The precise meaning of "basis" is unimportant at this point. We can think of a basis as a set of elements whose linear combinations (sums and integrals) span the space in question and are "essentially" linearly independent. A precise definition is given in (1.17)ff.

Grassmann parameter $\mathbf{g}$. This represents that parallel translate $L(0,\mathbf{g})$ of L that passes through the origin. Planes with the same Grassmann parameter form a **spread**.

Spread parameter s. This represents the linear coordinate $L\cap L^{\perp}$ on the line $L^{\perp}$ through the origin orthogonal to L.

We thus write a hyperplane as $L(s,\mathbf{g})$ where the Grassmann parameter $\mathbf{g}\in\mathbf{P}^{n-1}$ is identified with $L(0,\mathbf{g})$ while the spread parameter belongs to $\mathbb{R}^1$.

There is some ambiguity in defining s, because of the nonorientability of $\mathbf{P}^{n-1}$, or, what is the same thing, the use of unoriented planes L, but this will not give us any serious trouble because we can pass from $\mathbf{P}^{n-1}$ to the $n-1$ sphere S^{n-1} which is the same as orienting the planes $L(0,\mathbf{g})$. Using S^{n-1} allows for a consistent definition of s on all lines through the origin. The spread parameter s is the coordinate on $L(0,\mathbf{g})^{\perp}$ (oriented) of $L(s,\mathbf{g})\cap L(0,\mathbf{g})^{\perp}$. This means that we regard $L(0,\mathbf{g})^{\perp}$ as a *cross-section* $S_{\mathbf{g}}$ of the *spread* of all $L(s,\mathbf{g})$ for $\mathbf{g}$ fixed. In other terms, we choose a *base point* on each $L(s,\mathbf{g})$; this is the point closest to the origin. These base points form the cross-section.

In the present situation all the $S_{\mathbf{g}}$ are isomorphic in a natural way. We denote by S some fixed manifold isomorphic to the $S_{\mathbf{g}}$.

We shall put the cross-section in a more conceptual light presently. In addition to the parameters $s,\mathbf{g}$, we shall sometimes make use of an L *parameter* λ which is a convenient parameter on L. Since the $L(s,\mathbf{g})$ are isomorphic, the isomorphism allows for a consistent definition of λ which we regard as an "individualization" of the points on $L(s,\mathbf{g})$. Under these conditions we can use (s,λ) as coordinates in the whole space.

It is sometimes convenient to think of $\mathbb{R}^n$ as a cylinder with base $L(0,\mathbf{g})$. Hence we refer to (s,λ) as *cylindrical coordinates*. Such coordinates appear in detail in Section 10.2.

In general, we define the **nonparametric Radon transform**

$$\mathbf{R}(s,\mathbf{g})=\int_{L(s,\mathbf{g})} f. \tag{1.3}$$

Radon's theorem described above shows that the hyperplane Radon transform $\mathbf{R}$ is injective. Let us examine its image. With our usage of $\mathbf{g}\in S^{n-1}$ (rather than $\mathbf{P}^{n-1}$) we have now

$$L(-s,-\mathbf{g})=L(s,\mathbf{g}) \tag{1.4}$$

so that

$$\mathbf{R}f(-s,-\mathbf{g})=\mathbf{R}f(s,\mathbf{g}). \tag{1.5}$$

The Radon transform is not surjective on functions satisfying (1.5). There are additional conditions, called *moment conditions*, which seem to have been discovered independently by several authors including Cavalieri, Gelfand, and Helgason (see [94]). They are dealt with in detail in Chapter 2.

The manifold of affine hyperplanes has the structure of a general Möbius band. For $n = 2$ it is the usual Möbius band as can be seen as follows. We start with the product of the half-circle $0 \leq \theta \leq \pi$ with the line $\{s\}$. To obtain the space of affine lines we must identify $(s, 0)$ with $(-s, \pi)$. This is exactly the usual definition of the Möbius band.

For $n > 2$ the manifold of affine hyperplanes is obtained by a construction of a similar nature. We start with the product of a hemisphere $\{\mathbf{g}\}$ with the line $\{s\}$ and then identify $(s, \mathbf{g})$ with $(-s, -\mathbf{g})$ when $\mathbf{g}$ is in the boundary of the hemisphere, i.e. $\mathbf{g} \in S^{n-2}$.

We have termed (1.3) the "nonparametric" Radon transform because the explicit parameter on L does not appear. There are "reasonable" ways of introducing a linear parameter λ on $L(s, \mathbf{g})$. If we have one parameter λ^0 on $L(0, \mathbf{g}^0)$ then we can use the rotation group which acts transitively on the Grassmannian to define corresponding parameters on the $L(0, \mathbf{g})$ and then use translation to define the parameters on all $L(s, \mathbf{g})$.

This procedure is somewhat vague because it has many nonunique stages, so we introduce a better method.

The parameters $s, \mathbf{g}$ can be thought of in terms of equations that define $L(s, \mathbf{g})$. If $\mathbf{g} \in S^{n-1}$ is thought of as a unit vector in $\mathbb{R}^n$ then

$$L(s, \mathbf{g}) = \{x | \mathbf{g} \cdot x = s\}. \tag{1.6}$$

Thus $s, \mathbf{g}$ represent parameters in the equation defining $L(s, \mathbf{g})$ as an algebraic variety.

In general, an algebraic variety V is defined in terms of

(1) The equations of V.
(2) The points of V.

Remark. *For linear V, (1) can be thought of as a covector description and (2) as a vector description.*

To interpret $V = L(s, \mathbf{g})$ in terms of (2), we choose one fixed hyperplane L^0 with a fixed basis; a point on L^0 is defined by the vector λ of coefficients in this basis. We then map L^0 into $\mathbb{R}^n$ by

$$\lambda \mapsto \boxed{a}\lambda + b$$

where b is a point of $\mathbb{R}^n$ and $\boxed{a}$ is an $n \times (n-1)$ matrix. The image of this map is (generically) an affine hyperplane with the parameter λ. Of course, we have not completely avoided the ambiguity problem mentioned above because the same hyperplane corresponds to many $(\boxed{a}, b)$. But, as we shall see in Chapter 6, the ambiguity is "controlled."

In terms of this map, we define the **parametric Radon transform**

$$F(\boxed{a}, b) = \int f(\boxed{a}\lambda + b)\, d\lambda. \tag{1.7}$$

In a somewhat different form, the parametric Radon transform was introduced by John [99].

When $\mathbf{g}^0$ is fixed the $L(s, \mathbf{g}^0)$ form a decomposition of $\mathbb{R}^n$ into disjoint sets. Such a decomposition is called a *spread*. (Other types of spreads will be introduced in our work.) For each s, $L(s, \mathbf{g}^0)$ is called a *leaf* of that spread. In the present situation there are many spreads, one for each $\mathbf{g} \in \mathbf{G}$.

Spreads belong to the domain of geometry. There is an associated analysis which can sometimes penetrate problems that lie beyond the power of geometry (at least as we understand it now). With each spread we associate **spread functions**. These are the functions which are constant on every leaf of the spread.

Spread functions seem to go beyond geometry. Finite, positive, integral, linear combinations of the δ functions of leaves of the spread correspond to the geometric union of these leaves. But continuous combinations with complex coefficients seem difficult to interpret geometrically, as do certain natural operations on functions.[2]

Remark. *Most of our work is based on analysis and function theory. It would be of great interest if the reader could translate some of our function arguments into geometry.*

To set our ideas in a suitable analysis framework we assume that the functions f that we deal with belong to some topological vector space $\mathcal{W}$. The geometric hyperplanes L are embedded in the dual space $\mathcal{W}'$ by thinking of L as δ_L.

Remark. *The spaces $\mathcal{W}$ that we study consist of functions or distributions which have suitable decrease at infinity. Thus Radon and Fourier transforms are defined on $\mathcal{W}$. The elements of $\mathcal{W}'$ are "large" at infinity. This allows the differential operators we study to have large kernels on $\mathcal{W}'$.*

We come to a crucial idea that pervades much of our work: rather than think of $\mathcal{W}'$ as an entity, we think of it as being broken into pieces $\{\mathcal{W}'(\mathbf{g})\}_{\mathbf{g}\in G}$. The $\mathcal{W}'(\mathbf{g})$ are linear subspaces which span $\mathcal{W}'$; in fact $\mathcal{W}'$ is a sort of integral of the spaces $\mathcal{W}'(\mathbf{g})$ (see Section 1.8). Each $\mathcal{W}'(\mathbf{g})$ consists of the spread functions for $\mathbf{g}$ in $\mathcal{W}'$. $\mathcal{W}'(\mathbf{g})$ can be thought of as the set of limits of linear combinations of $\{\delta_{L(s,\mathbf{g})}\}_s$. The idea of breaking an object into smaller pieces appears in many branches of mathematics. Perhaps the most clear-cut example is combinatorial topology. The pieces have a simpler structure than the original object and one is led to a combinatorial problem of piecing together these simpler objects.

Individual functions are broken down by decomposition into combinations of elements of a suitable basis or, in a different vein, into their restrictions to various subdomains.

[2]We shall often use the terms "function," "measure" and "distribution" interchangeably. Sometimes the word "function" can apply to distributions and we often use the suggestive notation $\int u(x)f(x)\,dx$ for the value of the distribution u applied to f.

In statistical mechanics the pieces refer to "coarse grains" as contrasted with "fine grains" which are the individual objects (analogs of elements of a space). We shall adopt this language. Sometimes we shall decompose the coarse grains into finer grains which are still subsets and not individual objects. These may be termed "semi-coarse grains."

Remark. *In Section 1.4 we shall introduce the fundamental principle which can often be used to relate the coarse grain decomposition of* $\mathcal{W}'$ *into* $\{\mathcal{W}'(\mathbf{g})\}$ *to the decomposition of* $\mathbb{R}^n$ *or* $\mathbb{C}^n$ *into suitable algebraic subvarieties; such decompositions can be regarded as instances of the Radon ansatz. In a similar vein there are interesting decompositions of some algebraic varieties into subvarieties—providing a unification of the concepts of semi-coarse grains with the Radon ansatz (see e.g. Chapter 6).*

Although decompositions play a crucial role in mathematics one should not lose sight of the opposite construction which we call *hierarchy* (see [57]). In this setting we place our given object in a larger system (hierarchy). The structure of the hierarchy then sheds light on its components.

In writing (1.3) we have tacitly assumed that L is provided with some well-defined measure. Of course since we are in euclidean space there is a natural measure associated to each reasonable geometric object. The same is true if we replace $\mathbb{R}^n$ by any Riemannian manifold. The globally defined Riemann metric ds^2 provides a consistent way of associating measures, i.e. elements of $\mathcal{W}'$, to all geometric objects. In this way geometry is embedded in analysis.

There are other aspects of the spreads we have defined which broaden the horizons of the Radon transform.

(1) Differential equations
(2) Groups

(1) Differential equations

To understand the relation to differential equations, consider the simplest example of the (geometric) Radon transform: lines in $\mathbb{R}^2$. A spread $\mathbf{g}$ is the set of all lines parallel to a given L_0 passing through the origin. If we denote by $\partial(\mathbf{g})$ the directional derivative in the direction L_0 then for any $L \in \mathbf{g}$ we have

$$\partial(\mathbf{g})\delta_L = 0. \tag{1.8}$$

In fact, $\{\delta_L\}_{L\in\mathbf{g}}$ forms a basis for all solutions of this differential equation; such solutions form the spread functions for $\mathbf{g}$. Equation (1.8) is the (covector) description of the spread functions for $\mathbf{g}$ while the basis $\{\delta_L\}$ provides a vector description.

We have associated with the leaf L a *base point* which is the point on $L \cap L_0^\perp$; it defines the spread parameter s. The leaf L and hence the solution $\delta_{L(s,\mathbf{g})}$ of (1.8) is determined by s, i.e. by $L \cap L_0^\perp$. $S = L_0^\perp$ is a *parametrization surface* (PS)

for the equation $\partial(\mathbf{g})f = 0$, meaning, roughly, that solutions of the equation are parametrized by functions on the parametrization surface.

In case we are dealing with a spread $\mathbf{g}$ of hypersurfaces parallel to a given hypersurface L_0 through the origin, the equation for spread functions

$$\overrightarrow{\partial(\mathbf{g})}f = 0 \tag{1.9}$$

is to be interpreted as a system of equations $\partial_j(\mathbf{g})f = 0$ where $\{\partial_j(\mathbf{g})\}$ is a basis for directional derivatives along L_0. (We shall often use the notation $\overrightarrow{\partial(\mathbf{g})}$ to indicate that this is a system of equations.)

The same definition applies to $\mathbf{g}$ and L_0 of any dimension $l \leq n-1$. Now $\mathbf{G} = \{\mathbf{g}\}$ is the Grassmannian $G(n,l)$ of l planes in $\mathbb{R}^n$ through the origin. Again $L_0^\perp$ is a parametrization surface for (1.9) and $\{\delta_{L(s,\mathbf{g})}\}_{s\in L_0^\perp}$ forms a basis for solutions of (1.9).

It is clear where these examples lead. We invert our procedure and start from a differential equation or system $\vec{P}(D)f = 0$ like (1.9). This defines an "analytic spread" which is the space of solutions. In the "geometric case" the solutions had a given basis $\{\delta_{L(s,\mathbf{g}^0)}\}_s$. Is there an analog for general partial differential equations?

We approach this problem by searching for a subset $S = S_{\vec{P}(D)}$ of $\mathbb{R}^n$ (or, more generally, of the manifold on which our equation is given) which is a parametrization surface for $\vec{P}(D)$.

Let S be a "smooth" subset of $\mathbb{R}^n$ and let $\{h_j(x)\}_{j=1,\dots,r}$ be differential operators (which play the role of normal derivatives). A *parametrization problem* (PP) for $(\vec{P}(D); S, \{h_j\})$ is the determination of the kernel and image of the map

$$\alpha: \quad f \to \{h_j(x,D)f|_S\}$$

for f in the kernel $\mathcal{W}'(\vec{P}(D))$ of $\vec{P}(D)$ in $\mathcal{W}'$. The image of α is the set of *parametrization data*. Any set of r functions $\{g_j\}$ on S is called *potential parametrization data*.

Actually what we have described should more properly be called a geometric PP.

Other ways of parametrizing solutions are discussed in Section 1.4.

In the most favorable situation the set of parametrization data can be identified with a space of the form $[\mathcal{W}'(S)]^r$ where $\mathcal{W}'(S)$ can be thought of as the space of restrictions of functions in $\mathcal{W}'$ to S. The topology of $\mathcal{W}'(S)$ can generally be defined in a natural way from the topology of $\mathcal{W}'$. When the map from $\mathcal{W}'(\vec{P}(D))$ to $[\mathcal{W}'(S)]^r$ is a topological isomorphism we say the PP is *semi well posed*. (The concept of "well posed" is introduced in Section 1.4.)

It is allowed that $r = \infty$, in which case we call the PP *exotic*.

The primary example of the PPs we deal with is the Cauchy problem (CP) which means that S is a plane or a "suitable" type of surface[3] so we shall often use the word "Cauchy" for "parametrization."

It is sometimes convenient to regard parametrization surfaces in the framework of bases (see (1.17)ff). In fact the injectivity of α means that $\{h_j(D)\delta_s\}$ forms a basis for the dual of the kernel of $\vec{P}(D)$. We regard this as a geometric basis. Other interesting bases, e.g. those defined by eigenvalue problems, appear in our work. Some of our most important results depend on the interrelation of bases.

A PP for $\vec{P}(D)$ is called *hyperbolic* if for each $s^0 \in S$ and each $j^0 \in \{1, \dots, r\}$ there is a unique solution $\eta(s^0, j^0)$ of $\vec{P}(D)\eta(s^0, j^0) = 0$ whose parametrization data is given by

$$\left[h_j(D)\eta(s^0, j^0)\right](s) = \begin{cases} \delta_{s=s_0}, & \text{if } j = j_0 \\ 0, & \text{otherwise.} \end{cases}$$

Such an η is called a *null solution* of $\vec{P}(D)$. When fundamental solutions exist the null solutions can often be expressed in terms of them. The null solutions form a basis for all of $\mathcal{W}'(\vec{P}(D))$ because their Cauchy data (CD) form a basis for all CD.

As mentioned above we can think of $\mathcal{W}'(\vec{P}(D))$ as the generalization of the space of spread functions. The null solutions play the roles of the $\delta_L(s, \mathbf{g})$ and so are the "leaves" of the spread. We are led to define the Radon transform associated to $\vec{P}(D)$ by

$$\mathbf{R}_{\vec{P}(D)}u = \text{restriction of } u \text{ to } \mathcal{W}'(\vec{P}(D)). \tag{1.10}$$

In (1.10) we consider u as an element of $\mathcal{W}''$ so that $\mathbf{R}_{\vec{P}(D)}u$ is an element of $[\mathcal{W}'(\vec{P}(D))]'$.

Since the null solutions form a basis for $\mathcal{W}'(\vec{P}(D))$ we can write $\mathbf{R}_{\vec{P}(D)}u$ in the form

$$\mathbf{R}_{\vec{P}(D)}u(s, j) = \int u\eta(s, j).$$

(The integral is only a suggestive way of expressing $u \cdot \eta(s, j)$.)

We regard the subspaces $\mathcal{W}'(\vec{P}(D))$ as coarse grains of $\mathcal{W}'$. In the case when we have a hyperbolic parametrization surface S then each point $s^0 \in S$ defines a semi-coarse grain, namely the space of solutions of $\vec{P}(D)f = 0$ whose parametrization data is supported at the point s^0. In this way $\mathcal{W}'(\vec{P}(D))$ is decomposed into these semi-coarse grains.

[3]We do not know a "good" general definition of "suitable." In any case local PP are generally CP.

We shall meet interesting situations in which these semi-coarse grains are infinite dimensional.

We can form the space $\mathcal{W}(\vec{P}(D))$ which is the dual of $\mathcal{W}'(\vec{P}(D))$. Via the parametrization data we can regard

$$\mathcal{W}(\vec{P}(D)) \approx \mathcal{W}^r(S)$$

where $\mathcal{W}(S)$ is some suitable space of "functions" on S.

In case there is more than one parametrization surface, say S_1 and S_2 for $\vec{P}(D)$, with corresponding $\{h_j^1(D)\}$, and $\{h_j^2(D)\}$, the differential equation gives a way of *intertwining* parametrization data between S_1 and S_2. We start with ξ_1 as data on S_1, then solve $\vec{P}(D)f = 0$ with f having data ξ_1 on S_1. The data ξ_2 of f on S_2 is said to correspond to ξ_1 by intertwining (via $\vec{P}(D)$).

There is an essential difference between our original geometric discussion of the Radon transform and the present differential equations approach. For, a differential equation corresponds to a *single* spread. To understand this point, suppose $\vec{P}(D)$ has a hyperbolic PP. Then we have defined the leaves of the spread related to $\vec{P}(D)$ as the set of null solutions; they coincide with the geometric leaves in the usual Radon transform via (1.9). $\mathbf{R}_{\vec{P}(D)}$ is generally not injective; one needs many spreads, i.e. many $\vec{P}(D)(\mathbf{g})$, to produce injectivity.

(2) Groups

Instead of thinking of the spread of parallel planes $\{L(s, \mathbf{g})\}_s$ in terms of differential equations, we can think of it in terms of groups. For the usual Radon transform $L(0, \mathbf{g}) = L_0(\mathbf{g})$ is a subgroup of $\mathbb{R}^n$ as is $L_0^\perp$. The spread is defined by

$$L(s, \mathbf{g}) = s + L_0(\mathbf{g})$$

for $s \in L_0^\perp$. $L_0^\perp$ is a complement for L_0, meaning $L_0 \oplus L_0^\perp = \mathbb{R}^n$.

This is one spread and that is all that the group $\mathbb{R}^n$ can provide. But if we add the action of the rotation group $\mathcal{O}$, which means we pass from $\mathbb{R}^n$ to the affine group $\mathcal{A}^n$ which is the semi-direct product of $\mathcal{O}$ with $\mathbb{R}^n$, then we can obtain the other spreads from $L_0(\mathbf{g})$.

These ideas are expanded in Chapter 7.

Our above discussion of spreads centered on a single basis for the spread functions. Often we can produce a second basis for the same parametrization surface.

In case the cross-section or parametrization surface S is linear or has some group structure then, in addition to the basis $\{\delta_s\}_{s\in S}$, a natural basis for functions or distributions on S can be constructed from representation functions. The interrelation of the bases leads to a relation on the solutions of $\vec{P}(D)f = 0$ produced by expressing the parametrization data of f in terms of the two bases.

Sometimes it is advantageous to iterate the process or combine it with intertwining. Instead of going from a relation on S to a relation amongst solutions

on all of $\mathbb{R}^n$, we only go via intertwining to some intermediate surface S_1 which is a parametrization surface for another differential operator ∂_1. We then use ∂_1 and the relation on S_1 to derive a relation on all of $\mathbb{R}^n$.

Many aspects of this process appear in this work. Among them are:

(α) The projection–slice theorem of the Radon transform (Section 1.7).
(β) The Poisson summation formula (Sections 1.7, 8.2).
(γ) The relation between the Plancherel measures for compact and noncompact forms of semi-simple Lie groups (Section 8.4).
(δ) Selberg's trace formula (Section 8.2).

Solutions of $\vec{P}(D)f = 0$ are determined by their data g on parametrization surfaces S. We can think of a function g on S as a measure on all of $\mathbb{R}^n$ by identifying g with $g\delta_S$. In this way we have a new type of Radon transform of f, namely the Radon transform of g. We give a detailed examination of this process in Chapter 10. In particular this idea enables us to compute periods of Eisenstein and Poincaré series, extending the work of Hecke and Siegel [142].

Instead of differential operators, we can use more general operators such as differential difference operators. To get a feeling for the nature of these ideas, let us consider the case when $n = 1$ and

$$\partial(\mathbf{g}^0) = \tau - I \tag{1.11}$$

where $\tau f(x) = f(x+1)$ and I is the identity operator. Solutions of $\partial(\mathbf{g}^0)H = 0$ are periodic with period 1. A Cauchy surface S is the interval $[0,1]$; the CP is hyperbolic. Some care must be taken because 0 and 1 are identified so that CD must satisfy $h(0) = h(1)$. The null solutions (leaves of the spread defined by (1.11)) are of the form $\eta_s = \sum_n \delta_{s+n}$ for $s \in [0,1)$. The associated Radon transform of a function f which is small at infinity is given by

$$\mathbf{R}f(s) = \eta_s \cdot f = \sum f(s+n)$$

using the basis $\{\delta_s\}_{s\in S}$. If we relate this basis to the basis $\{\exp(2\pi inx)\}$ we obtain the usual Poisson summation formula (PSF).

Remark. *We have defined* $\mathbf{R}$ *by integrating over all of* $\mathbb{R}^n$*. If we "cut off" to a bounded domain* Ω *then we obtain a theory which has its roots in the Euler–Maclaurin sum formula (see Chapter 8).*

The Poisson summation formula deals with relations obtained from a single spread. We have noted that, generally, spreads correspond to a "cutting up" of the underlying space. If we have two or more cuttings then it is reasonable to expect relations amongst them. Such relations come, for example, when some function is a spread function for two spreads.

Suppose we have two geometric spreads $\{L^1(s^1, \mathbf{g}^1)\}_{s^1}$, $\{L^2(s^2, \mathbf{g}^2)\}_{s^2}$ and a common spread function

$$f = \sum c_j^1 \delta_{L^1(s_j^1, \mathbf{g}^1)} = \sum c_k^2 \delta_{L^2(s_k^2, \mathbf{g}^2)}.$$

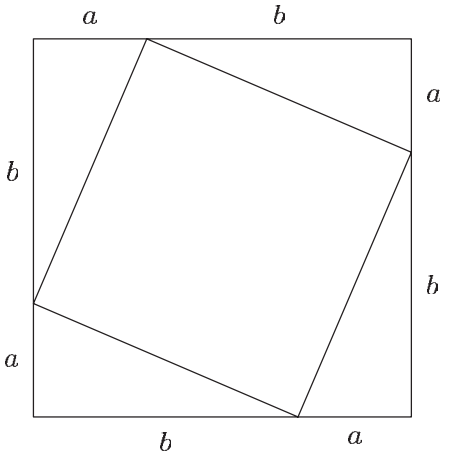

Figure 1.1

If h is a suitable function on $\mathbb{R}^n$ this leads to

$$\sum c_j^1 \int\limits_{L^1(s_j^1,\mathbf{g}^1)} h = \sum c_k^2 \int\limits_{L^2(s_k^2,\mathbf{g}^2)} h. \tag{1.12}$$

For the best known example of this idea, replace $\mathbb{R}^n$ by a square of side $a+b$ and let $f \equiv 1$. The decompositions $\mathbf{g}^1$ and $\mathbf{g}^2$ are given by Figure 1.1.

$\mathbf{g}^1$ is the trivial decomposition with a single s^1 consisting of the whole square of side $a+b$. $\mathbf{g}^2$ consists of the four triangles and the square of side c. Identity (1.12) for $h \equiv 1$ is the Pythagorean theorem.

In our applications to Radon transforms, we must deal with identities involving more than two spreads. Such identities yield the moment conditions (details in Chapter 2).

Let us return to the general structure of the Radon transform. We are now ready for the *reconstruction problem*, which means the determination of a function f from any of its various Radon transforms. In this book we shall concentrate on the theoretical problem of the complete determination of f from a sufficient amount of data. In addition to the complete reconstruction of f from $\mathbf{R}f$ we shall be interested in determining properties of f such as growth and regularity from those of $\mathbf{R}f$.

We shall deal with the important practical problem of *tomography*, which is the reconstruction of some properties of f from incomplete data in the appendix. One of the most significant problems is the determination of functions on a set Ω from data on the exterior of Ω since it may be impractical to physically penetrate Ω.

There are many ways of determining f from its Radon transform. If we were in a Hilbert space setting and $\{\delta_{L(s,\mathbf{g})}\}$ were replaced by a complete orthonormal

set $\{\phi_\alpha\}$ then we could determine f from the "Fourier coefficients" $\{f_\alpha\}$ where

$$f_\alpha = \int f\phi_\alpha$$

by

$$f = \sum f_\alpha \bar{\phi}_\alpha.$$

We refer to this type of reconstruction of f from $\{f_\alpha\}$ as *Hilbert reconstruction.*

There does not seem to be any natural Hilbert space in which $\{\delta_{L(s,\mathbf{g})}\}$ form a complete orthonormal set. Nevertheless we claim that the $\delta_{L(s,\mathbf{g})}$ are "almost orthogonal."

To understand what this means, for each $(s, \mathbf{g})$ let $f(s, \mathbf{g})$ be a smooth non-negative function of compact support supported on a very small neighborhood of $L(s, \mathbf{g})$ with $\int |f|^2 = 1$. If $s_1 \neq s_2$ then if the neighborhoods are thin enough $\int f(s_1, \mathbf{g}) f(s_2, \mathbf{g}) = 0$. On the other hand if $\mathbf{g}_1 \neq \mathbf{g}_2$ then $\int f(s_1, \mathbf{g}_1) f(s_2, \mathbf{g}_2)$ is small, but, in fact, it is clear that this integral gets larger when $\mathbf{g}_2 \to \mathbf{g}_1$ and $s_2 \to s_1$.

All this supports the idea that $\{\delta_{L(s,\mathbf{g})}$ behave like an "approximate orthogonal set." This leads us to a form of Hilbert reconstruction in which we think of $\{\delta_{L(s,\mathbf{g})}\}$ as an "approximately orthonormal basis" for all "functions" on $\mathbb{R}^n$. Thus

$$\mathbf{R}f(s^0, \mathbf{g}^0)\delta_{L(s^0,\mathbf{g}^0)}$$

is sort of the best approximation to f that we can obtain using the basis element $\delta_{L(s^0,\mathbf{g}^0)}$; that is,

$$\mathbf{R}f(s^0, \mathbf{g}^0)\delta_{L(s^0,\mathbf{g}^0)}$$

is essentially the component of f in this basis. (A detailed discussion of bases is found in (1.17) ff. We say "sort of" because f belongs to a space $\mathcal{W}$ of functions which are small at infinity and $\delta_{L(s,\mathbf{g})}$ must be considered as an element of the dual space $\mathcal{W}'$.) For $\mathbf{g}^0$ fixed the coarse grain of f, which is the function $\mathbf{R}f(s, \mathbf{g}^0)$ on the cross-section $S_{\mathbf{g}^0}$, can be used to define the function

$$\mathbf{R}^*\mathbf{R}f(\mathbf{g}^0) = \int \mathbf{R}f(s, \mathbf{g}^0)\delta_{L(s,\mathbf{g}^0)} ds.$$

Thus $\mathbf{R}^*\mathbf{R}f(\mathbf{g})$ is the spread function for $\mathbf{g}$ whose "value" on each leaf is $\mathbf{R}f(s, \mathbf{g})$.

Put in other terms, $\mathbf{R}^*\mathbf{R}f(\mathbf{g}^0)$ can be thought of as the "approximate projection" of f on the space of spread functions for $\mathbf{g}^0$. (Usually f is a function and $\mathbf{R}^*\mathbf{R}f(\mathbf{g}^0)$ is a measure.)

$\mathbf{R}^*$ can also be interpreted as a sort of adjoint of $\mathbf{R}$. This is discussed in detail in Section 1.7.

We can regard the above interpretation of $\mathbf{R}^*\mathbf{R}$ in another light. It seems quite clear that the integral over $L(s, \mathbf{g}) \cap L(s', \mathbf{g}')$ is largest when $s = s'$, $\mathbf{g} = \mathbf{g}'$. Although it may be infinite in other circumstances, the infinity is largest when

the two planes coincide. $\mathbf{R}^*\mathbf{R}$ can be understood in terms of "renormalization," a process which is common in quantum field theory.

We mentioned above that, in physical problems, the quantity $\mathbf{R}f(s,\mathbf{g})$ is significant because it represents the average of f over all the points on the affine hyperplane defined by $(s,\mathbf{g})$, which is the best information we can glean from our experiment. Such an experiment could not lead to a better approximation to f than $[\mathbf{R}f(s,\mathbf{g})]\,\delta_{L(s,\mathbf{g})}$.

We can now try to put all the $\mathbf{R}^*\mathbf{R}f(\mathbf{g})$ together to reconstruct f. The simplest process would be to form $\int \mathbf{R}^*\mathbf{R}f(\mathbf{g})d\mathbf{g}$. This is not quite f because of various impediments, most important being the overlap of leaves in the various spreads. Nevertheless, this integral is "close to" f; one has to apply a power of the Laplacian to obtain f from it (Section 2.1).

Let us summarize the reconstruction process for generic functions f:

(1) Integrate f over $L(s,\mathbf{g})$ to obtain the value $\mathbf{R}f(s,\mathbf{g})$ at the point s of the cross-section $S_{\mathbf{g}}$.
(2) Replace the value $\mathbf{R}f(s,\mathbf{g})$ by $\mathbf{R}f(s,\mathbf{g})\delta_{L(s,\mathbf{g})}$.
(3) Form the integral

$$\int \mathbf{R}f(s,\mathbf{g})\delta_{L(s,\mathbf{g})}\,ds = (\mathbf{R}^*\mathbf{R}f)(\mathbf{g})$$

which is the spread function (projection on the space of $\mathbf{g}$ spread functions) for the spread $\mathbf{g}$ associated by $\mathbf{R}^*\mathbf{R}$ to f.
(4) Combine the various $\mathbf{R}^*\mathbf{R}f(\mathbf{g})$ by some integral–differential operation.

Suppose we are given a "nice" decomposition of $\mathbb{R}^n$ (or some other manifold) into sets (slices) $S_{\mathbf{g}}$. If f is a function we can form its restrictions $f_{\mathbf{g}}$ to the $S_{\mathbf{g}}$. These now form the coarse grains of f. Then we can obtain f in an obvious way from the slices $f_{\mathbf{g}}$ or from the $\{S_{\mathbf{g}}\}$ decomposition $\{f_{\mathbf{g}}\delta_{S_{\mathbf{g}}}\}$.

We refer to the latter as a *slice decomposition* and to the former (i.e. (4)) as a *projective decomposition*. Actually we can regard the slice decomposition as a projective decomposition. The projective decomposition is defined using the nonlocal operator of integration on leaves to define $\mathbf{R}f(s,\mathbf{g})\delta_{L(s,\mathbf{g})}$ while the slice decomposition is defined using the local operator, namely the identity. The projection assigns a single number to a leaf while the slice assigns a function on the leaf. Thus one slice decomposition generally suffices to determine the function while for projective decompositions many spreads are needed.

In general when we have a "basis" $\mathbf{A} = \{a_s\}$ for a subspace $\mathcal{W}_0'$ of a (dual) space $\mathcal{W}'$ then we can form the $\mathbf{A}$ transform of functions in $\mathcal{W}$, namely $\mathbf{A}f(s) = a_s \cdot f$. We can then construct the projection

$$\mathbf{A}^*\mathbf{A}f = \int \mathbf{A}f(s)\tilde{a}_s\,ds \tag{1.13}$$

where ds is some fixed measure and $\tilde{a}_s$ is suitably chosen. In this way we have "projected" (or "sliced") f into $\mathcal{W}'$. We call $\mathbf{A}^*\mathbf{A}$ the *projection associated* to $\mathbf{A}$.

For example, let $\mathcal{W}$ be the space of functions (of x) on a linear subspace L of $\mathbb{R}^n$ and $\{a_s\} = \{\exp(is\cdot x)|_L \delta_L\}_{s\in\hat{L}}$ where $\hat{L}$ is the dual of L. Then $\mathbf{A}f = \mathcal{F}(f|_L)$ is the Fourier transform of $f|_L$ and (using $\tilde{a}_s = \bar{a}_s\delta_L$ in the definition of $\mathbf{A}^*$) $\mathbf{A}^*\mathbf{A}f = f\delta_L$.

In the case of hyperbolic operators discussed above we also used restrictions $h_j(D)f|_S$ defined by local operators $h_j(D)$. Such restrictions can also be regarded in the frame of slice decompositions when we have many $S(\mathbf{g})$.

However, the reconstruction of functions from slice decompositions, especially if the only operator is the identity, is clearer than from projective decompositions. Thus we often seek a change of basis to convert a projective decomposition into a slice decomposition.

One of the main tools we shall use in this book to convert projective decompositions into slice decompositions is the Fourier transform, i.e. basis of exponentials. For the ordinary Radon transform it takes the following form. After this change of basis, the slices are the planes $\hat{L}_{\mathbf{g}}^{\perp}$ through the origin. ($\hat{L}_{\mathbf{g}}^{\perp}$ is the orthogonal plane to $L_{\mathbf{g}}$ thought of in the Fourier transform space.) With each such slice we define $\mathbf{P}_{\mathbf{g}}$ as the operator which multiplies a function $\hat{f}(\hat{x})$ by $\delta_{\hat{\mathbf{g}}^\perp}$ which is the δ function of $\hat{L}_{\mathbf{g}}^{\perp}$. The Fourier transform of $\mathbf{P}_{\mathbf{g}}\hat{f}$ is

$$\widehat{\mathbf{P}_{\mathbf{g}}\hat{f}} = \delta_{\hat{\mathbf{g}}^\perp}\widehat{\hat{f}(\hat{x})} = \delta_{L_{\mathbf{g}}} * f = \mathbf{R}^*\mathbf{R}(\mathbf{g})(f).$$

Polar coordinates tell us how to write the identity operator as an integral of $\mathbf{P}_{\mathbf{g}}$. This shows us how to write a function f in terms of $\{\mathbf{R}^*\mathbf{R}(\mathbf{g})f\}$ (see Chapter 2).

A point which is crucial in this connection is that we can obtain *special Radon transforms* if the support of $\hat{f}$ is not all of $\mathbb{R}^n$ but is some subset V which is a union of planes $\hat{L}_{\mathbf{g},V}^{\perp}$ through the origin. Then the above reconstruction of f from $\{\mathbf{R}^*\mathbf{R}(\mathbf{g})f\}$ shows how to write f in terms of these special functions $\mathbf{R}^*\mathbf{R}(\mathbf{g},V)f$ whose Fourier transform is contained in $\hat{L}_{\mathbf{g},V}^{\perp}$ (see example below).

In particular when V is an algebraic variety then the condition support $\hat{f} \subset V$ is equivalent to f being the solution of a system $\vec{P}(D)f = 0$ of partial differential equations (see Section 1.4). When V is the union of planes $\{\mathbf{g}\}$, e.g. if V is a cone, then the Radon "coarse grain" decomposition of $\hat{f}$ into $\{\delta_{\hat{\mathbf{g}}^\perp}\hat{f}\}$ shows how to write solutions of $\vec{P}f = 0$ in terms of special solutions.

These considerations depend on the fact that f is not too large at infinity so that $\mathbf{R}^*\mathbf{R}f(\mathbf{g},V)$ is defined. For large functions f we can still define $\hat{f}$ and its support is $\subset V$ when $\vec{P}(D)f = 0$ (see Section 1.4). But $\hat{f}$ is defined in a somewhat nebulous fashion; in particular it is not uniquely determined from f although f is uniquely determined by $\hat{f}$. Nevertheless $\hat{f}$ is defined by a "nice" measure on the *complex* algebraic variety V. Thus it makes sense to multiply

by $\delta_{\hat{\mathbf{g}}}$ which is now the δ function of the complexification of $\hat{L}_{\mathbf{g}}^{\perp}$. In this way we obtain a (nonunique) decomposition of solutions of $\vec{P}(D)f = 0$ into special solutions.

The simplest example is $\vec{P}(D)f = \Delta f$, where Δ is the Laplacian. For $n = 2$

$$V = \{\hat{x}_1^2 + \hat{x}_2^2 = 0\} = \{\hat{x}_1 = i\hat{x}_2\} \cup \{\hat{x}_1 = -i\hat{x}_2\}.$$

Our decomposition corresponds to writing a harmonic function as a sum of a holomorphic function (support $\hat{f} \subset \{\hat{x}_2 = i\hat{x}_1\}$) and an antiholomorphic function (support $\hat{f} \subset \{\hat{x}_2 = -i\hat{x}_1\}$). Holomorphic functions are constant in $\mathbb{C}^2$ on the sets $\{x_1 + ix_2 = \text{const.}\}$. Thus they are spread functions for these "complex spreads." The decomposition is not unique because of the constants.

In Section 1.4 we shall explain how to form an analogous decomposition of harmonic functions when $n > 2$. For $n = 3, 4$ the idea is due to Whittaker and Bateman. Many other decompositions occur in this work.

In Section 9.3 we describe the lack of uniqueness in such representations. In certain cases our ideas are related to those of Penrose on twisters (see [32]).

This example shows that for slice decompositions it is often useful to use, as a basis for functions on each slice, the restrictions of a global basis to the slice. For the usual Radon transform or for the special Radon transforms described above the global basis is the exponentials. (Recall that the slices are in $\hat{x}$ space.)

Remark. *These decompositions into coarse grains which are solutions of simpler equations can be regarded as an analog of the Radon ansatz in $\hat{x}$ space.*

There are other choices for bases. For the hyperplane Radon transform $\hat{L}_{\mathbf{g}}^{\perp}$ is a line so we can use the restrictions of $\{|\hat{x}|^{is}\}$ as a basis; the decomposition in this basis is the *Mellin transform*. (Some modification is needed if $\hat{f}(-\hat{x}) \neq \hat{f}(\hat{x})$ and there are some problems at the origin.)

When each of the slices is a Lie group or a homogeneous space G/H of a Lie group G by a fixed closed subgroup H such that G behaves nicely with respect to the Fourier transform then it is natural to decompose functions on the slices under the action of G. This means that we use representation functions of G as a basis on G/H and hence on each slice.

There is a major difference between the bases $\{\exp(ix \cdot \hat{x})\big|_{\hat{L}_{\mathbf{g}}^{\perp}}\}$ and $\{|\hat{x}|^{is}\}$ introduced in the above illustrations. In the case of $\{|\hat{x}|^{is}\}$ we use a single set of functions on $\mathbb{R}^n$ whose restrictions to any slice form a basis on the slice. This is not true for the basis $\{\exp(ix \cdot \hat{x})\}$; we use $\hat{x} \in \hat{L}_{\mathbf{g}}^{\perp}$ so the set of $\hat{x}$ depends on $\mathbf{g}$. When we use a set of global functions whose restrictions to a slice are linearly independent and define a basis on the slice, we call the basis *harmonic*. Such bases play an important role in our work (see Chapter 3).

Let us illustrate the usage of harmonic bases. Suppose

$$P(D) = \frac{\partial^m}{\partial x_1^m} + P_1\left(\frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_n}\right)\frac{\partial^{m-1}}{\partial x_1^{m-1}} + \cdots + P_m\left(\frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_n}\right).$$

Let V be defined by

$$(i\hat{x}_1)^m + P_1(i\hat{x}_2, \dots, i\hat{x}_n)(ix_1)^{m-1} + \cdots + P_m(i\hat{x}_2, \dots, i\hat{x}_n) = 0.$$

Let $\mathcal{W}'$ be a suitable space of functions or distributions whose dual $\mathcal{W}$ consists of functions which are small at infinity (AU space in the language of Section 1.4). We call $\mathcal{W}'(P)$ the kernel of $P(D)$ in $\mathcal{W}'$. In Section 1.4 we shall show that the Fourier transform $\hat{\mathcal{W}}(P)$ of the dual of $\hat{\mathcal{W}}'(P)$ consists of functions $\hat{F}$ which can be considered as a holomorphic function on V. Let $\{\mathbf{g}\} = \{\mathbf{g}(V)\}$ be the slices

$$\mathbf{g}(\hat{x}_2^0, \dots, \hat{x}_n^0) = V \cap \{\hat{x}_2 = \hat{x}_2^0, \dots, \hat{x}_n = \hat{x}_n^0\}.$$

Each slice consists of m points (with multiplicity) in the complex $\hat{x}_1$ plane.

The harmonic basis is

$$\{\hat{u}_j(\hat{x})\} = \{1, \hat{x}_1, \dots, \hat{x}_1^{m-1}\}.$$

The Lagrange interpolation formula says that this is a basis. (Some care must be taken when roots of $P(\hat{x}_1, \hat{x}_2^0, \dots, \hat{x}_n^0)$ coincide.) In fact we have for $\hat{F} \in \hat{W}(P)$

$$\delta_{\mathbf{g}}\hat{F} = \sum \hat{f}_j(\mathbf{g})\delta_{\mathbf{g}}\hat{u}_j(\hat{x}). \tag{1.14}$$

Clearly $\mathbf{g}(V) \subset \mathbf{g}(\mathbb{C}^n)$ which is the complex line containing $\mathbf{g}(V)$; the union of the $\mathbf{g}(\mathbb{C}^n)$ is $\mathbb{C}^n$. The $\hat{f}_j$ of (1.14) are the Lagrange interpolation coefficients of $\hat{F}$ on the line $\mathbf{g}(\mathbb{C}^n)$, so they depend on $\hat{x}_2, \dots, \hat{x}_n$. If we take the Fourier transform of (1.14) in all variables we find

$$F = \sum f_j(x_2, \dots, x_n)\frac{\partial^j}{\partial x_1^j}\delta_{x_1=0}. \tag{1.15}$$

This means that every element of the dual of the space of solutions of $P(D)f = 0$ can be expressed in terms of functions (distributions) on $x_1 = 0$ and their x_1 derivatives of order $< m$.

By duality this says that every solution of $P(D)f = 0$ can be expressed in terms of its CD on $x_1 = 0$. (This is clarified in Section 1.7; it is treated in detail in Chapter IX of Ehrenpreis [33] and appears in several places in this book.)

Let us descend from the "clouds" to more pedestrian questions.

The injectivity of the Radon transform means that for no suitable function $f \not\equiv 0$ can $\delta_{L(s,\mathbf{g})} \cdot f$ vanish for all $s, \mathbf{g}$. To put this in more precise terms,

$$\{\delta_{L(s,\mathbf{g})}\}$$

is a *spanning set* for $\mathcal{W}'$. Moreover, the finite linear combinations of the $\delta_{L(s,\mathbf{g})}$ are linearly independent in $\mathcal{W}'$ if $\mathcal{W}$ is large enough. But the infinite linear combinations are not linearly independent. For, given any $\mathbf{g}_0$ we have

$$\int \delta_{L(s,\mathbf{g}_0)}\, ds = dx. \tag{1.16}$$

Thus the left side of (1.16) is independent of $\mathbf{g}_0$. This is an infinite linear relation amongst the $\delta_{L(s,\mathbf{g})}$.

The moment conditions (which will be discussed in various places in this book) generate all nontrivial infinite relations amongst $\{\delta_{L(s,\mathbf{g})}\}$.

The above shows that $\{\delta_{L(s,\mathbf{g})}\}$ is not a basis for $\mathcal{W}'$ in the usual sense, meaning a spanning set $\{w'_u\}$ such that

$$\lim_k \sum_u a_u^k w'_u = 0$$

implies $a_u^k \to 0$ for all u. (For each k the sums are finite.)

We shall sometimes use the term "basis" in the somewhat imprecise form of "spanning set with 'few' relations."

The category of bases which fits $\{\delta_{L(s,\mathbf{g})}\}$ best comes under the appellation of *direct integral*. Von Neumann [125] introduced this concept in the framework of Hilbert spaces. Let $\{\mathcal{H}_\alpha\}$ be Hilbert spaces depending on a parameter α defined on some space A with a fixed positive measure $d\alpha$. Then

$$\mathcal{H} = \int_\oplus \mathcal{H}_\alpha \, d\alpha$$

consists of all *cross-sections* h which are functions defined on A with $h(\alpha) \in \mathcal{H}_\alpha$ such that

$$\|h\|^2 = \int \|h(\alpha)\|^2 d\alpha < \infty.$$

In our theory the Hilbert spaces $\mathcal{H}_\alpha$ are replaced by locally convex topological vector spaces such as spaces of distributions $\mathcal{W}'(\mathbf{g})$. The measure $d\alpha$ becomes some fixed measure $d\mathbf{g}$. The direct integral

$$\mathcal{W}'_1 = \int_\oplus \mathcal{W}'(\mathbf{g}) \, d\mathbf{g}$$

consists of all functions w' such that $w'(\mathbf{g}) \in \mathcal{W}'(\mathbf{g})$ and such that the integral

$$\int w'(\mathbf{g}) \, d\mathbf{g} \tag{1.17}$$

defines a suitable type of distribution.

To make contact with the Radon transform we want to show how $\{\delta_{L(s,\mathbf{g})}\}_s$ can be interpreted as a basis for $\mathcal{W}'(\mathbf{g})$ using the concept of direct integral. It is somewhat simpler to explain how $\{\delta_s\}$ is a basis for function spaces $\mathcal{W}(S)$. Given any function f on S we can write

$$f = \int f(s)\delta_s \, ds$$

meaning that, when thought of as distributions,

$$f \cdot u = \left[\int f(s)\delta_s \, ds \right] \cdot u = \int f(s)u(s) \, ds.$$

Thus

$$\mathcal{W}(S) = \int_{\oplus} \mathcal{W}_{0s} \, ds \tag{1.18}$$

where $\mathcal{W}_{0s}$ is the space of constants times δ_s.

Actually in (1.18) we take only the cross-sections $w_0(s) = w(s)\delta_s$ for which w is a function belonging to $\mathcal{W}(S)$. Thus the space of cross-sections may be restricted by regularity and growth just as in von Neumann's theory the cross-sections have finite norm.

Remark. *The basis elements δ_s may not belong to $\mathcal{W}(S)$. It is only certain cross-sections (1.18) which lie in $\mathcal{W}(S)$. Borrowing the term from von Neumann we shall refer to such cross-sections as wave packets.*

When dealing with spaces of distributions such as $\mathcal{W}'(\mathbf{g})$ we may have to introduce derivatives of δ_s with respect to s. Thus, instead of (1.18) we can write

$$\mathcal{W}'(S) = \int \mathcal{W}'_{0s} \, ds \tag{1.18*}$$

where now

$$w'_0(s) = \sum w_{0j}(s)\delta_s^{(j)}$$

with w_{0j} suitable distributions. (All this assumes that the distributions in $\mathcal{W}'(S)$ are of finite order.) In this sense $\{\delta_s^{(j)}\}$ is considered a basis for $\mathcal{W}'(S)$. (Actually, $\{\delta_s\}$ is a basis because $\delta_s^{(j)}$ is a limit of linear combinations of $\delta_{s'}$.)

If we replace δ_{s^0} by $\delta_{L(s^0,\mathbf{g})}$ then we have a "direct integral" representation of the space of spread functions $\mathcal{W}'(\mathbf{g})$ in terms of the one-dimensional spaces spanned by the $\delta_{L(s^0,\mathbf{g})}$.

A central point in the theory of the Radon transform is

$$\mathcal{W}' = \int_{\tilde{\oplus}} \mathcal{W}'(\mathbf{g}) \, d\mathbf{g}. \tag{1.19}$$

We have written $\tilde{\oplus}$ rather than $\oplus$ because the right side of (1.19) represents all elements of $\mathcal{W}'$ though not uniquely. The moment conditions show that this integral is not direct and, in fact, they describe the relations (see Section 2.1 for details).

We shall discuss spreads in more detail in Section 1.3. In all our examples the spreads are all isomorphic, meaning all $\mathcal{W}'(\mathbf{g})$ are isomorphic to some fixed space $\mathcal{W}'(S)$. In this case we can also write

$$\mathcal{W}' = \mathcal{W}'(S)\tilde{\otimes}\mathcal{W}'(\mathbf{G}) \tag{1.19*}$$

where $\mathcal{W}'(\mathbf{G})$ is some suitable space of functions or distributions on $\mathbf{G}$. (The reason for writing $\tilde{\otimes}$ instead of $\otimes$ will be explained presently.)

In the usual definition of the tensor product of functions of x and functions of y, we start with product functions $f(x)h(y)$ then form the closure of finite linear combinations. When we write $f(x)h(y)$ it is tacitly assumed that f is extended to a function of (x, y) by making it constant in y, and similarly for h. Whenever we have a coordinate system, say x, y (x and y may be multidimensional variables), and we want to realize a tensor product of functions or distributions on x with functions or distributions on y as a space of functions or distributions on x, y, we need some idea of extension; that is, a way of extending functions or distributions on x or on y to corresponding objects on (x, y). For the usual tensor product we then form the linear space generated by products of these extensions.

In the case of (1.19*) the extension from S to $\mathbb{R}^n$ depends on $\mathbf{g} \in \mathbf{G}$. For each $\mathbf{g}$ we extend the distribution $U \in \mathcal{W}'(S)$ to the spread distribution $U_{\mathbf{g}}$ for $\mathbf{g}$ which equals U on the cross-section $S_{\mathbf{g}}$ corresponding to $\mathbf{g}$. This means that for suitable functions f on $\mathbb{R}^n$

$$U_{\mathbf{g}} \cdot f = U \cdot \mathbf{R}f(\mathbf{g})$$

so that

$$U_{\mathbf{g}} = \mathbf{R}'_{\mathbf{g}} U.$$

We identify $U_{\mathbf{g}}$ with the tensor $U \otimes \delta_{\mathbf{g}}$. We can think of $\{\delta_{\mathbf{g}}\}$ as a basis for $\mathcal{W}'(\mathbf{G})$. Our definition of $U \otimes \delta_{\mathbf{g}}$ extends by linearity to define the tensor product (1.19*).

The contrast between the usual tensor product of functions of x with functions of y and (1.19*) can be viewed as follows. The tensors $f(x) \otimes \delta_y$ (resp. $\delta_x \otimes g(y)$) satisfy differential equations which are independent of y (resp. x) whereas the differential equation satisfied by $U \otimes \delta_{\mathbf{g}}$ depends on $\mathbf{g}$. The former is an example of *harmonicity*, which is discussed in detail in Chapter 3.

Remark. *Our definition of tensor product does not treat the factors in a symmetric fashion.*

We call the type of extension of U to $U_{\mathbf{g}}$ *projection extension* as it is dual to the Radon projection. The corresponding tensor product (1.19*) is called the **projection tensor product**.

Relation (1.19*) is not a true tensor product but rather a *tensor product with amalgamation* for there are identities amongst the linear combinations of spread functions. For the simplest identity we note that if $U = ds$ then $U \otimes \delta_{\mathbf{g}_0} = dx$ is independent of $\mathbf{g}_0$.

Thus (1.19*) is the quotient of the ordinary tensor product by a set of identities. For this reason we shall call it a *supra tensor product*. Because of these identities we write $\tilde{\otimes}$ instead of $\otimes$. These identities correspond to the moment conditions. For more details see Sections 1.2 and 2.1.

We can also define a tensor product which is close to the usual tensor product but is modified because of nontrivial intersections. As we mentioned above (following (1.13)) in the example of the hyperplane Radon transform, the $\hat{L}_{\mathbf{g}}^{\perp} = S_{\mathbf{g}}$ are lines through the origin and hence intersect at the origin. Thus the **slice tensor product**

$$\mathcal{W} = \mathcal{W}(S) \underset{\sim}{\otimes} \mathcal{W}(\mathbf{G}) \tag{1.20}$$

will be called *sub tensor product* or *tensor product with identification*; it is a subspace of the usual tensor product. In the case of the hyperplane Radon transform it is the subspace of functions of $(s, \mathbf{g})$ whose value at $\{(0, \mathbf{g})\}$ is independent of $\mathbf{g}$. There are also compatibility conditions for s derivatives at $\{(0, \mathbf{g})\}$.

We could define (1.20) as the subspace of functions on $S \times \mathbf{G}$ which define functions on $\mathbb{R}^n$. From our point of view it is more natural to identify $f \underset{\sim}{\otimes} \delta_{\mathbf{g}}$ with the function which is f on $S_{\mathbf{g}}$ and 0 off $S_{\mathbf{g}}$ and then define (1.20) as the limit of sums

$$\sum f_j \underset{\sim}{\otimes} \delta_{\mathbf{g}_j}$$

for which the f_j agree on the intersection of the $S_{\mathbf{g}_j}$. The value of the sum at the points of intersection is the common value of the f_j.

Sometimes we require stronger conditions such as compatibility of derivatives on the intersections.

For both $\underset{\sim}{\otimes}$ and $\tilde{\otimes}$ the use of the basis $\{\delta_{\mathbf{g}}\}$ for $\mathbf{G}$ shows how to identify $\underset{\sim}{\otimes}$ or $\tilde{\otimes}$ with a supra or sub direct integral over $\mathbf{G}$ (see Section 1.8 for details). This basis also allows us to define these direct integrals even when the $S_{\mathbf{g}}$ are not isomorphic.

The terms "projection" and "slice" are consistent with our previous explanations. For $\tilde{\oplus}$ we identify $U \tilde{\otimes} \delta_{\mathbf{g}}$ with the spread function on the spread defined by $\mathbf{g}$ and equal to U on the parametrization surface $S_{\mathbf{g}}$. The definition of $U \tilde{\otimes} \delta_{\mathbf{g}}$ involves a nonlocal extension from $S_{\mathbf{g}}$. Conversely $f \underset{\sim}{\otimes} \delta_{\mathbf{g}}$ equals f on $S_{\mathbf{g}}$ and vanishes off $S_{\mathbf{g}}$. This is a purely local extension from $S_{\mathbf{g}}$—hence the term "slice."

In both cases we treat the tensor product in a nonsymmetric manner.

As mentioned above it is analysis which allows us to obtain a deeper understanding of the Radon transform. One aspect of analysis which enables us to piece together spread functions in a consistent manner depends on the

Projection–slice ansatz. *There exists a function $\phi(x, \tilde{x})$ where $x \in \mathbb{R}^n$ and $\tilde{x}$ is a variable in some finite dimensional manifold such that for any $\mathbf{g}$, if $\tilde{S}_{\mathbf{g}}$ denotes the set of $\tilde{x}$ for which $\phi(x, \tilde{x})$ is a spread function for $\mathbf{g}$, then*

$$\{\phi(x, \tilde{x})\}_{\tilde{x} \in \tilde{S}_{\mathbf{g}}}$$

is a basis for $\mathcal{W}'_{\mathbf{g}}$.

Remark. *This property of ϕ can be regarded as an analog of the fundamental principle (see Section 1.4) for the integral transform defined by ϕ. The standard*

ϕ arise from eigenfunction expansions; in this work ϕ is usually taken as the exponential function, except in Chapter 7. (Spread functions are the analog of solutions of a differential equation.)

$\phi(x, \tilde{x})$ is called a *projection–slice function*; it allows us to deal with all spreads simultaneously. For the usual Radon transform we can use

$$\phi(x, \tilde{x}) = e^{ix\cdot\tilde{x}}. \tag{1.21}$$

If we have such a projection–slice function then we can form the ϕ transform of suitable functions f

$$\tilde{f}(\tilde{x}) = \int f(x)\phi(x, \tilde{x})\, dx. \tag{1.22}$$

Suppose that for each $\mathbf{g}$ the measure $dx = ds\, d\lambda$ where $d\lambda$ is the measure on the slices of $\mathbf{g}$ used to define $\mathbf{R}$ and ds is a measure on $S_{\mathbf{g}}$. Then by Fubini's theorem

$$\tilde{f}(\tilde{x}) = \int \mathbf{R}f(s, \mathbf{g})\phi(s, \tilde{x})\, ds \tag{1.23}$$

for $\tilde{x} \in \tilde{S}_{\mathbf{g}}$.

Thus when restricted to $\tilde{x} \in \tilde{S}_{\mathbf{g}}$ the right side of (1.23) may be thought of as the $\phi_{\mathbf{g}}$ transform of $\mathbf{R}f$ on the spread of $\mathbf{g}$. Here

$$\phi_{\mathbf{g}}(\tilde{x}) = \phi(s, \tilde{x}).$$

Equation (1.23) is a general form of what we shall refer to as the **projection–slice theorem**. Think of $\mathbf{R}f(s, \mathbf{g})$ as the Radon projection of f on $S_{\mathbf{g}}$, and the restriction of $\tilde{x}$ to $\tilde{S}_{\mathbf{g}}$ as a slice. Thus (1.23) can be stated as:

The restriction of the ϕ transform of f to $\tilde{S}_{\mathbf{g}}$ is the same as the $\phi_{\mathbf{g}}$ transform of the Radon projection of f on $S_{\mathbf{g}}$.

Even if we do not have a projection–slice function $\phi(x, \tilde{x})$ we still have a (watered-down) version of the projection–slice theorem as long as we have a Fubini decomposition

$$dx = ds\, d\lambda \tag{1.24}$$

for any spread $\mathbf{g}$. Let $u \in \mathcal{W}'_{\mathbf{g}}$ be a spread function. Then Fubini's theorem gives

$$\int f(x)u(x)\, dx = \int \mathbf{R}f(s, \mathbf{g})u(s)\, ds. \tag{1.25}$$

We shall sometimes refer to (1.25) as the weak projection–slice theorem.

This form of the projection–slice theorem is valid in the general hyperbolic set-up described in (1.9)ff. above. Suppose u is a solution of $\partial(\mathbf{g})u = 0$. Then u can be expressed in terms of its CD by

$$u(x) = \sum \int u(s)\eta_j(s, x)\, ds. \tag{1.26}$$

To see this write, as in (1.17)ff.,

$$h_j(\mathbf{g})u\big|_S = \int [h_j(\mathbf{g})u\big|_S(s)]\delta_s \, ds. \tag{1.27}$$

This is the expression of a function on S as an integral of $\{\delta_s\}_{s\in S}$, meaning its expression in terms of the basis $\{\delta_{s\in S}\}$. The $\eta_j(s,x)$ are the null solutions; they (as functions of x) are the solutions of $\overrightarrow{\partial(\mathbf{g})}$ for which

$$h_k(\mathbf{g})\eta_j(s,x)\big|_{x\in S} = \begin{cases} \delta_s & k=j \\ 0 & \text{otherwise.} \end{cases}$$

Thus the right side of (1.26) is an element of the kernel of $\overrightarrow{\partial(\mathbf{g})}$ whose CD is the same as that of h. Equation (1.26) leads to

$$\int f(x)u(x)\,dx = \sum \int \mathbf{R}_{\partial(\mathbf{g})} f(s,\mathbf{g})u(s)\,ds \tag{1.28}$$

which is the general form of (1.25).

We deal with the projection–slice theorem in detail in Section 1.7.

Now is the time to show the power of analysis. We return to the ordinary hyperplane Radon transform. The geometry leads to the basis $\{\delta_{L(s,\mathbf{g}_0)}\}$ for $\mathcal{W}'_{\mathbf{g}_0}$. (For fixed $\mathbf{g}_0$ this is an actual basis.) But one can find many other bases. For example, we could take polynomials which are constant in the direction of $\mathbf{g}_0$ (more precisely, for each k a basis for such polynomials of degree k), or exponentials, etc.

Why are polynomials better than $\{\delta_{L(s,\mathbf{g}_0)}\}$? For any individual spread, δ functions work better than polynomials. But the set of all polynomials forms a ring while the linear combinations of $\{\delta_{L(s,\mathbf{g})}\}$ do not because we cannot square a δ function. Moreover if $\mathbf{g} \neq \mathbf{g}'$ then

$$\delta_{L(s,\mathbf{g})}\delta_{L(s',\mathbf{g}')} = c\delta_{L(s,\mathbf{g})\cap L(s',\mathbf{g}')} \tag{1.29}$$

for a suitable c depending on the angle of intersection. Note that $L(s,\mathbf{g})\cap L(s',\mathbf{g}')$ is a plane of codimension 2 so if we wanted to have a "pseudo-ring" structure we would have to consider simultaneously planes of all dimensions $\leq n-1$.

We have not carried out any algebraic program of this sort. Nevertheless we can sometimes bypass the algebraic ring structure by using the naive principle that "multiplication is a shorthand for repeated addition." This will be clarified below.

To illustrate the significance of the change of basis from $\{\delta_{L(s,\mathbf{g})}\}$ to polynomials, let us examine the proofs of the injectivity of the Radon transform using each of these bases. For simplicity we study the case $n=2$.

Now, the $L(s,\mathbf{g})$ represent geometric objects so it is reasonable to attempt to prove the injectivity of $\mathbf{R}$ on geometric objects by geometric means. Suppose

that $\Omega \subset \mathbb{R}^2$ is a nice domain, and let $\chi = \chi_\Omega$ be its characteristic function. It is clear that

$$\mathbf{R}\chi(s, \mathbf{g}) = \text{length } L(s, \mathbf{g}) \cap \Omega. \tag{1.30}$$

Thus we want to determine Ω by the length of its intersections with all lines.

Suppose first that Ω is a compact strictly convex domain with smooth boundary containing 0 in its interior. We pick some spread and pick an L which is far enough away so that it does not meet Ω, move this L in its spread until it becomes tangent to Ω, say at x_L. We will recognize this point because if we move L further the length of intersections with Ω becomes nonzero. In this way we have determined the conjugate diagram of Ω, meaning the function on the sphere which maps L into $|x_L|$. It is standard that the conjugate diagram of Ω determines Ω itself.

To get more insight into the geometry, we move L beyond the point of tangency; the size of the intersections $L \cap \Omega$ determines the curvature of the boundary. Thus the Radon transform determines the curvature of the boundary of Ω at each point. It is classical that this also determines Ω, but only up to an affine transformation; the additional information the Radon transform gives easily removes the ambiguity of the affine transformation.

Next we pass to the case of nonconvex Ω. We assume for simplicity that the boundary of Ω is real analytic. The above argument for convex sets breaks down completely for sets which are shaped as in Figure 1.2.

We now have two possible options:

(1) Move L parallel to itself.
(2) Rotate L through a point $p \in L$ an amount ϵ to L_ϵ.

Let us explain how to determine the local structure of Ω near the a_j using these motions.

Since the boundary of Ω is real analytic there are only a finite number, say N, of a_j. As a first approximation to the boundary of Ω near a_j we can assume

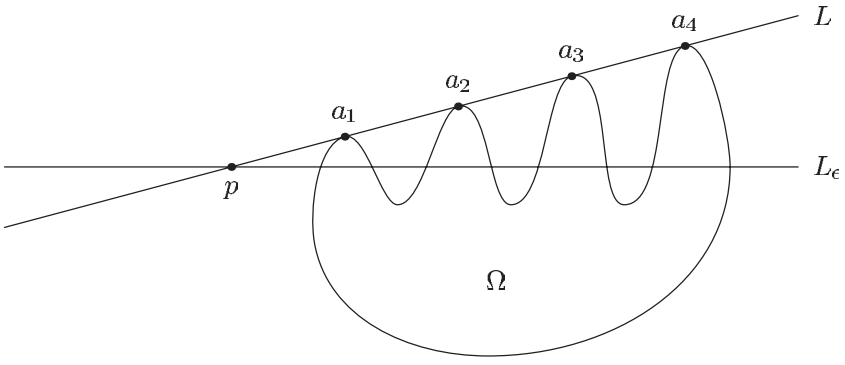

Figure 1.2

it is circular. This means that we have N centers b_j and N radii r_j which we have to determine. But we have great freedom in that we can translate L and rotate it through a variable point $p \in L$.

An easy calculation can be used to determine the b_j and r_j provided that we have some bound on N. This gives the quadratic approximation of Ω near $\{a_j\}$. A similar method gives the whole power series approximation to Ω near $\{a_j\}$. Since the boundary of Ω is real analytic this determines Ω.

If we do not have a bound on N the calculations are more complicated and we have not carried them out. (Bounding N can be regarded as the next stage after convexity.)

Problem 1.1 *Carry out this idea when N is not given an a priori bound and when boundary Ω is not assumed to be real analytic.*

There is another geometric approach to the injectivity and, in fact, to the whole theory of the Radon transform. Suppose we start with $n = 2$ and a function $\tilde{f}$ of compact support. By translation we can assume that support $\tilde{f} \subset$ first quadrant. Let us approximate $\tilde{f}$ by a measure supported on a grid, say

$$\tilde{f} \sim f = \sum f_{ij}\delta_{ij},$$

where (i, j) represent the grid points.

We form

$$\int f(x, 0)\, dx = \sum f_{i0} = h_0.$$

Then we can solve for f_{00} in terms of h_0 (in particular in terms of $\mathbf{R}f$) and f_{i0} with $i > 0$. Next we form

$$\int f(x, j)\, dx = \sum_i f_{ij} = h_j.$$

This enables us to solve for f_{0j} for all j in terms of h_j and f_{ij} for $i > 0$. Similarly using, for $k > 0$,

$$\int f(k, y)\, dy = \sum_i f_{ki} = u_k$$

we can solve for f_{k0} in terms of u_k and f_{ki} for $i > 0$ for all k.

This means that for all j, k including $(0, 0)$ we can solve for all f_{0j} and f_{k0} in terms of $\{h_j\}$, $\{u_k\}$, and $\{f_{jk}\}$ with neither j nor $k = 0$.

We want to iterate the process and solve for f_{jk} with j or $k = 1$ in terms of $\{f_{jk}\}$ and $\mathbf{R}f$ with $j > 1$ and $k > 1$ and then to continue in this way so as to obtain an explicit inversion of the Radon transform for functions defined on a lattice.

To accomplish this we start with a positive integer α; we form the Radon transform of $\{f_{ij}\}$ on the spread of lines of slope α. We call this

$$\mathbf{R}f(s, \alpha) = \sum_i f_{i,s+\alpha i}$$

for s an integer. Starting with $\alpha = 1$ we can solve for f_{ij} when either i or $j = 1$ in terms of $\mathbf{R}f(s,1)$ and $\{f_{ij}\}$ when both $i, j \geq 2$ (except for f_{00} which was considered above).

It is now possible to continue in this manner so as to obtain an explicit expression for $\{f_{ij}\}$ in terms of $\mathbf{R}f$.

It is easy to construct analogous procedures when $n \geq 3$.

In case $\tilde{f}$ is a continuous function, we approximate $\tilde{f}$ by its values f on a grid (times suitable normalizing constants to make $\mathbf{R}\tilde{f}$ approximated by $\mathbf{R}f$); then we let the mesh of the grid approach 0. It now requires estimates on the degree of approximation to determine if the inversion formula for $\mathbf{R}f$ approaches a limit which is an inversion formula for $\mathbf{R}\tilde{f}$.

We have not carried out these estimates but we believe it would be instructive if the reader were to make the appropriate calculations.

We now give a proof of the same result using the basis of polynomials. (We allow any $n \geq 2$.)

We assume that f belongs to a topological vector space $\mathcal{W}$ whose dual $\mathcal{W}'$ contains all polynomials as a dense subspace. If $\mathbf{R}f(s, \mathbf{g})$ vanishes identically in s then it follows from Fubini's theorem that $q \cdot f = \int qf = 0$ for any q which is a spread polynomial for $\mathbf{g}$.

Denote by $\sigma(\mathbf{g})$ the (linear) spread function for $\mathbf{g}$ whose value on any leaf is s. If $\partial/\partial\alpha$ represents a directional derivative then we sometimes use the notation $\alpha^{\perp}$ for $\sigma(\mathbf{g})$ where $\mathbf{g}$ is the spread whose leaves are orthogonal to α.

A spread polynomial for $\mathbf{g}$ is a linear combination of the $\sigma^m(\mathbf{g})$. The injectivity of $\mathbf{R}$ is a consequence of the fact that every polynomial is a linear combination of spread polynomials, which we can now verify.

We make the following important observation which we discussed above: for each $\mathbf{g}$ in the Grassmannian $\mathbf{G}$ of hyperplanes in $\mathbb{R}^n$ the spread functions $u_{\mathbf{g}}$ are characterized as being solutions of the system of differential equations

$$\frac{\vec{\partial} u_{\mathbf{g}}}{\partial \mathbf{g}} = 0. \tag{1.31}$$

This means that if $\partial/\partial\alpha_1, \ldots, \partial/\partial\alpha_{n-1}$ forms a basis for directional derivatives along $\mathbf{g}$ then

$$\frac{\partial u_{\mathbf{g}}}{\partial \alpha_j} = 0$$

for all j.

A sum P of functions each of which depends only on $\mathbf{g}_k^{\perp}$ (for distinct $\mathbf{g}_k, k = 1, \ldots, r$, with basis for directional derivatives $\{\partial/\partial\alpha_{kj}\}$) satisfies

$$\left(\frac{\vec{\partial}}{\partial \mathbf{g}_1} \frac{\vec{\partial}}{\partial \mathbf{g}_2} \cdots \frac{\vec{\partial}}{\partial \mathbf{g}_r} \right) P = 0 \tag{1.32}$$

where the product means the set of all products $\partial^r/\partial\alpha_{1j_1}\dots\partial\alpha_{rj_r}$. Also an arbitrary P of degree $< r$ satisfies (1.32). Thus we want to show that the polynomial solutions of (1.32), each of which depends on a single $\mathbf{g}_k^\perp$ span all solutions. We verify this first when $n = 2$. Then there is only one α_j for each $\mathbf{g}_j$ so (1.32) is a single equation rather than a system. We claim that, for $n = 2$, the polynomial solutions of (1.32) are exactly the linear combinations of the $\alpha_j^{\perp p}$.

Let us examine this result algebraically. Let P be a homogeneous polynomial of degree m in x, y satisfying (1.32). We want to write

$$P = \sum b_j \alpha_j^{\perp m}$$

where b_j are constants and

$$\alpha_j^\perp = x\cos\theta_j + y\sin\theta_j$$

is the linear spread function for $\mathbf{g}_j$ whose directions numbers are $(\sin\theta_j, -\cos\theta_j)$. We have

$$\alpha_j^{\perp m} = \sum \binom{m}{l} \cos^l\theta_j \sin^{m-l}\theta_j x^l y^{m-l}.$$

We are left with the problem of showing that the $\alpha_j^{\perp m}$ span the kernel of (1.32). This is proven by analysis below; we leave the details of an algebraic proof to the reader.

However, if $r = m + 1$ then every polynomial of degree m satisfies (1.32). In this case it remains to show that

$$\det\left\{\binom{m}{l}\cos^l\theta_j\sin^{m-l}\theta_j\right\} \neq 0$$

which is equivalent (if no $\theta_j = 0$ which we can assume by rotating the axes) to

$$\det\{\tan^l\theta_j\} \neq 0.$$

The nonvanishing is Vandermonde if we choose $m+1$ distinct angles $\theta_j \in [0, \pi)$. We conclude that if $\mathbf{g}_1, \dots, \mathbf{g}_{m+1}$ are $m+1$ distinct spreads then every polynomial of degree m is a linear combination of monomials of degree $\leq m$ in the $\alpha_j^\perp$.

It is, however, more incisive to give an analytic proof. We apply the *fundamental principle* (see Section 1.4 for details). The algebraic variety related to (1.32) is defined by

$$V = \{\hat{x} \,|\, \hat{\alpha}_1(\hat{x}, \hat{y}) \dots \hat{\alpha}_r(\hat{x}, \hat{y}) = 0\}$$

where

$$\hat{\alpha}_j(\hat{x}) = \hat{x}\sin\hat{\theta}_j - \hat{y}\cos\hat{\theta}_j.$$

A holomorphic function $f(\hat{x})$ belongs to the ideal of V if and only if

$$\frac{\partial^p}{\partial\hat{\alpha}_j^{\perp p}}\delta \cdot f = \frac{\partial^p f}{\partial\hat{\alpha}_j^{\perp p}}(0) = 0 \tag{1.33}$$

for all j, p. For, (1.33) for fixed j implies that f vanishes on $\{\hat{\alpha}_j(\hat{x}) = 0\}$ and hence is divisible by $\hat{\alpha}_j$. This means that $\{\partial^p \delta_0 / \partial \hat{\alpha}_j^{\perp p}\}$ span the orthogonal complement of the space of holomorphic functions on V. By the Fourier transform (see Section 1.4) the monomials $\{\alpha_j^{\perp p}(x)\}$ span the solutions of (1.32). It then follows from homogeneity considerations that every homogeneous polynomial solution of (1.32) of degree p is a finite linear combination of monomials $\alpha_j^{\perp p}$.

We now pass to $n > 2$. The number of monomials of degree m is

$$N_{mn} = \binom{n+m-1}{m} = \sum_{j=0}^{m} \binom{n+j-2}{j}$$

since a polynomial P of degree m in n variables is a sum of $\{x_n^{m-j} Q_j\}$ where Q_j is a polynomial in $x_1, \ldots, x_{n-1}$ of degree j. Suppose we have already established the fact that Q_j can be written as a linear combination of powers of $N_{j,n-1}$ linear spread functions $\alpha_{jk}^{\perp}$. Then we can write

$$x_n^{m-j} Q_j = \sum a_{jk} \alpha_{jk}^{\perp j} x_n^{m-j}$$

where a_{jk} are constant. Observe that $\alpha_{jk}^{\perp j} x_n^{m-j}$ is a polynomial of degree m in the 2 plane spanned by $\alpha_{jk}^{\perp}$ and x_n. Thus by our result for $n = 2$ we can write

$$\alpha_{jk}^{\perp j} x_n^{m-j} = \sum \tilde{a}_{jkp} \tilde{\alpha}_{jkp}^{\perp m}$$

where $\tilde{\alpha}_{jkp}^{\perp}$ are defined by $m+1$ directions in the $\alpha_{jk}^{\perp}$, x_n plane and $\tilde{a}_{jkp}$ are constants.

Putting all this together we have expressed P as a sum of monomials in the directions of the $\tilde{\alpha}_{jkp}^{\perp}$.

From this it follows that every homogeneous polynomial of degree m is a linear combination of monomials of degree m in suitable directions. Thus certain sets $\mathbf{g}_1, \ldots, \mathbf{g}_{N_{mn}}$ have the property that $\{\mathbf{g}_j^{\perp m}\}$ span the space $\mathbf{P}^m$ of homogeneous polynomials of degree m. It follows that the map $\{\mathbf{g}_1, \ldots, \mathbf{g}_{N_{mn}}\} \mapsto \{\mathbf{g}_j^{\perp m}\}$ thought of as a map of $\mathbf{G}^{N_{mn}} \mapsto \mathbf{P}^m$ is generically of maximal rank.

We have proven

Lemma 1.1 *Let P be any polynomial of degree m and $\mathbf{g}_1, \ldots, \mathbf{g}_{N_{mn}}$ generic spreads. Then P can be written in the form*

$$P = P_{\mathbf{g}_1} + \cdots + P_{\mathbf{g}_{N_{mn}}} \tag{1.34}$$

where $P_{\mathbf{g}_j}$ is a spread polynomial of degree $\leq m$. For $n = 2$ "generic" means "distinct."

Remark 1 *For $n > 2$ we cannot replace "generic" by "distinct" because, for example, the directions $\mathbf{g}_j^{\perp}$ might all lie in a plane so any sum of spread functions would be constant in directions orthogonal to the plane.*

Remark 2 *The count of the directions $\alpha^{\perp}_{jk}$ as we pass from $\mathbb{R}^{n-1}$ to $\mathbb{R}^n$ is somewhat unclear. For example, to pass from $\mathbb{R}^2$ to $\mathbb{R}^3$ we start with Q_j and express it in terms of $j+1$ polynomials which are powers $(\alpha^{\perp}_{jk})^j$ depending on single directions $\alpha^{\perp}_{jk}$. Since, for $n=2$, any $j+1$ directions work, we can subsume all the $\alpha^{\perp}_{jk}$ in the $m+1$ directions $\alpha^{\perp}_{mk}$. For each of these directions we multiply by x_n^{m-j} so we need $m+1$ directions in the $\alpha^{\perp}_{mk}$, x_n plane. All the directions are distinct except for 1 which we can choose as the x_n axis. We are thus left with $m(m+1)+1$ distinct directions. Since $m(m+1)+1 > N_{m3} = \binom{m+2}{m}$ for $m>1$ there is a redundancy.*

Remark 3 *Lemma 1.1 is purely qualitative in nature. A quantitative version, i.e. one that provides bounds in the change of basis from monomials to spread polynomials, is proven using Fourier series and spherical harmonics in Chapters 3 and 4.*

Once we have Lemma 1.1 we can appeal to Weierstrass' theorem on polynomial approximation. The polynomials are dense in various function spaces for example, the space of all continuous functions. Thus if f is a function of compact support we cannot have

$$\int P(x)f(x)\,dx = 0$$

for all polynomials P. By Lemma 1.1 this means that the Radon transform of f cannot vanish identically.

More subtle forms of Weierstrass' theorem (see e.g. [72], Chapter XIII of FA[4]) show that a similar result holds if f is assumed to vanish at infinity "almost exponentially."

Lemma 1.1 also shows that we do not need all the spreads. For $n=2$ any infinite set will do. For arbitrary n we can choose the spreads in an arbitrary neighborhood of a fixed spread.

Lemma 1.1 shows how the space of polynomials decomposes into a sum of spaces of spread polynomials. A discussion of such "coarse grain" decompositions under the nomenclature "direct integral" or "tensor product" is given in Section 1.8.

It is important to observe that, although the spread polynomials corresponding to different spreads are distinct, there are linear relations among spread functions coming from different spreads. In fact, this is an immediate consequence of Lemma 1.1 since the same P can be represented in the form (1.34) for any generic set of directions.

This means that there are nontrivial finite relations in the monomial "basis." In contrast, the geometric basis $\{\delta_{L(s,\mathbf{g})}\}$ has only infinite (limit) relations. We shall see in Chapter 2 that the finite polynomial relations coming from Lemma 1.1 are equivalent to the infinite geometric relations.

[4]The book [33] will be referred to as FA.

In passing it is interesting to examine a proof of Weierstrass' theorem in the framework of hyperbolicity. For $n = 1$ it is based on Taylor's formula with remainder. An analysis of this formula shows that it is a restatement of the fact that d^N/dx^N is an hyperbolic operator, meaning that it has a good CP with CD prescribed at the origin. In fact, Taylor's formula is equivalent to an explicit formula for a suitable fundamental solution for d^N/dx^N. (See Section 1.8 for details.)

For $n > 1$ we can either use a similar method or else use the ring structure on polynomials to reduce Weierstrass' theorem to $n = 1$.

Now, for $n = 1$ the affine hyperplanes are the points so the Radon transform of f is f. Thus the basis $\{\delta_x\} = \{\delta_L\}$ works better than the monomial basis. But the ring structure makes it easy to pass from $n = 1$ to $n > 1$ using the basis of monomials. The reader might try to use Lemma 1.1 to bypass the ring structure.

We can think of a spread polynomial, especially if it is positive, as being in the realm of geometry. The real departure from geometry comes from the fact that negative numbers are forced upon us in Lemma 1.1. Thus

$$2x^2y^2 = x^4 + y^4 - (x+y)^2(x-y)^2$$

which shows that negative coefficients enter even when expressing a positive product of spread functions in terms of spread functions.

We have just given a proof of injectivity using duality: that is, the density of the linear combination of $\{\delta_{L(s,\mathbf{g})}\}$ in the dual space $\mathcal{W}'$. More generally, suppose that $\mathcal{W}'$ is provided with a decomposition $\{\mathcal{W}'(\mathbf{g})\}$ into coarse grains each of which may be equipped with a basis $\{v(s,\mathbf{g})\}_s$. Injectivity of the associated Radon transform

$$\mathbf{R}f(\mathbf{g}) = \text{restriction of } f \text{ to } \mathcal{W}'(\mathbf{g})$$

means the linear combinations of the elements of the various $\mathcal{W}'(\mathbf{g})$ are dense in $\mathcal{W}'$. (f is thought of as an element of $\mathcal{W}''$.)

Besides combinatorial methods such as the one given above, one of the most powerful tools for proving density, when there is some associated group structure, is the *Tauberian principle*, which is formulated in Section 1.6.

Beyond the density question is the explicit decomposition of $\mathcal{W}'$ as a direct integral (sub or supra) of the $\mathcal{W}'(\mathbf{g})$. This yields an expression for $f \in \mathcal{W}''$ in terms of $\{\mathbf{R}f(\mathbf{g}, s)\}$ which can be regarded as an *explicit inversion formula*.

In the "group case" such explicit formulas depend on concrete versions of the Tauberian principle.

As we have noted above in case of the usual Radon transform, there may be (infinite) linear relations amongst the $\{v(\mathbf{g}, s)\}_{\mathbf{g},s}$. These lead to corresponding relations on the Radon transforms $\mathbf{R}f(\mathbf{g}, s)$, i.e. on the range of $\mathbf{R}$.

Injectivity of Radon transforms, explicit inversions, and range characterizations are some of the classical problems we study in this work.

If we think of leaves L as manifolds, then the measure δ_L is only one of the geometric constructs we can associate to L. We can also form tangent and cotangent bundles and various other jet bundles.

Our view of these jet bundles goes as follows. Suppose we have a jet bundle $\vec{j}$ which is defined in some natural manner on all L. For example, if $\mathcal{L} = \{L\}$ is the family of lines we can take the unit tangent vector at each point. Thus to each $p \in L$ we associate the components $\vec{j}(p)$ of the jet at p. If $\vec{f}$ is a vector function on $\mathbb{R}^n$ (with the same number N of components as $\vec{j}$) then we can form the Radon transform

$$(\overrightarrow{\mathbf{R}}_{\vec{j}}\vec{f})(L) = \int_L \vec{j}(p) \cdot \vec{f}(p). \tag{1.35}$$

What is the analog of the injectivity problem for this jet Radon transform? If L are lines and $\vec{j}$ is the tangent bundle then the vanishing of $\vec{\mathbf{R}}_{\vec{j}}\vec{f}$ is equivalent (if $\vec{f}$ is small at infinity) to $\vec{f}$, thought of as a 1-form, being closed or cohomologous to zero. Put in other terms, the vanishing of $\vec{\mathbf{R}}_{\vec{j}}\vec{f}$ is equivalent to the possibility of satisfying the equation

$$dh = \vec{f}.$$

This is an interesting relation between the geometry expressed by (1.35) and the analysis as expressed by the possibility of solving $dh = \vec{f}$. Can we find other instances where the possibility of solving equations

$$\overrightarrow{P}(x, D)h = \vec{f}$$

is expressed in terms of the vanishing of the Radon transform (1.35) for a suitable $\vec{j}$ which depends on $\overrightarrow{P}(x, D)$? This problem is studied in Chapter 8.

1.2 Parametric Radon transform

In most of our considerations up to now the parameter λ on L has played essentially no role, because we considered the $L(s, \mathbf{g})$ merely as sets, equipped with the euclidian measure. To introduce the L coordinate in a manner which is consistent with the fact that all $L(s, \mathbf{g})$ are isomorphic, in fact to $\mathbb{R}^{n-j}$, means that we start with $\Lambda = \mathbb{R}^{n-j}$ and study linear maps of Λ into $\mathbb{R}^n$.

To understand the resulting structure, let us begin with $j = n - 1$, i.e. lines in $\mathbb{R}^n$. In the previous section we discussed hypersurfaces in $\mathbb{R}^n$ but the ideas carry over to linear affine varieties of any dimension. We map Λ into $\mathbb{R}^n$ by

$$\lambda \mapsto a\lambda + b = L(a, b)(\lambda) \tag{1.36}$$

where $a, b \in \mathbb{R}^n$. Of course when $a = 0$ the line degenerates to a point. The actual (geometric) line $L(a, b) = L(a, b)(\Lambda)$, which is the image of Λ, depends only on b and on a modulo (nonzero) scalar multiplication. Moreover the parameters are related by

$$L(a, b) = L(a, b + a\lambda^0) \tag{1.37}$$

for any $\lambda^0 \in \Lambda$. Thus the original $2n$ parameters a, b are reduced to $2n-2$ which is the correct number for lines in $\mathbb{R}^n$ ($n-1$ parameters for lines through the origin and, for each such line, $n-1$ parameters of translation).

In terms of the coordinates $(s, \mathbf{g})$, $\mathbf{g}$ can be identified with $a/|a|$ modulo $\pm$ if $a \neq 0$ and s is the projection of b on the hyperplane orthogonal to a.

For a nice function (smooth and small at infinity) we define the parametric Radon transform

$$F(a,b) = \mathbf{R}_{\mathbf{P}} f(a,b) = \int f(a\lambda + b) d\lambda = |a|^{-1} \mathbf{R} f(s, \mathbf{g}). \tag{1.38}$$

We now change our viewpoint and study F as a function of a, b rather than $s, \mathbf{g}$. In this vein we make the important observation

For any $\lambda^0 \in \mathbb{R}, c^0 \in \mathbb{R}^n$ the function $f(a\lambda^0 + b)$ is constant on the set $\{(a,b) | a\lambda^0 + b = c^0\}$.

This observation suggests that we should introduce spreads $\gamma(\lambda)$ in the parameter (a, b) space. (The concept of spreads will be formalized in Section 1.3.) The leaves $\gamma(\lambda; c)$ of $\gamma(\lambda)$ are defined by

$$\gamma(\lambda; c) = \{(a,b) | a\lambda + b = c\}. \tag{1.39}$$

These spreads serve a different role from spreads like $\mathbf{g}$ of the previous section. For there the leaves were sets on which we integrated, whereas here the leaves are defined using the individualization of the points on the integration sets.

We remarked that the spread $\{L(s, \mathbf{g})\}_s$ or rather $\{\delta_{L(s,\mathbf{g})}\}_s$ is related to a system of differential equations, namely the $\delta_{L(s,\mathbf{g})}$ are annihilated by the tangential derivatives to the leaves (see (1.31)). The same is true of the leaves $\gamma(\lambda; c)$. In fact, for fixed λ we have

$$\left(\frac{\partial}{\partial a_j} - \lambda \frac{\partial}{\partial b_j} \right) \delta_{\gamma(\lambda;c)} = 0. \tag{1.40}$$

By differentiating the j-th equation with respect to b_k and the k-th equation with respect to b_j we deduce

$$\left(\frac{\partial^2}{\partial a_j \partial b_k} - \frac{\partial^2}{\partial a_k \partial b_j} \right) \delta_{\gamma(\lambda;c)} = 0 \tag{1.41}$$

which is an equation independent of λ.

Since any function $f(a\lambda + b)$ is a limit of linear combinations of δ functions along the sets where $f(a\lambda + b)$ is constant (see the discussion in (1.17)ff.) it follows that

$$\left(\frac{\partial^2}{\partial a_j \partial b_k} - \frac{\partial^2}{\partial a_k \partial b_j} \right) F(a,b) = 0 \tag{1.42}$$

for any j, k (for any $n \geq 2$). Of course, this equation is obvious from the definition (1.38) but the previous derivation puts it in a light which is useful in our work.

We call (1.42) the *enveloping equation* of the equations (1.40) because, as we shall see in Chapter 6, the space of solutions of (1.42) (in suitable spaces) coincides with the closure of the linear combinations (or rather, integrals) of the solutions of (1.40) for varying λ. Thus the second-order equation (1.42) is equivalent to the set of first-order equations (1.40) in much the same way as the Laplace equation in $\mathbb{R}^2$ is equivalent to the Cauchy–Riemann and anti-Cauchy–Riemann equations. (In all these considerations λ must be regarded as a projective variable so there is a contribution coming from $\lambda = \infty$.)

Remark. *In conformity with the nomenclature of Section 1.1 we regard the decomposition of solutions of (1.42) into components which are solutions of (1.40) as a form of the Radon ansatz.*

We have seen that functions of the form $f(a\lambda + b)$ are solutions of (1.42). A particular class of solutions is of the form $h((a\lambda + b) \cdot u)$ for any $u \in \mathbb{R}^n$ and any function h of one variable. One choice of h is the exponential function. We shall see that all solutions are integrals of the form

$$\int e^{(a\lambda+b)\cdot u}\, d\nu(u) \tag{1.43}$$

for suitable measures ν.

Remark. *We shall develop equations like (1.42) in a more systematic way in Section 6.3.*

Equations (1.42) cannot completely characterize parametric Radon transforms $F(a, b)$ of the form (1.38) because (1.42) would still hold if the measure $d\lambda$ were replaced by any measure, as is clear from the definition. The measure $d\lambda$ is the Haar measure on the line of integration. Translation invariance can be expressed as

$$F(a, b + \lambda^0 a) = F(a, b) \tag{1.44}$$

for any λ^0. In infinitesimal form (1.44) becomes

$$(a \cdot \nabla_b)F = 0. \tag{1.45}$$

We can derive (1.45) from a more general point of view. The Haar invariance of $d\lambda$ can be stated in terms of distributional derivatives as

$$\frac{d}{d\lambda}(d\lambda) = 0 \tag{1.46}$$

meaning $\int h'(\lambda)d\lambda = 0$ for any test function (smooth and small at infinity) h. In particular

$$\begin{aligned} 0 &= \int f(a\lambda + b)\frac{d}{d\lambda}(d\lambda) \\ &= -\int \frac{\partial}{\partial\lambda} f(a\lambda + b)d\lambda \\ &= -(a \cdot \nabla_b)F \end{aligned}$$

which is (1.45).

We refer to equations (1.42) and (1.45) as the *John equations*. We shall usually be more precise and refer to (1.42) as the John equation or the *first John equation* and to (1.45) as the *invariance equation* or the *second John equation*. Actually, John derived equations using a different parametrization of the set of lines so, in fact, equation (1.45) does not appear in his equations. (It is "factored out" by his parametrization.)

Remark. *In keeping with the ideas of the previous section we could replace the measure $d\lambda$ by other measures which satisfy ordinary differential equations. The measure $e^{\lambda}d\lambda$ has been used with great success (see [4] and [5]). We shall return to this point presently.*

In Chapter 6 we prove that, for suitable F, the John equations imply that F is of the form $\mathbf{R_P}(f)$.

All the above deals with lines. The extension of these ideas to planes of any dimension is given in the next section. But what happens for curved varieties?

To orient our ideas, let us begin with parabolas in R^2. We write them in the form

$$\lambda \mapsto (a_1\lambda^2 + b_1, a_2\lambda + b_2) \tag{1.47}$$

meaning that the parametric Radon transform takes the form

$$F(a,b) = \mathbf{R_P}f(a,b) = \int f(a_1\lambda^2 + b_1, a_2\lambda + b_2)d\nu(\lambda) \tag{1.48}$$

for suitable measures ν.

We shall derive John equations for several different measures ν.

Proceeding along our previous lines we introduce the spreads $\gamma(\lambda)$ with leaves

$$\gamma(\lambda; c) = \{(a,b) | a_1\lambda^2 + b_1 = c_1, a_2\lambda + b_2 = c_2\}. \tag{1.49}$$

The leaves are linear and the associated system of differential equations is

$$\begin{aligned} \left(\frac{\partial}{\partial a_1} - \lambda^2\frac{\partial}{\partial b_1}\right)\delta_{\gamma(\lambda;c)} = 0 \\ \left(\frac{\partial}{\partial a_2} - \lambda\frac{\partial}{\partial b_2}\right)\delta_{\gamma(\lambda;c)} = 0. \end{aligned} \tag{1.50}$$

To eliminate λ from these equations, we differentiate the second equation with respect to a_2 which gives, on using the second equation to evaluate $(\partial/\partial a_2)\delta_{\gamma(\lambda;c)}$,

$$\left(\frac{\partial^2}{\partial a_2^2} - \lambda^2 \frac{\partial^2}{\partial b_2^2}\right) \delta_{\gamma(\lambda;c)} = 0.$$

We differentiate this equation with respect to b_1 and the first equation of (1.50) twice with respect to b_2. We arrive at the enveloping equation which is now third order:

$$\left(\frac{\partial^3}{\partial a_1 \partial b_2^2} - \frac{\partial^3}{\partial a_2^2 \partial b_1}\right) F = 0 \tag{1.51}$$

which is clear from (1.48). As before, solutions of (1.51) are limits of linear combinations (integrals) of solutions of the first-order systems (1.50).

For the invariance equation we try $d\nu(\lambda) = d\lambda$ and integrate by parts as in (1.46)ff. We are led to

$$\int \left(2\lambda a_1 \frac{\partial f}{\partial b_1} + a_2 \frac{\partial f}{\partial b_2}\right) (a_1\lambda^2 + b_1, a_2\lambda + b_2)\, d\lambda = 0. \tag{1.52}$$

The problem is that $2\lambda a_1 \partial/\partial b_1 + a_2 \partial/\partial b_2$ is not an operator in a, b on functions of the argument $(a_1\lambda^2 + b_1, a_2\lambda + b_2)$.

We could try the measure $d\nu(\lambda) = d\lambda/\lambda$ which satisfies

$$\left(\frac{d}{d\lambda}\lambda\right)\left(\frac{d\lambda}{\lambda}\right) = 0. \tag{1.53}$$

Integration by parts now yields the operator

$$2\lambda^2 a_1 \frac{\partial}{\partial b_1} + \lambda a_2 \frac{\partial}{\partial b_2} = 2a_1 \frac{\partial}{\partial a_1} + a_2 \frac{\partial}{\partial a_2} \tag{1.54}$$

on functions of the argument $(a_1\lambda^2 + b_1, a_2\lambda + b_2)$ (compare (1.50)).

The methods of Section 6.3 can be used to prove the analog of John's theorem for this case.

Going back to (1.52) we could differentiate with respect to b_2 (under the integral sign) obtaining

$$2\lambda a_1 \frac{\partial^2}{\partial b_1 b_2} + a_2 \frac{\partial^2}{\partial b_2^2} = 2a_1 \frac{\partial^2}{\partial a_2 \partial b_1} + a_2 \frac{\partial^2}{\partial b_2^2} \tag{1.55}$$

which gives the second John's equation

$$\left(2a_1 \frac{\partial^2}{\partial a_2 \partial b_1} + a_2 \frac{\partial^2}{\partial b_2^2}\right) F = 0.$$

We have not investigated the sufficiency of John's condition in this case.

What is the ultimate we can go to in this direction?

We can study the general parametric curve of degree m

$$\lambda \mapsto a\lambda = \left(\sum a_{1j}\lambda^j, \sum a_{2j}\lambda^j, \ldots, \sum a_{nj}\lambda^j\right). \tag{1.56}$$

All sums are for $j = 0, 1, \ldots, m$. This leads to the general spreads $\gamma(\lambda)$ where

$$\gamma(\lambda; c) = \left\{a | \sum a_{ij}\lambda^j = c_i\right\} \tag{1.57}$$

for all i. The differential equations satisfied by $\delta_{\gamma(\lambda;c)}$ are generated by

$$\left(\frac{\partial}{\partial a_{ij}} - \lambda^j \frac{\partial}{\partial a_{i0}}\right) \delta_{\gamma(\lambda;c)} = 0. \tag{1.58}$$

The enveloping equations are generated by

$$\left(\frac{\partial^r}{\partial a_{i_1 j_1} \ldots \partial a_{i_r j_r}} - \frac{\partial^r}{\partial a_{i_1 k_1} \ldots \partial a_{i_r k_r}}\right) F(a) = 0 \tag{1.59}$$

whenever $\sum j_l = \sum k_l$. Here F is the general parametric Radon transform of degree m

$$F(a) = \mathbf{R}_{\mathbf{P}} f(a) = \int f(a\lambda)\, d\nu(\lambda) \tag{1.60}$$

for some measure ν.

As in the previous cases, the solutions of the system (1.59) are integrals of solutions of (1.58).

To study the system (1.59) we search for Cauchy surfaces; that is, submanifolds S of $\{a\}$ on which the solutions are determined and such that (1.59) induces "few" equations on S. For example, if $m > 1$ we can choose for S the surface defined by $a_{ij} = 0$ when $j > 1$. On S we are left with the usual John equation (1.42). These lead formally to the construction of f and ν satisfying (1.60) when $a \in S$. Then we define $F(a)$ for all a by (1.60); it clearly satisfies (1.59) and restricts to the given data on S.

For $n = m = 2$ we can choose S as the surface defined by $a_{11} = a_{22} = 0$. In this case (1.56) reduces to the map (1.47). The equations on S are given by (1.51). (In the notation of (1.47), b_i corresponds to a_{i0}, while a_1, a_2 correspond to a_{12}, a_{21} respectively.)

What happens to the second John equation, i.e. the invariance equation? If $d\nu(\lambda) = d\lambda$ we can integrate by parts as usual to obtain

$$\int \left(\sum_{ij} j\lambda^{j-1} a_{ij} \frac{\partial}{\partial a_{i0}}\right) f(a\lambda)\, d\lambda = 0. \tag{1.61}$$

Note that we can replace $\lambda^{j-1}\partial/\partial a_{i0}$ by $\partial/\partial a_{i,j-1}$ in accordance with (1.58). Thus we can write (1.61) in the form

$$\left(\sum_{ij} j a_{ij} \frac{\partial}{\partial a_{i,j-1}}\right) F(a) = 0. \tag{1.62}$$

When $m = 1$ this reduces to the usual condition (1.45).

Equations (1.59) and (1.62) constitute the general John equations. Using the ideas of Section 6.3 we can show that they characterize functions F which are parametric Radon transforms of degree m corresponding to the measure $d\lambda$.

The parametrization that we have examined starts with the real line $\mathbb{R}^1$ and maps of $\mathbb{R}^1 \mapsto \mathbb{R}^n$. There is no essential difficulty in replacing $\mathbb{R}^1$ by $\mathbb{R}^l$. But one might study more subtle parametrizations in which $\mathbb{R}^1$ is replaced by a different manifold, e.g. S^l. For an even more daring approach, think of a "general" manifold as being made up of euclidian cubes pieced together and use this for a parametrization.

We leave these ideas for the interested reader.

1.3 Geometry of the nonparametric Radon transform

For a general geometric framework of the nonparametric Radon transform we start with a manifold M and a collection of submanifolds $\mathcal{L} = \{L\}$ of fixed dimension l. Each L is equipped with some fixed measure $d\mu$. In order to be able to introduce analysis into the theory in a manner similar to the case of hypersurfaces we want the $\{L\}$ to be organized into spreads. We formalize our above concept of spread more precisely:

Definition *A spread* $\mathbf{g}$ *is a subcollection* $\{L(s, \mathbf{g})\}$ *of* $\mathcal{L}$ *such that for each* $\mathbf{g}$

$$\bigcup_s L(s, \mathbf{g}) = M \tag{1.63}$$

and

$$L(s, \mathbf{g}) \cap L(s', \mathbf{g}) = \emptyset \quad \text{for } s \neq s'. \tag{1.64}$$

$L(s, \mathbf{g})$*, the leaves of the spread, are equipped with fixed measures* $\mu(s, \mathbf{g})$.

The significance of spreads is that for $l = n - 1$ the leaves can be the level sets of a function, while, in general, the leaves can be the simultaneous level sets of $n - l$ functions.

We require the spreads to satisfy the following conditions:

A *The spreads are isomorphic in the sense that for fixed* $\mathbf{g}$ *the set of leaves* $\{L(s, \mathbf{g})\}_s$ *can be parametrized by a single topological space* $S = \{s\}$ *which is independent of* $\mathbf{g}$*. For any* $\mathbf{g}$ *the leaves* $L(s, \mathbf{g})$ *depend continuously on* s*.*

$\mathbf{G} = \{\mathbf{g}\}$ *is called the Grassmannian and* s *is the spread parameter.*

B Fubini property. *There exist fixed measures ds on S and dx on M such that for any* $\mathbf{g}$

$$dx = ds\, d\mu(s, \mathbf{g}). \tag{1.65}$$

Here $d\mu(s, \mathbf{g})$ is the given measure on $L(s, \mathbf{g})$.

We define the Radon transform by

$$\mathbf{R}f(s, \mathbf{g}) = \int_{L(s,\mathbf{g})} f\, d\mu. \tag{1.66}$$

For (1.65) to be meaningful M has to be locally the cartesian product of $L(s, \mathbf{g})$ and S. This means that if p^0 is a point in $L(s^0, \mathbf{g})$ then for s near s^0 the $L(s, \mathbf{g})$ are all locally isomorphic, and M has a product structure of $L(s, \mathbf{g}) \times S$ near p^0. The dx measure of the local product set is the product of the ds and the $d\mu$ measures.

Putting things another way, if f is small at infinity

$$\int_M f(x)\, dx = \int_S \mathbf{R}f(s, \mathbf{g})\, ds \tag{1.65*}$$

for any spread $\mathbf{g}$. It is (1.65*) which is the form of the Fubini property that we need.

As discussed above, a *spread function* for $\mathbf{g}$ is a function u which is constant on the leaves of $\mathbf{g}$. The spread functions for $\mathbf{g}$ correspond to functions of s. To any function $v(s)$ we can associate the spread function which is the constant $v(s)$ on the leaf $L(s, \mathbf{g})$ of $\mathbf{g}$.

In the examples which we study the leaves $L(s, \mathbf{g})$ have "nice" cross-sections $S_{\mathbf{g}}$ which are isomorphic. We can then rewrite (1.65*) in the form

$$\int_M f(x)h(x)\, dx = \int_{S_{\mathbf{g}}} h(s)\mathbf{R}f(s, \mathbf{g})\, ds_{\mathbf{g}} \tag{1.65**}$$

for any spread function h. $ds_{\mathbf{g}}$ is the measure ds expressed on $S_{\mathbf{g}}$.

Now, it may be that the spread $\mathbf{g}$ is a *multiplicity spread*. To understand this, consider the affine hyperplanes $L(s)$ in $\mathbb{R}^n$ orthogonal to S, which is the x_n axis, and passing through $x_n = s$. Geometrically this is a single spread. To make up for the lack of other spreads we use as *multiplicities* the functions $\exp(i\hat{\lambda} \cdot \lambda)$ where λ is a parameter on $L = L(0)$ and $\hat{\lambda}$ is a dual parameter. In general we introduce an *attenuation* which is a function $\psi(s, \mathbf{g})$ on $L(s, \mathbf{g})$. Then the ordinary Radon transform (1.66) becomes the **attenuated Radon transform**

$$\mathbf{R}^{\psi} f(s, \mathbf{g}) = \int_{L(s,\mathbf{g})} \psi(s, \mathbf{g}; \lambda) f(\lambda)\, d\mu. \tag{1.67}$$

For a multiplicity spread there are several attenuations. They form the multiplicities and define a **multiplicity Radon transform**. Actually (1.67) can

be generalized to the case where ψ is an operator, i.e. more general than a multiplication operator.

For each function $u(s)$ and each ψ we have an extension

$$u \mapsto u(s)\psi(s, \mathbf{g}; \lambda) \tag{1.68}$$

to $\mathbb{R}^n$ (or the manifold M under consideration). We can associate ψ *spread functions* with ψ; these are functions h_ψ which are constant multiples of ψ on each $L(s, \mathbf{g})$. The analog of (1.65**) is

$$\int_M f(x)h_\psi(s)\,dx = \int_{S_\mathbf{g}} h_\psi(s)\mathbf{R}^\psi f(s, \mathbf{g})\,ds_\mathbf{g}. \tag{1.65***}$$

We envisage the situation in which the functions ψ are part of a spanning set $\{\phi(x, \hat{x})\}$ as prescribed by the projection–slice ansatz. For this reason we think of $\mathbf{R}^\psi(s, \mathbf{g})$ for fixed $\mathbf{g}$ as the ψ projection of f on functions on S (or $S_\mathbf{g}$). When $\psi = 1$ we shall say "projection."

The spread functions belong in a natural manner to the dual space $\mathcal{W}'$ of the space of functions f. When the spaces $\mathcal{W}_{\psi,\mathbf{g}}$ of spread functions corresponding to each $\mathbf{g} \in \mathbf{G}$ and each attenuation ψ are isomorphic then we can decompose $\mathcal{W}'$ into a projection tensor-like product as discussed above; this is a tensor product with amalgamation, meaning a quotient of the tensor product.

There is a dual concept for slice tensor product when we have multiplicity. A subset U_γ of M can be given multiplicities which are generally differential operators, say $\{\partial_{\xi(\gamma)}\}_{\xi\in\Xi_\gamma}$. We have an associated restriction map

$$f \to \{\partial_{\xi(\gamma)} f|_{U_\gamma}\}. \tag{1.69}$$

Restrictions of the form (1.69) are important for CPs for hyperbolic differential equations (as in Section 1.1) and for general PPs (as in Section 1.7). When $\{\partial_{\xi(\gamma)}\}$ are essentially independent of γ and the U_γ are isomorphic then we are in the situation of a slice tensor product which is a subset of the tensor product defined by compatibility.

Actually in some cases the operators $\{\partial_{\xi(\gamma)}\}$ can be other than differential operators. For the example of $\exp(i\hat{\lambda} \cdot \lambda)$ discussed above, under Fourier transformation the operators become translations.

With this generality the existence of a projection–slice function is essentially the same as finding a basis in which the projection tensor product becomes the dual of a slice tensor product (see Section 1.2).

One of the examples that is important for us is $M = V$ where V is a complex algebraic hypersurface. Suppose V is defined in $\mathbb{C}^n$ by an equation

$$P(\lambda, s) \equiv \lambda^d + P_1(s)\lambda^{d-1} + \cdots + P_d(s) = 0. \tag{1.70}$$

We have written $x = (\lambda, s)$ where λ is a single complex variable and $s \in \mathbb{C}^{n-1}$. We assume that P is irreducible; otherwise we have to use the theory of multiplicity varieties (see Section 1.4).

We define a single multiplicity spread whose leaves are

$$L(s_0) = \{s = s_0\} \cap V. \tag{1.71}$$

These leaves consist of d points (counting multiplicity).

Since we have a single multiplicity spread, the multiplicity of this spread must be d. By the Lagrange interpolation formula, a basis for functions on the leaves consists of the monomials $1, \lambda, \dots, \lambda^{d-1}$. This means that we can choose these monomials to define the multiplicity. Thus for a function f on V we define

$$^{j}\mathbf{R}f(s) = \sum (\lambda_k(s))^j f(\lambda_k(s), s) \tag{1.72}$$

where λ_k are the roots of $P(\lambda, s) = 0$.

There is some difficulty in formula (1.72) when roots $\lambda_k(s)$ coalesce. This difficulty can be overcome by treating such $L(s)$ as multiplicity varieties. Thus V represents the locus $\{P = 0\}$. But at multiple points (λ, s) some λ derivatives of P vanish. If we include derivatives of f corresponding to these multiplicities then we obtain d values of f for each s.

All this is treated in detail in FA and, in various forms, in several places in this work. No essential difficulties are introduced so, at this point, we shall ignore this phenomenon.

The matrix $\{(\lambda_k(s))^j\}$ is the Vandermonde matrix. This means that we can recover $f(\lambda, s)$ (at generic s and, using the above remarks about multiplicity, at all s) in terms of the $^{j}\mathbf{R}f$. In fact we can find explicit formulas

$$f(\lambda_k(s), s) = \sum v_{kj}(s)\ {}^{j}\mathbf{R}f(s). \tag{1.73}$$

It is shown in Chapter IX of FA (and will be clarified in Section 1.7) that the Fourier transform represents this projection tensor product as the slice tensor product which corresponds to the CP on the hyperplane $\lambda = 0$ for the differential operator $P(\partial/\partial\lambda, \partial/\partial s)$.

The geometric projection given by the spreads $s = \text{const.}$ represents only one type of projection that is meaningful for algebraic varieties. We could project V onto any plane Q from any point $p \notin V \cup Q$. One particular case that appears in our work (Section 7.3) is for $V =$ the real half-hyperbola $x_1^2 - x_2^2 - x_3^2 = 1$, $x_1 > 0$, Q the tangent plane at $x = (1, 0, 0)$, and $p = 0$.

Let us give some examples of geometric families which are organized into spreads.

(1) $\mathcal{L} = \{l$ planes in $\mathbb{R}^n\}$. A spread consists of the family of planes parallel to a given l plane L_0 through the origin. $\mathbf{G}$ is the Grassmann variety $\mathbf{G}(n, l)$ of l planes $L_0(\mathbf{g})$ through the origin in $\mathbb{R}^n$. $S_{\mathbf{g}} = L_0(\mathbf{g})^{\perp}$ is a cross-section for the spread $\mathbf{g}$. The spread parameter tells us where L meets $L_0^{\perp}$ which is the $S_{\mathbf{g}}$ coordinate of the base point on L. The measure is the euclidean measure. The continuity and Fubini properties are clear. We deal with the difference between

oriented and unoriented planes as in the case $l = 1$ (see (1.4) and the discussion there).

The leaves $L(s, \mathbf{g}^0)$ of a fixed spread can be realized as the simultaneous level set of $n - l$ linear functions on $\mathbb{R}^n$.

Note that the number of parameters s is $n - l$ while the dimension of $\mathbf{G}$ is

$$\dim \mathbf{G} = l(n - l) \tag{1.74}$$

since $\mathbf{G}$ can be realized as $\mathcal{O}(n)/\mathcal{O}(l) \times \mathcal{O}(n - l)$ where $\mathcal{O}(k)$ is the orthogonal group in k variables. Thus when $l \neq 0, n-1$ the dimension of $\mathbf{G}$ plus $n-l$ is greater than that of $\mathbb{R}^n$. This means that there are many compatibility conditions for the corresponding Radon transform $\mathbf{R}^l f(s, \mathbf{g})$ (admissibility problem). Moreover, we could search for subsets of $\{(s, \mathbf{g})\}$ of dimension n for which we have "almost" an isomorphism. These questions are discussed in Chapter 2.

There is a parametric Radon transform $\mathbf{R}^l_{\mathbf{P}}$ associated to $\mathbf{R}^l$. As explained above, the parametric description of planes L involves maps of a fixed plane into L. Thus the points of L are "individualized" by the map.

We map $\Lambda = \mathbb{R}^l$ into $\mathbb{R}^n$ by

$$\lambda \mapsto a\lambda + b. \tag{1.75}$$

Now a is an n by l matrix and $b \in \mathbb{R}^n$. (Points in $\mathbb{R}^l$ and $\mathbb{R}^n$ are considered as column vectors.) For a "nice" function f on $\mathbb{R}^n$ we can form

$$F(a, b) = \mathbf{R}^l_{\mathbf{P}} f(a, b) = \int f(a\lambda + b)\, d\lambda. \tag{1.76}$$

This integral is defined only when rank $a = l$. However, it defines a distribution of a, b on $\mathbb{R}^{n(l+1)}$ when $l \leq n - 1$ (Section 6.1).

As in Section 1.2 we define spreads $\gamma(\lambda^0)$ for each λ^0 whose leaves are given by

$$\gamma(\lambda^0, c^0) = \{(a, b) | a\lambda^0 + b = c^0\}. \tag{1.77}$$

As in (1.40) the δ functions of these leaves satisfy

$$\left(\frac{\partial}{\partial a^i_j} - \lambda^0_i \frac{\partial}{\partial b_j} \right) \delta_{\gamma(\lambda^0; c^0)} = 0. \tag{1.78}$$

(In our notation i is the column index and j is the row index.) Since this holds for all i, j we can eliminate $\{\lambda_j\}$ to obtain the John equations

$$\left(\frac{\partial^2}{\partial a^i_j \partial b_m} - \frac{\partial^2}{\partial a^i_m \partial b_j} \right) F(a, b) = 0 \tag{1.79}$$

$$\left(\frac{\partial^2}{\partial a^i_j \partial a^k_m} - \frac{\partial^2}{\partial a^i_m \partial a^k_j} \right) F(a, b) = 0. \tag{1.80}$$

These equations can also be seen directly from (1.76). They are valid when $l \leq n-2$ (Section 6.1).

As in the case $l = 1$ the John equations are equivalent to the expression of F as an integral (over λ) of spread functions, i.e. solutions of the first-order systems (1.78).

We must add an equation which defines the measure $d\lambda$. This is easily seen to be the *invariance equation*

$$F(a, b + \lambda^0 a^i) = F(a, b) \tag{1.81}$$

for any i and any scalar λ^0. In infinitesimal form (1.81) becomes

$$(a^i \cdot \nabla_b) F(a, b) = 0 \tag{1.82}$$

for all i. Under suitable regularity and growth conditions (1.79), (1.80), and (1.82) imply that F is of the form $\mathbf{R}_{\mathbf{P}} f$.

Equations (1.79) and (1.80) are valid even if we attenuate the Radon transform by replacing the measure $d\lambda$ by another (suitable) measure $d\nu(\lambda)$. Of course the invariance equation is no longer valid. In fact we show in Section 6.2 that we can determine both f and ν from the corresponding attenuated Radon transform.

(2) Let $\mathcal{L}$ be the family of level sets of all homogeneous polynomials of degree m. The measure $\mu(P_s)$ on $\{P(x) = s\}$ is the euclidean measure divided by $|\text{grad } P|$.

To organize $\mathcal{L}$ into spreads we denote by $p = \{x^m\}$ some listing of the

$$N = \binom{n+m-1}{m} \tag{1.83}$$

monomials of degree m. Then any homogeneous polynomial of degree m can be written in the form

$$p(x) = \hat{p} \cdot p \tag{1.84}$$

where $\hat{p}$ is a nonzero vector of constants with N complex numbers as components. The space of spreads can be identified with the projective space $\mathbf{P}^{N-1}$ since the spreads are

$$\{\hat{p} \cdot p = s\}. \tag{1.85}$$

Thus the Grassmann parameter is $\mathbf{g} = \hat{p}/|\hat{p}|$ modulo $\pm$ and the spread parameter is $s \in \mathbb{R}^1$. The measure μ was chosen so as to have the Fubini property.

We define the *nonlinear Radon transform*

$$\mathbf{R}_m f(\hat{p}, s) = \int_{\hat{p} \cdot p = s} f \, d\mu. \tag{1.86}$$

For $m = 1$ we are back at the hyperplane Radon transform. But there are major differences between $m = 1$ and $m > 1$. For one thing $N > n$ when $m > 1$. Thus,

as in the case of the l plane Radon transform with $l < n-1$, the transform $\mathbf{R}_m$ is overdetermined, meaning that $\mathbf{R}_m f$ satisfies many relations. (These are discussed fully in Chapter 5.)

$\mathbf{R}_m$ can be thought of in a different manner. We consider the *Albanese map*

$$A : x \mapsto (p(x), x) = A(x) \tag{1.87}$$

which is a map of $\mathbb{R}^n$ into an algebraic variety $\mathcal{V}_m \subset \mathbb{R}^N \times \mathbb{R}^n$. The spread (1.85) is defined as the set of projections on $\mathbb{R}^n$ of intersections of $\mathcal{V}_m$ with the planes $\hat{p} \cdot p = s$ in $\mathbb{R}^N$; that is, it is the projection on $\mathbb{R}^n$ of the set of intersections of $\mathcal{V}_m$ with the product of $\mathbb{R}^n$ with the leaves of a hyperplane spread in the p coordinate.

Thought of in this way there is a natural generalization. We replace $\mathcal{V}_m$ by an arbitrary real algebraic variety $\mathcal{V} \subset \mathbb{R}^N \times \mathbb{R}^n$. When $\mathcal{V}$ is defined by a map of $\mathbb{R}^n \to \mathbb{R}^N$ then the projection on the second factor of the intersection of $\mathcal{V}$ with the product of $\mathbb{R}^n$ with leaves of some spread in $\mathbb{R}^N$ defines a spread in $\mathbb{R}^n$.

Instead of requiring that $\mathcal{V}$ be a one-sheeted covering of $\mathbb{R}^n$ we can allow it to be a covering with a finite, constant sheet number k'. Then we obtain a slight generalization of spreads for which the leaves of a spread cover each point $x \in \mathbb{R}^n$ exactly k' times (counting multiplicity).

More precisely we want $\mathcal{V}$, or, rather, the complexification $\mathcal{V}^{\mathbb{C}}$ of $\mathcal{V}$, to define an **harmonic correspondence**. Roughly speaking this means that there are algebraic subvarieties $V_1^{\mathbb{C}} \subset \mathbb{C}^N$ and $V_2^{\mathbb{C}} \subset \mathbb{C}^n$ such that $\mathcal{V}^{\mathbb{C}}$ is a q_1 sheeted covering of $V_1^{\mathbb{C}}$ and a q_2 sheeted covering of $V_2^{\mathbb{C}}$. Moreover there are Lagrange-like interpolation formulas (as in (1.73)) which identify suitable holomorphic functions on $\mathcal{V}^{\mathbb{C}}$ (e.g. polynomials) with q_1 tuples of holomorphic functions on $V_1^{\mathbb{C}}$ and with q_2 tuples of holomorphic functions on $V_2^{\mathbb{C}}$.

Such a correspondence gives a way of transporting information between V_1 and V_2.

In the case of $\mathcal{V} = \mathcal{V}_m$, $V_2 = \mathbb{R}^n(\mathbb{C}^n)$, the variety V_1 is the Plücker variety in $\mathbb{R}^n(\mathbb{C}^n)$, meaning the variety defined by the relations amongst the homogeneous monomials of degree m. For example,

$$(x_1 x_2)^2 = x_1^2 x_2^2 \tag{1.88}$$

is a Plücker relation for $m = 2$. This is discussed in detail in Chapter 5.

We shall also deal with *exotic correspondences* in which $\dim V_1^{\mathbb{C}} < \dim \mathcal{V}^{\mathbb{C}}$. This is compensated for by having $q_1 = \infty$.

The variety $\mathcal{V}$ gives a method of *intertwining*, i.e. transforming, spreads on V_1 into spreads on V_2. For example, for the variety $\mathcal{V}_m$, we can start with the spreads defined by hyperplanes in $\mathbb{R}^N$. We then intersect with the Plücker variety V_1 to obtain spreads in V_1. We then project after multiplying with $\mathbb{R}^n$. This is the intertwining to $V_2 = \mathbb{R}^n$.

A hyperplane $H_{\hat{p}^0}$ in $\mathbb{R}^N$ is defined by a vector $\hat{p}^0 \in \mathbb{R}^N$; $H_{\hat{p}^0}$ consists of all $\hat{p} \in \mathbb{R}^N$ with $\hat{p}^0 \cdot \hat{p} = 0$. The spread is defined by the affine hyperplanes

$$\hat{p}^0 \cdot \hat{p} = s. \tag{1.89}$$

The intersection of (1.89) (times $\mathbb{R}^n$) with $\mathcal{V}_m$ consists of pairs $\hat{p}, x$ such that $\hat{p} = p(x)$ and

$$\hat{p}^0 \cdot p(x) = s. \tag{1.90}$$

Hence we arrive at the same spreads as in (1.86).

From either point of view we can form the spread functions

$$h(\hat{p}^0 \cdot p(x)) \tag{1.91}$$

for any function h. As usual it is natural to introduce a basis in $\{h\}$. We could use monomials, but it seems more convenient to use exponentials. As we have emphasized it is this passage from the geometry (meaning the basis $\{\delta_s\}$) to the "linear combination" basis $\{\exp(i\hat{s}s)\}$ which allows us to use the power of analysis.

The introduction of the basis $\{\exp(i\hat{s}s)\}$ leads to the **nonlinear Fourier transform**

$$F(\hat{p}, \hat{x}) = \int e^{i\hat{p}\cdot p(x) + i\hat{x}\cdot x} f(x)\, dx. \tag{1.92}$$

(The extra term $i\hat{x} \cdot x$ is added for convenience.) Since $(\hat{p}, \hat{x})$ consists of $N + n$ variables there are many relations that such F must satisfy; these relations are differential equations. The nonlinear Fourier transform also has relations to several other branches of mathematics. It is developed in detail in Chapter 5.

Remark. *It is often difficult to determine properties of the Radon transform from the geometry. Given a set of smooth curves $\{\gamma\}$ in $\mathbb{R}^2$ one might expect that a reasonable condition for the injectivity of the Radon transform defined by integration over $\{\gamma\}$ with the euclidean measure is that for each x and each direction α there is a γ which passes through x and is tangent to α. However, this is not sufficient for injectivity as is seen from the example of $\{\gamma\}$ being all circles in $\mathbb{R}^2$ of fixed radius. An f whose integral over all such γ vanishes satisfies $f * \delta_{\gamma^0} = 0$ for any one fixed γ^0 since all γ are translates of γ^0. But $\hat{\delta}_{\gamma^0}$ is a Bessel function so if $\hat{f}$ is a sum of δ functions supported by the zeros of this Bessel function then $f * \delta_{\gamma^0} = 0$.*

Nevertheless injectivity holds locally, meaning for f of compact support. Similar local results apply when $\{\gamma\}$ are the geodesics of a Riemannian metric [124].

1.4 Parametrization problems

By a parametrization problem (PP) we mean the search for a way of parametrizing solutions of a system of partial differential equations $\vec{P}(x, D)f = 0$. There are classical parametrization problems such as the Dirichlet problem, where solutions

on a domain Ω are parametrized by their boundary values and some normal derivatives, and the Cauchy problem where they are parametrized by the values of suitable derivatives on a manifold which lies inside Ω.

In both the Cauchy and Dirichlet Problems the solutions are parametrized by a finite number of data on a manifold S of dimension r; this data is called parametrization data (PD) and S is called a parametrization surface (PS). If the $P_j(D)$ have constant coefficients then the variety V of common zeros of the $P_j(\hat{x})$ has complex dimension equal to r. According to the *fundamental principle* of FA (discussed below) this allows a parametrization of solutions by exponentials whose frequencies belong to V. Hence, again, the parametrization is in terms of functions or measures depending on r variables.

We can think of PS as a function-theoretic analog of *fundamental domain*. A fundamental domain $\mathcal{D}$ is defined by choosing a cross-section for the action of a discrete group Γ on a manifold M. In function-theoretic terms, $\mathcal{D}$ is a "minimal" subset of M on which all Γ invariant functions are determined.

Actually more is true: Γ invariance still has some relevance on $\mathcal{D}$ in the sense that functions on $\mathcal{D}$ are the restrictions of Γ invariant functions on M if and only if they satisfy Γ invariance on the boundary of $\mathcal{D}$. For example, if $M = \mathbb{R}^1$ and Γ is the group of translations by integers then $\mathcal{D} = [0, 1]$. Γ invariant functions f on $\mathbb{R}^1$ satisfy $f(0) = f(1)$. f also satisfies differential identities at $\{0, 1\}$ if $f \in C^\infty$.

Γ invariance is an equation. For other equations such as differential equations $\vec{P}(x, D)f = 0$ there are often proper subsets (PS) of M on which solutions are determined by their values and the values of some of their derivatives (PD). It may be the case that the equation $\vec{P}f = 0$ still implies some conditions on the PD; this is the analog of the compatibility condition on $\mathcal{D}$.

Let $\mathcal{W}$ be a suitable space of functions on M. We denote by $\mathcal{W}'(\vec{P})$ the kernel of $\vec{P}$ on $\mathcal{W}'$. If $f \in \mathcal{W}$ then we have defined (see (1.10)) the Radon transform of f related to $\vec{P}$ as the restriction of f, thought of as an element of $\mathcal{W}''$, to $\mathcal{W}'(\vec{P})$. If the range of $\vec{P}'$ is closed then $\mathbf{R}f \in \mathcal{W}''/\vec{P}'(\mathcal{W}'')^r$ where r is the number of components of $\vec{P}$.

If S is a PS for $\vec{P}$ then $\mathcal{W}'(\vec{P})$ can be identified with a space of functions or (generalized) distributions on S. Hence the same is true for its dual $\mathcal{W}''/\vec{P}'(\mathcal{W}'')^r$.

Now let us reverse the logic. $\mathcal{W}$ is a space of functions on M, meaning it is described in terms of suitable integrals of the basis $\{\delta_m\}_{m\in M}$ (see (1.17)ff.). A PS picks out a subbasis (perhaps adding derivatives) which describes the quotient $\mathcal{W}/\vec{P}'\mathcal{W}^r$.

If $\vec{P}$ is replaced by a finitely generated discrete group Γ acting on functions then a PS for Γ is a way of describing

$$\mathcal{W}/\sum_{\gamma_j\in\Gamma}(\gamma_j - \text{ id})\mathcal{W}$$

in terms of the basis $\{\delta_m\}$. Here $\{\gamma_j\}$ is a finite set of generators of Γ. In this case PS is the same as fundamental domain.

Analogous concepts appear in other contexts. For example, if we use the basis $\{\exp ix \cdot \hat{x}\}$ for $x \in M = \mathbb{R}^n$ then the fundamental principle (which will be discussed presently) asserts that for suitable spaces $\mathcal{W}$ the quotient space $\mathcal{W}/\vec{P}'\mathcal{W}^r$ can be identified with a space of functions on the algebraic variety $V \subset \mathbb{C}^n$ which is the set of common solutions of $P_j(i\hat{x}) = 0$. (There may also be some differential conditions.)

A similar situation appears in our discussion of the projection–slice ansatz (Section 1.1).

Remark. *The PS, or fundamental domain, or fundamental principle represent, as much as possible, the realization of the quotient $\mathcal{W}/\vec{P}'\mathcal{W}^r$ in a suitable basis as a construct on a subbasis. Thus quotient spaces are replaced by "close to subspaces." This process is an abstract form of balayage (discussed below).*

We shall also study PPs for which the data is given on manifolds of dimension less than r. This is compensated for by using infinitely many data. Sometimes we use manifolds of dimension greater than r; in this case the data satisfies differential identities. When the dimension of the parametrization manifold is $\neq r$ we refer to the PP as "exotic."

The first example of this "exotic" PP occurs for the operator $\bar{\partial}$ on $R^{2n} = C^n$. We can parametrize solutions by their data at a point if the data is

$$\{(\partial^{|\alpha|} f/\partial z_1^{\alpha_1} \dots \partial z_n^{\alpha_n})(0)\} \quad \text{with} \quad |\alpha| = \alpha_1 + \cdots + \alpha_n \tag{1.93}$$

for all α. This example illustrates that when we use infinitely many data we may need growth conditions in the "data parameter" α.

For a more interesting situation consider the wave operator on R^{n+1} $(n > 1)$

$$\Box \equiv \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x_1^2} - \cdots - \frac{\partial^2}{\partial x_n^2}. \tag{1.94}$$

We can use the t axis as a parametrization set. This is one dimensional as contrasted with the usual n-dimensional parametrization sets.

We refer to this problem as the *Watergate problem* (WP) and the data as *Watergate data* (WD). The reason for this terminology is that the author first encountered this problem during the Watergate hearings. What people did in Watergate was sit at their phones and attempt to amass enough information to know everything. Thus the data is at $x = 0$ for all t.

In general, we shall use the terminology "Watergate problem" whenever the data is given on manifolds of dimension less than r.

What is the WD corresponding to the t axis? The answer is

$$\left\{h_j\left(\frac{\partial}{\partial x}\right) f\Big|_T\right\} \tag{1.95}$$

for $j = 0, 1, \ldots$. Here $\{h_j\}$ is a basis for all homogeneous harmonic polynomials on $\mathbb{R}^n$. For example, if $n = 2$ then (1.95) becomes

$$\left\{\left(\frac{\partial}{\partial x_1} \pm i\frac{\partial}{\partial x_2}\right)^j f\Big|_T\right\} \tag{1.96}$$

for $j = 0, 1, \ldots$.

In order to put these concepts in a suitable framework, let us recall some ideas from FA. In dealing with boundary value problems we sometimes have to treat classes of functions which are large at infinity (or even only locally defined).

Let $\mathcal{W}$ be a reflexive topological vector space of functions or distributions (of x). We say that $\mathcal{W}$ is *analytically uniform* (AU) if

(i) $\exp(i\hat{x}\cdot x) \in \mathcal{W}'$ for all $\hat{x} \in C^n$. The linear combinations of $\exp(i\hat{x}\cdot x)$ are dense in $\mathcal{W}'$.
(ii) The map $\hat{x} \to \exp(i\hat{x}\cdot x)$ is holomorphic from $C^n \to \mathcal{W}'$.

By (i) and (ii) we can define the Fourier transform of any $w \in \mathcal{W}$ by

$$\hat{w}(\hat{x}) = w \cdot \exp(i\hat{x}\cdot x). \tag{1.97}$$

$\hat{w}$ is clearly an entire function.

(iii) There exists a family $K = \{k(\hat{x})\}$ of continuous positive functions on C^n (they may take the value ∞) so that $\hat{\mathcal{W}}$ consists of all entire functions $F(\hat{x})$ for which

$$\|F\|_k = \sup\frac{|F(\hat{x})|}{k(\hat{x})} < \infty \tag{1.98}$$

for all $k \in K$.

K is called an *analytically uniform structure* for $\mathcal{W}$.

(iv) The semi-norms in (1.98) define the same topology on $\hat{\mathcal{W}}$ as that obtained by Fourier transformation from $\mathcal{W}$.

Examples of AU spaces are:

(a) $\mathcal{E}'(\Omega)$ = dual of C^∞ functions on the convex domain Ω.
(b) $\mathcal{D}(\Omega) = C^\infty$ functions of compact support $\subset \Omega$.
(c) (When x is a complex variable) $\mathcal{H}'(\Omega)$ = dual of holomorphic functions on the convex domain Ω.

There are two ways of combining AU spaces to obtain new AU spaces (see p. 402 of FA):

(α) AU *union*. Given AU spaces $\mathcal{W}_1$ and $\mathcal{W}_2$ we define the AU union $\mathcal{W}_1 \overset{a}{\cup} \mathcal{W}_2$ to have the AU structure

$$k_\cup(x) = \min[k_1(\hat{x}), k_2(\hat{x})]$$

for each pair $k_1 \in K_1, k_2 \in K_2$ where K_j is an AU structure for $\mathcal{W}_j$.

(β) AU *intersection*. $\mathcal{W}_1 \overset{a}{\cap} \mathcal{W}_2$ is defined by the AU structure

$$k_\cap(x) = \max[k_1(\hat{x}), k_2(\hat{x})].$$

For example, if $\mathcal{W}_1' = \mathcal{E}(\Omega_1)$ and $\mathcal{W}_2' = \mathcal{E}(\Omega_2)$ are the spaces of C^∞ functions on convex sets Ω_1, Ω_2 respectively then

$$\left(\mathcal{W}_1 \overset{a}{\cup} \mathcal{W}_2\right)' = \mathcal{E}(\Omega_1 \cap \Omega_2)$$

and

$$\left(\mathcal{W}_1 \overset{a}{\cap} \mathcal{W}_2\right)' = \mathcal{E}(\text{convex hull of } \Omega_1 \cup \Omega_2).$$

(We should observe that our notation is somewhat bizarre because $\mathcal{W}_1 \overset{a}{\cup} \mathcal{W}_2 = \mathcal{W}_1 \cap \mathcal{W}_2$. The reason for this is that we originally (in FA) emphasized $\mathcal{W}'$ rather than $\mathcal{W}$.)

We set $\mathcal{W} = \mathcal{W}_1 \overset{a}{\cap} \mathcal{W}_2$. Observe that $\mathcal{W}_j \subset \mathcal{W}$. Thus for any $f \in \mathcal{W}'$ the *restriction* of f to $\mathcal{W}_j'$ defined by $S_j \to f \cdot S_j$ is continuous on $\mathcal{W}_j$ and so defines an element $\rho_j f \in \mathcal{W}_j'$.

For a pair $(f_1, f_2) \in \mathcal{W}_1' \oplus \mathcal{W}_2'$ to be of the form

$$\rho f = (\rho_1 f, \rho_2 f)$$

it is certainly necessary that (f_1, f_2) satisfies the compatibility condition

$$Sf_1 = Sf_2$$

for any $S \in \mathcal{W}_1 \overset{a}{\cup} \mathcal{W}_2$. We shall meet many conditions, e.g. the edge-of-the-wedge theorem (see Chapters 5 and 9) in which the converse is essentially true. In those examples $\mathcal{W}_j'$ are spaces of solutions of systems of partial differential equations on suitable convex sets Ω_j. The above compatibility condition says the solutions are equal on the intersection of the convex sets. Edge-of-the-wedge theorems say that compatibility implies that there is a "solution" on the convex hull of $\Omega_1 \cup \Omega_2$ which restricts to each F_j on Ω_j.

The compatibility defined by $\mathcal{W}_1 \overset{a}{\cup} \mathcal{W}_2$ works well when we deal with two spaces. But when there are more than two spaces a more elaborate compatibility is required.

We can define $\mathcal{W} = \overset{a}{\cap} \mathcal{W}_j$ as before. We again ask the question as to when $\{f_j\} \in \sum_\oplus \mathcal{W}_j'$ defines an element $f \in \mathcal{W}'$. By this we mean that

$$\rho_j f = f_j;$$

that is, f restricts to f_j on each $\mathcal{W}_j$.

Suppose f restricts to f_j on each $\mathcal{W}_j$. If $S_j \in \overset{a}{\cup} \mathcal{W}_j$ and

$$\sum S_j = 0$$

as an element of $\mathcal{W}$ then

$$0 = \left(\sum S_j\right) \cdot f = \sum S_j \cdot f_j.$$

Thus $\{f_j\}$ must vanish on such $\{S_j\}$.

We define $\overset{sa}{\cup}\mathcal{W}_j$ as this set of $\{S_j\}$. This is the natural compatibility space for $\{\mathcal{W}_j'\}$; we should point out that $\overset{sa}{\cup}\mathcal{W}_j$ is not an AU space. We refer to the associated identities as *strong AU compatibility*. One noteworthy example where strong compatibility plays a role is the Korevaar–Wiegerinck theorem discussed below.

What is crucial for us is that if an AU space $\mathcal{W}$ satisfies another property called *localizable* (in which case $\mathcal{W}$ is called LAU) then the fundamental principle holds. This is the extension to partial differential equations of Euler's theorem expressing the solutions of linear partial differential equations with constant coefficients in terms of exponential polynomial solutions.

Let

$$\vec{P}(D) = (P_1(D), \ldots, P_r(D)) \tag{1.99}$$

where the $P_i(D)$ are linear partial differential operators with constant coefficients. We say $\vec{P}(D)f = 0$ if $P_j f = 0$ for all j. We denote by V the complex algebraic variety of common zeros of the $P_j(i\hat{x})$. (There is a problem of multiplicities which is complicated; we shall ignore it in this book by assuming that the $P_j(i\hat{x})$ generate the ideal of polynomials which vanish on V.) We define $\hat{\mathcal{W}}(V)$ as the space of entire functions on V which satisfy the growth conditions (1.98) on V.

Theorem 1.2 (Fundamental principle (FP)) *The dual of the kernel $\mathcal{W}'(\vec{P}(D))$ of $\vec{P}(D)$ in $\mathcal{W}'$ is $\mathcal{W}/\vec{P}'(D)(\mathcal{W})^r$. The restriction map*

$$F \quad \to \quad F|_V \tag{1.100}$$

is a topological isomorphism of $\hat{\mathcal{W}}/\vec{P}'(\hat{x})(\hat{\mathcal{W}})^r$ onto $\hat{\mathcal{W}}(V)$. Any $f \in \mathcal{W}'(\vec{P}(D))$ has the Fourier representation

$$f(x) = \int_V e^{i\hat{x}\cdot x} \frac{d\mu(\hat{x})}{k(\hat{x})} \tag{1.101}$$

for some measure μ on V whose total variation is bounded and some $k \in K$.

Theorem 1.2 and its ramifications are expounded in FA.

In (1.101) we use all of V. However, this is not necessary. The reason is that there are inequalities which show that bounds on parts of V already imply similar bounds on all of V. The simplest such inequality is the maximum modulus theorem which shows that any compact set can be removed from any component of V which is of positive dimension. More subtle results come from application of Phragmén–Lindelöf theorems.

A closed set $\sigma \subset V$ is called $\mathcal{W}$ *sufficient for* V if the topology of $\hat{\mathcal{W}}(V)$ can be defined by means of the semi-norms

$$||F||_k = \sup_{\hat{x}\in\sigma} \frac{|F(\hat{x})|}{k(\hat{x})}. \tag{1.102}$$

For example, when $V = \mathbb{C}^n$, the union of the real and imaginary spaces is $\mathcal{E}'$ sufficient for $\mathbb{C}^n$ (see Chapter 3), as is the set of complex lattice points

$$\{k + il\}_{k,l\in\mathbb{Z}^n}$$

(see [149]).

If σ is sufficient for V then the support of the "Fourier transforms" μ/k in (1.101) can be chosen to be contained in σ so that we can modify (1.101) to

$$f(x) = \int_\sigma e^{i\hat{x}\cdot x} \frac{d\mu(\hat{x})}{k(\hat{x})}. \tag{1.101*}$$

We can regard the FP as giving a parametrization of the solutions of $\overrightarrow{P}(D)f = 0$, namely by points in V. We regard this as the fundamental parametrization as we can often use the FP to study other PPs.

One of the important PPs is the Cauchy Problem (CP). We now present a general method for giving explicit solutions to Cauchy-like problems which we term the *Watergate method* (WM).

Suppose $r = 1$ so V is a hypersurface. Let Y be a hyperplane. We write $x = (t, y)$ where t is the coordinate in $T = Y^\perp$, with a similar notation for $\hat{x}$. For $\hat{y} \in \hat{Y}$ we denote by $V_{\hat{y}}$ the set of $\hat{t}$ for which $(\hat{t}, \hat{y}) \in V$; we call $V_{\hat{y}}$ the points in V which lie above $\hat{y}$. ($\hat{Y}$ is the same as Y except $\hat{Y}$ lies in the Fourier dual space.) For the CP we want Y to be a noncharacteristic hyperplane for $P(D)$, meaning that for generic $\hat{y} \in \hat{Y}$ the number of points in $V_{\hat{y}}$ is the degree N of P and, moreover,

$$|\hat{t}| \le c(1 + |\hat{y}|) \tag{1.103}$$

for $\hat{t} \in V_{\hat{y}}$.

We introduce the Cauchy map[5]

$$\mathcal{C}' : f \mapsto (\{h_j(\partial/\partial t)f|_Y\}, P(D)f) \tag{1.104}$$

of

$$\mathcal{W}' \to [\mathcal{W}'(Y)]^N \oplus \mathcal{W}'. \tag{1.105}$$

Here h_j are polynomials representing normal derivatives; we always take $h_0 \equiv 1$. $\mathcal{W}'(Y)$ is the space of restrictions of $f \in \mathcal{W}'$ to Y. The CP is called *well posed* if $\mathcal{C}'$ is a topological isomorphism. Of course this depends on the space $\mathcal{W}'$.

[5]We have written $\mathcal{C}'$ rather than $\mathcal{C}$ for the Cauchy map since the kernels of the differential operators lie in $\mathcal{W}'$, which is a space of functions which are large at infinity.

The adjoint of $\mathcal{C}'$ is given by

$$\mathcal{C}(\{w_j\}, w) = \sum h_j'(\partial/\partial t)w_j + P'(D)w. \tag{1.106}$$

Here $w_j \in \mathcal{W}(Y)$ which is considered as a subspace of $\mathcal{W}$ by using the adjoint of restriction. The Fourier transform $\hat{\mathcal{C}}$ of $\mathcal{C}$ is (up to unimportant factors of $\pm i$)

$$\hat{\mathcal{C}}(\{\hat{w}_j\}, \hat{w}) = \sum \hat{h}_j(i\hat{t})\hat{w}_j(\hat{y}) + \hat{P}(i\hat{x})\hat{w}(\hat{x}). \tag{1.107}$$

Since we want $\hat{\mathcal{C}}$ to be a topological isomorphism we want to construct $\{\hat{w}_j\}, \hat{w}$ so that $\hat{\mathcal{C}}(\{\hat{w}_j\}, \hat{w})$ is equal to an arbitrary $F(\hat{t}, \hat{y}) \in \hat{W}$. We can analyze this equation by examining it on the variety $V : \{\hat{P}(i\hat{x}) = 0\}$. For each fixed $\hat{y}$ there are N points $(\hat{t}_k(\hat{y}), \hat{y})$ in V, counting multiplicity, above $\hat{y}$. We have assumed that the generic multiplicity is 1. Thus for generic $\hat{y}$ we are left with N equations in the N unknowns $\{\hat{w}_j(\hat{y})\}$:

$$\sum \hat{h}_j(i\hat{t}_k(\hat{y}))\hat{w}_j(\hat{y}) = F(\hat{t}_k(\hat{y}), \hat{y}). \tag{1.108}$$

If

$$\det\left[\hat{h}_j(it_k(\hat{y}))\right] \not\equiv 0$$

we can solve for $\hat{w}_j(\hat{y})$ generically. We then need special arguments to show things behave correctly at the points where the determinant vanishes, and also to show that $\hat{w}_j \in \hat{\mathcal{W}}(\hat{Y})$.

Remark. *Equation (1.108) shows how the Fourier transform of the dual of the CD corresponds to the function $F \in \hat{\mathcal{W}}(V)$. By duality this means that a well-posed PP defines a Fourier transform from the PS to V. This idea is pursued in Section 1.7.*

The same formalism applies when $r > 1$, but the CP is much more complicated. In general we may not be able to give all data on a single Y. But in all of our examples we shall be able to use a single Y whose dimension = dim V.

The CP has a certain structure which distinguishes it from other parameterization problems. When C is an isomorphism, $U \in W$ can be written in the form

$$U = \sum h_j'(\partial/\partial t)w_j + P'(D)w$$

as in equation (1.106). In particular we would choose $U = Q\delta_y$ where Q is a differential operator and $y \in Y$. Writing, more precisely, $w_j(Q, y), w(Q, y)$ for such U we have, for any $f \in \mathcal{W}'$,

$$(Qf)(y) = \sum w_j(Q, y) \cdot h_j(\partial/\partial t)f + w(Q, y) \cdot P(D)f. \tag{1.109}$$

In general we define a *Cauchy problem* for any system $\vec{P}(x,D)$ to consist of a *Cauchy surface (CS)* Y and differential operators h_j which cannot be expressed in terms of tangential derivatives to Y such that for any differential operator Q and any $y \in Y$ we can find $w_j(Q,y) \in \mathcal{W}(Y)$ and ${}_kw(Q,y) \in \mathcal{W}$ such that

$$(Qf)(y) = \sum w_j(Q,y) \cdot h_j f + \sum {}_kw(Q,y) \cdot P_k(x,D)f \tag{1.110}$$

for any $f \in \mathcal{W}'$. (P_k are the components of $\vec{P}$.)

In particular, if $\vec{P}f = 0$ then $(Qf)(y)$ can be expressed in terms $\{h_j f\}$ which is the *Cauchy data (CD)* of f.

This definition is meaningful on manifolds. It will be explored in Section 9.4.

In Chapter IX of FA the case $N < \infty$,

$$h_j(\partial/\partial t) = \partial^j/\partial t^j,$$

which corresponds to the usual CP is studied in detail. In Chapters 3 and 4 we introduce the idea of *harmonicity*, which leads to other possibilities and also allows $N = \infty$ in special situations. When $N = \infty$ we call the CP *exotic*.

One of these special situations is the WP for the wave equation with data on the time axis as described above. (Note that we are now calling t the variable on the CS (the time axis) and y the orthogonal variable. This differs from the notation we used above.) Now $V_{\hat{t}}$ is a sphere. Thus in equation (1.108) $F(\hat{y}(\hat{t}),\hat{t})$ is a function on this sphere and we have to expand it in terms of the functions $\{h_j(\hat{y})\}$. The natural choice of $\{h_j\}$ is the homogeneous harmonic polynomials as their restrictions to any $V_{\hat{t}}$ is the set of spherical harmonics.

For generic $\hat{t}$ the $\{h_j|_{V_{\hat{t}}}\}$ define a basis and the bases are "harmonious," meaning that they are defined in terms of the family $\{h_j(\tilde{y})\}$ which is independent of $\hat{t}$.

The theory of harmonics and harmonious bases is exposed in Chapters 3 and 4. In fact the usual Cauchy basis $\{\hat{t}^k\}$ used when $N < \infty$ is a harmonious basis.

Remark. *For $N = \infty$ the sum in (1.108) is not in $\hat{\mathcal{W}}$ for most interesting $\mathcal{W}$, e.g. $\mathcal{W} = \mathcal{E}'$ = space of distributions with compact support. This means that when $N = \infty$ we can only expect $\hat{\mathcal{C}}$ to be a topological isomorphism from suitable sequences $\{\hat{w}_j\}$ onto $\hat{W}(V)$. Such sequences are characterized in Chapters 3 and 4 in certain "group" cases like the sphere, which is a homogeneous space of the orthogonal group.*

We should emphasize that the examples we gave using coordinates (t,y) are only an illustration of a general theory. The main point is that we have a fibering of V by subvarieties (generically) $V_{\hat{y}}$ and we have a harmonious basis that works for the generic $V_{\hat{y}}$. We do not need the $V_{\hat{y}}$ to be the intersections of V with linear planes (slices); what is needed is the harmonious basis. These may exist when the $V_{\hat{y}}$ are of positive dimension if they are homogeneous spaces of a fixed "nice"

Lie group. The h_j are chosen so that the restrictions of the $h_j(\hat{t})$ to $V_{\hat{y}}$ define a group representation theoretic basis for functions on $V_{\hat{y}}$.

This is exactly the situation we encountered in our discussion of the parametric Radon transform. The enveloping equation (1.42) defines a variety V which is fibered by varieties associated to first-order equations. (These equations will be studied in terms of group theory in Section 1.6.) The measure $d\lambda$ used in the parametric Radon transform determines a behavior on these fibers; this is the content of the second John equation. Thus $d\lambda$ is one component of a harmonious basis. We could replace $d\lambda$ by other measures $d\nu$ as in the remarks following (1.82) and obtain different behaviors on the fibers; they correspond to other "basis elements." Sometimes we can describe ν explicitly in terms of the fiber behavior. (See Section 6.2 for other examples.)

The above discussion of the WM for solving the CP and some other PPs is essentially rooted in the Fourier transform in all variables. A somewhat different approach starts with the Fourier transform in y only. This involves serious technical difficulties; it is treated to some extent in Section 9.4.

The algebraic problem which forms the basis for the WM as illustrated in (1.106)ff. gives way to a system of partial differential equations (we allow $r \geq 1$)

$$C_H(\{\hat{w}_j\}, \vec{w}_H) = \sum h_j\left(\frac{\partial}{\partial t}\right)\hat{w}_j(\hat{y}) + \vec{P}_H\left(\frac{\partial}{\partial t}, \hat{y}\right) \cdot \vec{w}_H(t, \hat{y}). \qquad (1.107H)$$

The subscript H refers to the Fourier transform in y only.

We have written H for *holonomic*. The reason for this terminology is that when $N < \infty$, for each fixed $\hat{y}$ the system of partial differential equations defined by $\vec{P}_H\left(\frac{\partial}{\partial t}, \hat{y}\right)$ is a holonomic system [102] which means that it behaves like an ordinary differential equation in many respects.

Slices define coarse grains or semi-coarse grains (see Section 1.1). Thus the slice V defines the coarse grain of solutions of the system of equations defined by V. The slices $V_{\hat{t}}$ define the semi-coarse grains of solutions corresponding to V whose Fourier transforms have support in $V_{\hat{t}}$. A harmonious basis is a set of functions, usually of $\hat{t}^{\perp}$, whose restrictions to each $V_{\hat{t}}$ form a basis for $\hat{\mathcal{W}}'(V_{\hat{t}})$.

We conclude this section with an abstract definition of a well-posed parametrization problem (WPPP). Our definition is a take-off on H. Weyl's method [158] of orthogonal projection.

For a general (geometric) parametrization problem for $\overrightarrow{P}(D)$ we start with some submanifold T of $\mathbb{R}^n$ and N differential operators $h_j(\partial/\partial x)$. The parametrization problem is related to the map

$$\alpha' : f \mapsto \left(\{h_j(\partial/\partial x)f\}|_T, \overrightarrow{P}(D)f\right). \qquad (1.111)$$

This maps $\mathcal{W}'$ into ${\mathcal{W}'}^N(T) \oplus (\tilde{\mathcal{W}}^r)'$ where $(\tilde{\mathcal{W}}^r)'$ is the subspace of $(\mathcal{W}^r)'$ satisfying the same compatibility conditions as the $P_k(D)$.

It is proven in FA that $\vec{P}(D)\mathcal{W}' = (\tilde{\mathcal{W}}^r)'$. Thus the dual $\tilde{\mathcal{W}}^r$ of $(\tilde{\mathcal{W}}^r)'$ is the quotient of $\mathcal{W}^r$ by the kernel of $\vec{P}'$.

The adjoint α is given by

$$\alpha(\{w_j\}, \vec{w}) = \sum h_j(\partial/\partial x)w_j + \overrightarrow{P}'(D) \cdot \vec{w} \tag{1.112}$$

as in (1.104), (1.106). We say that we have a WPPP if α (or α') is a topological isomorphism.

If α defines a WPPP then given any f there is a unique φ for which $\alpha'\varphi = (0, \overrightarrow{P}(D)f)$. This is the φ with 0 parametrization data and having the same image as f under $\overrightarrow{P}(D)$. We set

$$\varphi = R\overrightarrow{P}(D)f.$$

R is clearly a continuous linear map.

By construction

$$\overrightarrow{P}(D)f = \overrightarrow{P}(D)\varphi = \overrightarrow{P}(D)R\overrightarrow{P}(D)f.$$

Since $\overrightarrow{P}(D)f$ is an arbitrary element of $(\tilde{\mathcal{W}}^r)'$ we conclude

$$R \text{ is a right inverse for } \overrightarrow{P}(D).$$

From an abstract point of view, a WPPP is defined by a continuous right inverse for $\overrightarrow{P}(D)$.

This statement needs some clarification. $\vec{P}$ is a column vector with r components so R is a row vector with r components and $\vec{P}R$ is an $r \times r$ matrix. This matrix acts as the identity on $(\tilde{\mathcal{W}}^r)'$.

Let us examine the same idea from the dual point of view. By (1.112) given any $S \in \mathcal{W}$ we can express it uniquely in the form

$$S = \sum h_j(\partial/\partial x)w_j + \overrightarrow{P}'(D) \cdot \vec{w}. \tag{1.113}$$

We set

$$\vec{w} = \tilde{R}'S. \tag{1.114}$$

$\vec{w}$ can be described as that element of $\tilde{\mathcal{W}}^r$ for which $\overrightarrow{P}' \cdot \vec{w}$ and S agree on functions with vanishing parametrization data.

It follows from (1.113) that if S is of the form $\overrightarrow{P}' \cdot \vec{w}$ then

$$\tilde{R}'S = \tilde{R}'\overrightarrow{P}'\vec{w}. \tag{1.115}$$

This means that $\tilde{R}'$ is a left inverse for $\overrightarrow{P}'$ acting on $\tilde{\mathcal{W}}^r$.

Actually $\tilde{R}'$ is the adjoint R' of R, meaning that $\tilde{R}'S$ and $R'S$ agree on $(\tilde{\mathcal{W}}^r)'$; that is, on all $\overrightarrow{P}f$ for $f \in \mathcal{W}'$. For,

$$\begin{aligned} R'S \cdot \overrightarrow{P}f &= S \cdot R\overrightarrow{P}f \\ &= S \cdot \varphi \end{aligned}$$

by the above definitions.

Moreover,

$$\begin{aligned} \tilde{R}'S \cdot \overrightarrow{P}f &= \tilde{R}'S \cdot \overrightarrow{P}\varphi \quad \text{since} \overrightarrow{P}f = \overrightarrow{P}\varphi \\ &= \vec{w} \cdot \overrightarrow{P}\varphi \qquad \text{by (1.114)} \\ &= S \cdot \varphi \end{aligned}$$

by (1.113) and the fact that φ has vanishing parametrization data. Thus our assertion $\tilde{R}' = R'$ is established.

For another point of view, consider the exact sequence

$$0 \to \tilde{\mathcal{W}}^r \xrightarrow{\overrightarrow{P}'(D)} \mathcal{W} \to \mathcal{W}/\overrightarrow{P}'(D)\tilde{\mathcal{W}}^r \to 0. \tag{1.116}$$

Proposition 1.3 *$\overrightarrow{P}'(D)$ has a continuous left inverse if and only if the exact sequence splits, that is, there is a continuous map ρ' of $\mathcal{W}/\overrightarrow{P}'(D)\tilde{\mathcal{W}}^r$ into a subspace $\mathcal{Q} \subset \mathcal{W}$ for which*

$$\mathcal{W} = \overrightarrow{P}'(D)\tilde{\mathcal{W}}^r \oplus \mathcal{Q}.$$

The $\tilde{\mathcal{W}}^r$ and $\mathcal{Q}$ components of $w \in \mathcal{W}$ depend continuously on w.

Proof Suppose first that $\overrightarrow{P}$ has a continuous right inverse R or, what is the same thing, $\overrightarrow{P}'$ has a continuous left inverse R'. Write $S \in \mathcal{W}$ in the form

$$S = \overrightarrow{P}'R'S + (1 - \overrightarrow{P}'R')S. \tag{1.117}$$

We set

$$\mathcal{Q} = (1 - \overrightarrow{P}'R')\mathcal{W},$$

(1.117) means that $\overrightarrow{P}'\tilde{\mathcal{W}}^r + \mathcal{Q} = \mathcal{W}$. Since R' is a left inverse of $\overrightarrow{P}'$ the map $\rho' = (1 - \overrightarrow{P}'R')$ is defined on $\mathcal{W}/\overrightarrow{P}'(D)\tilde{\mathcal{W}}^r$.

It remains to show that the sum is direct. Suppose that

$$\overrightarrow{P}'A + (1 - \overrightarrow{P}'R')B = 0.$$

Applying R' and using $R'\overrightarrow{P}' = 1$ yields $A = 0$. Hence also the $\mathcal{Q}$ component $(1 - \overrightarrow{P}'R')B$ vanishes. Thus the sum is direct.

Conversely, suppose that the exact sequence splits. Write $S \in \mathcal{W}$ as

$$S = \overrightarrow{P}' A + B.$$

We now define R' by $A = R'S$. Since $\vec{P}$ maps $\mathcal{W}'$ onto $(\tilde{\mathcal{W}}^r)'$ it follows that $\overrightarrow{P}'$ is one–one on $\tilde{\mathcal{W}}^r$. Since the sum is direct, A is uniquely determined by S; thus R' is defined. It is continuous by hypothesis.

To show that $R'P' = 1$ we write $B = \vec{P}'R'B + C$. Since B is in the complement of the image of $\vec{P}'$ we deduce that $R'B = 0$. We now apply R' to $S = \vec{P}'R'S + B$ to conclude that $R'\vec{P}' = 1$ on the image of R'. Our definition of R' gives the range of R' as $\{A\}$ which is all of $\tilde{\mathcal{W}}^r$. Thus $R'\vec{P}$ is the identity.

Remark. *For the space $\mathcal{E}$ of C^∞ functions not all $\overrightarrow{P}(D)$ have continuous right inverses. They exist for hyperbolic $\overrightarrow{P}(D)$, because the CP is well posed. They exist for some other $\overrightarrow{P}(D)$ but the situation is not well understood. (See the profound works of Meise, Taylor and Voigt [120], [121], [122].) In terms of the FP, R' gives a continuous linear extension of functions in $\hat{\mathcal{W}}(V)$ to $\hat{\mathcal{W}}$.*

These concepts are discussed in detail in Chapter 4.

Suppose we are given N functions $\vec{g} = (g_1, \ldots, g_N)$ on T and we want to know if there is an f in some space $\mathcal{W}'$ with $\alpha'(f) = (\vec{g}, 0)$. We call $\vec{g}$ *potential PD* (or *potential CD*). The question of whether potential CD is actual CD is examined in Section 1.5 and is studied in detail in Chapters 5 and 9.

1.5 Differential equations

We have already met differential equations in some of our above considerations such as the parametric Radon transform $\mathbf{R_P}$. Let us examine the John equations (1.42) or, more generally, (1.79) and (1.80) from a somewhat different point of view.

The FP as described in the previous section associates to the John equations the variety V of vectors $\hat{a}^i, \hat{b}$ which are parallel. For, equation (1.80) leads to

$$\hat{a}^i_j \hat{b}_m = \hat{a}^i_m \hat{b}_j, \tag{1.118}$$

which means that $\hat{a}^i$ is parallel to $\hat{b}$. As we shall see later the "main part" V^0 of V is defined by $\hat{b} \neq 0$, hence $\hat{a}^i = \lambda_i \hat{b}$. (Actually the FP does not apply in its usual form because the spaces we deal with are not AU. This point is clarified in Chapter 6.)

We start with the Fourier representation (1.101) of F. (We absorb $k(x)$ in μ, writing μ for μ/k.) Assume that the support of μ can be chosen in V^0. We regard V^0 as a fiber space with base $\{\hat{b}\}$ and fiber $\{\lambda\}$. We evaluate the integral (1.101) by first integrating over the fibers and then over the base. (In the present

notation $x = (a^1, \dots, a^l, b)$ and $\hat{x} = (\lambda_1 \hat{b}, \dots, \lambda_l \hat{b}, \hat{b})$.) We obtain

$$F(a,b) = \int h(a,b;\hat{b})\, d\nu(\hat{b}) \tag{1.119}$$

where

$$h(a,b;\hat{b}) = \int e^{i \sum \lambda_j a^j \cdot \hat{b} + ib \cdot \hat{b}}\, d\mu(\lambda, \hat{b}). \tag{1.120}$$

(Integration is only over λ.) The measure ν is obtained from μ (formally) in an obvious manner.

Solutions of the John equations for which support $\mu \subset V - V^0$ do not play an important role in our work. In Chapter 6 we shall show how such solutions can be eliminated.

Remark. *At first sight it might seem puzzling to use euclidean Fourier analysis in the space defined by* $\{a^i\}, \{b\}$*; the natural group structure of* $\{a, b\}$ *would seem to involve a more complicated, nonabelian group. We shall return to this point in Section 1.7 and in Chapter 6.*

Let us examine the invariance equation (1.81) from our Fourier analysis viewpoint. It is more convenient to use the infinitesimal form of invariance, i.e. equation (1.82). The Fourier transform of (1.82) is, formally,

$$(\hat{b} \cdot \nabla_{\hat{a}^i}) \hat{F} = 0 \tag{1.121}$$

where we have written $\hat{F}$ for μ. Equation (1.121) means that $\hat{F}$ is invariant under the addition to any $\hat{a}^i$ of an arbitrary multiple of $\hat{b}$. This indicates that

$$\hat{F}(\hat{x}) = \hat{F}(\lambda_1 \hat{b}, \dots, \lambda_l \hat{b}, \hat{b}) \tag{1.122}$$

is independent of λ as long as $\hat{b} \neq 0$ (which we assume for the present).

With this structure of $\mu = \hat{F}$ let us return to the evaluation of F as in (1.119) and (1.120). Since μ is independent of λ we can evaluate the (inverse) Fourier transform of μ by integrating first over $\hat{b}$ and then over λ, i.e. the opposite of (1.119), (1.120). λ invariance means that (formally) μ is of the form $\hat{F}(\hat{b})\, d\lambda\, d\hat{b}$; call $f(b)$ the Fourier transform of $F(\hat{b})$. Then

$$\begin{aligned} F(a,b) &= \iint e^{i \sum \lambda_j a^j \cdot \hat{b} + ib \cdot \hat{b}} \hat{F}(\hat{b})\, d\hat{b}\, d\lambda \\ &= \int f\left(\sum \lambda_j a^j + b\right) d\lambda. \end{aligned} \tag{1.123}$$

We have thus given a formal derivation of the sufficiency of the John equation for F to be of the form $\mathbf{R_P} f$. The same method applies to the other examples discussed in Section 1.1.2.

Of course, we are left with the crucial problem of showing why we can replace V by V^0. This is discussed in Chapter 6.

Remark 1 *In the case of the attenuated Radon transform, suitable attenuations imply a fixed behavior of $\hat{F}$ on the fibers (replacing constancy which is related to the measure $d\lambda$). Thus we are successful if this fixed behavior determines F on a fiber from its value at any point, e.g. $\lambda = 0$. Put in other terms, the Fourier transform of the attenuation often defines a differential equation on each fiber. Our success depends on the differential equation being of first order.*

Remark 2 *The above discussion shows that one natural setting for the parametric Radon transform lies in the fiber decomposition of algebraic varieties coupled with equations on the fibers.*

Let us examine the simplest case of the FP, meaning $P_j \equiv 0$, i.e. there is no equation, so $V = \mathbb{R}^n$ and the solutions of the equations are arbitrary functions on $\mathbb{R}^n$. The Fourier transform of a function (measure) supported on a line $\hat{S}_{\mathbf{g}}$ through the origin is readily seen to be a spread function for the spread of hyperplanes orthogonal to $S_{\mathbf{g}}$. Thus the geometric decomposition of $\mathbb{R}^n$ into lines through the origin corresponds to the nonparametric hyperplane Radon transform. ($\hat{S}_{\mathbf{g}}$ and $S_{\mathbf{g}}$ are the same, i.e. paired by $x \cdot \hat{x}$, except that $\hat{S}_{\mathbf{g}}$ is in $\{\hat{x}\}$ space and $S_{\mathbf{g}} \subset \{x\}$.) For $\mathbf{R}^l$ we decompose $\mathbb{R}^n$ into the planes of dimension $n - l$ through the origin. Of course this decomposition has much overlap which must be taken into account (see Chapter 2).

We can also decompose $\mathbb{R}^n$ into the union of spheres $\hat{S}_r$, center origin, radius r. Now, a function or measure $\hat{k}$ has its support on $\hat{S}_{\hat{r}}$ (complex) if and only if its Fourier transform k is an eigenfunction of the Laplacian with eigenvalue $-\hat{r}^2$ (FP). With this structure it makes sense to call $\hat{r}^2$ the Grassmann parameter. The spread functions are the eigenfunctions of Δ with eigenvalue $-\hat{r}^2$.

We have discussed a semi-coarse grain decomposition of solutions of partial differential equations for which V is a fiber space; the semi-coarse grains correspond to the fibers of V. Can we find a natural associated fine grain decomposition? This would correspond to a basis for functions on the fibers.

One natural basis is $\{\delta_{\hat{x}}\}_{\hat{x} \in \phi}$ for each fiber ϕ of V. When ϕ is an algebraic variety defined by equations $P(\phi)(ix) = 0$ then, by Fourier transform, this becomes the basis $\{\exp(ix \cdot \hat{x})\}_{\hat{x} \in \phi}$ for the kernel of the associated (by FP) differential equation $P(\phi)(D)f = 0$. In x space this basis does not have geometric significance.

A different basis corresponds to taking a multiplicity Radon transform on ϕ. Since ϕ is an algebraic variety, say irreducible of dimension m, we can find a principal noncharacteristic. This is a linear variety $\hat{S}$ (through the origin) of dimension m such that each slice $\hat{s} + \hat{S}^{\perp}$ meets ϕ generically in a fixed finite

number of points for any $\hat{s} \in \hat{S}$; there is also a condition at infinity. We can associate to such an $\hat{S}$ a CP with data on S (sometimes we have to put further conditions on some linear subvarieties; such complications do not appear in this work although they can be dealt with).

Thus in a hyperbolic situation, as in Section 1.1, $\{\delta_s\}_{s\in S}$ and certain derivatives constitute a basis for the kernel of $\partial(\phi)$ in the sense that the solutions of the CP with CD at a point (null solutions) form a basis for solutions.

A detailed treatment of the CP is found in Chapter IX of FA. Various aspects appear throughout the present book.

Other PPs lead to other bases for the kernel of $P(\phi)(D)$.

For the above-described spread defined using the spread in $\{\hat{x}\}$ by spheres centered at the origin there is another significant basis, the basis of spherical harmonics. The Fourier transform of spherical harmonics on the sphere of radius $\hat{r}$ forms the basis

$$\{J_\alpha(r\hat{r})P_j^\alpha(\theta)\} \tag{1.124}$$

for the same spread in (r, θ). The J_α are suitable Bessel functions and the P_j^α are spherical harmonics. (θ is a parameter on the unit sphere.) A detailed study of harmonic polynomials and related questions is treated in Chapters 3 and 4.

Instead of spheres we could use other decompositions of $\mathbb{C}^n$ by algebraic families of varieties $V_{\mathbf{g}}$. Our theory works well when $V_{\mathbf{g}}$ are homogeneous spaces of "nice" Lie groups.

It is important to characterize the expansion coefficients of various classes $\mathcal{W}$ of functions of x in the bases such as (1.124). This is carried out in Chapters 3 and 4.

These spreads are defined as solutions of differential equations rather than geometrically. We have noted that the geometric spreads of Radon have differential equations definitions. Let $L(\mathbf{g}, 0)$ be a plane of dimension l through the origin. Let $\alpha_1, \ldots, \alpha_l$ be l independent directions in $L(\mathbf{g}, 0)$. Then the spread functions h corresponding to $\mathbf{g}$ are the solutions of the system of equations

$$\frac{\partial h}{\partial \alpha_1} = \frac{\partial h}{\partial \alpha_2} = \cdots = \frac{\partial h}{\partial \alpha_l} = 0. \tag{1.125}$$

Of course, as emphasized, the solutions of (1.125) admit a geometric basis. This seems tied to the fact that the equations are first order. There have been many efforts in the literature at "geometrizing" partial differential equations; perhaps the most profound originates in the work of Éli cartan [26].

For another example, let us study harmonic functions, i.e. solutions of Laplace's equation

$$\Delta f = 0. \tag{1.126}$$

By the FP such functions are Fourier transforms of measures $d\mu(\hat{x})$ on the complex cone $\hat{S}_0^{\mathbb{C}} = \{\hat{x} \big| \sum \hat{x}_j^2 = 0\}$. This cone has a natural (semi-coarse grain) decomposition into the (complex) lines in it. Let us restrict our considerations

to the lines

$$\hat{x} = \lambda(\hat{\omega}_1, \ldots, \hat{\omega}_{n-1}, i) \tag{1.127}$$

where $\hat{\omega}$ is a point in $\hat{S}_1^{\mathbb{C}}$ (the complex unit sphere in $n-1$ variables) and λ is an arbitrary complex number.

The decomposition leads to

$$f(x) = \int_{\hat{S}_1^{\mathbb{C}}} h(x; \hat{\omega})\, d\nu(\hat{\omega}) \tag{1.128}$$

where $h(x; \hat{\omega})$ is the Fourier transform of μ (which we can assume is smooth) over the line (1.127). (We shall discuss the missing lines below.) It is easily seen that h extends to a holomorphic function of complex x since the "Fourier transform" $d\mu(\hat{x})$ of f is exponentially decreasing due to the ellipticity of Δ. Thus $h(x; \hat{w})$ is a spread harmonic function for the (geometric) spread

$$x \cdot (\hat{\omega}, i) = 0. \tag{1.129}$$

This spread is geometric if we think of x as a complex parameter and h as a holomorphic function of x. It is not completely geometric because the δ function of the leaves of the spread corresponding to (1.129) are not holomorphic, and hence do not correspond to actual spread functions. Rather they form a basis in the "wave packet" sense (see (1.17)ff.) for holomorphic spread functions. In particular we can write

$$h(x; \hat{\omega}) = k[x \cdot (\hat{\omega}, i)] \tag{1.130}$$

for a suitable holomorphic function k.

One might be puzzled by our assertion (1.128) because we have ignored the part of $\hat{S}_0^{\mathbb{C}}$ where $\hat{x}_n = 0$. It turns out that this set can be ignored. Even more is true: we do not need the whole complex sphere $\hat{S}_1^{\mathbb{C}}$; its real part $\hat{S}_1^{\mathbb{R}}$ will suffice. In the language of FA this means that the set

$$\{\lambda(\hat{\omega}_1, \ldots, \hat{\omega}_{n-1}; i)\}_{\lambda \in \mathbb{C}, \hat{\omega} \in \hat{S}_1^{\mathbb{R}}} \tag{1.131}$$

is **sufficient** for solutions of (1.126) (see Section 1.4). The proof that (1.131) is sufficient is given in Chapter 4.

In particular, for $n = 2$ we obtain the decomposition of a harmonic function into holomorphic and antiholomorphic parts. For $n = 3$, by means of (1.128), we arrive at Whittaker's representation

$$f(x_1, x_2, x_3) = \int_0^{2\pi} h(x_1 \cos t + x_2 \sin t + i x_3, t)\, dt \tag{1.132}$$

of harmonic functions. Here h is a holomorphic function of its first argument.

Instead of using lines, we can use planes of higher dimension when $n > 3$. Suppose $n = 2m$ is even. Then we can decompose the cone $\hat{x}^2 = 0$ into complex planes of dimension m. To see this, note that the m vectors

$$(\overbrace{0,0;0,0;\ldots;0,0}^{j-1};1,i;0,0;\ldots;0,0) = v_j \tag{1.133}$$

lie on $\hat{S}_0^{\mathbb{C}}$ and are orthogonal in the usual inner product. Thus they span a complex linear space $\hat{L}(\hat{0},\hat{0}) \subset \hat{S}_0^{\mathbb{C}}$ of dimension m.

We apply the complex orthogonal group to obtain other subspaces $\hat{L}(\hat{\mathbf{g}},0)$. As in the case of lines we need only the real orthogonal group.

In case $n = 2m+1$ is odd we can proceed as before except that we set the final component of the v_j equal to i and we replace $1,i$ in (1.133) by $1,0$.

Naturally, many other choices for parametrizing this set of m planes are possible. For example, when $n = 4, m = 2$ we can use

$$u_1 = (1,i,t,it), \quad u_2 = (t,-it,-1,i) \quad t \text{ real.} \tag{1.134}$$

This spans a one real parameter family of planes which covers a sufficient part of $\hat{S}_0^{\mathbb{C}}$. The analog of (1.128) for planes becomes Bateman's representation of harmonic functions

$$f(x) = \int h[(x_1+ix_2)+t(x_3+ix_4), t(x_1-ix_2)-(x_3-ix_4), t]\,dt \tag{1.135}$$

where h is a suitable holomorphic function of its first two arguments.

Remark. *Actually Bateman's representation is not complete. We must add a component at "infinity" (see Chapter 4).*

We can think of $\mathbb{R}^4 = \{(a,b)\}$ where a,b are the parameters of the parametric representation of lines in $\mathbb{R}^2$ as in (1.36). The John equations (1.42) reduce in this case to a single ultrahyperbolic equation. In the complex domain there is no difference between ultrahyperbolic and Laplace equations.

More precisely, Bateman's 2 planes are parametrized by $\{t\}$; they are spanned by the vectors u_1, u_2 of (1.134). John's fibers are parametrized by $\{\lambda\}$ and are spanned by the vectors

$$(a,b) = (\lambda,0,1,0) \quad \text{and} \quad (a,b) = (0,-\lambda,0,-1). \tag{1.136}$$

The change of variables

$$\begin{aligned} a_1 &= (x_3 - ix_4)/2 \\ a_2 &= (-x_1 - ix_2)/2 \\ b_1 &= (x_1 - ix_2)/2 \\ b_2 &= (x_3 + ix_4)/2 \end{aligned} \tag{1.137}$$

interchanges the two sets of two planes with $t \longleftrightarrow \lambda$.

Remark 1 *The complex dimension of $\hat{S}_0^{\mathbb{C}}$ is $n-1$. Therefore for $n=2m$ (resp. $2m+1$) we need an $m-1$ (resp. m) parameter family of m planes. Thus it is only for $n=3,4$ that the number of parameters is 1. This accounts for the simplicity of (1.132) and (1.135) as compared to the higher dimensional analogs.*

Remark 2 *For $n=4$ we can think of $\hat{S}_0^{\mathbb{C}}$ as the null quaternions. It seems reasonable, in light of the situation for $n=2$, that our decomposition should be related to a decomposition into quaternionic holomorphic and antiholomorphic functions; we do not understand this relation.*

The ideas we have presented bear a close relation to the Penrose transform (see [32]).

We have noted that the function $h(x,\hat{\omega})$ of (1.128) is holomorphic in x. The integral in (1.128) is taken over $S_{\hat{1}}$. $S_{\hat{1}}$ is a circle for the Whittaker and Bateman representations. We can think of $S_{\hat{1}}$ as a cycle in the Riemann sphere $S_{\hat{1}}^{\mathbb{C}}$. If $h(x;\hat{\omega})\,d\nu(\hat{\omega})$ were holomorphic in $\hat{\omega}$ in a neighborhood of $S_{\hat{1}}$ then we could deform the contour in (1.128) to a homologous cycle.

Penrose's idea is to regard $h(x;\hat{\omega})\,d\nu(\hat{\omega})$ as a differential form in $\hat{\omega}$ with values in holomorphic functions of x. This interpretation makes sense within the framework of Dolbeault cohomology (see [162]). $h(x;\hat{\omega})\,d\nu(\hat{\omega})$ becomes a 1-cocycle which can be integrated over 1-cycles.

Finally, in addition to proving existence for Bateman–Whittaker representations, Penrose shows that the lack of uniqueness in Bateman's representation comes from cocycles which are cohomologous to zero. Moreover Penrose's results apply to any simply connected region in x whereas the ideas using the FP apply only to convex sets.[6]

A different form of the uniqueness result is given in Chapter 9.

We can introduce another relation between differential equations and the Radon transform. Let $n=2m$ and regard $\mathbb{R}^n=\mathbb{C}^m$. Let h be a function (measure) which is exponentially small at infinity and supported by the Cauchy–Riemann variety

$$V_{CR}=\{x_1+ix_{m+1}=0,\dots,x_m+ix_n=0\}. \tag{1.138}$$

Form the complex hyperplane Radon transform

$$\mathbf{R}^C h(s,\mathbf{g})=h_{\mathbf{g}}(s). \tag{1.139}$$

Now $\mathbf{g}$ is a complex hyperplane, $s\in\mathbf{g}^{\perp}$, and we integrate over complex hyperplanes.

[6]The classes of harmonic functions that Penrose considers are more restrictive than those of the present work. For this reason it appears that the Penrose theory does not give the Bateman–Whittaker representation for arbitrary harmonic functions.

Since support $h \subset V_{CR}$ the Fourier transform $\hat{h}$ of h is a holomorphic function of m complex variables. As in the case of the real Radon transform the projection–slice theorem (see Section 1.7) asserts that the Fourier transform $\hat{h}_{\mathbf{g}}(\hat{s})$ of $h_{\mathbf{g}}(s)$ (in s) is the restriction (slice) of $\hat{h}$ on $\mathbf{g}^{\perp}$ and hence is holomorphic in $\hat{s}$, i.e.

$$\hat{h}\big|_{\mathbf{g}^{\perp}} = \hat{h}_{\mathbf{g}}. \tag{1.140}$$

Remark. *It is important to realize that h and $h_{\mathbf{g}}$ are analytic functionals and, as such, have nonunique representations as measures (see Chapter 9 for details). When we write support $h \subset \sigma$ we mean there is a representative measure for h with support contained in σ.*

We can now ask the question: when is a set $\{h_{\mathbf{g}}\}$, with $\hat{h}_{\mathbf{g}}(\hat{s}_{\mathbf{g}})$ holomorphic, of the form $h_{\mathbf{g}} = \mathbf{R}^{C}h(\mathbf{g})$ with support $h \subset V_{CR}$? Put in other terms, when does a set of holomorphic functions $\{\hat{h}_{\mathbf{g}}(\hat{s}_{\mathbf{g}})\}$ form the slices of an entire function?

What is needed is a compatibility condition. We formulate it as

COMP ORIGIN. *There is a formal power series $\hat{h}^0$ whose restriction to each $\hat{\mathbf{g}}^{\perp}$ is the formal power series at the origin of $\hat{h}_{\mathbf{g}}$.*

The restriction of a formal power series to a line L is defined as follows. Let $(\omega_1, \ldots, \omega_n)$ be direction cosines for L. Then substitute $\hat{x} = (\hat{s}\omega_1, \ldots, \hat{s}\omega_n)$ to obtain a formal power series in $\hat{s}$.

The existence of the $\hat{h}_{\mathbf{g}}$ can be thought of as saying that a suitable analytic wave front set of $\hat{h}^0$ is empty (see Section 5.3). The emptiness of the wave front set implies that $\hat{h}^0$ defines an entire function.

The sufficiency of COMP ORIGIN for the existence of an entire $\hat{h}$ is due to Korevaar and Wiegerinck [106–108]. Actually we do not need all complex lines through the origin—it suffices to use *real complex lines*; they are the complexifications of real lines, i.e. they are complex lines with real direction numbers. The fact that real complex lines suffice in this context is remarkable because the union of such lines has real dimension $n+1$ so, for $n \geq 2$, it is only a small part of $\mathbb{C}^n$.

COMP ORIGIN is also the crucial point for the study of the surjectivity of the ordinary Radon transform. By Fourier transformation it becomes moment conditions on $\mathbf{R}f$ or $\mathbf{R}^{\mathrm{C}}h$. Naturally there are analogs of COMP ORIGIN for varieties whose intersections are of positive dimension (see Chapter 2 for details).

We shall be interested in analogous questions in which the functions $\hat{h}$ and $\{\hat{h}_{\mathbf{g}}\}$ are restricted by growth conditions, e.g. entire functions of exponential type, entire functions of order ρ, etc. In general let Φ be a suitable function of $|\Re\hat{x}|, |\Im\hat{x}|$. Let $\mathcal{H}(\Phi)$ be the space of entire functions satisfying

$$\hat{h}(\hat{x}) = \mathcal{O}(e^{\Phi(c\hat{x})}) \tag{1.141}$$

for some constant c. We call $\mathcal{H}(\Phi)$ a *KW* (Korevaar–Wiegerinck) *space* if whenever $\{\hat{h}_{\mathbf{g}}\}$ is a collection of functions on the complex lines $\hat{\mathbf{g}}^{\perp}$ through

the origin such that each $\hat{h}_{\mathbf{g}}$ satisfies the growth condition induced by $\mathcal{H}(\Phi)$ on $\hat{\mathbf{g}}^{\perp}$ (the constants in $\mathcal{O}$ and c not uniform in $\mathbf{g}$) and COMP ORIGIN then there is an $\hat{h} \in \mathcal{H}(\Phi)$ whose restriction to each $\hat{\mathbf{g}}^{\perp}$ is $\hat{h}_{\mathbf{g}}$.

The Korevaar–Wiegerinck theorem can be regarded from two other viewpoints:

(1) COMP ORIGIN can be regarded as the strong slice compatibility for the spaces $\mathcal{H}(\Phi, \hat{\mathbf{g}}^{\perp})$ of functions which are holomorphic on the lines $\hat{\mathbf{g}}^{\perp}$ (and bounded by Φ).
(2) We can reverse our perspective and regard the formal power series $\hat{h}^0$ as the primary object.

To clarify (2), think of $\hat{h}^0$ as "potential exotic CD" for the Cauchy–Riemann equations on $\mathbb{C}^n = \mathbb{R}^{2n}$. Since support $\hat{h}^0$ is the origin we need infinitely many derivatives to "compensate" for the diminuation to 0 of the dimension n of the standard CS (see the discussion below). We want to know whether $\hat{h}^0$ is "actual exotic CD", meaning there is a holomorphic $\hat{h}$ whose exotic CD is $\hat{h}^0$. The Radon ansatz (Section 1.1) suggests that we examine restrictions of $\hat{h}^0$ to lower dimensional sets $\{L\}$. The Korevaar–Wiegerinck theorem chooses $\{L\}$ as the set of (real) lines through the origin.

Remark. *The formulation of COMP ORIGIN and its relation to uniqueness has its origin in Chapter XIII of FA where "holomorphic" is replaced by "quasianalytic."*

It is possible to generalize this construction. Instead of the Cauchy–Riemann system we could start with a general system $\overrightarrow{P}(D)\hat{h} = 0$ of linear, constant coefficient, partial differential equations. ($\overrightarrow{P} = (P_1, \ldots, P_r)$ where $\overrightarrow{P}\hat{h} = 0$ means $P_j\hat{h} = 0$ for all j.) The FP (Section 1.4) asserts that a solution $\hat{h}$ of $\overrightarrow{P}\hat{h} = 0$ is the Fourier transform of h whose support lies in the variety V defined by $\{P_j(ix) = 0\}$. The role of the complex lines $\hat{\mathbf{g}}^{\perp}$ is played by *mouths* for $\overrightarrow{P}$. These are suitable submanifolds (which we again denote by $\hat{\mathbf{g}}^{\perp}$) of $\mathbb{R}^n$ on which $\overrightarrow{P}$ induces equations $\overrightarrow{P}_{\hat{\mathbf{g}}^{\perp}}$ (see Chapter X of FA).

It is reasonable to ask when a collection of functions $\{\hat{h}(\hat{s}_{\mathbf{g}^{\perp}})\}$ supported on a sufficiently large set of mouths $\hat{\mathbf{g}}^{\perp}$ and satisfying the mouth equations

$$\overrightarrow{P}_{\hat{\mathbf{g}}^{\perp}}\hat{h}(\hat{s}_{\mathbf{g}^{\perp}}) = 0 \tag{1.142}$$

fit together to form a solution $\hat{h}$ of $\overrightarrow{P}\hat{h} = 0$.

We have not studied this question in detail.

Instead of taking complex lines through the origin we could use complex lines parallel to the coordinate planes. Again such L are mouths for $\bar{\partial}$. Actually we

do not need all such lines but only lines of the form

$$L(\hat{x}_1^0, \ldots, \hat{x}_{j-1}^0, \hat{x}_{j+1}^0, \ldots, \hat{x}_n^0) = \{\hat{x}_1 = \hat{x}_1^0, \ldots, \hat{x}_{j-1} = \hat{x}_{j-1}^0, \hat{x}_j \text{ arbitrary}, \hat{x}_{j+1} = \hat{x}_{j+1}^0, \ldots, \hat{x}_n = \hat{x}_n^0\} \tag{1.143}$$

for

$$\hat{x}_1^0, \ldots, \hat{x}_{j-1}^0, \hat{x}_{j+1}^0, \ldots, \hat{x}_n^0 \text{ real.} \tag{1.144}$$

A theorem of Hartogs and of Bernstein [18] (see Section 5.2) asserts that if $\hat{h}$ is a function on $\mathbb{R}^n$ which extends to a holomorphic function on each $L(\hat{x}_1^0, \ldots, \hat{x}_{j-1}^0, \hat{x}_{j+1}^0, \ldots, \hat{x}_n^0)$ then $\hat{h}$ actually extends to be entire. (There is also a local form of this theorem.)

This is another instance of the Radon ansatz. $\hat{h}$ is regarded as "potential ordinary CD" for the Cauchy–Riemann system $\bar{\partial}$. We want to know whether $\hat{h}$ is "actual CD," meaning there is a holomorphic function $\hat{H}$ whose CD on $\mathbb{R}^n$ is $\hat{h}$. The Radon ansatz suggests that we study the restrictions of $\hat{h}$ to lower dimensional sets $\{L\}$. For the Hartogs–Bernstein theorem $\{L\}$ is the set of lines parallel to the coordinate axes.

Since the $L(\hat{x}_1^0, \ldots, \hat{x}_{j-1}^0, \hat{x}_{j+1}^0, \ldots, \hat{x}_n^0)$ are not planes through the origin, the translation of this result into Fourier–Radon language is of a different nature from our previous concepts of the Radon transform.

To simplify the notation, suppose $n = 2$. We are given holomorphic functions $\hat{h}_{\hat{x}_1^0}(\hat{x}_2)$ and $\hat{h}_{\hat{x}_2^0}(\hat{x}_1)$ (of one complex variable) for $\hat{x}_j^0$ real with $\hat{h}_{\hat{x}_1^0}(\hat{x}_2^0) = \hat{h}_{\hat{x}_2^0}(\hat{x}_1^0)$; we seek an entire function $\hat{h}$ of two complex variables which restricts to them on each complex line of the form $\hat{x}_1 = \hat{x}_1^0$ or $\hat{x}_2 = \hat{x}_2^0$. We want to represent $\hat{h}$ as the Fourier transform of h whose support lies in V_{CR}. Now, as before, $\hat{h}_{\hat{x}_j^0} = \hat{h}\delta_{\hat{x}_j = \hat{x}_j^0}$ so that

$$h_{\hat{x}_j^0} = h * [\exp(ix_j\hat{x}_j^0)\delta_{x_j}]. \tag{1.145}$$

In contrast to the previous notions of Radon transform in x space, we have only two geometric spreads, namely the lines parallel to the coordinate axes. However there is a crucial difference between the present situation and the previous one. The spreads in Radon transform are used to define averages (projections) over families of parallel lines; hence we obtain only one number for each leaf. Korevaar–Wiegerinck and Hartogs–Bernstein theorems deal with the actual values (slices) of functions on such families. This is more potent information.

For the Hartogs–Bernstein theorem in $\mathbb{C}^n$ we need n spreads, namely the lines parallel to the coordinate axes in $\mathbb{R}^n$. Each leaf is of real dimension 2 and a spread has $n-1$ real parameters. Thus the total number of real parameters is $n+1$. The Radon ansatz suggests that we should have a $2n$ parameter family of conditions.

The apparent contradiction is resolved by the fact that holomorphic functions of n variables ($n > 1$) and, more generally, solutions of overdetermined systems

of partial differential equations, have extension properties. They are discussed in some detail in Chapter 9.

Let us contrast the cases of lines through the origin with lines parallel to the coordinate axes, which is the contrast between the Korevaar–Wiegerinck and the Hartogs–Bernstein theorems in terms of viewpoint (1). In the former case we start with functions h_λ holomorphic on each complex line λ through the origin; all the (infinite) compatibility, namely COMP ORIGIN, is given at the origin. In the latter case we start with functions h_L holomorphic on the complex lines (1.144) parallel to the coordinate axes. Compatibility is given by equality at the points of intersection of these lines; the set of these intersections forms $\mathbb{R}^n$. In both cases the conclusion is the existence of an extension of the given data to a holomorphic function on a suitable open set in $\mathbb{C}^n$. (In Hartogs' work more complicated families of lines are possible; we shall not deal with that situation as the technique is quite different from the one we shall use.)

One might wonder if there is an intermediate situation. For example, in the case of lines parallel to the coordinate axes can we give all data on one line? To clarify this question, let us suppose $n = 2$. (We write $\hat{x}, \hat{y}$ for coordinates.) Let us use only lines parallel to the $\hat{x}$ axis but let us use all their derivatives on the $\hat{y}$ axis. This means that we assume that we are given functions $\hat{h}_{\hat{y}^0}$ on $\hat{y} = \hat{y}^0$ ($\hat{y}^0$ real) which are holomorphic in $\hat{x}$. Moreover for each k the derivatives

$$\frac{d^k}{d\hat{x}^k}\hat{h}_{\hat{y}^0}(0) = \hat{\phi}_k(\hat{y}^0) \tag{1.146}$$

are entire functions of $\hat{y}^0$.

It turns out that in this case we *cannot* assert the existence of an entire function $h(\hat{x}, \hat{y})$ whose restriction to each $\hat{y} = \hat{y}^0$ is $\hat{h}_{\hat{y}^0}$. To see this, let

$$\hat{h}(\hat{x}, \hat{y}) = \sum \hat{u}_j(\hat{y})\frac{\hat{x}^j}{j!}. \tag{1.147}$$

The $\hat{u}_j$ are entire functions which are uniformly bounded on the real axis but grow at $\hat{y} = i$ like $(j!)^2$. It is clear that the functions $\hat{h}_{\hat{y}^0} = \hat{h}(\hat{x}, \hat{y}^0)$ satisfy all the above hypotheses. But if, for example, $\hat{u}_j \equiv 0$ except for a lacunary sequence of j then the function $\hat{h}(\hat{x}, \hat{y})$ could not be regular at $\hat{y} = i$ for any $\hat{x}$.

By a slight variation of the construction we can construct $\hat{h}$ which is not holomorphic for any nonreal $\hat{y}$.

It is clear from this example that what is needed is a growth condition in k on the functions $\hat{\phi}_k$ of (1.146) for complex $\hat{y}^0$. Of course, there is the trivial condition that the series $\sum \hat{\phi}_k(\hat{y})\hat{x}^k/k!$ converges in the space of entire functions. Unfortunately this seems to be the best we can do.

We can, however, replace the origin in the Korevaar–Wiegerinck theorem by a sphere. Let S^{n-1} be the unit sphere in $\mathbb{R}^n$. Suppose that for each real hyperplane $\hat{L}(1, \mathbf{g})$ tangent to S^{n-1} we are given a function $\hat{h}_{\mathbf{g}}$ which extends to an entire

function on the complexification of $\hat{L}(1, \mathbf{g})$. Suppose that the $\hat{h}_{\mathbf{g}}$ agree on the intersection of the $\hat{L}(1, \mathbf{g})$. Then there is an entire function $\hat{h}$ which restricts to $\hat{h}_{\mathbf{g}}$ on each $\hat{L}(1, \mathbf{g})$.

For $n = 2$ this result is due to [4]. It was extended to $n > 2$ in [5].

One might think that the difference between this result and the negative result discussed above is that the sphere is curved (and in a suitable limit sense the origin in the Korevaar–Wiegerinck theorem is curved) while the lines $\hat{y} = \hat{y}^0$ are flat. However, this is not the case. To show this let us replace the tangent hyperplanes to the sphere by the normals to the S^{n-1}.

Thus, suppose that for each $\mathbf{g}$ we are given a function $\hat{h}_{\mathbf{g}}$ on the line $\hat{L}(0, \mathbf{g})$ through the origin. (We can assume $\hat{h}_{\mathbf{g}}$ is even so that we are really on the projective space rather than on the sphere.) We assume that the $\hat{h}_{\mathbf{g}}(\hat{s})$ are entire functions and that for each j the functions $d^j \hat{h}_{\mathbf{g}}(1)/d\hat{s}^j$ qua functions of $\mathbf{g}$ extend to be entire functions on the complexification $(S^{n-1})^{\mathbb{C}}$ of S^{n-1}. Then we cannot conclude the existence of an entire $\hat{h}$ which restricts to each $\hat{h}_{\mathbf{g}}$ on $\hat{L}(\mathbf{g}, 0)^{\perp}$.

Actually, this situation is simpler than the one encountered in (1.147). For, let v be an entire nonconstant function on $(S^{n-1})^{\mathbb{C}}$ which is invariant under $\hat{x} \to -\hat{x}$. Extend v along the real rays through S^{n-1} to be constant. Then the resulting function $\hat{h}$ restricts to an entire function (even a constant) on each $\hat{L}(0, \mathbf{g})^{\perp}$, and all s derivatives are entire on the cross-section $(S^{n-1})^{\mathbb{C}}$ but $\hat{h}$ is not defined at the origin unless $v = \text{const}$.

With a slight modification we could find an $\hat{h}$ which is defined and has N continuous derivatives at the origin, for any $N > 0$. We cannot make $\hat{h} \in C^{\infty}$ because that would force $\hat{h}$ to be entire by the Korevaar–Wiegerinck theorem.

These examples show that *regularity* on a cross-section does not suffice; what is needed is *equality* on enough intersections.

As the situation seems confusing, let us attempt to clarify it. We restrict our consideration to $n = 2$.

Remark. *The elements of the dual spaces $\hat{\mathcal{E}}$, $\hat{\mathcal{D}}'$, etc., are analytic functionals and, as such, can be represented in many ways by functions. We are going to clarify the above ideas using the heuristic of representing analytic functionals as functions and, in fact, representing the Fourier transform of holomorphic functions by functions which decrease exponentially in appropriate regions. The justification of these heuristics can be given using the ideas developed in Chapters 5 and 9.*[7]

One significant difference between the geometries related to d and $\bar{\partial}$ is that the $\hat{y}$ axis is a CS for $\partial/\partial\hat{x}$ ($\hat{x}$ a real variable) in the usual sense. If we now consider $\hat{x} = \hat{u} + i\hat{v}$ and $\hat{y} = \hat{s} + i\hat{t}$ as complex variables then it is a CS for $\partial/\partial\bar{\hat{x}}$ only in the exotic sense. This means that we need infinitely many CD, namely $\{\partial^j \hat{h}(0, \hat{y})/\partial\hat{x}^j\}$ for all j. Moreover there is a growth condition in j associated

[7]The reader not interested in dealing with these heuristics can skip to the paragraph before (1.153), beginning "The results . . .".

with the CD. Similarly, in the case of lines through the origin there is an exotic CP for $\bar\partial$ with CS the origin and CD given by $\{\partial^{j_1+j_2}\hat h(0,0)/\partial\hat x^{j_1}\partial\hat y^{j_2}\}$ with obvious growth conditions.

In the Korevaar–Wiegerinck (KW) theorem we are given functions $\hat h_{\mathbf g}$ on each line $\hat L(0,\mathbf g)$ which are holomorphic on $\hat L(0,\mathbf g)$ and satisfy COMP ORIGIN. Thus we have estimates for $\partial^j\hat h_{\mathbf g}(0)/\partial\hat s^j$ in all directions. They can be pieced together using a quantitative variant of Lemma 1.1 to give estimates for all mixed derivatives $\partial^{j_1+j_2}\hat h(0,0)/\partial\hat x^{j_1}\partial\hat y^{j_2}$.

But in the case dealt with in (1.146)f., for the exotic CP with data on the $\hat y$ axis the only estimates we know involve

$$\partial^{j_1+j_2}\hat h(0,\hat y)/\partial\hat x^{j_1}\partial\hat y^{j_2}$$

for fixed j_1 and varying j_2 (holomorphicity on the $\hat y$ axis) or for $j_2=0$ and varying j_1 (holomorphicity in $\hat x$ for any fixed $\hat y$). This is not enough to enable us to apply the method of Lemma 1.1 to give the suitable growth of $\{\partial^{j_1+j_2}\hat h(0,\hat y)/\partial\hat x^{j_1}\partial\hat y^{j_2}\}$ for varying j_1, j_2.

Let us now pass to the above examples involving the sphere S^{n-1}. It would appear, at least locally, that the CP structure with CD on S^{n-1} should be similar to (1.146). In fact, this is true when we use normals to S^{n-1}, i.e. radial lines. But tangent planes behave differently. The essential difference is that normals intersect at the origin, which is far from S^{n-1}, while neighboring tangent planes intersect close to S^{n-1}. Suppose for simplicity that $n=2$. Then two close tangent lines L_1, L_2 intersect very close to S^1. Since L_1 and L_2 have different directions we can use a subtle quantitative form of Lemma 1.1 to estimate the mixed derivatives $\partial^2/\partial\hat x\partial\hat y$ in terms of second derivatives on L_1 and L_2 separately. Using many close complex lines allows for estimates on $\partial^{j_1+j_2}/\partial\hat x^{j_1}\partial\hat y^{j_2}$. The estimates allow us to extend from S^{n-1}. We then need a special argument to show we can extend to the origin. This argument is quite complicated and will not be pursued further.

For the Hartogs–Bernstein theorem we use the fact that $\mathbb R^m$ is a CS of the usual kind for the Cauchy–Riemann system. The separate analyticity is a regularity condition. We shall discuss this in more detail presently.

Let us examine these questions in x space. For the situation related to (1.146) we start with $\hat h$ holomorphic on lines parallel to the $\hat x$ axis; that is, $\hat h\delta_{\hat y=\hat y^0}$ is holomorphic in $\hat x$ for all $\hat y^0$. This means that for all $\hat y$,

$$\text{support } [h * e^{iy\hat y}\delta_y] \subset \{u+iv=0\}. \tag{1.148}$$

(Recall that δ_y is the δ function of the complex y axis; a heuristic form of the projection–slice theorem suggests that $\delta_y = \hat{\delta}_{\hat x}$. x and y are regarded as complex variables $x=u+iv, y=s+it$.) Since the linear combinations of functions $\{\exp(iy\hat y)\}_{\hat y}$ are dense in most spaces the linear combinations of $\{\exp^{(iy\hat y)}\delta_y\}_{\hat y}$ are dense in all functions on the y axis. In particular we can approximate $\delta_{x=0}\times\delta_{y=y^0}$

for any y^0. Expression (1.148) gives

$$\text{support } h \subset \{u + iv = 0\}. \tag{1.149}$$

For the behavior in $s + it$ we want to use the holomorphicity of $\{\partial^j \hat{h}(0, \hat{y})/\partial \hat{x}^j = \hat{h}\partial^j/\partial \hat{x}^j \delta_{\hat{y}}\}$. The Fourier transform of this is

$$\text{support } [h * x^j \delta_x] \subset \{s + it = 0\}. \tag{1.150}$$

We should like to conclude that

$$\text{support } h \subset \{s + it = 0\} \tag{1.151}$$

which together with (1.149) means that support $h \subset V_{CR}$.

The passage from (1.150) to (1.151) is very puzzling. For each fixed y^0 (1.149) asserts that support of $h(x, y^0)$ is contained in $\{u + iv = 0\}$. If, for example, $\hat{h} \in \mathcal{E}$ then, from the AU structure of $\mathcal{E}$ (Section 1.4), this implies an exponential decrease of $h(x, y^0)$ in x. In general polynomials are dense in the duals of spaces of functions which decrease exponentially (see Chapter VIII of FA). This should mean that the polynomials in x are dense in the dual of the space of $h(x, y^0)$ so (1.150) should imply (1.151).

In $(\hat{x}, \hat{y})$ the argument would go roughly as follows. Since $\hat{h}(\hat{x}, \hat{y})$ is holomorphic in $\hat{x}$ for each $\hat{y}$ we could expand

$$\hat{h}(\hat{x}, \hat{y}) = \sum \frac{\partial^j}{\partial \hat{x}^j} \hat{h}(0, \hat{y}) \frac{\hat{x}^j}{j!}. \tag{1.152}$$

Each of the functions $\partial^j \hat{h}(0, \hat{y})/\partial \hat{x}^j$ is holomorphic in $\hat{y}$ so we should be able to apply $\partial/\partial \bar{\hat{y}}$ to (1.152) and conclude that $\hat{h}$ is holomorphic in $\hat{y}$ for all $\hat{x}$.

The problem with applying $\partial/\partial \bar{\hat{y}}$ to (1.152) is that (1.152) is only valid for real $\hat{y}$. If we knew that the series in (1.152) converged uniformly on compact sets of the complex $\hat{y}$ plane then certainly $\bar{\partial} \hat{h} = 0$.

It is more difficult to analyze our arguments in (x, y) space. The main point is that h is not a function but an analytic functional; that is, an element of the dual of the space of holomorphic functions (see Section 1.4 for details). As such the values $h(x, y)$ are not completely meaningful.

It may be possible to use our results (see Chapter 9) describing the lack of uniqueness in h to make these ideas precise. We have not explored this possibility.

There is a connection between these ideas and the notion of *analytic wave front set.*[8] From a formal point of view the ray $\rho^+(\mathbf{g}^0)$, which is the set of points in the real line $\mathbf{g}^{0\perp}$ with $s = s_{\mathbf{g}^0} > 0$, is not in the analytic wave front set of $\hat{h}$ if $\mathbf{R}\hat{h}(\hat{s}, \mathbf{g})$ is holomorphic in $\Im \hat{s}_{\mathbf{g}} > 0$ for $\{\mathbf{g}\}$ in a neighborhood of $\mathbf{g}^0$. Equivalently, by the projection–slice theorem,

$$h(s, \mathbf{g}) = \widehat{\mathbf{R}\hat{h}}(s, \mathbf{g}) = \mathcal{O}(e^{-\alpha s}) \quad s \to +\infty.$$

[8] A detailed study of wave front sets is contained in Chapters 5 and 9.

In the definition of wave front set, and in the notion of Radon transform in general, we should think of $\mathbf{g}$ as a *covector* as it is related to hyperplanes. In contrast to this, the Hartogs–Bernstein theorem deals with *directions* rather than codirections. The hypothesis in the Hartogs–Bernstein theorem is that $\hat{h}$ is holomorphic in the $\hat{x}_j$ directions (actually, in an affine sense, since we require analyticity on lines parallel to the coordinate axes). Formally it means that

$$h(x) = O\left(e^{-\alpha |x_j|}\right) \quad \text{uniformly in } (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n).$$

Thus the wave front set corresponds to exponential decrease of h on rays through the origin while the hypothesis of Hartogs–Bernstein corresponds to exponential decrease in "large" cones; the cones are almost half spaces.

The results of Korevaar–Wiegerinck and Hartogs–Bernstein can be put in a general differential equations framework. Let us first give a convenient restatement of their results. (We shall work with viewpoint (2).) We call M_{KW} the union of real complex lines through the origin (complexifications of real lines) and M_{HB} the union of the complex lines $L(x_1^0, \ldots, x_{j-1}^0, x_{j+1}^0, \ldots, x_n^0)$ (see (1.144)). Suppose we start with potential CD for $\bar{\partial}$ which is a formal power series $\hat{f}$ at the origin (resp. a function $\hat{f}$ on $\mathbb{R}^n$) which has a holomorphic extension $\tilde{f}$ to M_{KW} (or to M_{HB}). This means that $\tilde{f}$ is holomorphic on every complex line in M_{KW} (or M_{HB}). We conclude that $\tilde{f}$ extends to an entire function F, i.e. the potential CD is actual CD.

From the point of view of the general theory of partial differential equations $\mathbb{R}^n$ is a CS for the Cauchy–Riemann equations. The above results assert that if we can extend $\hat{f}$ which is defined on this CS to a function $\tilde{f}$ on a suitable $M \supset \mathbb{R}^n$ with $\dim M = n+1$ in such a way that $\tilde{f}$ satisfies the induced Cauchy–Riemann equations on M, then $\hat{f}$ actually is the CD of a solution F of the Cauchy–Riemann equations on all of $\mathbb{R}^{2n}$. (For the Hartogs–Bernstein theorem $M = M_{HB}$. The induced Cauchy–Riemann equation means holomorphicity on each complex line parallel to a coordinate axis.)

The general type of result this suggests is of the following structure. We start with a system of partial differential operators

$$\overrightarrow{P}(x, D) = (P_1(x, D), \ldots, P_r(x, D)). \tag{1.153}$$

We write $\overrightarrow{P}\hat{F} = 0$ if each $P_j\hat{F} = 0$. Let M^0 be a uniqueness CS for $\overrightarrow{P}$, meaning that there are linear differential operators $h_1(x, D), \ldots, h_N(x, D)$ such that the map

$$C' : \hat{F} \to (h_1\hat{F}|_{M^0}, \ldots, h_N\hat{F}|_{M^0}) \tag{1.154}$$

is injective on the kernel of $\overrightarrow{P}$ in some suitable function space. (Thus M^0 is a multiplicity slice.)

Now, we start with a function $\vec{\hat{f}}$ (potential CD) with N components on M^0. We want to know if $\vec{\hat{f}}$ is in the image of C' on some space of functions on all of

$\mathbb{R}^n$ (or the analogous local question). Note that

$$M_{HB} = (\mathbb{R}^m + \hat{T}_1) \cup (\mathbb{R}^m + \hat{T}_2) \cup \ldots \cup (\mathbb{R}^m + \hat{T}_m) \tag{1.155}$$

where $\hat{T}_j$ is the $\hat{x}_{m+j}$ axis. The general result of this structure is: we can find k lines $\hat{L}_1, \ldots, \hat{L}_k$ so that if $\vec{\hat{f}}$ extends to (is a CD of) a solution on

$$M_{\{\hat{L}_j\}} = (M^0 + \hat{L}_1) \cup \ldots \cup (M^0 + \hat{L}_k), \tag{1.156}$$

meaning to a solution of whatever equations $\overrightarrow{P}$ induces on each $M^0 + \hat{L}_j$, then $\vec{\hat{f}}$ extends to a solution on all of $\mathbb{R}^n$.

These ideas are studied in detail in Chapters 5 and 9.

We mentioned the results of [4] regarding the replacement of the origin in the KW theorem by the unit sphere S in $\mathbb{R}^m$. From viewpoint (2) the problem can be formulated as follows. We are given a function $\hat{f}$ which is potential CD for $\bar{\partial}$; but $\hat{f}$ is defined only on the exterior of S. The result is that if $\hat{f}$ extends to an entire function on the complexification of each real tangent plane to S then $\hat{f}$ extends to an entire function.

A tangent plane is a tangent sphere of radius ∞.

Problem 1.2 *What happens for tangent spheres of finite radius?*

Although we cannot answer this question there is a variation which has been studied under the name "strip problem" and is dealt with in detail in Section 9.5. S is now the unit circle in the complex plane so the unit exterior tangent spheres (circles $L(\hat{t})$) cover an annulus U. $\hat{f}$ is defined on U and its restriction to each $L(\hat{t})$ extends to a holomorphic function $F_{\hat{t}}$ in the interior of $L(\hat{t})$. Is $\hat{f}$ holomorphic in U?

In the literature U is replaced by the strip $|\Im \hat{x}| \leq 1$ and $\{L(\hat{t})\}$ become the real circles radius 1 centered on the real axis. Variations of this problem for operators other than $\partial/\partial\bar{z}$ are studied in Section 9.5.

1.6 Lie groups

There is a natural relation between Lie groups and the usual nonparametric Radon transform $\mathbf{R}$. Let G be the affine group of $\mathbb{R}^n$, meaning the group generated by rotations and translations. G acts on $\mathbb{R}^n$ in the usual way. The isotropy group of the origin is K = rotations around the origin. Thus $\mathbb{R}^n$ can be identified with G/K.

Let H be the subgroup of G consisting of translations by a fixed $n-1$ parameter subgroup. We want to examine the right action of H on G/K. This means that a function f on G/K gets mapped into the right H invariant function

$$\mathbf{R}_H f(g) = \int f(gh)\, dh \tag{1.157}$$

where dh is a suitably normalized Haar measure on H.

(1.157) can be expressed in integral-geometric terms. Since f is a function on $G/K = \mathbb{R}^n$, $f(gh)$ is the value of f at $gh \cdot \mathcal{O}$. ($\mathcal{O} = p_K$ is the K fixed point on G/K.) This means that we translate the origin by h and then apply g. Thus the integral in (1.157) is the integral of f over $gH \cdot \mathcal{O}$ which is the affine hyperplane obtained by applying g to $H \cdot \mathcal{O}$, that is, rotating $H \cdot \mathcal{O}$, and then translating the result.

We can reverse the construction: start with a function u on G/H and form

$$\mathbf{R}_K u(g) = \int u(gk)\, dk. \tag{1.158}$$

The geometry related to G/H is somewhat more complicated than that of G/K. However, we could replace H by $\tilde{H}$ which is the subgroup of G leaving $H \cdot \mathcal{O}$ fixed without making any essential changes. It is clear that

$$\tilde{H} = G(H) \times K(H^\perp)$$

where $G(H)$ is the affine group of H and $K(H^\perp)$ is the orthogonal group of $H^\perp$. Thus $\tilde{H}$ differs from H by a "compact factor." By (1.157) $\mathbf{R}_H f$ is invariant under right action by $\tilde{H}$.

$\tilde{H}$ has the advantage that $G/\tilde{H}$ is the space of hyperplanes while G/H has some additional structure. For example, we could prescribe orthonormal frames on H and on $H^\perp$ (when $\dim H^\perp = 1$ a frame on $H^\perp$ is an orientation of H). G acts simply transitively on the space of H with these framed structures.

As mentioned, the functions we deal with do not depend on the frames.

We are thus led to the notion of double fibration, pursued in great detail by Helgason and by Gelfand and his school [94,77]. Let G be a connected Lie group and let K, H be closed subgroups; call $L = K \cap H$. The double fibration takes the form shown in Figure 1.3.

We go from G/K to G/H by right integration over H/L and vice versa. Such right integration is the prototype of

Intertwining: linear transformations of functions which commute with (left) G action. We denote these transformations respectively by $\mathbf{R}_H, \mathbf{R}_K$.

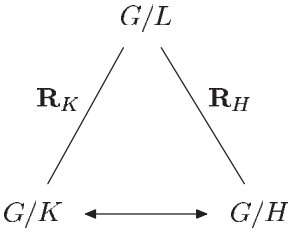

Figure 1.3

We can think of the intertwinings in terms of *incidence relations*. In terms of the above example of the affine group, $\mathbf{R}_K u(g)$ is the integral of u over the set of all hyperplanes incident to the point $g \cdot K \cdot p_H$ while $\mathbf{R}_H f(g)$ is the integral of f over all points incident to the hyperplane $g \cdot (H \cdot \mathcal{O})$ (see (1.157) and (1.158)). (p_H is the H invariant point on G/H.) We normalize p_H to contain the origin so $k \cdot p_H$ is the set of all hyperplanes incident to the origin and $g \cdot k \cdot p_H$ is the set of hyperplanes incident to $g \cdot \mathcal{O}$.

From the perspective of group theory we can express $\mathbf{R}_H$ and $\mathbf{R}_K$ as group convolution. Thus, for any closed subgroup B of G we can write (assuming G is unimodular)

$$\mathbf{R}_B f = f * \delta_B \tag{1.159}$$

where the convolution of two functions is

$$(f_1 * f_2)(g) = \int f_1(g g_1^{-1}) f_2(g_1)\, dg_1. \tag{1.160}$$

Convolution equations on G are best analyzed using group representations. Much progress has been made recently on the study of group representations. We shall, however, restrict ourselves to situations where representation theory is understood and not too complicated. The simplest examples are for G abelian or compact. Another class of groups where representation theory is fairly well understood is real semi-simple Lie groups.

We shall study representations ρ of G on a complex Hilbert space $\mathcal{H}$. Given $\alpha, \beta \in \mathcal{H}$ we can form the matrix coefficient

$$u_{\alpha\beta}(g) = (\rho(g)\alpha, \beta). \tag{1.161}$$

The closure (in a suitable topology) of the linear span of $\{u_{\alpha\beta}\}$ is called the space of representation functions for ρ.

Let B be a closed subgroup of G. If there exists a vector $\alpha^0 \in \mathcal{H}$ which is B invariant (i.e. $\rho(b)$ invariant for all $b \in B$) then

$$\begin{aligned} u_{\alpha^0\beta}(gb) &= (\rho(gb)\alpha^0, \beta) \\ &= (\rho(g)\rho(b)\alpha^0, \beta) \\ &= u_{\alpha^0\beta}(g) \end{aligned}$$

for all β. The set of representation functions which are right B invariant forms a representation space for G. (G acts on such functions by left multiplication.)

Under conditions on G which apply to essentially all the groups we meet in this work there is a "good" Fourier analysis. In the language of the direct integral of Hilbert space (Section 1.1) this is formulated as

$$L_2(G) = \int_{\oplus} \mathcal{H}_\rho \, d\mu(\rho). \tag{1.162}$$

We regard the Hilbert space $L_2(G)$ as a representation space (left regular representation) for G. $\{\rho\}$ are irreducible unitary representations and $\mathcal{H}_\rho$ is an irreducible space of representation functions for ρ; roughly speaking, each ρ appears as many times as $\dim \mathcal{H}_\rho$. $d\mu(\rho)$ is called the *Plancherel measure* for G.

The direct integral provides the representation functions as a base for $L_2(G)$ in the sense of (1.17)ff. The representation functions may not belong to $L_2(G)$ if G is not compact but "wave packets" belong to $L_2(G)$.

The simplest examples of the direct integral decomposition of $L_2(G)$ occur when G is abelian or G is compact. For example, if $G = \mathbb{R}^n$ then $\mathcal{H}_\rho$ is one dimensional and is spanned by $\exp(i\rho \cdot x)$ where $\rho \in \mathbb{R}^n$. Of course $\exp(i\rho \cdot x) \notin L_2(\mathbb{R}^n)$ so one needs to form wave packets to obtain functions in $L_2(\mathbb{R}^n)$.

We have explained in Section 1.1 that there is an analogous theory for AU spaces; in this case we usually need nonunitary representations.

Under usual conditions the representation of G on $L_2(G/B)$ decomposes in a direct integral where the $\mathcal{H}_\rho$ consist of right B invariant representation functions. From a "generic" point of view each such $\mathcal{H}_\rho$ contains a B invariant vector α_ρ and $\mathcal{H}_\rho$ is spanned by $\{\rho(g)\alpha_\rho\}$. This is a form of the Frobenius reciprocity theorem since the representation of G on $L_2(G/B)$ is the *induced representation* on G from the trivial representation of B; the Frobenius reciprocity theorem is discussed in detail in several places in this work.

Observe that $L_2(G/B)$ is defined as the subspace of functions $f \in L_2(G)$ which satisfy, for every b,

$$f(g) = f(gb) = (f * \delta_{b^{-1}})(g), \tag{1.163}$$

i.e.

$$f * (\delta_1 - \delta_{b^{-1}}) = 0.$$

We have used the Frobenius reciprocity theorem to say that the solutions of this system of right convolution equations on G can be expressed in terms of solutions which are representation functions for irreducible representations.

How general is this result?

Before discussing this question let us reformulate the condition for the existence of representation functions which are solutions of a system of convolution equations $\{f * S_j = 0\}$.

If ρ is a representation of G on a complex Hilbert space $\mathcal{H}$ then ρ acts on functions or distributions $f(g)$ by

$$\rho(f) = \hat{f}(\rho) = \int f(g)\rho(g)\, dg. \tag{1.164}$$

$\rho(f)$ is an operator on $\mathcal{H}$. From the historical point of view (going back to finite groups) we can interpret $\rho(f)$ as follows. Think of f as expanded in terms of the basis $\{\delta_g\}$ as in (1.17)ff. Thus

$$f = \int f(g)\delta_g\, dg.$$

We define

$$\rho(\delta_g) = \rho(g) \tag{1.165}$$

and extend the definition by linearity.

The relation of representations to convolution equations stems from the formal property

$$\rho(u * v) = \hat{u}(\rho)\hat{v}(\rho) \tag{1.166}$$

for suitable functions or distributions. Thus the equations $f * S_j = 0$ become $\hat{f}(\rho)\hat{S}_j(\rho) = 0$.

Remark. *This argument assumes that $\hat{f}(\rho)$ and $\hat{S}_j(\rho)$ can be defined by (1.164) so they have to be small at infinity. In practice the $\hat{S}_j$ are small at infinity but $\hat{f}$ cannot be defined by (1.164). In this case some of what follows (up to the statement of the Tauberian principle) is heuristic.*

There is some slight complication because of our notation and the noncommutativity of G. Our notation for operator action is to put the operator on the left. This means that $\alpha \in \mathcal{H}$ is regarded as a column vector and $(\alpha, \beta) = \alpha^*\beta$ where $*$ denotes adjoint. It is sometimes convenient to replace $\hat{f}(\rho)\hat{S}_j(\rho) = 0$ by $\hat{S}_j^*(\rho)\hat{f}^*(\rho) = 0$.

We call $\{S_j\}$ *rich at* ρ if $\hat{S}_j^*(\rho)\alpha = 0$ for all j implies $\alpha = 0$. When this is the case $\hat{S}_j^*(\rho)\hat{f}^*(\rho) = 0$ implies $\hat{S}_j^*(\rho)(\hat{f}^*(\rho)\alpha) = 0$ for all j, α so $\hat{f}^*(\rho)\alpha = 0$ for all α, which means $\hat{f}(\rho) = 0$. If this happens for all irreducible ρ then $f = 0$ by the uniqueness of the Fourier transform.

Conversely if $\{S_j\}$ is not rich at ρ then there is an $\alpha \in \mathcal{H}$ with $\hat{S}_j^*(\rho)\alpha = 0$ for all j. We then form

$$u(g) = (\rho^{\#}(g)\alpha, \beta)$$

where $\rho^{\#}$ is the representation $\rho^{\#}(g) = \rho^*(g^{-1})$. Since $\rho(g)$ is invertible there is a β such that $u(g) \not\equiv 0$.

By definition $u(g)$ is a representation function. Moreover

$$\begin{aligned}(u * S_j)(g) &= \int (\rho^{\#}(gg_1^{-1})\bar{S}_j(g_1)\alpha, \beta)\, dg_1 \\ &= \int (\rho^{\#}(g_1^{-1})\bar{S}_j(g_1)\alpha, \rho(g^{-1})\beta)\, dg_1 \\ &= (\hat{S}_j^*(\rho)\alpha, \rho(g^{-1})\beta) \\ &= 0. \end{aligned} \tag{1.167}$$

We have shown that if $\{S_j\}$ is not rich at ρ then there exists a representation function u which is in the simultaneous kernel of the S_j. Conversely if $\{S_j\}$ is rich at every ρ then there are no solutions of $f * S_j = 0$ for all j. Thus richness is equivalent to the nonexistence of solutions f on G of $\{f * S_j\} = 0$.

It is important to have an analogous result for functions on G/B. We could add the convolution equation $\{f * (I - \delta_b)\} = 0$ (see (1.163)f.) but it is more convenient to work directly.

Note that if f is B invariant then

$$\hat{f}(\rho) = \int f(gb)\rho(g)\,dg = \hat{f}(\rho)\rho(b^{-1}). \tag{1.168}$$

This means that $\hat{f}(\rho)$ annihilates

$$\mathcal{H}_{\rho B} = \text{closure} \bigcup_b [\rho(b) - I]\mathcal{H}. \tag{1.169}$$

We only need to deal with α which are orthogonal to $\mathcal{H}_{\rho B}$. Such α are $\rho^*(B)$ invariant because

$$((\rho^*(b) - I)\alpha, \beta) = (\alpha, (\rho(b) - I)\beta) = 0.$$

Suppose f is B invariant and $f * S_j = 0$ for all j. Then, as before, $\hat{S}_j^*(\rho)\hat{f}^*(\rho) = 0$. We say $\{S_j\}$ is ρB *rich* if $\alpha \in \mathcal{H}_{\rho B}^{\perp}$ and $\hat{S}_j^*(\rho)\alpha = 0$ for all j implies $\alpha = 0$.

We claim that if f is right B invariant and $\{S_j\}$ is ρB rich and if $\hat{S}_j^*(\rho)\hat{f}^*(\rho) = 0$ for all j then $\hat{f}(\rho) = 0$. For we know that $\hat{S}_j^*(\rho)\hat{f}^*(\rho)\beta = 0$ for all j, β. But $\hat{f}^*(\rho)\beta \in \mathcal{H}_{\rho B}^{\perp}$ because for $\gamma \in \mathcal{H}_{\rho B}$ we have by (1.168) and (1.169)

$$(\hat{f}^*(\rho)\beta, \gamma) = (\beta, \hat{f}(\rho)\gamma) = 0.$$

Since $\hat{S}_j^*(\rho)(\hat{f}^*(\rho)\beta) = 0$ for all j, β it follows that $\hat{f}^*(\rho)\beta = 0$ for all β, which means that $\hat{f}^*(\rho) = 0 = f(\rho)$.

If, on the other hand, $\{S_j\}$ is not ρB rich then there is an $\alpha \in \mathcal{H}_{\rho B}^{\perp}$ for which $\hat{S}_j^*\alpha = 0$ for all j. As in the case $B = \{1\}$ this implies that

$$u(g) = (\rho^{\#}(g)\alpha, \beta)$$

is a nontrivial representation function which is a solution of $u * S_j = 0$ for all j. We have seen that $\alpha \in \mathcal{H}_{\rho B}^{\perp}$ is $\rho^*(B)$ invariant so u is right B invariant.

We explained that the Frobenius reciprocity theorem can be regarded as a description of the solutions of the convolution system $\{f * (\delta_1 - \delta_b) = 0\}_{b \in B}$ in terms of representation function solutions. We might expect some analog for more general convolution systems. The above discussion leads us to formulate the

Tauberian principle. *There is no right B invariant solution of the system of convolution equations $\{f * S_j = 0\}$ if $\{S_j\}$ is ρB rich for every irreducible representation ρ.*

Remark. *The definition of ρB rich indicates that we have to examine only those ρ which have B invariant vectors.*

In the simplest case $B = \{1\}$ and G is abelian. Then the Tauberian principle asserts that there is no nontrivial solution of $f * S_j = 0$ if the $\hat{S}_j$ have no common zero. This is the usual form of Wiener's tauberian theorem if $f \in L_1$.

We should emphasize that the Tauberian principle is a principle whose truth depends on G, B, and the space of f under consideration. It is valid for many groups and many spaces but it does not hold universally. As we mentioned, for the space $L_1(\mathbb{R}^n)$ its veracity is Wiener's Tauberian theorem. But it fails for $L_1(G)$ when, for example, $G = SL(2, \mathbb{R})$ is the group of 2×2 real matrices of determinant 1 as was shown by Ehrenpreis and Mautner [68,69].

The Tauberian principle tells us when a set of homogeneous convolution equations has a nontrivial solution. A secondary question is the determination of all solutions. This is

Spectral synthesis. *The right B invariant solutions of the system $\{f * S_j = 0\}$ can be represented as integrals of right B invariant representation function solutions.*

Spectral synthesis as well as the FP are fine grain decompositions of spaces of solutions of systems of equations. Sometimes there are related coarse (more properly "semi-coarse") grain decompositions. They occur naturally when $\{f * S_j = 0\}$ is an enveloping system (Sections 1.2 and 1.4) of systems $\{f * S_j^k = 0\}_k$. Then solutions of $\{f * S_j = 0\}$ can, under suitable circumstances, be written as sums or integrals over k of solutions of $\{f * S_j^k = 0\}$. (Each k corresponds to a coarse grain.) This coarse grain description is in the spirit of the Radon ansatz as we see from the Whittaker and Bateman representations of harmonic functions (Section 1.4).

Spectral synthesis is a group version of the FP. It is readily seen to be valid for compact groups. It is a step beyond the Tauberian principle and it does not even hold for $L_1(\mathbb{R}^n)$ unless the S_j are restricted.

Let A and B be two closed subgroups of G. We wish to study the Radon transform (intertwining) of functions on G/A with functions on G/B, given by the Radon transform as in (1.159), by means of group representations. In particular we want to examine how the Tauberian principle impinges on the injectivity of this intertwining.

According to the Tauberian principle, by (1.159) injectivity means that, for every irreducible representation ρ of G which contains an A invariant vector, the distribution δ_B is ρA rich. To verify this in any case requires a study of the representation theory of G.

For the original Radon transform $A = K, B = H$, and G is the affine group of $\mathbb{R}^n$. Since we start with functions on $\mathbb{R}^n = G/K$ we only examine representations having K invariant vectors. By the Tauberian principle (in this case Frobenius reciprocity) these occur in the decomposition of the representation of G on $L^2(G/K)$.

There are two general methods for the decomposition of representations:

(1) Find operators that commute with the action of G. Then diagonalize the algebra $\mathcal{A}$ of such operators meaning decompose the representation space into simultaneous eigenspaces (one dimensional representation spaces) for $\mathcal{A}$. (This works if $\mathcal{A}$ is abelian; otherwise we have to decompose the representation space into more general representation spaces of $\mathcal{A}$.)
(2) Start with an abelian subgroup E of G which is as "large as possible." Diagonalize E meaning decompose the representation space into eigenspaces of E. Then examine the action of G/E on these eigenspaces.

(1) For the affine group acting on $\mathbb{R}^n$ the only commuting operators are powers of the Laplacian Δ. For each complex number λ the representation ρ_λ is defined as the restriction of the action of G on the eigenspace

$$\mathcal{H}_\lambda = \{f \big| \Delta f = \lambda f\}. \tag{1.170}$$

Passing from $L_2(G/K)$ to $\mathcal{H}_\lambda$ does not appear to simplify the problem.

(2) We set $E = \mathbb{R}^n$. The eigenvectors of E are $\exp(i\hat{x} \cdot x)$ (as functions of x). If $\exp(i\hat{x} \cdot x)$ belongs to a representation space $\mathcal{H}$ for G then so does $\exp(i\hat{x} \cdot kx) = \exp(ik^*\hat{x} \cdot x)$ for all $k \in K$. Thus $\mathcal{H}$ contains $\{\exp(i\hat{x} \cdot x) \big| |\hat{x}|^2 = \lambda\}$ and ρ_λ is the representation of G on this space defined by the standard action of G on $\{x\} = \mathbb{R}^n$. The FP (Section 1.4) identifies $\mathcal{H}$ with $\mathcal{H}_\lambda$ defined by (1.170). The point of approach (2) is that it provides $\mathcal{H}_\lambda$ with the fine grain special basis $\{\exp(i\hat{x} \cdot x)\}$; this was not the case for approach (1).

It remains to compute $\rho_\lambda(\delta_H) = \hat{\delta}_H(\rho_\lambda)$. By definition (see (1.164))

$$\hat{\delta}_H(\rho_\lambda)e^{i\hat{x}\cdot x} = \int \delta_H(g)\rho_\lambda(g)e^{i\hat{x}\cdot x}\, dg = \int e^{i\hat{x}\cdot hx}\, dh. \tag{1.171}$$

The integral vanishes unless $\hat{x} \in H^\perp$ since for $\hat{x}, x$ fixed, we can regard $\exp(i\hat{x}\cdot hx)$ as an additive character of H which is trivial only when $\hat{x} \cdot H = 0$. Note that $hx = h + x$. (The integral over a group of any nontrivial character vanishes.)

We have to verify that $\hat{\delta}_H$ is $\rho_\lambda(K)$ rich. According to the definition this amounts to showing that if α is $\rho_\lambda^*(K)$ invariant and $\hat{\delta}^*(\rho_\lambda)\alpha = 0$ then $\alpha = 0$. Now α lies in the representation space for ρ_λ and is $\rho_\lambda^*(K)$ invariant. We have seen that the representation space is defined by an orbit of K, i.e. it is the linear span of $\{\exp(ik^*\hat{x} \cdot x)\}$. (There is some ambiguity here because this is only true if the functions are small at infinity; otherwise the complex group $K^{\mathbb{C}}$ may play a role. We shall not enter into this question.)

The $\rho_\lambda^*(K)$ invariant elements in this representation space are of the form

$$\alpha = c \int e^{ik^*\hat{x}\cdot x}\, dk. \tag{1.172}$$

From our above description of $\hat{\delta}_H(\rho_\lambda)$ we see that

$$\hat{\delta}_H(\rho_H) \cdot \alpha = ce^{ik^*(H)\hat{x}\cdot x}$$

where $k^*(H)\hat{x}$ is the point on $K^*\hat{x}$ which is orthogonal to $H \cdot \mathcal{O}$. Hence $\hat{\delta}_H(\rho_H) \cdot \alpha = 0$ implies $c = 0$, which means $\alpha = 0$.

This completes the proof of injectivity.

This proof of the injectivity shows how the Tauberian principle can be used. Actually the Tauberian principle is easier to apply when both subgroups of G are compact.

Problem 1.3 *Study the Tauberian principle for the intertwining from G/H to G/K.*

The Tauberian principle depends on analysis. We want to look at intertwining from a more geometric viewpoint. For this reason we analyze the possibility of introducing spreads into the double fibration formalism.

Let K, H be closed subgroups of the Lie group G. We study the intertwining from G/K to G/H. In conformity with our discussion of the relation of the Radon transform to hyperbolic equations one reasonable spread consists of the orbits of H on G/K (leaves). We postulate the existence of a subgroup S of G such that these orbits are $\{sHp_K\}_{s\in G}$ where p_K is the K fixed point (base point) on G/K. This suggests the decomposition

$$G = SHK. \tag{1.173}$$

The disjointness of the orbits means that we cannot have $s_1Hp_K \cap s_2Hp_K \neq \emptyset$ unless $s_1 = s_2$. Suppose

$$s_1h_1k_1 = s_2h_2$$

then

$$s_2^{-1}s_1h_1k_1 = h_2.$$

Thus our condition is

The H component in the decomposition $G = SHK$ is unique.

In the case of the usual Radon transform $S = H^\perp$ and the uniqueness is clear. But suppose we reverse things and examine the intertwining from G/H to G/K. Actually, as we mentioned above, it is more convenient to replace H by $\tilde{H}$ where $\tilde{H}$ is the stabilizer of $H \cdot p_K$ in G. As we have seen $G/\tilde{H}$ is the space of hyperplanes.

Let $p_H = H$ be the"base hyperplane." KH is the set of hyperplanes through the origin. It is not hard to see that there is no subgroup S of G for which $\{sKp_H\}$ is a spread.

We now present another example of intertwining which fits into the spread formalism. This example is relevant to several problems in this work.

Let $G = SL(2, \mathbb{R})$, the group of real 2×2 matrices of determinant 1, and let K be the rotation subgroup. The quotient space $Y = G/K$ can be identified with the Poincaré upper half-plane (Section 10.1).

The action of G is fractional linear transformation

$$gz = \frac{az+b}{cz+d} \qquad \text{for } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

and $p_K = i$ is the fixed point of K. For H we can take either

$$N = \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \right\} \tag{1.174a}$$

or

$$A = \left\{ \begin{pmatrix} \cosh\zeta & \sinh\zeta \\ \sinh\zeta & \cosh\zeta \end{pmatrix} \right\}. \tag{1.174b}$$

In either case we set

$$S = \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \right\}.$$

For both choices of H in (1.174) we can write G uniquely in the form[9]

$$G = SHK$$

so that the conditions required for a good spread theory are met.

To verify our statement about $G = SHK$, we must know that SH acts simply transitively on the Poincaré half-plane. The N orbit of the base point $p_K = i$ is the line $\Im z = i$ and the A orbit of i is readily seen to be the upper half of the unit circle. S acts by positive scalar multiplication so in either case the S orbit of Hi covers the upper half-plane simply.

For $H = N$ we obtain the *horocyclic Radon transform* as the orbits gNi are classically called horocycles. The primitive spread is

$$\begin{aligned} \{sni\}_n &= \left\{ \begin{pmatrix} a & an \\ 0 & a^{-1} \end{pmatrix} \cdot i \right\} \\ &= \{a^2 n + a^2 i\} \end{aligned} \tag{1.175}$$

which is the set of lines in the upper half-plane parallel to the real axis.

For $H = A$, which is the *geodesic Radon transform*, the primitive spread is

$$\begin{aligned} \{sAy_0\} &= \left\{ \begin{pmatrix} a\cosh\zeta & a\sinh\zeta \\ a^{-1}\sinh\zeta & a^{-1}\cosh\zeta \end{pmatrix} \cdot i \right\} \\ &= a^2 \left(\frac{i\cosh\zeta + \sinh\zeta}{i\sinh\zeta + \cosh\zeta} \right). \end{aligned} \tag{1.176}$$

[9]Technically, when thinking of G as the group of holomorphic transformations of the Poincaré half-plane, we usually replace G by $G/\pm I$. Then S is replaced by S^+ for which $a > 0$. We shall ignore the difference here as it creates only trivial discrepancies.

(The geodesics in hyperbolic geometry are semicircles centered on the real axis.) (1.176) is the set of semicircles in the upper half-plane with center 0 and radius a^2.

We have explained that the intertwining from $G/\tilde{H}$ to G/K (G = affine group, $\tilde{H}$ = stability group of $H \cdot \mathcal{O} = \mathbb{R}^{n-1}$ in G) cannot be formulated in terms of spreads. The existence of a spread formalism for intertwining from G/K to G/H depends on the decomposition

$$G = SHK. \tag{1.177}$$

The leaves of the primitive spread are the H orbits on G/K. It seems more natural to search for a subgroup S of G which is a cross-section of the H orbits in the sense that the orbits are $\{Hsp_K\}_s$. The orbits form the leaves of a spread because distinct orbits do not intersect. This leads us to a second relation between G/K and G/H which we term

Transformation:

$$T_s f(g) = \int f(ghs)\, dh. \tag{1.178}$$

T_s transforms functions on G/K into functions on G/H although, in general, it does not commute with G.

If we reverse our notation and replace $K \to H$ and $p_K \to p_{\tilde{H}}$ then $\{Ksp_{\tilde{H}}\}_s$ are the K orbits on $G/\tilde{H}$ which are the hyperplanes at distance $|s|$ from the origin. (S is $H^\perp$ as before.) $Ksp_{\tilde{H}}$ are the leaves of a spread on $G/\tilde{H}$. We have thus produced a formalism in which a passage from $G/\tilde{H}$ to G/K has a spread formalism.

We now examine T_s from the group viewpoint; we start with (1.178). The fact that $HS = G/K$ means that

$$G = HSK. \tag{1.179}$$

If S is not a cross-section for the H orbits then

$$h_1 s_1 p_K = h_2 s_2 p_K \tag{1.180}$$

with $s_1 \neq s_2$. Thus

$$h_1 s_1 k_1 = h_2 s_2 k_2. \tag{1.181}$$

The fact that this is impossible means that

The S component in the decomposition (1.179) is uniquely determined by $g \in G$.

This is one spread. How do we find others? Again we postulate that they can be described by a closed subgroup B of G. Precisely they are of the form

$$bHs \cdot p_K = bHb^{-1}bsb^{-1}b \cdot p_K. \tag{1.182}$$

In the nicest situation B is a subgroup of K. Then the leaves of this spread are the orbits of bHb^{-1} and the parameter space is bSb^{-1}. Clearly

$$G = (bHb^{-1})(bSb^{-1})K. \tag{1.183}$$

Moreover the bSb^{-1} component is unique for any b if it is true for $b = 1$ (or for some b^0) because then

$$bh_1b^{-1}bs_1b^{-1}k_1 = bh_2b^{-1}bs_2b^{-1}k_2 \tag{1.184}$$

implies $s_1 = s_2$ if $B \subset K$.

These spreads are parametrized by $B/(\text{normalizer } H) \cap B$.

In the case of the passage from $G/\tilde{H}$ to G/K we have defined one spread as the set of orbits (leaves) of K acting on $G/\tilde{H}$. The other spreads are the orbits of xKx^{-1} for $x \in \mathbb{R}^n$. These orbits consist of the hyperplanes which are at a fixed distance from x. The spread parameter $s \in S$ has dimension 1. The Grassmannian $= \{x\}$ has dimension n so we have overdetermined data.

Sometimes there is no group S which parametrizes the H orbits. In this case we may use a larger group to define transformations which have "spread-like" structure and then "cut down" the excess by means of compatibility equations.

As an example of this let G be the affine group of $\mathbb{R}^n$. We set $H = K$ and $S = \mathbb{R}^n$ in (1.178). Since $G = \mathbb{R}^n K$ we can restrict g in (1.178) to be in $\mathbb{R}^n$ (which now plays the role of B). We obtain the expression (f is a function on $\mathbb{R}^n = G/K$)

$$J_1 f(x^1, x^2) = \int f(x^1 + kx^2)\, dk. \tag{1.185}$$

Another choice for H is the multiplicative group of the line. Actually in this case we replace G by G_1 which is the semi-direct product of $\mathbb{R}^N$ with the group $\mathbb{R}^\times$ of scalar multiplication. In (1.178) we set

$$\begin{aligned} H &= \mathbb{R}^\times = \text{scalar group} \\ S &= \mathbb{R}^n. \end{aligned} \tag{1.186}$$

Again we need take g only from $\mathbb{R}^n$. Then T_s becomes the parametric Radon transform

$$(J^1 f)(x^1, x^2) = \int f(x^1 + tx^2)\, dt. \tag{1.187}$$

Note that dt is not the Haar measure on $\mathbb{R}^\times$.

Remark. *We used the measure dt in (1.187) to conform to the standard notation for the parametric Radon transform. In Chapter 6 we clarify the change from dt/t to dt.*

We have already analyzed the equations (John equations) satisfied by the parametric Radon transform so let us examine J_1 from this point of view.

Since the Laplacian Δ commutes with rotation it is clear that

$$(\Delta_{x^1} - \Delta_{x^2})\, J_1 f = 0. \tag{1.188}$$

$J_1 f(x^1, x^2)$ is invariant under rotation in x^2, that is, it is a function of $x^1, r^2 = |x^2|$. If we express Δ_{x^2} in polar coordinates we obtain from (1.188)

$$\left[\Delta_{x^1} - \frac{\partial^2}{\partial (r^2)^2} - \frac{(n-1)}{r^2}\frac{\partial}{\partial r^2}\right] J_1 f(x^1, x^2) = 0. \tag{1.189}$$

Equation (1.189) is known as the Darboux equation. We shall see in Chapter 6 that equation (1.189) characterizes those functions $F(x^1, x^2)$ which are of the form $J_1 f$. (Do not confuse the superscript 2 in r^2 with the square as in $(r^2)^2$.)

We have thus established a close relation of John's equations with the Darboux equation.

For another example of an overdetermined parametrization, this time in the form (1.177), we set $S = G$. A point x in the "spread" gHp_K has coordinates g, h if

$$x = ghp_K. \tag{1.190}$$

g and h are highly nonunique.

Instead of using all of G we could use some "large" subgroup or subset G^0 of G with the proviso that $G^0 HK = G$.

In case $G = A$ is the general affine group (semi-direct product of the general linear group with $\mathbb{R}^n$), we can take $G^0 \sim A/\tilde{H}$.

If we write matrices of A in the form

$$\begin{pmatrix} a_1^1 & a_1^2 & \dots & a_1^n & b_1 \\ a_2^1 & a_2^2 & \dots & a_2^n & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_n^1 & a_n^2 & \dots & a_n^n & b_n \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \tag{1.191}$$

then, when H is the subgroup of $\mathbb{R}^n$ spanned by the first l coordinate vectors, $\tilde{H}$ consists of matrices of the form

$$\tilde{H} = \left\{ \begin{pmatrix} 1 & \dots & 0 & a_1^{l+1} & \dots & a_1^n & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots \\ 0 & \dots & 1 & a_l^{l+1} & \dots & a_l^n & 0 \\ 0 & \dots & 0 & a_{l+1}^{l+1} & \dots & a_{l+1}^n & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & \dots & 0 & a_n^{l+1} & \dots & a_n^n & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & 1 \end{pmatrix} \right\}. \tag{1.192}$$

Thus the parameter γ on $A/\tilde{H}$ can be identified with

$$\gamma = \begin{pmatrix} a_1^1 & a_1^2 & \dots & a_1^l & 0 & \dots & 0 & b_1 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ a_n^1 & a_n^2 & \dots & a_n^l & 0 & \dots & 0 & b_n \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 \end{pmatrix}. \tag{1.193}$$

With this notation the "intertwining" which stems from this variation of (1.177) becomes the parametric Radon transform

$$\begin{aligned} F(\gamma) &= \int f(a^1 h_1 + a^2 h_2 + \ldots + a^l h_l + b)\, dh \\ &= J^l f(a^1, \ldots, a^l; b) \end{aligned} \tag{1.194}$$

where we have written

$$a^1 = (a_1^1, \ldots, a_n^1), \ldots, b = (b_1, \ldots, b_n) \tag{1.195}$$

(thought of as column vectors). We extend the above definition to all the vectors $a^1, \ldots, a^l$ even if they are not linearly independent (see Chapter 6).

We wish to study the parametric Radon transform from a somewhat different viewpoint. Let us start with the case $n = 3, l = 1$. Instead of defining G^0 as before it is more transparent to replace G^0 by the group

$$G_0 = \{\gamma\} = \left\{ \begin{pmatrix} a_1 & 0 & 0 & b_1 \\ 0 & a_2 & 0 & b_2 \\ 0 & 0 & a_3 & b_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right\}. \tag{1.196}$$

G_0 acts transitively on $\mathbb{R}^3$, so a function f on $\mathbb{R}^3$ can be regarded as a function on G_0 which is invariant under right action by a subgroup K which is the isotropy group of some fixed point in $\mathbb{R}^3$. In particular if that point is the origin then

$$K_0 = \left\{ \begin{pmatrix} a_1 & 0 & 0 & 0 \\ 0 & a_2 & 0 & 0 \\ 0 & 0 & a_3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right\}. \tag{1.197}$$

The isotropy group K_λ of a point of the form $\overrightarrow{\lambda} = (\lambda, \lambda, \lambda) \in \mathbb{R}^3$ is the set of $\overrightarrow{a'}, \overrightarrow{b'} \in G_0$ satisfying

$$\lambda \overrightarrow{a'} + \overrightarrow{b'} = \overrightarrow{\lambda}, \tag{1.198}$$

i.e.

$$\overrightarrow{b'} = \overrightarrow{\lambda} - \lambda \overrightarrow{a'}. \tag{1.199}$$

For later purposes it is convenient to think of $(\lambda, \lambda, \lambda)$ as the result of operating on $(1, 1, 1)$ by scalar multiplication by λ.

We now use the idea expressed in Section 1.4 that redundancy in the parameters leads to differential equations. For each λ the redundancy is given by K_λ. This means that if a function $F(\gamma)$ on G_0 is right K_λ invariant then $F[\gamma(\overrightarrow{a'}, \overrightarrow{b'})] = F(\gamma)$ for $\gamma = (\overrightarrow{a}, \overrightarrow{b})$ and $(\overrightarrow{a'}, \overrightarrow{b'}) \in K_\lambda$. Setting $\overrightarrow{a}' = (a', 1, 1)$ and letting $\overrightarrow{b}'$ correspond to $\overrightarrow{a}'$ by (1.199) yields

$$F(a_1 a', a_2, a_3; a_1\lambda(1 - a') + b_1, b_2, b_3) = F(\overrightarrow{a}, \overrightarrow{b}). \tag{1.200}$$

Differentiation with respect to a' gives

$$a_1 \frac{\partial F}{\partial a_1} - \lambda a_1 \frac{\partial F}{\partial b_1} = 0. \tag{1.201}$$

Similarly, setting $\overrightarrow{a}' = (1, a', 1)$ or $(1, 1, a')$ we find

$$\begin{aligned} a_2 \frac{\partial F}{\partial a_2} - \lambda a_2 \frac{\partial F}{\partial b_2} &= 0 \\ a_3 \frac{\partial F}{\partial a_3} - \lambda a_3 \frac{\partial F}{\partial b_3} &= 0. \end{aligned} \tag{1.202}$$

When $a_1 a_2 a_3 \neq 0$ (as is the case of the affine group G_0) we deduce

$$\frac{\partial F}{\partial a_j} - \lambda \frac{\partial F}{\partial b_j} = 0 \tag{1.203}$$

for all j and all λ. This means that the vectors $\partial F/\partial \overrightarrow{a}$ and $\partial F/\partial \overrightarrow{b}$ are parallel. The Fourier transform (see below) asserts that this parallelism is equivalent to John's equations (1.42). In fact, the parallelism is independent of λ. This is in conformity with our remarks following (1.43) asserting that we could replace the measure $d\lambda$ of (1.38) by any measure and John's first equations would still hold.

Let us examine the invariance properties of John's operators $\partial^2/\partial a_k \partial b_m - \partial^2/\partial a_m \partial b_k$ which are thought of as operators on the affine group G_0. From the group multiplication formulas it follows easily that $(\vec{a}, \vec{b}) \to (\vec{a}', \vec{b}')(\vec{a}, \vec{b})$ sends

$$\begin{aligned} \frac{\partial}{\partial a_j} &\to \frac{1}{a'_j} \frac{\partial}{\partial a_j} \\ \frac{\partial}{\partial b_j} &\to \frac{1}{a'_j} \frac{\partial}{\partial b_j}. \end{aligned} \tag{1.204}$$

Thus John's operators are multiplied by $(a'_k a'_m)^{-1}$. In the terminology of [58] such operators are G_0 *conformal*. In any case the kernels of John's operators are G_0 invariant and that is all that we care about.

Actually the operators $a_j\partial/\partial a_j$ and $a_j\partial/\partial b_j$ are (left) G_0 invariant since they represent differentiation with respect to right action by

$$\begin{pmatrix} 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \ddots & & & & & \vdots \\ \vdots & & 1 & & & & \vdots \\ \vdots & & & a' & & & \vdots \\ \vdots & & & & 1 & & \vdots \\ \vdots & & & & & \ddots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & \ddots & & & & & & \vdots \\ \vdots & & \ddots & & & & & 0 \\ \vdots & & & \ddots & & & & b' \\ \vdots & & & & \ddots & & & 0 \\ \vdots & & & & & \ddots & & \vdots \\ \vdots & & & & & & \ddots & 0 \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 & 1 \end{pmatrix}$$

(which commute with left G_0 action). This means that we can write John's equation (1.51) in the group invariant manner

$$a_k a_m \left(\frac{\partial^2}{\partial a_k \partial b_m} - \frac{\partial^2}{\partial a_m \partial b_k} \right) F = 0. \tag{1.205}$$

Remark. *It is curious that for parametric hypersurfaces the first John operator plays an essential role in Capelli's identity [159]. This seems to indicate a deep relation between the parametric hypersurface Radon transform and invariants.*

Our formulation of John's equations differs somewhat from the original formulation of John or from the consequent ideas of Helgason and of Gelfand and his collaborators. To explain the difference of the approaches, consider lines in the plane. The usual parametrization is

$$y = mx + b.$$

This parametrization does not capture the lines corresponding to infinite slope, i.e.

$$x = \text{const}.$$

For l planes in $\{(x,y)\} = \mathbb{R}^n$ the John–Gelfand parametrization is

$$y = m \cdot x + b$$

where $x \in \mathbb{R}^l$, $y \in \mathbb{R}^{n-l}$, m is an $n-l$ by l matrix, and b is a vector in $\mathbb{R}^{n-l}$. This should be contrasted with our parametrization as described above. Because the John–Gelfand parametrization is not so complete or redundant as ours, there is no second equation.

1.7 Fourier transform on varieties: The projection–slice theorem and the Poisson summation formula

Fourier transform is generally thought of as a map from functions $f(x)$ on $\mathbb{R}^n$ to functions $\hat{f}(\hat{x})$ on $\mathbb{R}^n$. When dealing with large functions as in the dual of AU spaces (Section 1.4) $\hat{x} \in \mathbb{R}^n$ gets replaced by $\hat{x} \in \mathbb{C}^n$. We now study what happens if we cut down x and $\hat{x}$ to slices which we denote by S, V respectively. We use the appellation "Fourier transform on varieties" because, both in this section and in other instances, S and V are algebraic varieties.

The FP (Section 1.4) shows that solutions of systems of partial differential equations $\vec{P}(D)f = 0$ are represented by Fourier integrals with $\hat{x} \in V$ where V is the variety of (complex) common zeros of the components $P_j(i\hat{x})$ of $\vec{P}(i\hat{x})$. (We shall ignore multiplicities; they present no serious difficulties.)

On the other hand if S is a parametrization surface for a WPPP (Section 1.4) then solutions of $\vec{P}(D)f = 0$ correspond to (vector) functions on S, namely to the PD of f. The correspondence between $\mathrm{PD}(f)$ and $\hat{f}$ (whose support is on V) is the *Fourier transform for the varieties (S,V).*

Let $\mathcal{W}$ be a locally convex, reflexive, topological vector space of functions of $x \in \mathbb{R}^n$ (thought of as functions which are small at infinity). We call $\mathcal{W}'(\vec{P})$ the kernel of $\vec{P}$ in $\mathcal{W}'$. By the FP $u \in \mathcal{W}'(\vec{P})$ has a Fourier representation (formally)

$$u(x) = \int \hat{u}(\hat{x}) e^{ix\cdot\hat{x}}\, d\hat{x}. \tag{1.206}$$

Since $\vec{P}(D)u = 0$, u cannot be small at infinity so $\hat{u}$ is generally a distribution. As such we cannot multiply two elements $\hat{h}, \hat{h}_1 \in \hat{\mathcal{W}}'(\vec{P})$. But if $f \in \mathcal{W}$ then f is small at infinity so $\hat{f}$ is smooth. Thus $\hat{u}\hat{f}$ is defined and lies in $\hat{\mathcal{W}}'(\vec{P})$. Put in other terms,

$$u * f = \widehat{\hat{u}\hat{f}} \in \mathcal{W}'(\vec{P}). \tag{1.207}$$

In this way the solution u defines a map from $\mathcal{W}$ into $\mathcal{W}'(\vec{P})$.

We now concentrate on the slice S. Since $u*f \in \mathcal{W}'(\vec{P})$ it is determined by its values and suitable "normal derivatives" $\{h_j(D)(u*f)\}$ on the Parametrization

Surface S. There are two bases for the PD:

(1) The δ functions of points $s \in S$ and their normal derivatives (geometric basis).
(2) The restrictions of $\{\exp(ix \cdot \hat{x})\}_{\hat{x} \in V}$ and normal derivatives to S (analytic basis).

We can think of the passage from the δ function basis to the $\{\exp(is \cdot \hat{x})\}_{\hat{x} \in V}$ basis as a Fourier transform from S to V. If S and V are linear then this Fourier transform is close to the usual Fourier transform on S, and V can be thought of as $\hat{S}$. If S and V are not linear then the meaning of $\hat{S}$ and of the explicit Fourier transform from $\hat{S}$ to V is more complicated.

In any case, we write symbolically

$$\delta_{s^0} = \int_V e^{is^0 \cdot \hat{x}} \widehat{\delta_{s^0}(S, V)}(\hat{x})\, d\hat{x}. \tag{1.208}$$

$\widehat{\delta_{s^0}(S, V)}$ has its support on V. We define $\widehat{\delta_{s^0}(S, V)}$ as the Fourier transform of δ_{s^0} from S to V. Equation (1.208) means that

$$\eta(s^0, 0; x) = \int e^{ix \cdot \hat{x}} \widehat{\delta_{s^0}(S, V)}(\hat{x})\, d\hat{x}$$

is the null solution of $\vec{P}$ with PD $= (\delta_{s^0}, 0, \ldots, 0)$ (see Section 1.1). If h_j is a normal derivative operator on S then there are similar formulas for the null solutions $\eta(s^0, j; x)$ for which

$$h_{j'} \eta(s^0, j; x) = \delta_{jj'} \delta_{s^0}.$$

We define the Radon transform $\mathbf{R}_{\vec{P}}$ by

$$\mathbf{R}_{\vec{P}} f(s^0, j) = \eta(s^0, j; x) \cdot f(x) \tag{1.209}$$

for $f \in \mathcal{W}$. This is a concrete representation of f thought of as an element of $\mathcal{W}''$ restricted to $\mathcal{W}'(\vec{P})$ via the basis $\{\eta(s^0, j; x)\}$. In accordance with Section 1.1 we set

$$\mathbf{R}^*_{\vec{P}} \mathbf{R}_{\vec{P}} f(x) = \sum_j \int \eta(s^0, j; x) \mathbf{R}_{\vec{P}} f(s^0, j)\, ds^0. \tag{1.210}$$

Here ds^0 is a measure on S used to identify functions on S such as $\mathbf{R}_{\vec{P}}(s^0, j)$ with dual elements (distributions). Thus $\mathbf{R}^*_{\vec{P}} \mathbf{R}_{\vec{P}} f$ is the solution of the PP whose PD is $\mathbf{R}_{\vec{P}} f$ (thought of as a distribution). For $f \in \mathcal{W}, u \in \mathcal{W}'(\vec{P})$ we define

$$\mathbf{R}_{u\vec{P}} f = \text{PD of } u * f \tag{1.211}$$

so that

$$\begin{aligned} \mathbf{R}^*_{\vec{P}} \mathbf{R}_{u\vec{P}} f &= u * f \\ &= \text{solution of PP whose PD is } \mathbf{R}_{u\vec{P}} f. \end{aligned} \tag{1.212}$$

Let us apply the Parseval formula to definition (1.209) of $\mathbf{R}_{\vec{P}}$. This gives

Theorem 1.3 *PST and PSF.*

$$\begin{aligned}\mathbf{R}_{\vec{P}}f(s^0, j) &= \int \hat{\eta}(s^0, j; \hat{x})\hat{f}(\hat{x})\, d\hat{x} \\ &= \int_V \hat{\eta}(s^0, j; \hat{x})\hat{f}(\hat{x})\, d\hat{x}. \end{aligned} \tag{1.213}$$

Thus $\mathbf{R}_{\vec{P}}f$ is expressed in terms of the restriction of $\hat{f}$ to V. We can think of $\mathbf{R}_{\vec{P}}f$ as the transform of $\hat{f}\big|_V$ from V to S using the kernel η. A similar formula is valid for $\mathbf{R}_{u\vec{P}}$.

To obtain explicit results from Theorem 1.3 we need to know explicit formulas for η and $\hat{\eta}$ or for u and $\hat{u}$. We now give some examples to show the far-reaching nature of Theorem 1.3.

Example 1 *The PST for the hyperplane Radon transform.* We start with the system $\vec{P} = (\partial/\partial x_1, \ldots, \partial/\partial x_{n-1})$ so $V = \{\hat{x}_n \text{ axis}\}$. We can choose $S = \{x_n \text{ axis}\}$. The Fourier transform from S to V is the usual Fourier transform on S.

$$\eta(s; x) = \delta \text{ function of hyperplane } \{x_n = s\}.$$

$$\hat{\eta}(s; \hat{x}) = e^{is\hat{x}_n} \text{ times } \delta \text{ function } \{\hat{x}_n \text{ axis}\}.$$

Since $f \in \mathcal{W}$ we form

$$\begin{aligned}\mathbf{R}_{\vec{P}}f(s) &= \int \eta(s; x)f(x)\, dx \\ &= \int f(x^{\#}, s)\, dx^{\#} \\ &= \mathbf{R}f(s) \end{aligned} \tag{1.214}$$

where $\mathbf{R}$ is the usual hyperplane Radon transform and we have written $x^{\#}$ for $(x_1, \ldots, x_{n-1})$.

We now apply the Parseval formula to (1.214). We obtain (writing $\{\hat{x}_n\}$ for the $\hat{x}_n$ axis)

$$\begin{aligned}\mathbf{R}_{\vec{P}}f(s) &= e^{is\hat{x}_n}\delta_{\{\hat{x}_n\}} \cdot \hat{f} \\ &= \int e^{is\hat{x}_n}\hat{f}(0, \ldots, 0, \hat{x}_n)\, d\hat{x}_n \\ &= \int e^{is\hat{x}_n}\hat{f}\big|_V(\hat{x}_n)\, d\hat{x}_n. \end{aligned} \tag{1.215}$$

Put in other terms, Theorem 1.3 becomes

$$\mathbf{R}_{\vec{P}}f \textit{ is the Fourier transform from } V \textit{ to } S \textit{ of } \hat{f}\big|_V.$$

The PST is usually stated as

The Fourier transform on S of $\mathbf{R}_{\vec{P}} f$ is the restriction of $\hat{f}$ to V.

Example 2 *Partial differential equations.* We study the CP. This is well posed, meaning the null functions are ordinary distributions, when $\vec{P}$ is hyperbolic (Chapter VIII of FA). We shall work formally so the reader does not have to know what hyperbolicity means.

To formulate the projection–slice theorem (PST) we have to express the null solutions in terms of $\{\exp(ix \cdot \hat{x})\}_{\hat{x} \in V}$. This means solving the CP with CD $= \{\delta_{s^0}\, \delta_{jk}\}$ for all $s^0 \in S$ and all j, k. Let us explain how this is done when S is linear; we follow the lines of Chapter IX of FA and Section 1.4.

We use coordinates $s \in S$ and $t \in S^\perp$ with dual variables $\hat{s}, \hat{t}$. The well-posedness of the CP means that we can regard V (generically) as an N sheeted covering of $\hat{S}$. The points where the sheets coalesce are not important (see FA) so we shall ignore them here.

We can express a solution H of $\vec{P}H = 0$ in the form

$$H(s,t) = \int \sum_l e^{is\hat{s}+it\hat{t}_l}\, \hat{H}(\hat{s}, \hat{t}_l)\, d\hat{s} \tag{1.216}$$

for some "function" $\hat{H}(\hat{s}, \hat{t}_l)$ where $\hat{t}_l = \hat{t}_l(\hat{s})$ are the $\hat{t}$ coordinates of the points in V above $\hat{s}$. It follows that

$$\vec{\mathrm{CD}}(H)(s) = \left\{ \int \sum e^{is\hat{s}}\, \hat{h}_j(\hat{s}, \hat{t}_l) \hat{H}(\hat{s}, \hat{t}_l)\, d\hat{s} \right\}. \tag{1.217}$$

Since all of our calculations behave nicely with respect to translation in s we can restrict our considerations to $s = 0$.

Remark. *Equation (1.217) shows that $\vec{CD}$ can be regarded as (the Fourier transform of) a multiplicity Radon transform from functions on V to N tuples of functions on $\hat{S}$. The multiplicities are given by $\{\hat{h}_j\}$.*

This is in keeping with the general philosophy of PSTs. For we can regard taking CD as a multiplicity restriction (slice) so its Fourier transform is a multiplicity (projection) Radon transform.

We can rewrite (1.217) in the more suggestive form

$$\vec{\mathrm{CD}}(H)(s) = \int e^{is\hat{s}} \boxed{\hat{h}}(\hat{s}) \vec{\hat{H}}(\hat{s})\, d\hat{s} \tag{1.218}$$

where $\vec{\hat{H}}(\hat{s})$ is the column vector whose components are $\hat{H}(\hat{s}, \hat{t}_l(\hat{s}))$ and $\boxed{\hat{h}}(\hat{s})$ is the matrix whose $j\,l$ entry is $\hat{h}_j(\hat{t}_l(\hat{s}))$. (We denote a vector by $\vec{a}$ and $\boxed{a}$ is a matrix.)

$\eta(s^0, j; s, t)$ is the function in $\hat{\mathcal{W}}'(\vec{P})$ (it is a function of s, t) whose CD is $\delta_{s^0}\delta_{jk}$. By (1.218)

$$\delta_{jk}\delta_{s^0} = \int e^{is\hat{s}} \boxed{\hat{h}}(\hat{s})\vec{\eta}(s^0, j; \hat{s})\, d\hat{s} \tag{1.219}$$

or

$$\boxed{I}_{s^0} = \int e^{is\hat{s}} \boxed{\hat{\eta}}(s^0; \hat{s}) \boxed{\hat{h}}'(\hat{s})\, d\hat{s}. \tag{1.220}$$

We have written $\boxed{\hat{\eta}}(s^0; \hat{s})$ for the matrix whose rows are $\hat{\eta}(s^0, j; \hat{s}, \hat{t}^l(\hat{s}))$ (j fixed) in conformity with our notation for $\boxed{h}$. For this reason we have replaced $\boxed{\hat{h}}$ by $\boxed{\hat{h}}'$ and inverted the order of multiplication in (1.220). $\boxed{I}_{s^0}$ is the identity matrix in jk times δ_{s^0}.

From (1.220) we derive the explicit formula

$$\boxed{\hat{\eta}}(s^0; \hat{s}) = \boxed{\hat{h}}'^{-1}(\hat{s}) e^{-is^0\hat{s}}. \tag{1.221}$$

This gives

$$\begin{aligned} \mathbf{R}_{\vec{P}} f(s^0) &= \vec{\eta}(s^0) \cdot f \\ &= \int \boxed{\hat{\eta}}(s^0; \hat{s}) \vec{\hat{f}}(\hat{s})\, d\hat{s} \\ &= \int e^{-is^0\hat{s}} \boxed{\hat{h}}'^{-1}(\hat{s}) \vec{\hat{f}}(\hat{s})\, d\hat{s}. \end{aligned} \tag{1.222}$$

Here $f \in \mathcal{W}$ and $\vec{\hat{f}}(\hat{s})$ is the vector whose components are $\hat{f}(\hat{s}, \hat{t}_l)$. Equation (1.222) represents $\mathbf{R}_{\vec{P}} f$ as the Fourier transform of $\boxed{\hat{h}}'^{-1}(\hat{s})\vec{\hat{f}}(\hat{s})$ which depends only on $\hat{f}\big|_V$. This is of the same nature as the usual PSF (1.215) except for algebraic complications.

Example 3 *Standard PSF.* Instead of P being a partial differential operator,

$$P = \tau - I$$

where τ represents translation by 1. The solutions of $Pf = 0$ are periodic with period 1. S is the unit interval with endpoints identified. $V = \{2\pi \text{ integers}\}$.

Remark. *The significance of the identification of 0 with 1 in S will be discussed below.*

The null solutions are

$$\eta(s; x) = \sum \delta_{s+n}$$

so that

$$\mathbf{R}_P f(s) = \sum f(s+n).$$

To compute $\hat{\eta}$ we use Fourier analysis on the circle group S. The Fourier inversion formula for S gives[10]

$$\delta_{s^0} = \sum e^{ins^0} e^{-ins}. \tag{1.223}$$

This formula holds on S but it is clear that the right side of (1.223) is the PD of

$$\sum e^{ins^0} e^{-inx} = \hat{\eta}(s^0) \in \mathcal{W}'(P). \tag{1.224}$$

Theorem 1.3 thus becomes

$$\begin{aligned} \mathbf{R}_P f(s^0) = \eta(s^0) \cdot f &= \sum f(s^0+n) \\ &= \hat{\eta}(s^0) \cdot \hat{f} = \sum e^{ins^0} \hat{f}(n) \end{aligned} \tag{1.225}$$

which is the usual PSF.

Example 4 *Convolution equations.* Instead of partial differential equations we can study more general types of hyperbolic equations. We discussed the PSF related to the operator $P = \tau - I$. $f \to \mathbf{R}^* \mathbf{R} f$ maps f into the kernel of $\tau - I$. $\mathbf{R}^*$ is the operator defining the solution of the parametrization (Cauchy) problem for $\tau - I$ since it maps functions g on $[0,1]$ (satisfying $g(0) = g(1)$) onto periodic functions on the line.

We can regard $\tau - I$ as the operator of convolution by $\delta_1 - \delta_0$. There are other operators of convolution by kernels T, which are distributions of compact support, which are *Cauchy hyperbolic*. This means that the Cauchy map $\mathcal{C}'$ of C^∞ solutions of $T * f = 0$ into their CD (or PD), which are functions on some finite interval S, is essentially an isomorphism of the kernel of $T*$ with all C^∞ functions on S. By translating T we may assume that $S = [0, A]$ is the smallest interval containing support T. (We say "essentially" because, as we shall see presently, there may be compatibility relations like $f(0) = f(1)$ for $T = \delta_1 - \delta_0$. We denote these relations by $\mathbf{C}$.)

We can, to some extent, extend our previous treatment of partial differential equations to convolution equations. (We restrict this discussion to $n = 1$.)

As in the case of differential operators there are two bases for solution:

(a) $\{\eta(s,x)\}_{s \in S}$
(b) $\{e^{i\lambda_k x}\}_{\hat{T}(\lambda_k)=0}$.

$\eta(s,x)$ is the solution of $T * \eta = 0$ whose CD is δ_s. (Some modification must be made because of the compatibility $\mathbf{C}$. We clarify this below.) We assume for simplicity that the zeros of $\hat{T}$ are simple.

[10] We generally use a normalization which removes the factor 2π.

The existence of basis (b) depends on T. We need the analog of the FP (Section 1.4) for T. This requires

(1) The set of zeros $V = \{\lambda_k\}$ is an *interpolation set* for some given AU space $\mathcal{W}$. This means that any function H_V on V which satisfies the growth conditions of $\hat{\mathcal{W}}$ restricted to V can be extended to a function $H \in \hat{\mathcal{W}}$ (see [67]). We shall usually take $\mathcal{W} = \mathcal{E}'$.

(2) $\hat{T}$ is *slowly decreasing* (see [37]) which means that there are constants c, N_0, A' such that

$$|\hat{T}(\hat{x})| \geq c(1 + |\hat{x}|)^{-N_0} e^{-A'|\Im \hat{x}|} \tag{1.226}$$

at enough points $\hat{x}$. "Enough points" means that for any $\hat{y}$ there is an $\hat{x}$ with

$$|\hat{x} - \hat{y}| \leq c'(1 + |\hat{x}|)^{N_1} e^{A''|\Im \hat{x}|}$$

for suitable constants c', N_1, A''.

It is sometimes convenient to think of (a) as defining an exotic CP. If T has support at the origin, so $T * f = 0$ is an ordinary differential equation, then the interval $[0, A]$ collapses to the origin, but there are multiplicities which define the CP. Thus we can think of the interval $[0, A]$ as being the origin with "multiplicities" which are the points in $[0, A]$.

The CP for $T*$ is best described in terms of the Cauchy map $\mathcal{C}'$. The Fourier transform of the adjoint of $\mathcal{C}'$ is given by (compare (1.107))

$$\hat{\mathcal{C}}(\hat{u}) = \hat{\alpha} + \hat{T}\hat{\beta} \tag{1.227}$$

which is a map of $\hat{\mathcal{E}}'$ into $\hat{\mathcal{E}}'[0, A] \oplus \hat{T}\hat{\mathcal{E}}'$. This means

$$\hat{u} = \hat{\alpha} \quad \text{on } V. \tag{1.228}$$

A formal solution of this equation is

$$\hat{\alpha}(\hat{x}) = \sum \hat{u}(\hat{x}_\lambda) \frac{\hat{T}(\hat{x})}{(\hat{x} - \hat{x}_\lambda)\hat{T}'(\hat{x}_\lambda)}. \tag{1.229}$$

There is some difficulty in the convergence of this series unless $\hat{u}$ is small at infinity, which means u is sufficiently differentiable. There are various convergence schemes to overcome this problem. The reader can deal with them or else replace $\hat{\mathcal{E}}'$ by a space $\mathcal{W}$ for which the functions $\hat{u} \in \hat{\mathcal{W}}$ decrease like $|\hat{x}|^{-N}$ for a suitable N.

We translate T so $[0, A]$ is the convex hull of support T. By the Paley–Wiener theorem α as given by (1.229) has its support contained in $[0, A]$.

However, α is not uniquely determined from (1.227) and support $\alpha \subset [0, A]$. If there were another α_1 satisfying (1.227) then $\hat{\alpha} - \hat{\alpha}_1$ vanishes on V so is of the form $\hat{T}\hat{\beta}_1$. Thus

$$\alpha - \alpha_1 = T * \beta_1$$

has support in $[0, A]$. By the Titchmarsh convolution theorem [15] this means that support $\beta_1 =$ origin, i.e. $\beta_1 = \sum c_j \delta^{(j)}$.

Let f be a function (distribution) on $[0, A]$ which we want to be the CD of a solution of $T * f = 0$, that is, $(T * f) \cdot \beta = f \cdot (T * \beta) = 0$ for all $\beta \in \mathcal{W}$. This means (see (1.227)) that for any $u \in W$

$$f \cdot u = f \cdot \alpha. \tag{1.230}$$

Since α is not uniquely determined we need

$$f \cdot T * \beta_1 = 0 \tag{1.231}$$

for all β_1 as above. This set of compatibility conditions forms **C**. (They can be reduced to a finite number.)

For an explicit formulation of Theorem 1.3 for convolution equations we need to compute the Fourier transform $\hat{\eta}(s^0; \hat{x}_\lambda)$ of $\eta(s^0; x)$. In analog with Example 2 (see (1.219)) we might expect to be able to solve the system of equations

$$\sum \hat{\eta}(s^0; \hat{x}_\lambda) e^{i\hat{x}_\lambda s} = \delta_{s^0} \tag{1.232}$$

for $\hat{\eta}$. In Example 2 the analogous system was finite but here it is infinite and does not seem susceptible to direct analysis.

One could also attack the problem by constructing fundamental solutions $e^\gamma_{s^0}(x)$ for $T*$ by

$$e^\gamma_{s^0}(x) = \int_\gamma \frac{e^{i(x-s^0)\hat{x}}}{\hat{T}(\hat{x})}\, d\hat{x} \tag{1.233}$$

over suitable contours γ in the complex plane.

Problem 1.4 *Develop a PSF for $T*$.*

Example 5 *Abelian groups.* There is a PST for any locally compact, abelian topological group **G**. Let H be a closed subgroup of **G** and call $\hat{H}^\perp$ the subgroup of the character group $\hat{\mathbf{G}}$ which is trivial on H. We replace the exponential function $\exp(ix \cdot \hat{x})$ by $\chi(g, \hat{g})$ which is the character defined by $\hat{g} \in \hat{\mathbf{G}}$.

Call $\mathcal{W}$ the space of continuous functions on G which "vanish at infinity," meaning the closure in the uniform topology of the space of continuous functions of compact support. $\mathcal{W}'$ is the space of measures having total bounded variation. In analogy with the case of the circle group we define $\mathcal{W}'(P)$ as the space of H invariant measures of **G**. Thus $\hat{H}^\perp$ plays the role of V.

$\hat{H}^\perp$ is a (spectral) basis for $\mathcal{W}'(P)$. The geometric basis $\{s\}$ can be identified with the cosets $\{s\}$ of $S = G/H$. We are clearly in a hyperbolic situation, meaning the map from $\mathcal{W}'(P)$ to $\mathcal{W}'(S)$ is an isomorphism.

For f a continuous function of compact support on **G** we can define $\mathbf{R}f$ (f thought of as an element of $\mathcal{W}''$)=restriction of f to $\mathcal{W}'(P)$.

Since the set of characters in $\hat{H}^\perp$ forms an analytic basis for $\mathcal{W}'(P)$, $\mathbf{R}f$ is determined by

$$\mathbf{R}^A f(\chi) = \hat{f}(\chi) = f \cdot \chi \tag{1.234}$$

for $\chi \in \hat{H}^\perp$.

The δ functions of the cosets $\{s\}$ form a geometric basis for $\mathcal{W}'_P$, so $\mathbf{R}f$ is also determined by

$$\mathbf{R}^G f = \{\delta_s \cdot f\}. \tag{1.235}$$

Denote by s_0 the coset corresponding to H. The Fourier inversion formula for $\mathbf{G}/H$ can be written as the relation of the two bases on $\mathbf{G}/H$ (spectral-geometric identity)

$$\delta_{s_0} = \int_{\chi \in \hat{H}^\perp} \chi \, d\hat{H}^\perp \tag{1.236}$$

where $d\hat{H}^\perp$ is a suitably normalized invariant measure on $\hat{H}^\perp$.

Apply $\mathbf{R}^*$ to this formula, meaning extend everything to H invariant functions or measures on $\mathbf{G}$. This leads to

$$\eta(s^0, g) = \delta_H = \widehat{\delta_{\hat{H}^\perp}} \tag{1.237a}$$

which is the usual form of the PSF for $\mathbf{G}, H$.

We can formulate the PST according to Theorem 1.3: for $f \in \mathcal{W}$

$$\mathbf{R}f(s^0) = \int_V \hat{f}(\chi) \, d\chi. \tag{1.237b}$$

Example 6 *Nonabelian group.* Suppose $\mathbf{G}$ is compact and Γ is a closed subgroup. $\mathbf{G}/\Gamma$ is not generally a group since we do not assume that Γ is a normal subgroup.

Fourier transform involves the irreducible representations of $\mathbf{G}$, which are finite dimensional. The relation between Fourier analysis on $\mathbf{G}$ and Fourier analysis on (right) Γ invariant functions was studied in detail by Frobenius.

Formulas (1.237a) and (1.237b) involve sums (integrals) over a subgroup $\{\chi\}$ of the dual group. For nonabelian groups the points $\{j\}$ are replaced by conjugacy classes and the $\{\chi\}$ by sets of characters or other "special" representation functions.

As usual we consider two bases for functions (distributions) on G/Γ. The first (geometric) basis is $\{\sum_\gamma \delta_{x\gamma}\}_{x \in G/\Gamma}$ (assuming Γ is discrete; otherwise the sum is replaced by an integral). The second (analytic) basis is $\{\varphi(x, \lambda)\}$ where $\{\varphi(x, \lambda)\}$, for fixed λ, is a basis for (right) Γ invariant representation functions for $\mathbf{G}$ corresponding to the representation λ. (See Section 1.6 for details on representations.)

$\mathbf{G}$ acts on functions on $\mathbf{G}/\Gamma$ by left translation. We want to compute the character Ψ of this representation. Note that $g \sum \delta_{x\gamma} = \sum \delta_{gx\gamma}$ is a geometric basis element. If we compute Ψ using the geometric basis then, clearly, the

geometric basis element $\sum_\gamma \delta_{x\gamma}$ contributes a nontrivial amount to the value $\Psi(g)$ if

$$\sum \delta_{gx\gamma} = \sum_\gamma \delta_{x\gamma}. \tag{1.238}$$

This is equivalent to

$$gx = x\gamma^0,$$

i.e.

$$g = x\gamma^0 x^{-1}. \tag{1.239}$$

This means that Ψ is nontrivial on the conjugacy class of the elements of Γ.

Frobenius calls Ψ the *induced character* of the trivial character of Γ. More generally, let ρ be an N-dimensional representation of Γ. The *induced representation* $\rho\overset{\mathbf{G}}{\underset{\Gamma}{\uparrow}}$ is defined as follows. The representation space consists of functions on $\mathbf{G}$ with values in $\mathbb{C}^N$ satisfying

$$f(g\gamma) = f(g)\rho(\gamma). \tag{1.240}$$

$\mathbf{G}$ acts on this space by left translation.

Suppose $\mathbf{G}$ is finite. Let χ be the character of ρ. Define the function $\chi_{0\Gamma}$ on $\mathbf{G}$ to be χ on Γ and 0 off Γ. Frobenius' (geometric) formula for the induced character $\chi\overset{\mathbf{G}}{\underset{\Gamma}{\uparrow}}$ is

$$\chi\overset{\mathbf{G}}{\underset{\Gamma}{\uparrow}}(g) = \frac{1}{|\Gamma|}\sum_{\tilde{g}\in\mathbf{G}} \chi_{0\Gamma}(\tilde{g}g\tilde{g}^{-1}). \tag{1.241}$$

Remark. *The argument (1.238)f. leads to a proof of Frobenius' formula when $\chi = 1$. A slight modification proves the formula in general.*

Since any class function on Γ is a sum of characters, Frobenius' formula suggests that we define the *Frobenius induction* $\alpha(u)$ of any class function u on Γ by

$$\alpha(u)(g) = \frac{1}{|\Gamma|}\sum_{\tilde{g}\in G} u_{0\Gamma}(\tilde{g}g\tilde{g}^{-1}) \tag{1.241*}$$

where $u_{0\Gamma}$ is the function on $\mathbf{G}$ which is u on Γ and 0 off Γ.

To put Frobenius' formula in our framework, we introduce the Frobenius matrix $\boxed{F}$ which defines induction on characters. $\boxed{F}$ is a matrix with $|\Gamma|$ rows labeled by $\gamma \in \Gamma$ and $|\mathbf{G}|$ columns labeled by $g \in \mathbf{G}$. Note that $\chi_{0\Gamma}(\tilde{g}g\tilde{g}^{-1})$ is nonvanishing for some $\tilde{g}$ if and only if g is conjugate in $\mathbf{G}$ to some $\gamma \in \Gamma$. The γg entry in $\boxed{F}$ is the number of $\tilde{g}$ for which $\tilde{g}\gamma\tilde{g}^{-1} = g$. Thus row γ has nonvanishing elements at those g which are conjugate to γ in $\mathbf{G}$. For each such g we put, for that matrix element, the number $|\mathcal{C}_{\mathbf{G}}(\gamma)|$ where $\mathcal{C}_{\mathbf{G}}(\gamma)$ is the centralizer of γ in $\mathbf{G}$.

If f is a function on G, think of f as a column vector with $|G|$ entries. Thus

$$\left[\boxed{F}f\right](\gamma) = |\mathcal{C}_{\mathbf{G}}(\gamma)| \sum f(g) \tag{1.242}$$

where the sum is over g conjugate to γ. In particular if f is a class function then

$$\left[\boxed{F}f\right](\gamma) = |\mathbf{G}| f(\gamma) \tag{1.243}$$

so $\boxed{F}f$ is $|G|$ times the restriction of f to Γ.

On the other hand if u is a function on Γ thought of as a row vector with $|\Gamma|$ components, then $u\boxed{F}$ is the Frobenius induction $\alpha(u)$ defined by (1.241*) times $|\Gamma|$. This leads to the inner product

$$(u, f) = u\boxed{F}f/|\mathbf{G}||\Gamma| \tag{1.244}$$

which illustrates the fact that the adjoint of α as a map on class functions from $L_2(\Gamma)$ to $L_2(\mathbf{G})$ is the restriction from $\mathbf{G}$ to Γ. For, reading from left to right on the right side of (1.244) we obtain the inner product of $\alpha(u)$ and f on G, while reading from right to left gives the inner product on Γ of u with restriction to Γ of f.

(1.241) is a geometric formula for the induced character. There is also a representation-theoretic, i.e. Fourier transform (analytic), formula for the induced character. Let χ^0 be the character of an irreducible representation ρ^0 of $\mathbf{G}$. We can ask how many times ρ^0 occurs in $\rho\overset{\mathbf{G}}{\underset{\Gamma}{\uparrow}}$. This is given by the inner product

$$\begin{aligned} \chi^0 \cdot \chi\overset{\mathbf{G}}{\underset{\Gamma}{\uparrow}} &= \chi^0 \cdot \alpha\chi \\ &= \alpha'\chi^0 \cdot \chi. \end{aligned} \tag{1.245}$$

Hence this number is given by the number of times that ρ occurs in the restriction of ρ^0 to Γ. For a precise formulation we need to assume that $\mathbf{G}/\Gamma$ is compact or, at least, that it has finite volume.

In particular, if $\rho = I$ then (1.245) means that the number of times ρ^0 occurs in the representation of $\mathbf{G}$ on $L^2(\mathbf{G}/\Gamma)$ equals the number $m(\rho^0)$ of linearly independent vectors in the representation space for ρ^0 which are invariant under $\rho^0(\Gamma)$.[11]

We can deduce the following nonabelian PSF:

$$1\overset{\mathbf{G}}{\underset{\Gamma}{\uparrow}} = \sum m_j \chi_j. \tag{1.246}$$

[11]For noncompact semi-simple Lie groups this analog of the Frobenius reciprocity theorem was introduced in [70].

The sum is over all characters of irreducible representations of $\mathbf{G}$ that appear in $L^2(\mathbf{G}/\Gamma)$; the multiplicity of χ_j is m_j. The left side of (1.246), via (1.241*) is the geometric formula for the induced character as in (1.238)f.

The equality of (1.246) with (1.241) is (essentially) *Selberg's trace formula* (see Chapter 8). It provides a relation between the geometry of conjugacy classes and the analysis of characters in $L^2(\mathbf{G}/\Gamma)$.

Let us apply (1.246) to a class function $f(g)$. Using the fact that induction and restriction are adjoints we find

$$\sum f(\gamma) = \sum m_j \chi_j \cdot f, \tag{1.247}$$

i.e.

$$\sum \delta_\gamma = \sum m_j \chi_j. \tag{1.248}$$

In case G is abelian each element is a conjugacy class and each $m_j = 1$ so (1.248) reduces to the standard PSF (1.237a).

The relation of (1.247) to (1.237a) suggests that we should regard the irreducible characters χ for which the corresponding representation contains Γ invariant vectors as the analog of $\hat{\Gamma}^\perp$. This set of χ generalizes the abelian $V = \hat{H}^\perp$ so (1.247) is the nonabelian analog of (1.237b).

The Frobenius reciprocity theorem and the corresponding PSF relate the geometry of conjugacy classes to the analysis of irreducible characters. There is another ingredient which was promulgated by Selberg.

The Frobenius reciprocity theorem is derived using (1.238)f. We learn from this proof that whenever we have a transformation $\mathcal{T}$ of $\mathbf{G}/\Gamma$ we can try to obtain a geometric expression for the trace by using the δ functions of cosets as a basis for functions on $\mathbf{G}/\Gamma$.

Another basis that is useful for the computation of the trace is the basis of eigenfunctions of T.

We obtain transformations $\mathcal{T}$ when we can realize $\mathbf{G}/\Gamma$ as an orbit of $\mathbf{G}$ (or a significant part of an orbit) in some finite dimensional space on which $\mathbf{G}$ acts and, moreover, $\mathbf{G}/\Gamma$ is a PS for a $\mathbf{G}$ invariant system of partial differential equations. T represents the passage from $\mathbf{G}/\Gamma$ to another PS which, in favorable conditions, can be carried out geometrically by an integral kernel and analytically using eigenfunction expansions. This is detailed in Chapters 7 and 8.

Remark. *We do not seem to be able to obtain any more formulas for semisimple Lie groups* $\mathbf{G}$ *in this way than we can obtain by Frobenius reciprocity. The reason seems to be that the characters are eigenfunctions of the center of the enveloping algebra of* $\mathbf{G}$ *(polynomials in the Lie algebra) and the characters "encode" all information about eigenfunctions.*

In the Frobenius method the entities of prime importance for a representation are characters and their geometric counterparts which are class functions. When dealing with the rotation group other representation functions such as spherical

harmonics are of great interest. If G is a noncompact semi-simple Lie group then the spherical function often plays an important role in representation theory.

The spherical function (or some analog) is related to *parabolic induction*. In its simplest form it arises from the Iwasawa decomposition $\mathbf{G} = KA^+N$. Here K is a maximal compact subgroup of $\mathbf{G}$, A^+ is abelian, N is nilpotent, and A normalizes N. We start from a character $\hat{a}$ of A^+, then extend $\hat{a}$ to A^+N by making it trivial on N. Finally we form the representation

$$\rho = \rho_{\hat{a}} = \operatorname{ind} \begin{matrix} \mathbf{G} \\ \uparrow \\ A^+N \end{matrix} \hat{a}.$$

Formally the representation space is $L_2(\mathbf{G}/A^+N)$ which is essentially $L_2(K)$. (More precisely it is $L_2(K/M)$ where M = centralizer of A in K, but M plays no role in the present heuristics.)

The matrix coefficient

$$\phi(\rho, \mathbf{g}) = \rho(\mathbf{g})1 \cdot 1$$

is called the *spherical function* for ρ. K acts on $L_2(K)$ by left translation so the constant function 1 on K is fixed by $\rho(K)$, which gives

$$\phi(\rho, k\mathbf{g}\kappa) = \phi(\rho, \mathbf{g})$$

for any $k, \kappa \in K$. This bi K invariance characterizes spherical functions amongst all representation functions for ρ when ρ is irreducible.

It is easy to pass from characters to spherical functions. If χ is a character of any representation, we call

$$\phi(\mathbf{g}) = \int \chi(k\mathbf{g})\, dk. \tag{1.249}$$

ϕ is clearly left K invariant. To check right invariance we write

$$\begin{aligned} \phi(\mathbf{g}\kappa) &= \int \chi(k\mathbf{g}\kappa)\, dk \\ &= \int \chi(\kappa^{-1}\kappa k\mathbf{g}\kappa)\, dk \\ &= \int \chi(\kappa k\mathbf{g})\, dk \\ &= \phi(\mathbf{g}). \end{aligned}$$

Of course in both cases it may happen that $\phi \equiv 0$.

The inverse of (1.249) is somewhat more complicated and depends on specific information (see Section 8.2).

The characters of certain representations form a space which we can regard as the dual $\hat{\mathbf{G}}$ of $\mathbf{G}$. We shall not discuss the meaning of "certain." In this way

of looking at $\hat{\mathbf{G}}$ the class functions are (essentially) the functions on $\hat{\mathbf{G}}$. But we can also regard the spherical functions as parametrizing a space which is close to $\hat{\mathbf{G}}$. The space spanned by the spherical functions is (essentially) the space of doubly K invariant functions f on $\mathbf{G}$, meaning $f(k\mathbf{g}\kappa) = f(\mathbf{g})$ for all $k, \kappa \in K$. The map (1.249) defines an "almost" isomorphism between the spaces.

In particular we can apply left integration over K to (1.246) for certain interesting Γ using the Frobenius formula (1.241). Since Frobenius' expression for $\operatorname{ind}_{\Gamma}^{\mathbf{G}} 1$ is geometric this gives an expression for a geometric entity on $K\backslash\mathbf{G}/\Gamma$ in terms of an analytic (spectral) entity, namely the spherical functions $\phi(\rho, \mathbf{g})$. This relation is the original form of the Selberg trace formula. It is discussed in detail in Section 8.2.

The PSF and the associated PST deal with all of $\mathbb{R}^n$ or $K \setminus \mathbf{G}$. Are there results related to "part" of $\mathbb{R}^n$ or of $K \setminus \mathbf{G}$?

Let us begin with $\mathbb{R}^n$. There are two types of subsets that are of interest:

(1) Algebraic subsets A
(2) Domains Ω

The type of result we are searching for is the evaluation of sums $\sum f(m)$ for $m \in A$ or $m \in \Omega$.

There is an obvious approach to this question: namely, use the Poisson summation formula to write

$$\chi(\Omega) \sum_{\text{all } j} \delta_j = \sum_{j \in \Omega} \delta_j = \widehat{\chi(\Omega)} * \sum_{\text{all } \hat{j}} \delta_{\hat{j}}. \tag{1.250}$$

Here $\chi(\Omega)$ is the characteristic function of Ω. A similar formula can be derived for $\{m \in A\}$ but there are some technical difficulties because multiplication of distributions has to be defined carefully. Formula (1.250) has proved useful in many situations. For example, when evaluated on $f \equiv 1$ the left side of (1.250) counts the number of lattice points in Ω. When $n = 2$ and Ω is the ball radius r, center 0, formula (1.250) leads to the best known estimates for the error term in the Gauss lattice point problem.

When we replace Ω by A it seems difficult to apply (1.250). A different method, based on nonlinear Fourier analysis, is described in Chapter 5.

There is another approach to the case of domains which seems most successful when Ω is a convex polyhedron. Let us start with the interval $\Omega = [-N, N] \subset \mathbb{R}^1$. The sum

$$U = \sum_{m \in \Omega} \delta_m$$

is no longer invariant under translation τ by 1 but rather

$$(\tau - I)U = \delta_{N+1} - \delta_{-N}. \tag{1.251}$$

The Fourier transform of (1.251) is

$$\hat{U}(\hat{x}) = \frac{e^{i(N+1)\hat{x}} - e^{-iN\hat{x}}}{e^{i\hat{x}} - 1}. \tag{1.252}$$

Of course there are many solutions of (1.251) besides U since we can add any periodic function to any solution. Determining which solution is U corresponds to giving a proper interpretation to (1.252). The simplest interpretation is the holomorphic function which is the quotient; since U is of compact support it is this interpretation which is U.

The PSF involves the expansion of the denominator in powers of $\exp(i\hat{x})$. This seems quite reasonable because everything we have done is related to the additive structure of the line. Nevertheless, we shall appeal to the *multiplicative* structure and expand $(\exp(i\hat{x}) - 1)^{-1}$ in a Laurent series with remainder. There is a first-order pole at 0. The standard expansion [159] of $[\exp(i\hat{x}) - 1]^{-1}$ is

$$\frac{1}{e^{i\hat{x}} - 1} = \frac{1}{i\hat{x}} - \frac{1}{2} + \frac{B_1 i\hat{x}}{2!} - \frac{B_2 (i\hat{x})^3}{4!} + \cdots \tag{1.253}$$

where B_j are the Bernoulli numbers. Equation (1.252) becomes

$$\hat{U}(\hat{x}) = \frac{e^{i(N+1)\hat{x}} - e^{-iN\hat{x}}}{i\hat{x}} + Q_M(\hat{x})[e^{i(N+1)\hat{x}} - e^{-iN\hat{x}}] + R_M(\hat{x})[e^{i(N+1)\hat{x}} - e^{-iN\hat{x}}]. \tag{1.254}$$

Q_M is a polynomial of degree M and R_M is the remainder, which has simple poles at $\{2\pi j\}_{j\neq 0}$.

Note that $[e^{i(N+1)\hat{x}} - e^{-iN\hat{x}}]/i\hat{x}$ is the Fourier transform of the characteristic function of $[-N, N+1]$. Moreover $Q_M(\hat{x})e^{i(N+1)\hat{x}}$ is the Fourier transform of a distribution supported at $N+1$. Similarly the Fourier transform of $Q_M(\hat{x})e^{-iN\hat{x}}$ is supported at $-N$. Thus the Fourier transform of (1.254) is

$$\sum_{-N\leq m\leq N} \delta_m = \chi(-N, N+1) + B_M^+(N+1) + B_M^-(-N) + E_{M,N}. \tag{1.255}$$

$B_M^{\pm}$ are distributions of order $\leq M$ supported at $N+1$ and $-N$ whose coefficients are expressible in terms of the Bernoulli numbers. $E_{M,N}$ is the remainder, which has support in $[-N, N+1]$ and is "small."

Equation (1.255) is the *Euler–Maclaurin sum formula*.

We can apply a similar method to treat convex polygons in higher dimensions if their supporting planes have rational normals.

Instead of using finite intervals $[-N, N]$ we could have used a semi-infinite interval, e.g. $[0, \infty)$. This corresponds to a formula intermediate to those of Poisson and Euler–Maclaurin. There is no difficulty in extending the above method to half-lines, or half-planes.

Perhaps the most interesting example is found in Riemann's loop integral proof of the functional equation of the zeta function [150]. Riemann wrote, for $\sigma > 1$,

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{\hat{x}^{s-1}}{e^{\hat{x}} - 1}\, d\hat{x}. \tag{1.256}$$

This can be derived from the explicit formula for the Fourier transforms of $|x|^{-s}$ and $\sum_{n\geq 1} \delta_n = U^+$. The Fourier transform of $|x|^{-s}$ is $\Gamma(1-s)\hat{x}^{s-1}$ times a sum of powers of i. (In Chapter 7 we study this question in detail using analytic continuation.) The Fourier transform of U^+ is determined from the equation

$$(\tau - I)U^+ = \delta_0 \tag{1.257}$$

so

$$\hat{U}^+ = \frac{1}{e^{i\hat{x}} - 1}. \tag{1.258}$$

To derive (1.256) from these formulas we apply the Parseval identity to $\zeta(s) = U^+ \cdot |x|^{-s}$ and then reduce the integral over $\hat{x} \in \mathbb{R}^1$ to $\hat{x} \geq 0$. The powers of i change $\Gamma(1-s)$ to $1/\Gamma(s)$ because

$$\Gamma(s)\Gamma(1-s) = \pi \csc \pi s.$$

Remark. *The series (1.253) converges only for $|x| < 1$ because only the pole of $[\exp(i\hat{x}) - 1]^{-1}$ at $\hat{x} = 0$ is taken into account. We can obtain better approximations to U by using more poles of $[\exp(i\hat{x}) - 1]^{-1}$. This is carried out in Chapter 8. When applied to the zeta function this "approximate Euler–Maclaurin formula" leads to the approximate functional equation for the zeta function.*

We can also study the Euler–Maclaurin formula related to the Radon transform $\mathbf{R}_\partial$ for a "general" differential or convolution operator ∂. We start with a smooth convex domain Ω. We now cut off solutions e of ∂ by multiplying by $\chi(\Omega)$ which is the characteristic function for Ω. $\chi(\Omega)$ satisfies

$$\partial[\chi(\Omega)e] = b \tag{1.259}$$

where support b is close to the boundary of Ω. The closeness is determined by support ∂.

Using Fourier transform we obtain (for $\hat{x} \in \mathbb{C}^n$)

$$\widehat{\chi(\Omega)e}(\hat{x}) = \frac{\hat{b}(\hat{x})}{\hat{\partial}(\hat{x})}. \tag{1.260}$$

Now the set $\hat{\partial}(\hat{x}) = 0$ is a union of irreducible holomorphic hypersurfaces V_j. Suppose V_0 passes through the origin. Let $\hat{\partial}_0(\hat{x}) = 0$ define V_0, and let m be the highest power of $\hat{\partial}_0$ dividing $\hat{\partial}$. Then we can write

$$\hat{\partial}^{-1}(\hat{x}) = \sum_{j=0}^{m} \alpha_j(\hat{x})\hat{\partial}_0^{-j}(\hat{x}) + \hat{\partial}^*(\hat{x}). \tag{1.261}$$

The α_j are entire and $\hat{\partial}^*$ is regular near 0. Moreover $\alpha_m(0) \neq 0$.

Since $\hat{\partial}^*$ is regular near 0 we can expand it in a power series (with remainder). This process would seem to lead to an analog of the Euler–Maclaurin formula with $\alpha_j \partial_0^{-j}$ playing the roles of "integration." More precisely, we replace $\hat{\partial}^{-1}$ in (1.260) by its value in (1.261). But the problem arises that, although $\alpha_j \hat{\partial}_0^{-j} \hat{b}$ are entire, they may not be of the exponential type necessary to have support on a small neighborhood of Ω. (There are many interesting cases in which $\alpha_j \hat{\partial}_0^{-j} \hat{b}$ is of exponential type.)

However, if $\hat{\partial}$ is a polynomial then $\hat{\partial}_0$ and the α_j are polynomials so we obtain a "reasonable" Euler–Maclaurin theory.

Problem 1.5 *When $\hat{b}$ is an exponential polynomial are the supports of the Fourier transforms of $\alpha_j \hat{\partial}_0^{-j} \hat{b}$ in a small neighborhood of Ω?*

Let us now analyze the analogous question for more general groups $\mathbf{G}$. The simplest nonabelian example from our point of view is $\mathbf{G} = SL(2, \mathbb{R})$ which is the group of 2×2 real matrices of determinant 1. (We shall often ignore the subgroup $\{\pm I\}$ as it is not important for our considerations.) K is the orthogonal subgroup and $K \backslash \mathbf{G}$ is the hyperbolic plane. We choose Γ as a discrete group with compact fundamental domain $\mathcal{D}$ which is a noneuclidean polygon. Γ is generated by reflections in the sides of $\mathcal{D}$.

The most reasonable analog $\Omega = \Omega_N$ of $[-N, N]$ is a noneuclidean polygon which is a disjoint union $\Omega = \bigcup \gamma_m \mathcal{D}$ with $\gamma_m \in \Gamma$, $\{\gamma_m\}$ finite. We might want Ω to be convex in the hyperbolic geometry and also to be the fundamental domain for a subgroup $\tilde{\Gamma}$ of Γ.

We start with a Γ invariant geometric entity U on $K \setminus \mathbf{G}$ whose Fourier transform on G is known. Our objective is to study $\chi(\Omega)U$. This is no longer Γ invariant. One thing we need to know is

$$\text{support } (\gamma - I)[\chi(\Omega)U] \sim \text{bd } \Omega, \tag{1.262}$$

meaning that it covers only a small part of Ω, uniformly as γ varies through all reflections in the sides of Ω.

Assuming this holds we could apply Fourier analysis on $\mathbf{G}$ to obtain an analytic formula for $\chi(\Omega)U$. We are far from being able to carry out this program.

1.8 Tensor products and direct integrals

The PST deals with a single spread. We now put these "coarse grains" together to fill up the whole space. This combination process is a direct integral or a tensor product. (See (1.117)ff. for a detailed discussion.)

If we are given a decomposition of a space X into subsets (slices) $X_{\mathbf{g}}$ then we can think of functions f on X as being sums of functions $f_{\mathbf{g}}$ supported on $X_{\mathbf{g}}$ with compatibility conditions:

$$\textbf{COMP: } f_{\mathbf{g}} = f_{\mathbf{g}'} \quad \text{on } X_{\mathbf{g}} \cap X_{\mathbf{g}'}. \tag{1.263}$$

We can think of the decomposition $\{X_{\mathbf{g}}\}$ as giving a slice decomposition (semi-direct integral) of function spaces

$$\mathcal{W}(X) = \sum_{\underset{\sim}{\oplus}} \mathcal{W}(X_{\mathbf{g}}) \tag{1.264}$$

where $\underset{\sim}{\oplus}$ is the subdirect sum meaning the subspace of the direct sum defined by the compatibility conditions (1.263).

Suppose all the $X_{\mathbf{g}}$ are isomorphic. Then we can express $\underset{\sim}{\oplus}$ in terms of (slice) *subtensor products*. If $\mathbf{G}$ denotes the indexing set $\{\mathbf{g}\}$ then we write (1.264) as

$$\mathcal{W}(X) = \mathcal{W}(X_{\mathbf{g}_0}) \underset{\sim}{\otimes} \mathcal{W}(\mathbf{G}) \tag{1.265}$$

where $\mathcal{W}(\mathbf{G})$ is some space of functions on $\mathbf{G}$. We envisage the situation in which $\mathbf{G}$ is a manifold of positive dimension, in which case the tensor product (1.268) is identified with the subdirect integral $\int_{\underset{\sim}{\oplus}}$ when we use the basis $\{\delta_{\mathbf{g}}\}$ for $\mathcal{W}(\mathbf{G})$. Of course the $\delta_{\mathbf{g}}$ may not belong to $\mathcal{W}(\mathbf{G})$. It is only suitable integrals (wave packets) of the $\delta_{\mathbf{g}}$ that belong to $\mathcal{W}(\mathbf{G})$.

We shall introduce other bases in $\mathcal{W}(\mathbf{G})$. For example, if $\mathbf{G}$ is a sphere or, more generally, a homogeneous space of a Lie group, then it is natural to use representation functions which, for the sphere, are spherical harmonics (see Chapter 2).

The tensor products must be understood in the topological sense. In general there are many different topological tensor products but most of the spaces we deal with are nuclear so the space is pointwise unique (see [82, 46]). However, the topology is somewhat ambiguous. This is clarified in Chapter 2.

An equivalent formulation can be given using the functor Hom in place of $\otimes$. The direct integral leads naturally to

$$\mathcal{W}(X) = \mathrm{Hom}[\mathcal{W}'(\mathbf{G}), \underset{\sim}{\mathcal{W}}(X_{\mathbf{g}_0})] \tag{1.266}$$

when we use the basis $\{\delta_{\mathbf{g}}\}$ for $\mathcal{W}'(\mathbf{G})$.

Naturally the remarks we made regarding topology related to $\underset{\sim}{\otimes}$ have analogs for $\mathrm{Hom}(\underset{\sim}{\ ,\ })$.

All this applies to spaces of functions on $X = \mathbb{R}^n$ if the $X_{\mathbf{g}}$ are chosen to be the planes of dimension l through the origin. If $\mathcal{W}$ is a space of C^∞ functions then the spaces $\mathcal{W}(X_{\mathbf{g}_0})$ and $\mathcal{W}(\mathbf{G})$ should be spaces of C^∞ functions. We might also expect to have to replace COMP by **FORMAL COMP**, meaning, as in our discussion in Section 1.4, that there is a formal power series on $X_{\mathbf{g}} \cap X_{\mathbf{g}'} = X_{\mathbf{g}\mathbf{g}'}$ of the form $f_{\mathbf{g}\mathbf{g}'} = \sum f_k(y) t^k$ where $f_k(y)$ are C^∞ functions on $X_{\mathbf{g}\mathbf{g}'}$ (k being a multi-index) and t is a variable on $X^\perp_{\mathbf{g}\mathbf{g}'}$ such that for any $X_{\mathbf{g}''} \supset X_{\mathbf{g}\mathbf{g}'}$ the formal power series for $f_{\mathbf{g}\mathbf{g}'}$ restricted to $X_{\mathbf{g}''}$ agrees with that of $f_{\mathbf{g}''}$ along $X_{\mathbf{g}\mathbf{g}'}$. We

shall show, however, that COMP itself is usually sufficient when dim $X_{\mathbf{g}} > 1$. When dim $X_{\mathbf{g}} = 1$ the intersections are the origin and COMP ORIGIN (which is the same as FORMAL COMP) is needed (see Chapter 2).

The above discussion dealt with a slice tensor product. There is also a *projective tensor product* which is related to $\mathbf{R}$ or $\mathbf{R}^*\mathbf{R}$. Suppose that for each $\mathbf{g}$ in an appropriate Grassmannian $\mathbf{G}$ we have a space of functions $\mathcal{W}'(\partial(\mathbf{g}))$ which are solutions of equations $\partial(\mathbf{g})f_{\mathbf{g}} = 0$. We assume that we are in the hyperbolic situation so $\mathcal{W}'(\partial(\mathbf{g}))$ can be written in the form $\mathcal{W}'(S_{\mathbf{g}})$ (or $\mathcal{W}'^k(S_{\mathbf{g}})$) where $S_{\mathbf{g}}$ is a CS for $\partial(\mathbf{g})$.

We assume, as in the case of $\mathbf{R}^*\mathbf{R}$, that there is a map defined on a space $\mathcal{W}'$ which sends $f \to \{f_{\mathbf{g}}\}$. We define the *supra* (or *projective) tensor product* in such a way that

$$\mathcal{W}'(S_{\mathbf{g}_0})\tilde{\otimes}\mathcal{W}'(\mathbf{G}) = \mathcal{W}'. \tag{1.267}$$

The supra tensor product is formed by the sums of the $f_{\mathbf{g}}$; it is the quotient of the actual tensor product by a set of relations. (For $\mathbf{R}^*\mathbf{R}$ they are clarified in Chapter 2.) Besides the notion of compatibility the only difference between (1.267) and (1.265) is the type of identification of a tensor $a(s) \otimes \delta_{\mathbf{g}}$ with a function on $\mathbb{R}^n$ (or X). In (1.265) we use the identification $a(s)\chi(X_{\mathbf{g}})$, meaning the function $a(s)$ on $S_{\mathbf{g}}$ times the characteristic function of $X_{\mathbf{g}}$, while in (1.267) we identify it with the solution of the CP for $\partial(\mathbf{g})$ with CD given by $a(s)$ on $S_{\mathbf{g}}$.

Remark. *The prefixes "sub" and "supra" refer to subspaces or quotient spaces. Thus there are natural dualities between the sub and supra objects.*

In Sections 1.2 and 1.4 we have explained that solutions of the first John equation are integrals of solutions of first-order equations. The first John equation constitutes the enveloping equations of this set of first-order equations. A similar situation occurs in the Whittaker–Bateman representation of harmonic functions.

We can regard the equation $0f = 0$ as the enveloping equation of the systems $\partial(\mathbf{g})f = 0$ where $\partial(\mathbf{g})$ is the system defined by tangential derivatives to the l plane defined by $\mathbf{g}$ for $\mathbf{g} \in$ Grassmannian of l planes.

These examples indicate that the solutions of an enveloping equation of a system $\{\partial_j\}$ form a supra tensor product of the kernels of the ∂_j, assuming the kernels are isomorphic.

As in the case of $\mathbf{R}^*\mathbf{R}$ and enveloping equations, an element of a supra tensor product is generally a sum (integral) of elements in a fixed space. The slice tensor product, in its simplest form, represents functions f in terms of their restrictions. The components of the slices of f seem more intrinsically related to f than the components in a projective tensor product. Thus one of our tasks is to "transform" projective tensor products to slice tensor products.

There are two methods to perform this transformation:

(1) Direct methods
(2) Fourier analysis

Fourier methods, which constitute the main thrust of this book, change differentiation into multiplication and in so doing change projective tensor products into slice tensor products, provided that the operators $\partial(\mathbf{g})$ commute with translation (or commute with group operations if we are dealing with operators on groups). For, the differential equation $\partial(\mathbf{g})f_{\mathbf{g}} = 0$ is transformed into $\widehat{\partial(\mathbf{g})}\hat{f}_{\mathbf{g}} = 0$ which says (formally) that support $\hat{f}_{\mathbf{g}} \subset \{\widehat{\partial(\mathbf{g})} = 0\}$; this can be thought of as a slice related to $\mathbf{g}$.

There are several direct methods that work in various situations. It is essential to use them when the $\partial(\mathbf{g})$ do not commute with translation. However, these methods are generically ad hoc and more difficult than Fourier methods.

We have already met one such idea in Lemma 1.1. Another technique relates to a new proof of Weierstrass' theorem on polynomial approximation. Suppose the operators $\partial(\mathbf{g})$ commute. Then, under reasonable conditions the kernel of $\partial(\mathbf{g}_{1\dots m}) = \partial(\mathbf{g}_1)\partial(\mathbf{g}_2)\dots\partial(\mathbf{g}_m)$ consists of sums $f_{1\dots m} = \sum f_j$ where $\partial(\mathbf{g}_j)f_j = 0$. This means that $\partial(\mathbf{g}_{1,\dots,m})$ is the enveloping operator for $\partial(\mathbf{g}_1), \dots, \partial(\mathbf{g}_m)$. Now, $\partial(\mathbf{g}_{1\dots m})$ is an operator of high degree; suppose it is hyperbolic. Let $S_{1\dots m}$ be a CS for $\partial(\mathbf{g}_{1\dots m})$ and for each $\partial(\mathbf{g}_j)$. Since $\partial(\mathbf{g}_{1\dots m})$ is of high order, the CD involves many normal derivatives on $S_{1\dots m}$.

Our objective is to, somehow, transform the projective tensor product into a slice tensor product. Suppose we can find many $S = S_{1\dots m}$; we want them to be the slices. Our first task is to try to write $h(s)\delta_S$ as a suitable limit of functions $k^N_{1,\dots,m}(x)$ in the kernel of $\partial(\mathbf{g}_{1\dots m})$ as $N, m \to \infty$ suitably; here $h(s)$ is an "arbitrary" function on S. If we can accomplish this then we can write

$$\int h(s)f(s)\,ds = \lim \int f(x)k^N_{1,\dots,m}(x)\,dx. \tag{1.268}$$

Since h is arbitrary we can thus determine f on each S and since there are many S we can determine f.

The problem we face is in the same spirit as the one we met in the proof of Weierstrass' theorem on polynomial approximation in Section 1.1. Polynomials ($n = 1$) of degree $\le m$ form the kernel of d^{m+1}/dx^{m+1}; we want them to approximate δ_0. The spreads are the one-dimensional spaces spanned by the monomials. They are the solutions of $(xd/dx - n)f = 0$. The operator d^{m+1}/dx^{m+1} is the enveloping operator for $\{xd/dx - j\}_{j\le n}$.

One of the standard proofs of Weierstrass' theorem uses Taylor's theorem (with remainder) to approximate $t^{-1/2}\exp(-tx^2)$ by polynomials of degree m and to show that the remainder is small uniformly, say for $x \in [-1, 1]$. Taylor's formula with remainder is nothing else than the statement that the function $\chi_{[0,\infty)}x^m/m!$ is the fundamental solution for d^{m+1}/dx^{m+1}.

We conclude that the problem of directly transforming the projective tensor product into a slice tensor product involves estimates for fundamental (or null) solutions of $\partial(\mathbf{g}_{1\dots m})$ for properly chosen $\mathbf{g}_1, \dots, \mathbf{g}_m$.

We do not know how to carry out this program without using the Fourier transform. Certainly a direct method would be valuable especially for variable coefficient equations.

Problem 1.6 *Carry out this program.*

There is another direct method which depends on explicit calculations.

Consider the simplest example: lines in the plane. We identify the Grassmannian with the unit circle. Thus, for suitable f

$$\begin{aligned}\mathbf{R}^*\mathbf{R}f(0) &= \int \mathbf{R}^*\mathbf{R}f(0,\mathbf{g})\,d\theta \\ &= \iint f(r,\theta)\,dr\,d\theta \\ &= \iint f(x,y)\frac{dx\,dy}{r}. \end{aligned} \tag{1.269}$$

In general

$$\mathbf{R}^*\mathbf{R}f = f * \frac{1}{r}. \tag{1.270}$$

If we use l planes in $\mathbb{R}^n$ the measure on each such plane is of the form cr^{l-1}. It follows that

$$\mathbf{R}^*\mathbf{R}f = cf * r^{l-n}. \tag{1.271}$$

Now, r^{2k-n} is (up to a constant) the fundamental solution for Δ^k unless $2k = n$ when there is a log term. (Δ is the Laplacian.) Thus if $l = 2k$ we can obtain f from $\mathbf{R}^*\mathbf{R}$ by applying a simple operator, namely Δ^k. (Δ^k is a pseudodifferential operator when l is odd.)

Remark. $\mathbf{R}$ *and* $\mathbf{R}^*$ *are defined geometrically but* Δ^k *is a purely analytic object! Thus the inversion formula for the Radon transform leads to a new geometric insight into harmonicity. This is clarified in Section 2.1.*

In examples such as this, since $\mathbf{R}^*$ can be thought of as modified adjoint of $\mathbf{R}$, we can think of $\mathbf{R}$ as being "close to unitary."

In the case of a double fibration $\mathbf{R}^*\mathbf{R}$ is, at least formally, a convolution operator so it is often tractable by group representations.

There is a way of avoiding Δ^k and dealing with the inversion of the Radon transform in a purely geometric manner. For $k = 1$ the Laplacian can be defined geometrically

$$\Delta f(0) = \lim_{r\to 0} \frac{1}{r^2}\left[\int f(r,\theta)\,d\theta - f(0)\right]. \tag{1.272}$$

This generalizes the formula for the second derivative when $n = 1$. There are similar formulas for Δ^k; these involve limits of linear combinations of integrals over $k+1$ spheres (one sphere can be taken as the origin).

Problem 1.7 *Use the geometric definition of* Δ^k *to derive the inversion for* $\mathbf{R}$.

We can regard Lemma 1.1 as an illustration of a supra tensor product. Suppose $n = 2$. Let P be a polynomial of degree m and let $\mathbf{g}_1, \ldots, \mathbf{g}_{m+1}$ be $m+1$ distinct directions. Lemma 1.1 asserts that we can write

$$P = \sum P_{\mathbf{g}_j} \tag{1.273}$$

where $P_{\mathbf{g}_j}$ is a spread polynomial for $\mathbf{g}_j$ (meaning $P_{\mathbf{g}_j}$ is a constant in the direction of $\mathbf{g}_j$). This leads to the identities

$$\sum P_{\mathbf{g}_j} = \sum P_{\mathbf{g}'_j} \tag{1.274}$$

whenever $\mathbf{g}'_1, \ldots, \mathbf{g}'_{m+1}$ is another set of $m+1$ distinct directions.

If f is any function which is small at infinity we derive from (1.274)

$$\sum \int P_{\mathbf{g}_j}(s)\mathbf{R}f(s, \mathbf{g}_j)\, ds = \sum \int P_{\mathbf{g}'_j}(s)\mathbf{R}f(s, \mathbf{g}'_j)\, ds. \tag{1.275}$$

Equations (1.275) are called the *moment conditions*, a moment being the integral of a function times a polynomial. Equation (1.275) follow from (1.274) by applying Fubini's theorem to $\int P_{\mathbf{g}_j} f$ and integrating first in the directions of $\mathbf{g}_j$.

Identities (1.274) form a basis for all polynomial spread identities, so they define the amalgamation in

$$\mathbf{P} = \mathbf{P}(\mathbf{g}_0)\tilde{\otimes}\mathbf{P}(\mathbf{G}). \tag{1.276}$$

$\mathbf{P}$ is the space of polynomials, $\mathbf{P}(\mathbf{g}_0)$ is the space of spread polynomials for $\mathbf{g}_0$, and $\mathbf{P}(\mathbf{G})$ consists of finite sums of δ functions.

We have remarked that identities (1.274) form a basis for all polynomial amalgamations. It is remarkable that they also form a basis for amalgamations for tensor products of spaces of C^∞ functions which are small at infinity, when dealing with the hyperplane Radon transform. The analogous result for the l plane Radon transform with $l < n-1$ is not true. We determine the amalgamation in this case in Chapter 2 which deals with these questions in detail.

The examples of Radon transform that we have discussed thus far were related to the CP for hyperbolic operators. There are other interesting parametrization problems such as the Dirichlet problem (DP).

To illustrate the ideas, consider the Laplacian Δ operating on $\mathbb{R}^n$. Δ defines one spread and a corresponding Radon transform $\mathbf{R}_\Delta$. Any sphere S is a PS for solutions of Δ, since harmonic functions in $\mathbb{R}^n$ are uniquely determined by their restrictions to S. But the restrictions of such harmonic functions are very regular. In particular no δ_s is a restriction so this PP is not hyperbolic in the sense we have presented in Section 1.4.

Hyperbolicity uses the basis $\{\delta_s\}$ as its central ingredient. Indeed this is a geometric basis so hyperbolicity has geometric aspects. A natural basis related

to Δ is that given by spherical harmonics on S since they are related to the rotation group. Of course the expansion coefficients must decrease very rapidly in order to correspond to a solution on $\mathbb{R}^n$.

If we replace $\mathbb{R}^n$ by the ball Ω with $S = \text{bd } \Omega$ then the PP becomes hyperbolic; the δ_s are Dirichlet data. For the corresponding $\mathbf{R}^*$ the null solutions $\mathbf{R}^*\delta_s$ form the Poisson kernel. We can express δ_s in terms of spherical harmonics and thereby derive a PST by applying $\mathbf{R}^*$ to this identity.

All this refers to the single spread of harmonic functions. How do we construct other operators $\Delta(\mathbf{g})$ to obtain other coarse grains which fill the whole space? One obvious way is to replace the kernel of Δ by the eigenfunctions of Δ, i.e. solutions of $(\Delta - \lambda)h = 0$. But there is another possibility which, loosely speaking, corresponds to infinitesimal eigenvalues of arbitrary order. This latter idea is examined in detail in Chapters 3 and 4 under the title "harmonic functions."

Tensor products give a clear illustration of the distinction between the parametric and nonparametric Radon transform. We have already seen that the nonparametric Radon transform is "close to" a tensor product of the coarse grains which are the spreads. On the other hand, the parametric Radon transform appears as one factor of a tensor-like decomposition of the solutions of an enveloping equation (see Section 1.5).

We can put this tensor factorization in a more general setting. Let V be an algebraic variety and H a linear group that acts on V. Let $P_1, \ldots, P_r$ be generators for the ideal of V so the system of differential equations $\vec{P}(D)f = 0$ is H invariant if H is orthogonal for the same quadratic form which we use to define Fourier transformation.

Consider functions f in some space $\mathcal{W}$. Suppose we have a sufficient set (see Section 1.4) $V^0 \subset V$ on which the orbits of H are all isomorphic, meaning the isotropy subgroups of all $v^0 \in V^0$ are all conjugate in H. Then under reasonable conditions we can find a single set of polynomials $\{h_j(\hat{t})\}$ which forms a basis on every H orbit on V^0. $\{h_j\}$ is called a *harmonic basis*. Usually these h_j are representation functions for H. For example, if V is the light cone $\sum \hat{t}_j^2 = \hat{y}^2$ and H is the orthogonal group in $\hat{t}$ (note the change from the usual light cone notation) then the homogeneous harmonics in $\hat{t}$ provide such a basis. (The theory of harmonic bases is dealt with in detail in Chapters 3 and 4.)

If $\hat{Y}$ is a linear space representing a cross-section of the orbits then we might expect a tensor product decomposition

$$\mathcal{W}(V) \approx \mathcal{W}(\hat{Y}) \otimes \mathcal{W}(\hat{T}) \tag{1.277}$$

where $\mathcal{W}(\hat{T})$ is a suitable space of functions on the generic orbit. The passage from V^0 to V is accomplished by analyzing degenerate orbits and Phragmén–Lindelöf inequalities. The tensor product in (1.277) might be a sub or supra tensor product.

From another point of view we can regard (1.277) as the general form of John's first equation or rather, its solution (see Section 1.2 and Chapter 6).

When the fibers' o of V^0, which are the H orbits, are algebraic varieties then we can regard $\{P_j(\hat{x}) = 0\}$ as the enveloping equation of the equations defining the various orbits o.

The second John equation involves the behavior of solutions of $\vec{P}(D)f = 0$ under H. We noted in Section 1.6 that, when dealing with the usual Radon transform related to lines in $\mathbb{R}^n$, this equation amounted to H invariance. We should expect (and indeed this is so in many cases) that an H invariant f should be represented as the Fourier transform of an H invariant measure $\hat{f}$ on V^0. Such an $\hat{f}$ would be identified with a function of $\hat{y}$ in accordance with the tensor decomposition (1.277). In the case of John's equations this idea represents our approach to proving that John's equations determine the range of the parametric Radon transform.

A general John theorem depends on the

Orbit principle. *Suppose $\vec{P}(D)$ defines an H invariant system. Suppose there exists a sufficient set V^0 of the corresponding algebraic variety V on which the orbits of H are all isomorphic. Suppose f is a solution of $\vec{P}(D)f = 0$ which transforms according to an irreducible representation ρ of H. Then f has a Fourier representation by a measure $\hat{f}$ supported on V^0 which transforms on each orbit according to ρ. When a harmonic basis $\{h_j(\hat{t})\}$ exists then $\hat{f}$ depends only on those basis elements for which h_j belong to the representation ρ.*

The orbit principle is very delicate. Its validity depends on $\vec{P}, H,$ and $\mathcal{W}$. Important examples of the orbit principle are studied in Chapter 6.

If H is not compact (as in the usual parametric Radon transform) then, in the orbit principle, we cannot assume that f is small at infinity. In such cases we shall usually work in spaces like $\mathcal{S}'$.

Remark. *Often the most difficult part of establishing the orbit principle or the tensor product (1.277) is the establishment of the sufficiency of V^0.*

We have discussed the orbit principle and John's second equations in terms of H invariant functions or, more generally, functions which transform according to representations of H. More generally we can consider the parametric Radon transform with an arbitrary attenuation μ

$$\mathbf{R}^\mu f(a,b) = \int f(at+b)\, d\mu(t). \tag{1.278}$$

For varying μ, f we should expect $\{\mathbf{R}^\mu f\}$ to be essentially all solutions of the first John equation and so correspond, via the Fourier transform, to "all" suitable measures on V. Thinking in terms of the tensor product decomposition (1.277) leads to the idea that a generic μ^0, which corresponds to an explicit structure on $\hat{Y}$, provides an isomorphism between $\{f\}$ and $\{\mathbf{R}^{\mu^0} f\}$, or a second John equation. This is a more general form of the orbit principle. Thus $\mathbf{R}^\mu f$ should determine both μ and f (modulo changing μ to $c\mu$ and f to $c^{-1}f$). Indeed this is the case as shown in Chapter 6.

Conversely a fixed generic f^0 leads to an isomorphism between $\{\mu\}$ and $\{\mathbf{R}^\mu f^0\}$. Fixing f^0 and studying the correspondence between μ and $\mathbf{R}^\mu f^0$ when $n = 1$ is one of the central problems studied in connection with *wavelets.*

There is a curiosity regarding the restriction operators

$$\nu(\mathbf{g})\hat{f} = \hat{f}\delta_{L(0,\mathbf{g})} \tag{1.279}$$

where $\mathbf{g} \in G$ is an arbitrary line through the origin. The operators $\nu(\mathbf{g})$ satisfy many relations. For simplicity suppose $n = 2$.

If $(\mathbf{g}_1, \mathbf{g}_2)$ and $(\mathbf{g}_3, \mathbf{g}_4)$ are any pairs of lines which meet at the same nonzero angle then $\nu(\mathbf{g}_1)\nu(\mathbf{g}_2) = \nu(\mathbf{g}_3)\nu(\mathbf{g}_4)$. There are also three term relations amongst $\{\nu(\mathbf{g})\}$ and relations that arise from differentiation.

Let $\hat{u}(\mathbf{g})$ be functions defined for each $\mathbf{g}$ with support $\hat{u}(\mathbf{g}) \subset L(0, \mathbf{g})$. Suppose the $\hat{u}(\mathbf{g})$ satisfy the necessary conditions imposed on $\nu(\mathbf{g})\hat{f}$ by the operator relations for the operators $\nu(\mathbf{g})$. Then the moment conditions imply, assuming suitable regularity of $\hat{u}(\mathbf{g})$, that there is an $\hat{f}$ such that $\hat{u}(\mathbf{g}) = \nu(\mathbf{g})\hat{f}$. Put in other terms, the first cohomology group of $\{\nu(\mathbf{g})\}$ with coefficients in $\hat{W}$ vanishes.

Problem 1.8 *Do all cohomology groups vanish?*

Remark. *The cohomology group is defined in a nonstandard way because* $\{\nu(\mathbf{g})\}$ *is not a group since* ν^2 *is not defined. Nevertheless the triviality of the cohomology groups can be defined as they assert that certain relations amongst the* $\nu(\mathbf{g})$ *come from module-type relations.*

A study of this cohomology theory is presented at the end of Section 2.3.

2

THE NONPARAMETRIC RADON TRANSFORM

In Section 2.1 we prove the injectivity of the nonparametric Radon transform and we compute its range. The range conditions for the hyperplane Radon transform and $\mathbf{R}^l$ for $l < n-1$ are different. Explicit inversion formulas are given. We also study the spherical Radon transform $\mathbf{R}_\Omega$ in which integration is over all spheres tangent to the strictly convex compact hypersurface Ω; we show it is injective in odd dimensions.

Section 2.2 deals with the topology of the Radon transform. For $l < n-1$, $\mathbf{R}^l$ is a topological isomorphism onto the range in its natural tensor product topology. But in the hyperplane case the topological isomorphism fails for some spaces, e.g. $\mathcal{D}$. This result is analogous to a theorem of Hertle [96, 97].

Section 2.3 is a take-off on the theorem of Helgason that relates the convex hull of the support of a function f to the convex hull of the support of $\mathbf{R}^{n-1}f$ when the latter is compact. Variations of Helgason's theorem are discussed, in particular to the case when the support of $\mathbf{R}^{n-1}f$ is not compact and for $\mathbf{R}^l$ when $l < n-1$.

2.1 Radon transform and Fourier transform

This section is, for the most part, devoted to the classical Radon transform defined by integration over affine subspaces of $\mathbb{R}^n$. At the end of this section we replace affine subspaces by spheres.

We recall some of the constructions of Chapter 1. Let $\mathbf{G}(n,l)$ be the Grassmann variety of unoriented l planes through the origin in $\mathbb{R}^n$. For any $L \in \mathbf{G}(n,l)$ we denote by $L^\perp$ the orthogonal $n-l$ plane through the origin. It is standard (see [123]) that

$$\mathbf{G}(n,l) = \mathcal{O}(n)/\mathcal{O}(l) \times \mathcal{O}(n-l)$$

where $\mathcal{O}(n)$ is the orthogonal group.

For the (most) classical Radon transform $l = n-1$, so $\mathbf{G}(n,l) = \mathbf{P}^{n-1}$ is $n-1$ dimensional projective space.

Now let f be any smooth function on $\mathbb{R}^n$ which is small at infinity. We want to form the set of integrals of f over l planes. To do this in a systematic way we choose an l plane L through the origin, i.e. a point $\mathbf{g}$ in $\mathbf{G}(n,l)$, and consider the spread which is the set of l planes $\tilde{L} = L(s,\mathbf{g})$ not necessarily through the origin which are parallel to L. Each $\tilde{L}$ is uniquely determined by the point $\tilde{L} \cap L^\perp$. Thus the set of all planes is parametrized by $L \in \mathbf{G}(n,l)$ and a point $s \in S = L^\perp$,

which is the spread parameter. We call such planes *affine planes*; they form the natural, or tautological, bundle over the Grassmannian.

We set

$$\mathbf{R}^l f(s, \mathbf{g}) = \int_{L(s,\mathbf{g})} f \tag{2.1}$$

the integral being taken over the l plane which is parallel to L and intersects $L^\perp$ at s. $\mathbf{R}^l f$ is thus a function on the natural $n-l$ plane bundle over $\mathbf{G}(n,l)$. For fixed $\mathbf{g}$, $\mathbf{R}^l f$ is a function on $L^\perp = S_\mathbf{g}$. As mentioned in Chapter 1 we can use oriented planes if we set

$$\mathbf{R}f(s, \mathbf{g}) = \mathbf{R}f(-s, -\mathbf{g}).$$

In Chapter 1 we gave a general abstract definition of the Radon transform which places it in the framework of functional analysis. We consider f to be an element of a function space $\mathcal{W}$, so we can regard $f \in \mathcal{W}''$. Any closed subspace $\mathcal{W}'_P$ of $\mathcal{W}'$ defines a Radon transform $\mathbf{R}_P$ where

$$\mathbf{R}_P(f) = \text{restriction of } f \text{ to } \mathcal{W}'_P.$$

The spread $\mathbf{g} \in G(n,l)$ defines a closed subspace of $\mathcal{W}'$, namely the space $\mathcal{W}'_\mathbf{g}$ of spread functions. These are the functions[1] which are constant on the leaves of the spread. In the present case this means functions constant in the directions of L. Functions (distributions) in $\mathcal{W}'_\mathbf{g}$ are determined by their restrictions to the cross-section $S_\mathbf{g} = L^\perp$. Thus the restriction of f on $\mathcal{W}'_\mathbf{g}$ can be defined by the function $\mathbf{R}_\mathbf{g} f\big|_{L^\perp}$ which we can think of as $\mathbf{R}^G_\mathbf{g} f$, the geometric representation of $\mathbf{R}_\mathbf{g}$.

In Section 1.7 we defined $\mathbf{R}^{l*}$ when $\mathbf{g}$ is fixed; it is a sort of adjoint of $\mathbf{R}^l$. For functions h on $S_\mathbf{g}$, $\mathbf{R}^{l*}h$ is the function on $\mathbb{R}^n$ which is constant in directions orthogonal to $S_\mathbf{g}$ and which equals h on $S_\mathbf{g}$. In particular, $\mathbf{R}^{l*}\mathbf{R}^l f$ replaces f by the spread function for $\mathbf{g}$ which is equal to $\mathbf{R}f$ on $S_\mathbf{g}$. Put in other terms, $\mathbf{R}^*\mathbf{R}f$ is the solution of the Cauchy problem for the system of differential equations defined by directional derivatives on L which agrees with $\mathbf{R}f$ on the Cauchy Surface $S_\mathbf{g} = L^\perp$.

Remark. *We sometimes write* $\mathbf{g}$ *for* L *and* $\mathbf{g}^\perp$ *for* $L^\perp$.

We can express $\mathbf{R}^{l*}\mathbf{R}^l f$ (for $\mathbf{g}$ fixed) by means of convolution. In fact it is clear that

$$\mathbf{R}^{l*}\mathbf{R}^l f = \delta_L * f = \mathbf{R}^l f \times \delta_L. \tag{2.2}$$

Being a convolution it is natural to analyze $\mathbf{R}^{l*}\mathbf{R}^l$ by Fourier transformation. We have

$$\widehat{\delta_L * f} = \hat{f}\delta_{\hat{L}^\perp} = \hat{f}\big|_{\hat{L}^\perp}\delta_{\hat{L}^\perp}. \tag{2.3}$$

[1]Recall our convention of Chapter 1: we use the word "function" to include "measure" and "distribution" when no confusion is possible.

Our notation is such that $\hat{L}$ is the same plane as L but $\hat{L}$ is in the dual space. (More precisely, the quadratic form defining Fourier transform identifies $\hat{L}$ with L.)

We can think of the identity (2.3) in a somewhat different light. By comparing the third terms in (2.2) and (2.3) we see that $\hat{f}|_{\hat{L}^\perp}$ is the Fourier transform on $L^\perp$ of $\mathbf{R}^l f|_{L^\perp}$.

This is a form of the **projection–slice** theorem:

The n-dimensional Fourier transform of f restricted to the slice $\hat{L}^\perp$ is the function on $\hat{L}^\perp$ which is the $n-l$ dimensional Fourier transform of $\mathbf{R}^l f(s, L)$ (projection) in the spread parameter s.

Remark. *We call $\mathbf{R}^l(s, L)$ a projection because the value at the point s is obtained by applying a nonlocal operator to f, namely integration on the leaf through s. Restriction is called a slice because the value of the restriction at a point $\hat{s} \in \hat{L}^\perp$ depends on the value at $\hat{s}$ of a local operator applied to $\hat{f}_j$. In this case the local operator is the identity.*

Formula (2.3) suggests that we should analyze $\hat{f}$ using polar coordinates. There is a useful formalism related to polar coordinates and to Fubini's theorem. If we have any coordinate system (u, v) on a manifold and a measure $dx = d\mu(u)\, d\nu(v)$ then, in terms of the concept of wave packets associated to bases as described in (1.17)ff., we can write

$$dx = \int \delta_u \, d\mu(u)$$

where δ_{u^0} is the δ function of the manifold $u = u^0$, meaning the measure $d\nu(v)$ on this set. This expression is a formal way of writing Fubini's theorem

$$\begin{aligned} dx \cdot f = \int f(x)\, dx &= \iint f(u, v)\, d\mu(u)\, d\nu(v) \\ &= \int d\mu(u) \int f(u, v)\, d\nu(v) \\ &= \int (\delta_u \cdot f)\, d\mu(u). \end{aligned}$$

Note that the measures $\delta_{\hat{L}^\perp}$ have no common zeros so the Tauberian principle (Section 1.6) applies. In fact we can explicitly verify that the Tauberian principle is valid in this case; this gives the injectivity of $\mathbf{R}^l$. We can even go further and give an explicit inversion of $\mathbf{R}^l$. We start by expressing the euclidean measure as an integral of the $\delta_{\hat{L}^\perp}$. In terms of polar coordinates on each $\hat{L}^\perp$ we write

$$d\hat{x}_{\hat{L}^\perp} = \hat{r}_{\hat{L}^\perp}^{n-l-1}\, d\hat{r}_{\hat{L}^\perp} d\hat{\theta}_{\hat{L}^\perp}. \tag{2.4}$$

Here $d\hat{\theta}_{\hat{L}^\perp}$ is the standard measure on the unit sphere in $\hat{L}^\perp$. For points in $\hat{L}^\perp$ the polar coordinate $\hat{r}_{\hat{L}^\perp}$ coincides with the polar coordinate $\hat{r}$ of $\hat{\mathbb{R}}^n$. Thus if $\hat{\theta}$ denotes the angular coordinate in $\mathbb{R}^n$,

$$d\hat{x} = \hat{r}^{n-1}\, d\hat{r}\, d\hat{\theta} = c_0 \hat{r}^l\, d\hat{x}_{\hat{L}^\perp} d\hat{\mathbf{g}} \tag{2.5}$$

where $d\hat{\mathbf{g}}$ is the $SO(n)$ invariant measure on $\mathbf{G}(n,l)$.

We multiply (2.5) by $\hat{f}$ and take the inverse Fourier transform; that is, we multiply by $\exp(ix\cdot\hat{x})$ and integrate. We evaluate the Fourier transform by integration first over each fixed $\hat{L}^\perp$ and then over $\{\hat{L}^\perp\}$. This is the same as our above "basis" description of Fubini's theorem. By (2.3) the inverse Fourier transform of $\hat{f}\big|_{\hat{L}^\perp}\delta_{\hat{L}^\perp}$ is $\delta_L * f$. By comparing the first and last integrals arising from (2.5) we obtain

$$\begin{aligned} f(x) &= c_0 \int e^{ix\cdot\hat{x}} \hat{f}(\hat{x}) \hat{r}^l \delta_{\hat{L}^\perp}\, d\mathbf{g} \\ &= c_0 \Delta^{l/2} \int (\delta_L * f)(x)\, d\mathbf{g} \qquad (2.3) \quad \text{and } \hat{r}^l = \Delta^{l/2} \\ &= c_0 \Delta^{l/2} \int \mathbf{R}^*\mathbf{R} f(s,\mathbf{g})\, d\mathbf{g} \qquad (2.2) \\ &= c_0 \Delta^{l/2} \int \mathbf{R} f(\mathbf{g}^\perp \cdot x, \mathbf{g})\, d\mathbf{g} \qquad s = \mathbf{g}^\perp \cdot x \end{aligned} \tag{2.6}$$

where $\mathbf{g}^\perp$ is a unit vector orthogonal to $L\,(\mathbf{g})$ and Δ is the Laplacian on $\mathbb{R}^n$.

Note that $(\delta_L * f)(x)$, or any of the other forms of the integrand, depends only on the affine planes parallel to L incident to (passing through) x. ($\mathbf{R}^*\mathbf{R}f(s,L)$ has a nonzero value at x only when $s = g^\perp \cdot x$ for some $g^\perp \in \mathbf{g}^\perp$, for then the plane $L + sg^\perp$ passes through x. For such s the value at x is $\mathbf{R}f(g^\perp \cdot x, L)$.) When l is an even integer this shows that $f(x)$ is determined by the integrals of f on affine planes passing through an infinitesimal neighborhood of x. But when l is odd $\Delta^{l/2}$ is a pseudo-differential operator which is not local so we cannot make this assertion.

We should observe that the expression (2.5) for $d\hat{x}$ is subject to many variations. We can multiply $d\hat{x}_{\hat{L}^\perp}$ by a nonvanishing function $\hat{\alpha}(\hat{r})$. (We could also introduce functions $\hat{\alpha}(\hat{r},\hat{\theta})$ but we have not pursued this possibility.) Then formula (2.5) can be rewritten as

$$d\hat{x} = c_0 \hat{r}^l \hat{\alpha}^{-1}(\hat{r}) \hat{\alpha}(\hat{r}) d\hat{x}_{\hat{L}^\perp}\, d\hat{\mathbf{g}}^\perp. \tag{2.5*}$$

When l is odd one simple choice for $\hat{\alpha}(\hat{r})$ is $\hat{r}^{-1}$, in which case

$$d\hat{x} = c_0 \hat{r}^{l+1} \hat{r}^{-1} d\hat{x}_{\hat{L}^\perp} d\hat{\mathbf{g}}^\perp.$$

We write

$$\hat{f}\, d\hat{x} = c_0 \hat{r}^{l+1} \left[\hat{r}^{-1} \hat{f} \delta_{\hat{L}^\perp}\, d\hat{\mathbf{g}}^\perp\right]$$

in accordance with our above discussion. We replace $\hat{f}$ by $\exp(ix\cdot\hat{x})\hat{f}$ and integrate; (2.6) is modified to

$$f(x) = c_0\Delta^{(l+1)/2}\int \widehat{\hat{r}^{-1}\hat{f}\delta_{\hat{L}^\perp}}\, d\hat{\mathbf{g}}^\perp. \tag{2.6*}$$

In the case of the hyperplane Radon transform $\hat{L}^\perp$ is one dimensional so $\hat{r}^{-1}d\hat{x}_{\hat{L}^\perp}$ is the measure $\hat{r}^{-1}$ on this line. The one-dimensional Fourier transform of $|\hat{x}|^{-1}$ is $\log|x|$. (This is seen most easily from the fact that $\hat{x}\,|\hat{x}|^{-1}$ is the Heaviside function (-1 for $x<0$ and $+1$ for $x>0$) whose Fourier transform is x^{-1} agreeing with the derivative of $\log|x|$. The Fourier transform of multiplication by $\hat{x}$ is differentiation.) This means that the n-dimensional Fourier transform of $\hat{r}^{-1}\delta_{\hat{L}^\perp}$ is $\log r$ in the direction of $L^\perp$ and constant in the direction of L. The Fourier transform in (2.6*) is

$$(\widehat{\hat{r}^{-1}\delta_L} * f)(x) = \int \log|s'|[\mathbf{R}^*\mathbf{R}f(s-s',\mathbf{g})]\,ds'$$

where $s = s(x,L) = x\cdot\mathbf{g}^\perp$ so we can write

$$f(x) = c\Delta^{(l+1)/2}\int d\mathbf{g}\int \log|s'|[\mathbf{R}f(x\cdot\mathbf{g}^\perp - s',\mathbf{g})]\,ds'. \tag{2.6**}$$

For l odd (2.6**) represents a gain over (2.6) in that the pseudo-differential (nonlocal) operator $\Delta^{l/2}$ is replaced by the differential operator $\Delta^{(l+1)/2}$. But it also represents a loss because the integration in (2.6**) involves $\log|s'|$ instead of $\delta_{s'}$ which occurs in (2.6).

We have shown how to use $\hat{\alpha}(\hat{r}) = \hat{r}^{-1}$. Another simple choice is $\hat{\alpha}(\hat{r}) = \hat{r}$ as this changes l into $l-1$. For the hyperplane Radon transform, since $\hat{L}^\perp$ has dimension 1, we can write $\hat{r} = |\hat{x}|$ on $\hat{L}^\perp$. Since $|\hat{x}|$ is the fundamental solution for $d^2/d\hat{x}^2$, its Fourier transform is x^{-2}. This means that the factor $\log|s'|$ in (2.6**) is replaced by $(s')^{-2}$ and the power $(l+1)/2$ of Δ is replaced by $(l-1)/2$.

In his book John [100] replaces $(s')^{-2}$ by $(s')^{-1}$ with a consequent change of $\mathbf{R}f$ to $d(\mathbf{R}f)/ds'$. Naturally there are many choices for $\hat{\alpha}$ and many ways of writing the resulting formulas. Nevertheless it is impossible to write a purely local formula when n is even (see [77]) and, in fact, if $\mathbf{R}f(s,L)$ vanishes for all L when $|s|<1$ then f may not vanish in the unit disk.

When $l<n-1$ the situation is quite similar. We obtain local formulas when l is even but not when l is odd. The pseudo-differential operator $\Delta^{l/2}$ (l odd) can be replaced by $\Delta^{(l-m)/2}$ by introducing $\hat{\alpha}(\hat{r}) = \hat{r}^m_{\hat{L}^\perp}$ with m odd. Now $\hat{r}_{\hat{L}^\perp}$ represents an $n-l$ dimensional distance function. The $n-l$ dimensional Fourier transform (up to a constant)

$$\widehat{\hat{r}^m_{\hat{L}^\perp}} = \begin{cases} r_{L^\perp}^{-m-n+l}, & \text{if } -m-n+l\neq 0 \\ \log r_{L^\perp}, & \text{if } -m-n+l = 0. \end{cases}$$

(This can be verified by homogeneity considerations and the same type of argument used when dim $L^\perp = 1$.)

The dimension of $\{(s, L)\}$ is

$$\dim \{(s, L)\} = (l + 1)(n - l)$$

because of the standard result that $\mathbf{G}(n, n - l)$ has dimension $l(n - l)$. Observe that $(l + 1)(n - l) > n$ unless $l = n - 1$ which is the usual hyperplane Radon transform.

Because of this difference in dimension it seems reasonable to search for n-dimensional sets of $\{s, \mathbf{g}\}$ so that $\{\delta_{L(s,\mathbf{g})}\}$ forms a basis for $\mathcal{W}'$ (the admissibility problem). For that we need to find an l-dimensional family of $\{\hat{\mathbf{g}}^\perp\}$ which covers $\mathbb{R}^n$ with "little overlap." One interesting choice (actually, set of choices) can be made as follows. Choose a fixed plane $\hat{\mathbf{p}}$ of dimension $n - l - 1$. Then the family $\{\hat{\mathbf{g}}^\perp\}$ consists of all $n - l$ planes through the origin containing $\hat{\mathbf{p}}$. Any such $\hat{\mathbf{g}}^\perp$ is determined uniquely by a line through the origin, orthogonal to $\hat{\mathbf{p}}$, so the family $\{\hat{\mathbf{g}}^\perp\}$, or more precisely $\{\hat{\mathbf{g}}^\perp\}(\mathbf{p})$, is parametrized by the projective plane in $(\hat{\mathbf{p}})^\perp$ which is essentially the unit sphere S^l in $(\hat{\mathbf{p}})^\perp$.

We are in a situation which is analogous to that of the hyperplane transform in the plane orthogonal to $\hat{\mathbf{p}}^\perp$. For any $\hat{\theta}$ the restriction of $\hat{f}$ to the plane $\hat{\mathbf{p}}^\perp(\hat{\theta}) = \hat{\mathbf{p}}^\perp + \text{line}\,\hat{\theta}$ is the Fourier transform of $f * \delta_{\mathbf{p}(\theta)}$.

Call $d\hat{\theta}$ the invariant measure on S^l. Then instead of (2.5) we can decompose the measure $d\hat{x}$ in cylindrical coordinates, meaning $\hat{x}_{\hat{\mathbf{p}}}$ on $\hat{\mathbf{p}}$ and $\hat{\theta}$. We find

$$d\hat{x} = \hat{r}^l_{\hat{\mathbf{p}}^\perp}\, d\hat{r}_{\hat{\mathbf{p}}^\perp}\, d\hat{\theta}\, d\hat{x}_{\hat{\mathbf{p}}}. \tag{2.5**}$$

This leads to

$$f(x) = c_1 \Delta^{l/2}_{\mathbf{p}^\perp} \int (\delta_{\mathbf{g}} * f)(x)\, d\hat{\theta}. \tag{2.6***}$$

One might be puzzled at the appearance of the factor $\Delta^{l/2}$ in (2.6) or (2.6***). For, $\mathbf{R}$ is defined in an essentially geometric fashion and Δ is an analytic object. The relation between $\mathbf{R}$ and Δ lies at the heart of the potential theory. To understand this, suppose first that $n = 3$, $l = 1$. Since $\mathbf{R}$ and Δ are linear it suffices (formally) to take $f = \delta_{x^0}$. (Recall that $\{\delta_x\}$ is a basis.)

Of course we cannot take the 1 plane Radon transform $\mathbf{R}^1\delta_{x^0}$ directly because the integral is not defined. But if we think of δ_{x^0} as being the limit of $N^3\chi(\gamma(x^0, N^{-1}))$ where $\gamma(x^0, N^{-1})$ is the cube centered at x^0, sides of length $1/N$ which are parallel to the x_j coordinate axis, and χ denotes the characteristic function, then the integrals over lines M parallel to the x_j axis are N^2 if the j' coordinates of M are $\leq N^{-1}$ for all $j' \neq j$; the integrals vanish otherwise. Thus it makes sense to say that

$$(\mathbf{R}^{1*}\mathbf{R}^1\delta_0)(x_j \text{ axis}) = \delta_{x_j \text{ axis}}.$$

More generally,

$$(\mathbf{R}^{1*}\mathbf{R}^1\delta_0)(L) = \delta_L.$$

Hence, by integration,

$$\int (\mathbf{R}^{1*}\mathbf{R}^1\delta_0)(L)\, dL = \int \delta_L\, dL = cr^{-2}\, dr\, d\theta$$

because of the change of measure, as before. The last equality should be thought of in the framework of the theory of distributions, meaning that we apply both sides to a smooth function of compact support.

Note that r^{-2} is not harmonic outside the origin. But if we took the hyperplane Radon transform $\mathbf{R}^2$ then we would find by the same method

$$\int (\mathbf{R}^{2*}\mathbf{R}^2\delta_0)(L)\, dL = cr^{-1}\, dr\, d\theta \tag{2.7}$$

which is harmonic outside the origin and is, in fact, the fundamental potential.

Similar formulas are valid for arbitrary n and l.

We can think of $\mathbf{R}^{l*}\mathbf{R}^l\delta_{x^0}$ as a "radiation field" produced by a point source at x^0. In particular, when $n = 3, l = 2$ the radiation field is the fundamental solution for Δ.

Harmonicity in $\mathbb{R}^n$ can also be understood from the geometric definition of Δ:

$$\Delta f(x) = \lim_{r\to 0} r^{-2} \int_{S_r(x)} [f - f(0)] \tag{2.8}$$

where $S_r(x)$ is the sphere centered at x, radius r. (This formula can be verified either directly or using

$$\int_{S_r(x)} f = (f * \delta_{S_r(0)})(x) \tag{2.9}$$

and the fact that the Fourier transform of $\delta_{S_1(0)}$ is a suitable Bessel function.) Harmonicity is equivalent to the mean value property over spheres. (In fact, by a theorem of Delsarte [30] (see also [157]) the mean value property over spheres of two radii r_1, r_2 with arbitrary centers suffices to guarantee harmonicity if r_1/r_2 is not a zero of a suitable Bessel function.)

For integral k, $\Delta^k f(x)$ can also be given a geometric definition in terms of the limit of a well-defined linear combination of averages of f over $k+1$ spheres with center x. Similarly the kernel of Δ^k consists of all functions for which such linear combinations of averages vanish identically.

It is possible to verify geometrically that these spherical averages of $\mathbf{R}^{l*}\mathbf{R}^l\delta_0$ corresponding to Δ^k vanish when $l = 2k$ for $x \neq 0$.

Remark. *This method represents a subtle relation between the linear structure of $\mathbb{R}^n$ as given by l planes and the rotation structure. Both linear and rotation*

structures are embedded in the affine group A so a deep penetration of A should reveal the secrets as to why this method works.

A striking interplay of the linear and rotational structures is found in the interpretation of $\mathbf{R}^*\mathbf{R}$ given in Section 1.7. For any given spread $\mathbf{g}$, $\mathbf{R}^*\mathbf{R}f(\mathbf{g})$ represents the projection of f on the constant term in the linear Fourier analysis on the leaves of $\mathbf{g}$. Integration of these projections over the Grassmannian G (rotational structure) is "almost," i.e. modulo applying a power of Δ, the same as taking the value of f at the origin.

The method that we have given for dealing with the Radon transform uses the fiber structure of $\hat{\mathbb{R}}^n$ by $n-l$ planes through the origin with base $\mathbf{G}(n, n-l)$. This leads to the slice decomposition of functions on $\hat{\mathbb{R}}^n$ as integrals over the base of functions on the fibers (slice subdirect integral).

For $l = 1$ we could reverse matters and use the decomposition of $\hat{\mathbb{R}}^n$ by slices which are spheres $S_{\hat{r}}$ centered at the origin of radius $\hat{r}$. A cross-section is the real axis (modulo $\pm$). The Fourier transform of a measure supported on $S_{\hat{r}}$ is an eigenfunction of the Laplacian, i.e. a function ϕ satisfying

$$(\Delta - \hat{r}^2)\phi(x; \hat{r}) = 0.$$

As in the case of the hyperplane Radon transform we can regard the Fourier transform on a slice $S_{\hat{r}}$ as defining a spread $[\hat{r}]$ whose spread functions are eigenfuctions of Δ with eigenvalue $\hat{r}^2$. (This spread does not seem to have a geometric description in x space.) $[\hat{r}]$ can be regarded as a closed subspace of the dual space $\mathcal{W}'$ as in Section 1.7. Thus $[\hat{r}]$ defines a Radon transform on $\mathcal{W} = \mathcal{W}''$, namely

$$\mathbf{R}_{\hat{r}}f = \text{restriction of } f \text{ to the space of } [\hat{r}] \text{ spread functions.}$$

To define $\mathbf{R}^*_{\hat{r}}$ we need a basis for $[\hat{r}]$ spread functions. The fundamental principle provides us with a sort of basis $\{\exp(ix\cdot\hat{x})\}_{\hat{x}\in S_{\hat{r}}}$. We say "sort of" because, as pointed out in Section 1.4, this basis is highly overdetermined, meaning expansions in it are highly nonunique if the functions in $\mathcal{W}$ are so small at infinity that their Fourier transforms are holomorphic, in which case we need the complex sphere $S^{\mathbb{C}}_{\hat{r}}$. (See Chapter 9 for an analysis of the nonuniqueness.)

In any case $\mathbf{R}_{\hat{r}}f$ is determined by the function $\mathbf{R}^a_{\hat{r}}f(\hat{x}) = f\cdot\exp(ix\cdot\hat{x}) = \hat{f}(\hat{x})$ for $\hat{x} \in S^{\mathbb{C}}_{\hat{r}}$. We should like to define $\mathbf{R}^*_{\hat{r}}\mathbf{R}_{\hat{r}}f$ by first injecting $\mathbf{R}^a_{\hat{r}}f$ in $\mathcal{W}'$ and then applying $\mathbf{R}'$. This yields

$$\mathbf{R}^*_{\hat{r}}\mathbf{R}_{\hat{r}}f = \int_{S^{\mathbb{C}}_{\hat{r}}} \mathbf{R}^a_{\hat{r}}f(\hat{x})e^{-ix\cdot\hat{x}}\, d\hat{x}. \tag{2.10}$$

This integral certainly does not converge absolutely because $\mathbf{R}^a_{\hat{r}}f$ grows exponentially in imaginary directions. In fact, it does not seem possible to make sense of this integral for $f \in \mathcal{D}$ or $f \in \mathcal{E}'$. We must therefore restrict our theory to spaces like $\mathcal{S}$ for which the relevant sphere $\{\hat{x}^2 = \hat{r}^2\}$ is real and compact. In this case

we can define $\mathbf{R}^*\mathbf{R}f(x) = \int \exp(-ix\cdot\hat{x})\hat{f}(\hat{x})\,d\hat{x}$. This is the solution of $\Delta - \hat{r}^2$ whose value on $\exp(ix\cdot\hat{s})$ is $\hat{f}(\hat{s})$ in accordance with the general principles set forth in Section 1.7.

If we set $\mathcal{W} = \mathcal{S}$ then $S_{\hat{r}}$ is a real sphere (for $\hat{r}$ real). Thus we can use the basis of spherical harmonics on $S_{\hat{r}}$ instead of the δ functions of points. This gives us a new definition of $\mathbf{R}^*\mathbf{R}$.

We express

$$\delta_{\hat{x}^0} = \sum \bar{\phi}(j;\hat{x}^0)\phi(j;\hat{x})$$

on $S_{\hat{r}}$. Here $\phi(j;\hat{x})$ are the normalized spherical harmonics on $S_{\hat{r}}$. Taking the Fourier transform leads to the identity

$$e^{ix\cdot\hat{x}^0} = \sum \bar{\phi}(j,\hat{x}^0)J^j(x) \tag{2.11}$$

where $J^j = \hat{\phi}(j;\ )$ is a suitable Bessel function. (x is a variable in $\mathbb{R}^n$ but $\hat{x}$ is restricted to $S_{\hat{r}}$.) For example, when $n = 2$, $S_{\hat{r}}$ is the circle of radius $\hat{r}$ in the plane, $\phi(j;\hat{x}) = \delta_{S_{\hat{r}}}\exp ij\hat{x}$, and $\hat{\phi}$ is the standard Bessel function [11, vol. II].

There is a Poisson summation formula (PSF) associated to this decomposition. In order to formulate it we need a hyperbolic (geometric) PP for solutions of $\Delta - \hat{r}^2$. Unfortunately we do not know of any PP for solutions of $\Delta - \hat{r}$ in any reasonable space of functions on all of $\mathbb{R}^n$. We can, however, use the Dirichlet problem (DP) on S_1 for $\mathcal{W} = \mathcal{E}(B_1)$ where B_1 is the closed ball of radius 1 if $\hat{r}$ is not an eigenvalue of Δ for the DP, i.e. there is no solution whose DD $\equiv 0$.

We use the basis of null solutions $\eta(x^0;x)$ whose DD on S_1 is δ_{x^0}. Thus $\mathbf{R}f$ can be identified (see Section 1.7) with

$$\mathbf{R}^G f(x^0) = f\cdot\eta(x^0;x).$$

Again we have a second basis, namely $\{\phi(j,x)\}$. Expanding δ_{x^0} in terms of $\{\phi(j,x)\}$ and then applying $\mathbf{R}^*$, which sends a function h on S_1 into the solution of the DP whose DD is h, leads to a PSF which gives the decomposition of $\eta(x^0,x)$ under the orthogonal group. We leave the details to the reader.

Up to now we have discussed the injectivity of the Radon transform and the related inversions. We now pass to a study of its range.

The values $(\mathbf{R}^l f)(s,L)$ for fixed l are not independent. This is seen most readily from (2.3). For, the restrictions $\hat{f}|_{\hat{L}^\perp}$ are equal on the intersections of the $\hat{L}^\perp$. By the projection–slice theorem these restrictions $\hat{f}|_{\hat{L}^\perp}$ are the Fourier transforms of $(\mathbf{R}^l f)(s,L)$ in the variable s. Moreover, if $\hat{f} \in C^\infty$ then certain differential conditions must hold on these intersections. Let us begin with the hyperplane Radon transform. Precisely we have the following form of compatibility called COMP ORIGIN of Section 1.5:

Proposition 2.1 *Let $l = n-1$. Necessary and sufficient conditions that a function $\hat{h}(\hat{s},\hat{L}^\perp)$ defined for $\hat{L} \in \mathbf{G}(n,n-1)$ and $\hat{s} \in \hat{L}^\perp$ be of the form*

$\hat{h}(\hat{s}, \hat{L}^\perp) = \hat{f}|_{\hat{L}^\perp}(\hat{s})$ *for a Schwartz function* f *on* $\mathbb{R}^n$ *are:*

(1) $\hat{L}^\perp \to \hat{h}(\hat{s}, \hat{L}^\perp)$ *is the* C^∞ *map of the Grassmannian into the Schwartz space.*

(2) *COMP ORIGIN holds, meaning that there is a formal power series* $\hat{f}^0$ *at the origin in* $\mathbb{R}^n$ *whose restriction to each* $\hat{L}^\perp$ *is the formal power series of* $\hat{h}(\hat{s}, \hat{L}^\perp)$ *at the origin of* $\hat{L}$.

Condition (2) is equivalent to:

(2*) *Let* $\partial_{\hat{L}^\perp}$ *denote the directional derivative along* $\hat{L}^\perp$

$$\partial_{\hat{L}^\perp} = \sum (\cos \hat{\alpha}_j) \frac{\partial}{\partial \hat{x}_j} \tag{2.12}$$

where $\cos \hat{\alpha}_j$ *are the direction cosines of* $\hat{L}^\perp$. *Then for any* m *and any* $\hat{L}^\perp$ *we have*

$$(\partial^m_{\hat{L}^\perp} \hat{h})(\hat{s} = 0, \hat{L}^\perp) = \sum_{m_1 + \cdots + m_n = m} [\cos^{m_1} \hat{\alpha}_1 \ldots \cos^{m_n} \hat{\alpha}_n] \beta_{m_1 \ldots m_n} \tag{2.13}$$

for numbers $\beta_{m_1 \ldots m_n}$ *(which can be identified with the power series coefficients of* $\hat{f}^0$ *times suitable binomial coefficients).*

Notation.

$\hat{f}$ *is the Fourier transform of* f *on* $\mathbb{R}^n$.

$\hat{h}(\hat{s}, \hat{L}^\perp)$ *is the Fourier transform of* $h(s, L^\perp)$ *in the variable* s; $\hat{h}(\hat{x}, \hat{L}^\perp)$ *is the Fourier transform of* $h(x, L^\perp)$ *in the variable* x.

$\hat{L}$ *is the same (i.e. identified by the usual quadratic form) as* L *except* $\hat{L}$ *is in the Fourier transform space.*

$h(x, L^\perp)$ *is the extension of* $h(s, L^\perp)$ *which is constant in the directions of* L. *Thus* $\mathbf{R}^*[h(s, L^\perp)] = h(x, L^\perp)$.

$\hat{h}(\hat{x}, \hat{L}^\perp)$ *is the extension of* $\hat{h}(\hat{s}, \hat{L}^\perp)$ *obtained by multiplying* $\hat{h}(\hat{s}, \hat{L}^\perp)$ *by* $\delta_{\hat{L}^\perp}$. *Thus* $h(x, L^\perp)$ *and* $\hat{h}(\hat{x}, \hat{L}^\perp)$ *are Fourier transforms of each other.*

We begin with the proof of the necessity of conditions (1), (2*) Condition (1) is clear. We take the Fourier transform condition of (2*). The case $m = 0$ of (2.13) asserts that the values $\hat{h}(\hat{s} = 0, \hat{L}^\perp)$ must be independent of $\hat{L}^\perp$. By Fourier transform

$$\hat{h}(0, \hat{L}^\perp) = \int_{L^\perp} h(s, L^\perp)\, ds. \tag{2.14}$$

If

$$\hat{h}(\hat{s}, \hat{L}^\perp) = \hat{f}\big|_{\hat{L}^\perp}(\hat{s})$$

then our notation convention prescribes

$$\hat{h}(\hat{x}, \hat{L}^\perp) = \hat{f} \delta_{\hat{L}^\perp}(\hat{s}).$$

Thus $h(x, L^\perp)$ is of the form $f * \delta_L$. We write $x = (\lambda_L, s)$ where λ_L is a variable on L and s is a variable on $L^\perp$. (By our assumptions we have standard parameters

$s, \hat{s}$ for all $L^\perp$ and $\hat{L}^\perp$.) Then

$$h(s, L^\perp) = (f * \delta_L)(\lambda_L, s) = \int f(\lambda_L - \lambda'_L, s)\, d\lambda'_L$$

so that $\int h(s, L^\perp)\, ds = \int f(x)\, dx$ is manifestly independent of L, which verifies (2^*) for $m = 0$.

Next let m in (2.13) be arbitrary. We regard (2.13) as relating the derivatives of the various $\hat{h}(\hat{s}, \hat{L}^\perp)$ at $\hat{s} = 0$. The right side consists of two parts: the term $\beta_{m_1 \dots m_n}$ which is independent of $\hat{L}$ and the term $\cos^{m_1} \hat{\alpha}_1 \dots \cos^{m_n} \hat{\alpha}_n$ which depends on $\hat{L}$.

The Fourier transform of the left side of (2.13) is the m-th moment $\int s^m h(s, L^\perp)\, ds$. If, in accordance with (2.12) we write $s = \sum(\cos \hat{\alpha}_j) x_j$ on $L^\perp$, then (2.13) states that the m-th moments transform like m-th order symmetric tensors under the rotation group. Of course, if $h(x, L^\perp)$ is of the form $f * \delta_L$ then this tensor behavior is clear from the formula

$$\begin{aligned} \int s^m (f * \delta_L)(s)\, ds &= \int s^m\, ds \int f(s, \lambda_L - \lambda'_L)\, d\lambda'_L \\ &= \iint f(s, \lambda_L) s^m\, d\lambda_L\, ds = \int f(x) \tilde{s}^m_{L^\perp}\, dx \end{aligned} \tag{2.15}$$

upon applying Fubini's theorem to $s^m_{L^\perp}$.

In this formula $\tilde{s}_{L^\perp}$ is the function which is the coordinate $s_{L^\perp}$ on $L^\perp$ and constant in directions parallel to L.

This verifies the necessity condition in Proposition 2.1.

The moment condition can also be stated as

Theorem 2.2 *Necessary and sufficient conditions for the functions $h(x, \mathbf{g})$ which are constant in the directions of L to be of the form $f * \delta_L$ for an f in the Schwartz space $\mathcal{S}$ are:*

(1) *$h(s, \mathbf{g})$ is a Schwartz function in s for each $\mathbf{g}$.*
(2) *The $h(s, \mathbf{g})$ (thought of as elements of $\mathcal{S}(s)$) depend in a C^∞ manner on the Grassmannian variable $\mathbf{g}$.*
(3) *The moments $\int s^m h(s, \mathbf{g})\, ds$ form a homogeneous function of degree m in the direction cosines of $L^\perp$.*

We delay the proof of the sufficiency of the conditions in Proposition 2.1 and of Theorems 2.2 and 2.3 (below).

The use of COMP ORIGIN to piece together functions on lines has its origin in Chapter XIII of FA. The complete proof of Theorem 2.2 was given for the first time by Helgason [95]. Our above remarks show that Theorem 2.2 is equivalent to Proposition 2.1.

One might be tempted to give a proof in terms of h rather than $\hat{h}$. For, the tensor condition allows us to construct the values of the proposed function f on polynomials (i.e. the integrals of f times polynomials, polynomials being

considered as Schwartz distributions.) These values (moments) are the Fourier transforms of the power series coefficients of $\hat{f}^0$. But the polynomials are not dense in the space of Schwartz distributions so it seems quite difficult to carry out this proof of Proposition 2.1. The density of polynomials is equivalent to the quasianalyticity of the Fourier transform of the dual of the space of h or f (see Chapter XIII of FA). In the quasianalytic case it is possible to show directly that f exists.

It seems strange that the polynomial consistencies are sufficient for the existence of f even though the polynomials are not dense. This point is clarified in the remark following the statement of Theorem 2.3.

We have seen that one origin of the moment conditions is the tensorial behavior of powers of directional derivatives. There is a second origin which is based on the proof of injectivity of the Radon transform using Weierstrass' theorem as in Section 1.1. In particular, for $n = 2$ Lemma 1.1 says that we can express any polynomial of degree $\leq m$ in terms of polynomials of degree $\leq m$ of the form $P_1(\tilde{s}_{L_1^\perp}) + \cdots + P_{m+1}(\tilde{s}_{L_{m+1}^\perp})$ for any fixed $m+1$ distinct directions. If the original polynomial is homogeneous of degree m then all the other polynomials can be chosen to be homogeneous of degree m. In particular, suppose that the original polynomial is $\tilde{s}^m_{L^\perp}$. Then we obtain an expression of the moment $\int \tilde{s}^m_{L^\perp} f(x)\,dx$ in terms of moments of $f * \delta_{L_1}, \ldots, f * \delta_{L_m}$. Thus, if we use two different sets $\{L_j\}, \{L'_j\}$ of hyperplanes then there must be a relation between the moments of $\{f * \delta_{L_j}\}$ and $\{f * \delta_{L'_j}\}$.

There is an analog of the above for $n > 2$. Instead of $m + 1$ directions we need $N(m) = \binom{m+n-1}{m}$ directions, meaning polynomials which depend only on these directions. These directions must satisfy certain genericity conditions; for example, they cannot all be in a hyperplane. We obtain moment conditions as in the case $n = 2$.

We also have the algebraic result

Theorem 2.3 *The moment condition of Theorem 2.2 is equivalent to the identities obtained by expressing the $\tilde{s}^m_{L^\perp}$ for various L and m as linear combinations $\tilde{s}^m_{L^\perp}(\tilde{s}^m_{L^\perp_1}, \ldots, \tilde{s}^m_{L^\perp_N})$ for any fixed set of $N(m)$ generic directions and then identifying the consequent moment relations. By this we mean those moment identities that are obtained from $\int s^m_{L_j^\perp} h(s_{L_j^\perp}, L_j^\perp)\,ds_{L_j^\perp}$ and $\int s^m_{L_j'^\perp} h(s_{L_j'^\perp}, L_j'^\perp)\,ds_{L_j'^\perp}$ for different sets $\{L_j\}, \{L'_j\}$ of generic directions by writing $\tilde{s}^m_{L'^\perp_j} = \tilde{s}^m_{L'^\perp_j}(\tilde{s}^m_{L^\perp_1}, \ldots, \tilde{s}^m_{L^\perp_N})$.*

Remark. *We have pointed out in Section 1.8 that such identities can be thought of as amalgamations of the supradirect integral (1.268). It is an amazing consequence of Theorem 2.3 that all amalgamations are consequences of such polynomial identities—there are no additional transcendental relations.*

Actually this property of amalgamations (at least if they are finite), namely that they are all consequences of polynomial identities, is capable of a direct

proof. Thus, suppose we have a relation of the form

$$u = \sum_{j=1}^{M} u_{L^{\perp}{}_j} = \sum_{j=1}^{M} u_{L'^{\perp}_j}$$

where $u_{L^{\perp}{}_j}$, $u_{L'^{\perp}_j}$ are functions which are constant in the respective directions L_j, L'_j. To avoid trivial relations we assume that all L_j are distinct from all the L'_j. Then u satisfies two sets of differential equations

$$\left(\prod \frac{\partial}{\partial \vec{L}_j}\right) u = 0$$

$$\left(\prod \frac{\partial}{\partial \vec{L}'_j}\right) u = 0$$

where $\partial/\partial\vec{L}$ represents a system $\partial/\partial\vec{L} = \left(\partial/\partial L^1, \ldots, \partial/\partial L^{n-1}\right)$ with $L^1, \ldots, L^{n-1}$ being $n-1$ linearly independent directions in L. We have written $\prod \partial/\partial\vec{L}_j$ to mean the system consisting of all products of the form $\partial^M/\partial L_1^{k_1} \cdots \partial L_M^{k_M}$.

One way of analyzing these equations is by the Fourier transform. The fundamental principle of FA (see Section 1.4) asserts that u is the Fourier transform of a distribution supported by $(\cup \hat{L}_j^{\perp}) \cap (\cup \widehat{L'}_j^{\perp})$ which is the origin as long as all L_j are distinct from all the L'_j. Thus u is a polynomial which establishes our assertion regarding the amalgamations.

It would be interesting to give a direct (non-Fourier transform) proof that u are polynomials.

Proof of Proposition 2.1 It is clear that the functions $\hat{h}(\hat{s}, \hat{L}^{\perp})$ fit together to form a continuous function $\hat{f}(\hat{x})$ which is rapidly decreasing at infinity (i.e. $O(1 + |\hat{x}|)^{-t}$ for any t). It remains to show that $\hat{f}$ is C^{∞} and that its derivatives are also rapidly decreasing. If we think in terms of polar coordinates, the derivatives in $\hat{s}$ are the $\hat{r}$ derivatives except that $\hat{s}$ is a linear variable while $\hat{r}$ is a ray variable; the derivatives in the Grassmann variables $\hat{L}^{\perp}$ are θ derivatives. It follows by a simple change from polar to cartesian coordinates that $\hat{f}$ is C^{∞} except possibly at the origin and that derivatives of all orders in $\hat{x}$ are rapidly decreasing.

It remains to show that $\hat{f}$ is C^{∞} at the origin. The only difficulty involves the use of polar coordinates at the origin.

We can think of $\hat{f}$ as being a function on the cylinder whose base is the $n-1$ sphere and whose generator is $\hat{s} \in (-\infty, \infty)$. Moreover $\hat{f}(\hat{s}, \mathbf{g}) = \hat{f}(-\hat{s}, -\mathbf{g})$. (It is more convenient to use the sphere than the projective space.) By our hypotheses $\hat{f}$ is C^{∞} on this cylinder. We want to use the existence of $\hat{f}^0$ to "collapse" the sphere $\hat{s} = 0$ to a point.

In order to clarify the proof we start with $n = 2$. We can expand $\hat{f}$ in a Fourier series

$$\hat{f}(\hat{s}, \hat{\theta}) = \sum \hat{f}_k(\hat{s}) e^{ik\hat{\theta}}.$$

The functions $\hat{f}_k$ belong to $\mathcal{S}$ and, in fact, for any p

$$\{(1 + |k|^p)\hat{f}_k(\hat{s})\}$$

is a bounded set in $\mathcal{S}$. Put in other terms, the Fourier series of $\hat{f}$ converges to $\hat{f}$ in the topology of $\mathcal{S}$(cylinder).

Next let us examine $\hat{f}^0$ on the cylinder. In terms of polar coordinates in the plane

$$\hat{x}_1 = \hat{r}\cos\hat{\theta}, \quad \hat{x}_2 = \hat{r}\sin\hat{\theta}$$

so that

$$\hat{x}_1^{m_1}\hat{x}_2^{m_2} = \hat{r}^{m_1+m_2} \sum_{|j|\le m_1+m_2} c^j_{m_1 m_2} e^{ij\hat{\theta}} \tag{2.16}$$

where $c^j_{m_1 m_2}$ are expressed in terms of binomial coefficients. From this we can write $\hat{f}^0$ in the form

$$\begin{aligned} \hat{f}^0(\hat{x}_1, \hat{x}_2) &= \sum \gamma_{m_1 m_2} \hat{x}_1^{m_1}\hat{x}_2^{m_2} \\ &= \sum \gamma_{m_1 m_2} \hat{r}^{m_1+m_2} \sum_{|j|\le m_1+m_2} c^j_{m_1 m_2} e^{ij\hat{\theta}} \\ &= \sum \hat{f}^0_k(\hat{r}) e^{ik\hat{\theta}} \end{aligned} \tag{2.17}$$

where $\hat{f}^0_k$ is of the form

$$\hat{f}^0_k(\hat{r}) = \sum \alpha_{kp} \hat{r}^{|k|+2p}$$

as is clear from the definitions.

We can identify $\hat{r}$ with $\hat{s}$ for $\hat{s} \ge 0$.

Our hypothesis states that for any fixed $\hat{\theta}^0$ the formal power series of $\hat{f}(\hat{s}, \hat{\theta}^0)$ at $\hat{s} = 0$ agrees with that of $\hat{f}^0(\hat{r}, \hat{\theta}^0)$. Since $\hat{f}(\hat{s}, \hat{\theta})$ is C^∞ on the cylinder, the Fourier series of the $\hat{s}$ derivatives of $\hat{f}$ of any order converge uniformly in $\hat{\theta}$ as do their $\hat{\theta}$ derivatives. This means that

$$\frac{\partial^q \hat{f}}{\partial \hat{s}^q}(\hat{s}, \hat{\theta}) = \sum \hat{f}^{(q)}_k(\hat{s}) e^{ik\hat{\theta}}.$$

In particular the formal power series of $\hat{f}(\hat{s}, \hat{\theta}^0)$ at $\hat{s} = 0$ is given by

$$\hat{f}(\hat{s}, \hat{\theta}^0) \sim \sum_q \frac{\hat{s}^q}{q!} \sum_k \hat{f}^{(q)}_k(0) e^{ik\hat{\theta}^0}.$$

Comparing this with the above computation for $\hat{f}^0$ yields the result:

For any $\hat{\theta}^0$ and q,

$$\frac{1}{q!}\sum \hat{f}_k^{(q)}(0)e^{ik\hat{\theta}^0} = \sum \alpha_{k,(q-|k|)/2}e^{ik\hat{\theta}^0}.$$

For any q only a finite number of k, namely $|k| \le q$, appear in the sum. It follows from the uniqueness of Fourier series that

$$\hat{f}_k^{(q)}(0) = 0 \quad \text{unless } q = |k| + 2p \quad \text{for some } p. \tag{2.18}$$

We are now in a position to prove that $\hat{f}$ is C^∞ at $\hat{x} = 0$. This means that $\hat{f}$ is defined (i.e. independent of $\hat{\theta}$) and continuous at the origin and all of its $\hat{x}$ derivatives are bounded there.

By the above $\hat{f}_k(0) = 0$ for $k \neq 0$. Thus $\hat{f}(0, \hat{\theta}) = \hat{f}_0(0)$ and $\hat{f}$ is clearly continuous at $\hat{x} = 0$. When we take derivatives in polar coordinates

$$\frac{\partial}{\partial \hat{x}_1} = \cos\hat{\theta}\frac{\partial}{\partial \hat{s}} - \frac{\sin\hat{\theta}}{\hat{s}}\frac{\partial}{\partial \hat{\theta}}$$

$$\frac{\partial}{\partial \hat{x}_2} = \sin\hat{\theta}\frac{\partial}{\partial \hat{s}} - \frac{\cos\hat{\theta}}{\hat{s}}\frac{\partial}{\partial \hat{\theta}}$$

it is the denominator $\hat{s}$ which could cause trouble at $\hat{s} = 0$. For $\hat{x}$ derivatives of first order, the term $\hat{f}_0$, being constant, presents no problem because the $\hat{\theta}$ derivative vanishes. Also for $|k| > 1$ both $\hat{f}_k$ and $\hat{f}_k^{(1)}$ vanish at the origin by (2.18) so their $\hat{x}$ derivatives are continuous at the origin.

It follows that only the terms $\hat{f}_{\pm 1}(\hat{s})e^{\pm i\hat{\theta}}$ could cause trouble. We write

$$\hat{f}_{\pm 1}(\hat{s}) = \hat{f}'_{\pm 1}(0)\hat{s} + \hat{h}_{\pm 1}(\hat{s})$$

where $\hat{h}_{\pm 1}(s)$ vanishes to third order at the origin by (2.18). As above, the derivatives of $\hat{h}_{\pm 1}(\hat{s})\exp(\pm i\hat{\theta})$ are continuous at the origin.

We are left with

$$\hat{f}'_{\pm 1}(0)\hat{s}e^{\pm i\hat{\theta}} = \hat{f}'_{\pm 1}(0)(\hat{x}_1 \pm \hat{x}_2)$$

which is manifestly differentiable.

This completes the proof for differentiation of orders 0, 1. Although this method can be extended to higher order derivatives it becomes somewhat complicated so we introduce a slight variation.

If $m = (m_1, m_2)$ then, as above, when forming $\partial^{|m|}\hat{f}/\partial\hat{x}_1^{m_1}\partial\hat{x}_2^{m_2}$ the terms $\hat{f}_k$ for $|k| > |m|$ are unimportant. For $|k| \le m$ we subtract a polynomial $P_k(\hat{s})$ from $\hat{f}_k$ so $\hat{f}_k - P_k$ vanishes to order $> |m|$ at the origin.

By (2.18) P_k contains only the monomials $\hat{s}^a$ with $a = |k| + 2p$. The smoothness of $\hat{s}^a \exp(ik\hat{\theta})$ with $a \geq |k|$ follows from

$$\hat{s}^{|k|+2p} e^{ik\hat{\theta}} = (\hat{x}_1^2 + \hat{x}_2^2)^p \begin{cases} z^{|k|}, & \text{if } k \geq 0 \\ \bar{z}^{|k|}, & \text{if } k \leq 0 \end{cases}$$

with $z = x_1 + ix_2$.

This completes the proof of Proposition 2.1 for $n = 2$. For $n > 2$ we use spherical harmonics on the spheres $\hat{r} = \text{const.}$ and proceed as for $n = 2$. (Spherical harmonics are discussed in some detail in Chapters 3 and 4.)

We have seen that Proposition 2.1 and Theorem 2.2 are equivalent. We are left with

Proof of Theorem 2.3. Theorem 2.3 can be thought of in terms of "change of basis" from Theorem 2.2; we shall show that Theorems 2.2 (or, equivalently, Proposition 2.1) and 2.3 are equivalent. We are using two bases for homogeneous polynomials on $\mathbb{R}^n$ of degree m:

(1) $\tilde{s}^m_{L^\perp{}_1}, \ldots, \tilde{s}^m_{L^\perp{}_N}$.
(2) Monomials of degree m.

N is the binomial coefficient $N = \binom{m+n-1}{m}$ and the directions $L_j^\perp$ are suitably (generically) chosen.

Assume first that we know the moment relations for the $h(x, L)$ coming from the relations amongst various bases $\{\tilde{s}^m_{L_j}\}$. In the notation of Proposition 2.1 these are relations amongst power series coefficients of the $\hat{h}(\hat{s}, \hat{L}_j^\perp)$. We want to construct a formal power series $\hat{f}^0$ so that COMP ORIGIN is satisfied for $\{\hat{h}(\hat{s}, \hat{L}^\perp)\}$. The coefficients of $\hat{f}^0$ should be defined as the moments by monomials μ of the desired f. We can write μ (of degree m) as a linear combination of the $\tilde{s}^m_{L^\perp{}_j}$ corresponding to any basis of the form (1). The coefficient of $\hat{f}^0$ corresponding to μ is constructed from this linear combination by forming the $\tilde{s}^m_{L_j}$ moments of $h(x, L_j)$; according to our hypothesis it is independent of the basis.

It remains to show that the formal power series of $\hat{h}(\hat{s}, \hat{L}^\perp)$ agrees with that of $\hat{f}^0$. In particular we can choose $L_1 = L$ which verifies our assertion of COMP ORIGIN for $\{\hat{h}(\hat{s}, \hat{L}^\perp)\}$. Thus Theorem 2.3 implies Proposition 2.1 and hence Theorem 2.2.

The converse goes along the same lines. This completes our proof of the equivalence of Theorems 2.2 and 2.3 and hence establishes Theorem 2.3.

It should be pointed out that Fourier series are somewhat easier to handle than $\hat{f}$ directly. For if we simultaneously approximate all $\hat{h}(\hat{s}, \hat{\theta})$, considered as functions of $\hat{s}$, by their Taylor series at $\hat{s} = 0$ (which can be done uniformly in $\hat{\theta}$) then it is somewhat difficult to control the $\hat{\theta}$ derivatives.

Remark. *The advantage of using Fourier series to prove Proposition 2.1 illustrates the principle that what is required for a proof is the "right basis." In this case the Fourier basis is simpler than the basis $\{\delta_{\hat{\theta}}\}$ mainly because the moment conditions have a simpler interpretation in the Fourier basis.*

Problem 2.1 *Prove Theorem 2.2 directly; that is, without using the Fourier transform.*

In the course of the proof of Proposition 2.1 we have shown

Proposition 2.4 *Let the notation be as in Proposition 2.1. A necessary and sufficient condition for the existence of the Schwartz function $\hat{f}$ such that $\hat{h}(s, \hat{L}^{\perp}) = \hat{f}|_{\hat{L}^{\perp}}(s)$ for all L is (1) and*

(3) *Let $\{\hat{f}_k(\hat{s})\}$ be the Fourier series coefficients of the function defined by the $\hat{h}(\hat{s}, \mathbf{g})$ on the cylinder $\{\hat{s}\} \times \{\mathbf{g}\}$. The Taylor coefficients at the origin of $\hat{f}_k(\hat{s})$ of order m are nonvanishing only when m is of the form $|k| + 2p$. ($|k|$ is the order of the relevant spherical harmonic.)*

All the above concerns the hyperplane transform ($l = n - 1$). If $l < n - 1$ then $\hat{L}^{\perp}$ is no longer a line but is a linear space of dimension $n - l$. Two such $\hat{L}^{\perp}$ can intersect in a plane of dimension > 0, so the compatibility conditions might seem to be complicated. Actually they are somewhat simpler when $l > 1$ than for $l = 1$.

Let us start with $l = n - 2$. Then the $\hat{L}^{\perp}$ are of dimension 2. Any line $\hat{L}^1$ is the intersection of such $\hat{L}^{\perp}$. In fact, the set of $\hat{L}^{\perp}$ passing through $\hat{L}^1$ forms a projective space $\mathbf{P}^{n-2}$. Certainly for such $\hat{L}^{\perp}$ the functions $\hat{h}(\hat{s}, \hat{L}^{\perp})$ must agree on $\hat{L}^1$. Moreover we should expect that there should be differentiation conditions on $\hat{L}^1$ in the directions orthogonal to $\hat{L}^1$. These differentiation conditions should be analogous to those given in the above case of the hyperplane transform.

However, the differentiation conditions, except at the origin, are unnecessary. To see this, note first that if the $\hat{h}(\hat{s}, \hat{L}^{\perp})$ agree on the intersections of the $\hat{L}^{\perp}$ then they define a function $\hat{f}(\hat{x})$ which is easily seen to be continuous.

We claim that $\hat{f}$ is C^{∞} if the $\hat{h}(\hat{s}, \hat{L}^{\perp})$ are C^{∞} on $\mathbf{G}(n, 2) \times \{\hat{s}\}$ and satisfy a suitable compatibility condition at the origin. To verify this, fix one line $\hat{L}^1$. The $\hat{L}^{\perp} \supset \hat{L}^1$ fill up all of $\mathbb{R}^n$ with their only intersection being $\hat{L}^1$. Since $\hat{h}(\hat{s}, \hat{L}^{\perp})$ are C^{∞} in $\hat{s}, \hat{L}^{\perp}$, it is clear (e.g. using polar coordinates) that $\hat{f}$ is C^{∞} off $\hat{L}^1$. Since $\hat{L}^1$ is an arbitrary line through the origin, $\hat{f}$ is C^{∞} except possibly at the origin which is the intersection of the $\hat{L}^1$.

If the formal power series of $\hat{h}(\hat{s}, \hat{L}^{\perp})$ satisfies COMP ORIGIN then the same is true of $\{\hat{f}|_{\hat{L}^1}\}$. Thus, by Proposition 2.1, $\hat{f} \in C^{\infty}$.

The same argument clearly works with obvious modifications if $l < n - 2$.

Proposition 2.5 *Let $l < n - 1$. Suppose for each $\hat{L}^{\perp}$ of codimension l we have a function $\hat{h}(\hat{s}, \hat{L}^{\perp})$ on $\hat{L}^{\perp}$ belonging to $\mathcal{S}(\hat{L}^{\perp})$. Necessary and sufficient*

conditions that there exists a function $\hat{f}(\hat{x}) \in \mathcal{S}$ *whose restriction to each* $\hat{L}^{\perp}$ *is* $\hat{h}(\hat{s}, \hat{L}^{\perp})$ *are*

(1) $\hat{L}^{\perp} \to \hat{h}(\hat{s}, \hat{L}^{\perp})$ *is a* C^{∞} *map of* $\mathbf{G}(n, n-l)$ *into* $\mathcal{S}$.
(2) $\hat{h}(\hat{s}, \hat{L}_1^{\perp}) = \hat{h}(\hat{s}, \hat{L}_2^{\perp})$ *on* $\hat{L}_1^{\perp} \cap \hat{L}_2^{\perp}$ *for all* $\hat{L}_1^{\perp}, \hat{L}_2^{\perp}$.

Remark. *If we added the hypothesis*

(3) *The* $\hat{h}$ *satisfy COMP ORIGIN*

then the fact that $f \in \mathcal{S}$ *would follow from our above remarks. We shall show below that (3) is a consequence of (1) and (2) thereby completing the proof of Proposition 2.5.*

What is the translation of Proposition 2.5 into moment conditions? We claim that condition (2) means that

$$T \cdot h(s, L_1^{\perp}) = T \cdot h(s, L_2^{\perp}) \tag{2.19}$$

whenever $\hat{T} \in \mathcal{S}'(\hat{L}_1^{\perp} \cap \hat{L}_2^{\perp})$. Here $h(s, L^{\perp})$ is the Fourier transform of $\hat{h}(\hat{s}, \hat{L}^{\perp})$ on the plane $\hat{L}^{\perp}$. T is extended from $L_1^{\perp} \cap L_2^{\perp}$ to $L_1^{\perp}$ and to $L_2^{\perp}$ by making it constant in the orthogonal directions. Since the linear combinations of δ functions on $\hat{L}_1^{\perp} \cap \hat{L}_2^{\perp}$ are dense in $\mathcal{S}'(\hat{L}_1^{\perp} \cap \hat{L}_2^{\perp})$ we can restrict $\{T\}$ to be the set of exponentials whose frequency lies in $\hat{L}_1^{\perp} \cap \hat{L}_2^{\perp}$; that is,

$$e^{ix \cdot \hat{x}} \cdot h(s, L_1^{\perp}) = e^{ix \cdot \hat{x}} \cdot h(s, L_2^{\perp}) \tag{2.19*}$$

if $\hat{x} \in \hat{L}_1^{\perp} \cap \hat{L}_2^{\perp}$, which is exactly condition (2).

Thus Proposition 2.5 implies

Theorem 2.6 *Let* $l < n-1$. *Suppose that* $L^{\perp} \to h(s, L^{\perp})$ *is a* C^{∞} *map of* $\mathbf{G}(n, n-l)$ *into* $\mathcal{S}(\mathbb{R}^{n-l})$. *Then (2.19) is a necessary and sufficient condition that there is a function* $f \in \mathcal{S}(\mathbb{R}^n)$ *such that*

$$(f * \delta_L)(x) = h(x, L^{\perp}) \tag{2.20}$$

for all $L \in \mathbf{G}(n, l)$; *that is,* $h(x, L^{\perp}) = \mathbf{R}^* \mathbf{R} f(x, L)$.

Remark. *In case* $h(s, L^{\perp})$ *decreases exponentially in* s *for all* L, $\hat{h}(\hat{s}, \hat{L}^{\perp})$ *is holomorphic in* $\hat{s}$. *Thus the equality* $\hat{h}(\hat{s}, \hat{L}_1^{\perp}) = \hat{h}(\hat{s}, \hat{L}_2^{\perp})$ *on* $\hat{L}_1^{\perp} \cap \hat{L}_2^{\perp}$ *is a consequence of the equality of their formal power series at the origin. Hence (2.19) is unnecessary. Rather we need only the usual moment conditions*

$$\sum c_j \int \tilde{\lambda}_j^m h(s, L_j^{\perp})\, ds = 0 \tag{2.21}$$

whenever $\sum c_j \tilde{\lambda}_j^m = 0$ *(as in Lemma 1.1). Another way of looking at this is the fact that in the dual of the space of functions that decrease exponentially,*

the exponential is the limit of polynomials. Actually we can weaken "decrease exponentially" to "decrease close to exponentially," which corresponds to quasi-analyticity rather than holomorphicity. (Quasianalyticity is discussed in detail in Chapter 5.)

We want to investigate the relation between the moment conditions (2.21) and the intersection conditions (2.19) (assuming $\hat{h}$ are C^∞ in $\hat{L}^\perp$ and $\hat{s}$).

In the first place, the intersection conditions do not imply the moment conditions in case $l = n - 1$. The $\hat{L}^\perp$ are lines which meet only at the origin. The function $|\hat{x}|\hat{x}_1$ $(= |\hat{x}|^2 \cos\hat{\theta}$ for $n = 2)$ is C^∞ on every line through the origin, takes the common value 0 at the origin on each line, and is C^∞ on the cylinder $\{\hat{L}^\perp\} \times \{\hat{s}\}$, but is clearly not C^∞ in $\hat{x}$.

Proof of Proposition 2.5 We claim that for $l < n - 1$ the intersection conditions imply that $\hat{f}$ must be C^∞; we have already shown that $\hat{f}$ is C^∞ except possibly at the origin. Let us first show that $\hat{f}$ is C^1. We take cognizance of the fact that the relations among linear functions are eventually planar relations, i.e. they stem from relations on planes (dimension 2). To clarify this point suppose we have a relation

$$\sum \alpha_j \xi_j = \xi$$

where ξ_j are linear functions thought of as vectors in the dual space and α_j are constants (say $\alpha_1 \neq 0$). In the plane spanned by $\overrightarrow{u} = \xi_1$, $\overrightarrow{v} = \alpha_2\xi_2 + \cdots + \alpha_n\xi_n$ the relation becomes

$$\alpha_1 \overrightarrow{u} + \overrightarrow{v} = \xi.$$

Let us apply this idea to verify the first moment conditions. For example, to show that

$$\left(\sum \frac{\partial}{\partial \hat{x}_j}\right) \hat{f}(0) = \sum \frac{\partial \hat{f}}{\partial \hat{x}_j}(0) \tag{2.22}$$

we examine the plane $\hat{Q}$ containing $\hat{x}_1$ and $(\hat{x}_2 + \cdots + \hat{x}_n)$. Since $\hat{h}(\hat{L}^\perp, \hat{s})$ is C^∞ in any $n - l$ (≥ 2) plane $\hat{L}^\perp$, it is C^∞ in $\hat{Q}$ and $\hat{f} = \hat{h}$ there. Thus

$$\left(\sum_{j=1}^{n} \frac{\partial}{\partial \hat{x}_j}\right) f(0) = \frac{\partial f}{\partial \hat{x}_1} + \left(\sum_{j=2}^{n} \frac{\partial}{\partial \hat{x}_j}\right) f(0). \tag{2.22*}$$

We continue the process to write

$$\left(\sum_{j=1}^{n} \frac{\partial}{\partial \hat{x}_j}\right) f(0) = \frac{\partial f}{\partial \hat{x}_2} + \left(\sum_{j=3}^{n} \frac{\partial}{\partial \hat{x}_j}\right) f(0).$$

By iteration we arrive at (2.22).

The same argument works for any directional derivative. Now the argument used to prove Proposition 2.1 shows that $\hat{f} \in C^1$.

The fact that $\hat{f} \in C^1$ was proven by the author. This result was completed by Jan Boman (personal communication) who showed:

The hypotheses (1) and (2) of Proposition 2.4 imply that $\hat{f} \in C^\infty$.

Proof We already know that $\hat{f} \in C^1$. A standard argument shows that $\hat{f} \in C^2$ if we know that all derivatives of order 2 and 3 of $\hat{f}$ are bounded in a neighborhood of 0. By iteration we must prove that any derivative of $\hat{f}$ is bounded in a neighborhood of the origin.

Let $\partial = \partial^m / \partial \hat{x}_1^{m_1} \dots \partial \hat{x}_n^{m_n}$. By Lemma 1.1 we can express ∂ as a sum of directional derivatives of order m. We are thus reduced to the case $\partial = \partial^m / \partial \hat{x}_1^m$.

Let $\hat{x}$ be any point near 0. Let $\hat{L}^\perp$ be an $n-l$ plane containing the $\hat{x}_1$ axis and $\hat{x}$. Since $\hat{f} = \hat{h}(\cdot, \hat{L}^\perp)$ on $\hat{L}^\perp$ it follows from hypothesis (1) of Proposition 2.5 that $\partial \hat{f}(\hat{x}) = \partial \hat{h}(\hat{L}^\perp, \hat{x})$ is bounded independently of $\hat{x} \in \hat{L}^\perp$ near 0. Since $\{\hat{L}^\perp\}$ is compact we conclude that $\partial \hat{f}(\hat{x})$ is bounded independently of $\hat{x}$ near 0.

This completes the proof of Proposition 2.5 and hence of Theorem 2.6.

We have established the fact that the intersection conditions imply the moment conditions for $l < n-1$ but not for $l = n-1$. Conversely, it is possible to prove that the moment conditions do not imply the intersection conditions. We need:

There is a nontrivial function $u \in \mathcal{S}_0[0, \infty)$ *which is orthogonal to all polynomials.*

($\mathcal{S}_0[0, \infty)$ consists of those functions in $\mathcal{S}[0, \infty)$ all of whose derivatives vanish at 0.)

Proof For $\alpha \in \mathcal{S}_0[0, \infty)$, we define the Mellin transform

$$\underset{\vee}{\alpha}(s) = \int \alpha(t) t^{is} \, dt/t.$$

The Mellin transform of $\mathcal{S}_0[0, \infty)$ is the space of entire functions which are rapidly decreasing (faster than any polynomial) in directions parallel to the real axis. This is readily verified by making an exponential change of variable which replaces $\mathcal{S}_0[0, \infty)$ by the space of C^∞ functions on the whole line, all of whose derivatives are exponentially decreasing at $\pm\infty$. Orthogonality to polynomials means that the Mellin transform $\underset{\vee}{u}$ of u vanishes at the points $\{ik\}$ for k a negative integer. Clearly such functions $\underset{\vee}{u}$ exist. The existence of (many) u is thus assured by Mellin inversion.

Any such u defines radial functions $h(s, L)$ for all L. As all derivatives of u vanish at the origin, $h \in C^\infty$. It is clear that, in fact, $h(s, L)$ is C^∞ in s and L. Since the euclidean measure on a linear plane L is $r^{l-1} dr d\theta$ it follows from the fact that u is orthogonal to all polynomials that $\int_{L^\perp} \nu(s)^m h(s, L) ds = 0$

for any linear function $\nu(s)$ on $L^\perp$. Thus all the moment conditions hold trivially.

Given one function $h(s, L^\perp)$ we can construct many others, for example, by setting

$$\hat{h}^1(\hat{s}, \hat{L}^\perp) = \hat{h}(\hat{s}, \hat{L}^\perp)\hat{v}(\hat{s}, \hat{L}^\perp)$$

where $\hat{v}(\hat{s}, \hat{L}^\perp) \in \mathcal{D}(\hat{L}^\perp)$ for fixed $\hat{L}^\perp$ and $\hat{v}$ is a C^∞ function of $\hat{s}, \hat{L}^\perp$. All $\hat{s}$ derivatives of $\hat{h}^1$ vanish at $\hat{s} = 0$ so the smoothness conditions and moment conditions are satisfied by $h^1(s, L^\perp)$. But we can choose $\hat{v}$ so that, given $\hat{L}_1 \neq \hat{L}_2$, the functions $\hat{v}(\hat{s}, \hat{L}_1^\perp)$ and $\hat{v}(\hat{s}, \hat{L}_2^\perp)$ do not agree on $\hat{L}_1^\perp \cap \hat{L}_2^\perp$.

Thus the intersection conditions are not satisfied.

We have proven

Theorem 2.7 *For $l < n-1$ the smoothness condition and the moment conditions using polynomials are* not *sufficient to guarantee that $\{h(L^\perp, x)\}$ is of the form $\{f * \delta_L\}$.*

Theorem 2.7 shows that to get enough moment conditions for $l < n-1$ we must go beyond polynomials, e.g. to exponentials as in (2.19*).

In Chapter 6 we shall find another approach to the moment conditions via the equations satisfied by the parametric Radon transform.

The classical Radon transform involves integration over affine planes. It is natural to replace planes by other geometric entities. We shall meet several examples of such Radon transforms in later chapters. We present here a Radon transform which has been the object of much study (see e.g. [2]). Let Ω be a smooth strictly convex compact hypersurface in $\mathbb{R}^n$. The spherical Radon transform $\mathbf{R}_\Omega$ relative to Ω is defined by integration over all spheres tangent to Ω using the standard (euclidean) measure on the sphere. We shall work with exterior tangent spheres; the same ideas apply to the interior of Ω.

Let $p \in \Omega$ and call $\psi(p)$ the unit exterior normal to Ω at p. The circle of radius $|c|$ tangent to Ω at p has center at $p + c\psi(p)$. The equation of this circle is

$$c^2 = [x - (p + c\psi(p))]^2$$

or

$$c = \frac{x^2 + p^2 - 2x \cdot p}{2\psi \cdot (x - p)}. \tag{2.23}$$

For fixed p these circles do not form a spread since they all meet at p and only fill up the half-space exterior to the tangent plane to Ω at p.

We define a "cone-like" structure Γ_Ω in $\mathbb{R}^{n+1}$ (coordinates (t, x)). For each p and for $t \geq 0$ we erect the ray $\gamma(p)$ which lies in the 2 plane spanned by the t axis and $\psi(p)$ and bisects the angle formed by the positive t axis and the exterior normal ray to Ω at p. We then continue this ray to form a line. Γ_Ω is the union of these lines (generators). This means that from the point $p \in \Omega$ we form the line $\{t, p + t\psi(p)\}$.

We can describe Γ_Ω in a different fashion. We expand (contract) Ω in the direction of its normal at a fixed speed, say 1. While doing this we "lift" boundary Ω in the t direction with speed 1 (resp. -1).

In the language of Section 4.1 we "pull apart" the boundary of the expanding (contracting) Ω to form Γ_Ω.

A property of this expanding Ω is

Proposition 2.8 *The normals to the expanding (contracting) Ω stay fixed. Precisely, we expand Ω by $p \to p + t\psi(p)$ to form a new region Ω_t. The line normal to Ω at p is the same as the line normal to Ω_t at $p + t\psi(p)$.*

Proof Call L^p the support plane to Ω at p. We move L^p parallel to itself a distance t; call the translated plane L_t^p (see Figure 2.1). Then L_t^p meets Ω_t at $p + t\psi(p)$ as is clear from the definitions.

We claim that L_t^p cannot intersect Ω_t except at p. For Ω lies on one side of L^p. Since L^p and L_t^p are parallel planes whose distance apart is t, no point $q \in L_t^p$ except $p + t\psi(p)$ can be at a distance $\leq t$ from Ω. This shows that the plane L_t^p meets Ω_t exactly at $p + t\psi(p)$ so L_t^p is the plane of support at this point. It follows that $(L_t^p)^\perp = \psi$ which is the desired result.

The relation of Γ_Ω to the spherical Radon transform $\mathbf{R}_\Omega$ is the geometric property that if $q \in \Gamma_\Omega$ and we draw the translate $\Gamma + q$ of the usual light cone Γ by q then $(\Gamma + q) \cap \{t = 0\}$ is a sphere tangent to Ω. Moreover all such spheres are obtained in this manner (see Figure 2.2).

We shall now relate this geometry to the wave equation.

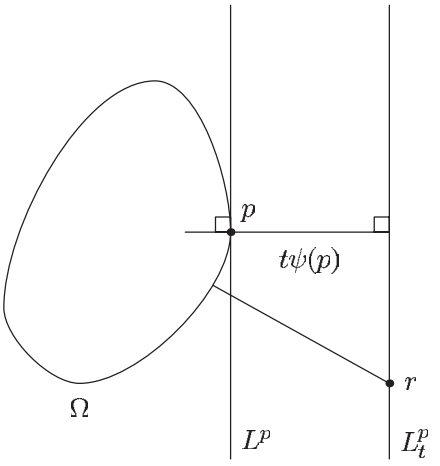

Figure 2.1

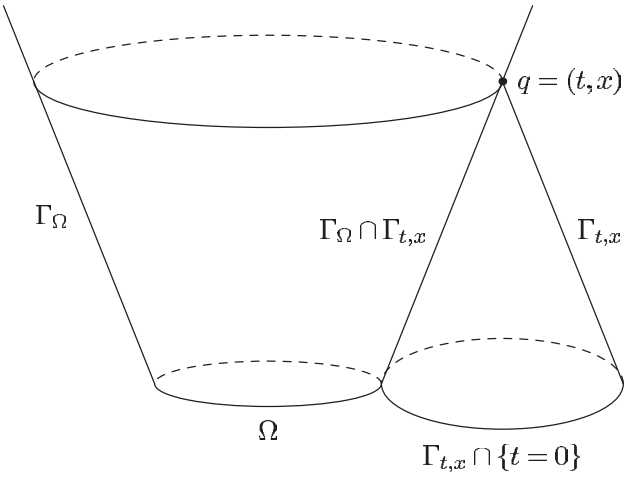

Figure 2.2

Let $f(x)$ be a given function. We form the solution $F(t, x)$ of the CP for the wave equation

$$\begin{aligned} \Box F(t,x) &= 0 \\ F(0,x) &= f(x) \\ \frac{\partial F}{\partial t}(0,x) &= 0. \end{aligned} \tag{2.24}$$

F is the even solution of the wave equation whose CD is $(f, 0)$.

We shall use some standard properties of solutions of the wave equation; they can be found in [29, 134].

Suppose n is odd. There is a Huygens principle for the wave equation. This means that the fundamental solutions $e^{\pm}$ of $\Box$ have supports on the forward (resp. backward) light cones $\Gamma^{\pm}$. (For n even the support is the interior of the (half) light cone.) i.e. call $e^{\pm}_{t,x}$ the translate of $e^{\pm}$ by (t, x). $F(t, x)$ is expressible in terms of the integral of f over the intersection of the support of $e^{\pm}_{t,x}$, i.e. $\Gamma_{t,x} = \Gamma^{\pm} + (t, x)$, with $\{t = 0\}$ [29, vol. II, p. 202 (11)]. (See Figure 2.2.) This intersection is a sphere and the measure is a constant times the standard measure on the sphere so, by our above remarks, $F(t, x)$ is defined by $\mathbf{R}_\Omega(f)$.

The problem of inversion of $\mathbf{R}_\Omega$ becomes the problem of retrieving f from the values of F on Γ_Ω. For this it is crucial that F is *even* in t. We form $e^{\pm}_{0,x}\chi^E_{\Gamma_\Omega}$ where $\chi^E_{\Gamma_\Omega}$ denotes the characteristic function of the exterior of Γ_Ω. (It follows from standard theory that this product is defined if $x \neq 0$.) Note that

$$\Box e^{\pm}_{0,x}\chi^E_{\Gamma_\Omega} = \delta_{0,x} + \zeta^{\pm}_x \tag{2.25}$$

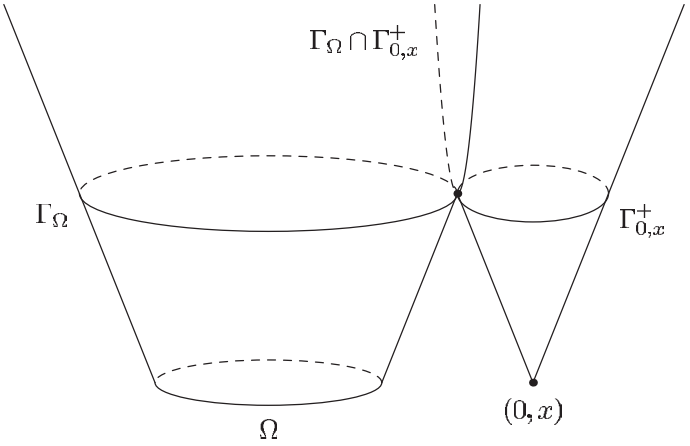

Figure 2.3

where $\zeta^\pm$ is a distribution supported by $\Gamma_\Omega \cap \Gamma_{0,x}$ (see Figure 2.3). Thus

$$\begin{aligned} 0 &= (\Box F) \cdot (e^\pm_{0,x} \chi^E_{\Gamma_\Omega}) \\ &= F \cdot [\Box (e^\pm_{0,x} \chi^E_{\Gamma_\Omega})] \\ &= f(x) + \zeta^\pm_x \cdot F. \end{aligned} \tag{2.26}$$

It remains to show that $\zeta^\pm_x \cdot F$ can be evaluated in terms of $F\big|_{\Gamma_\Omega}$. This means that $\zeta^\pm_x$ is a (finite) linear combination of $\{\partial\mu\}$ where μ is a measure on Γ_Ω and ∂ is a tangential derivative. (If ζ_x involved normal derivatives we could not accomplish the goal of expressing $f(x)$ in terms of $\mathbf{R}_\Omega(f)$.)

There is a secondary problem which we shall deal with below: the support of $\zeta^\pm_x$ is noncompact so the meaning of $\zeta^\pm_x \cdot F$ and the inversion must be clarified (see Figure 2.3).

Let $(t,p) \in \Gamma_\Omega$ and call $\psi(p)$ the unit vector in $\{t = \text{const.}\}$ normal to the intersection of Γ_Ω with the plane $\{t = \text{const.}\}$. Since the Laplacian $\Delta = \Delta(x)$ is rotation invariant we can write

$$\Delta = \frac{\partial^2}{\partial u^2} + \Delta_y$$

where u is in the direction of $\psi(p)$ and Δ_y is the Laplacian in the plane orthogonal to $\psi(p)$.

This gives

$$\Box = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial u^2} - \Delta_y.$$

The t, u plane contains the line $\gamma(p)$ which was constructed above to define Γ_Ω. In fact, $\gamma(p)$ is a translation of the line $t - u = 0$. Setting

$$t = v + w$$

$$u = v - w$$

we have

$$\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial u^2} = \frac{\partial^2}{\partial v \partial w}. \tag{2.27}$$

v is the variable along $\gamma(p)$ and w is the orthogonal variable. Thus $\partial^2/\partial v \partial w$ involves only first-order derivatives orthogonal to Γ_Ω. Since by (2.25) we are differentiating the "jump functions" $e^{\pm}_{0,x}\chi^E_{\Gamma_\Omega}$ the normal derivative produces a measure on Γ_Ω.

We now turn to Δ_y. We have shown in Proposition 2.8 that the tangent plane to $\Gamma_\Omega \cap \{t = \text{const}\}$ at a point p is orthogonal to $\psi(p)$. Hence Δ_y (which is a differential operator in $\{t = \text{const.}\}$) involve derivatives in directions tangent to Γ_Ω. Thus, again $\Delta_y(e^{\pm}_{0,x}\chi^E_{\Gamma_\Omega})$ does not involve normal derivatives. (The reader unfamiliar with this fact can examine the operator $\partial^2/\partial x^2$ at the point $(0, 1)$ of the circle $x^2 + y^2 = 1$ in the x, y plane. In terms of polar coordinates

$$\frac{\partial}{\partial x} = \cos\theta \frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}$$

so

$$\frac{\partial^2}{\partial x^2} = \cos^2\theta \frac{\partial^2}{\partial r^2} - 2\sin\theta\cos\theta \frac{\partial^2}{\partial r \partial \theta} + \frac{\sin\theta\cos\theta}{r^2}\frac{\partial}{\partial\theta} + \frac{\sin\theta\cos\theta}{r}\frac{\partial}{\partial r} - \frac{\sin\theta\cos\theta}{r^2}\frac{\partial}{\partial\theta} + \frac{\sin^2\theta}{r^2}\frac{\partial^2}{\partial\theta^2}.$$

$\cos\theta = 0$ at the point of tangency of $\{y = 1\}$ with the circle so we are left with only first r derivatives which, as before, lead to measures on the circle.)

There is, however, a secondary problem: namely, support $\zeta^{\pm}_x$ is generally noncompact so we have to show that $\zeta^{\pm}_x \cdot F$ is defined and solves the inversion problem.

Actually the situation is much worse because about half of the light rays emanating from $(0, x)$ do not meet Γ_Ω. (This is easily seen if Ω is the unit sphere.) Thus the natural inversion using $\zeta^{\pm}_x$ cannot be applied.

We are saved because, as noted above, F is even in t. When the ray $\rho \in \Gamma^+_{0,x}$ does not meet Γ_Ω it is true that $-\rho$ meets $-\Gamma_\Omega$.

We now present the geometric insight into this result. To avoid notational complications suppose $n = 2$. (Although there is no Huygens principle when $n = 2$ the geometry of the cones is independent of the parity of n.)

Let us examine the intersection of rays of $\Gamma_{0,1,0}$ with Γ. Let ρ be the ray in $\Gamma_{0,1,0}$ with direction cosines $(1, c, s)$ where $c = \cos\theta$ and $s = \sin\theta$. A point on

ρ is of the form $(\lambda, 1+\lambda c, \lambda s)$. Note that for $\lambda = 0$ the point $(0,1,0)$ lies in the exterior of Γ. The same persists if $\cos\theta > 0$ on the whole ray $\rho = \{\lambda > 0\}$. But on $-\rho = \{\lambda < 0\}$ this is no longer the case and, in fact, $-\rho$ intersects Γ when $\lambda = -1/2c$. When $c = 0$ the rays $\{\lambda, 1, \pm\lambda\}$ do not meet Γ.

For $\Omega = \{x^2 + y^2 = 1\}$ the situation is more favorable:

$$\Gamma_\Omega = \{(|t|+1)^2 = x^2 + y^2\}.$$

We now examine light rays emanating from $(0,2,0)$. They are of the form $\rho = (\lambda, 2+\lambda c, \lambda s)$. They again start in the exterior of Γ but they meet Γ when

$$(|\lambda|+1)^2 = 4 + 4\lambda c + \lambda^2,$$

i.e.

$$|\lambda| - 2\lambda c = \tfrac{3}{2}.$$

The solutions are

$$\lambda = \begin{cases} \frac{3}{2(1-2c)} & \lambda > 0 \\ -\frac{3}{2(1+2c)} & \lambda < 0. \end{cases}$$

There exists at least one intersection for all c.

It would be difficult to extend this analytic argument to general convex sets Ω so we transfer to geometry.

Let ρ be a ray lying in $\Gamma^+_{0,1,0}$; call $P(\rho)$ the projection of ρ on the x, y plane. Figure 2.4 illustrates the types of intersections of $P(\rho)$ with the projections of

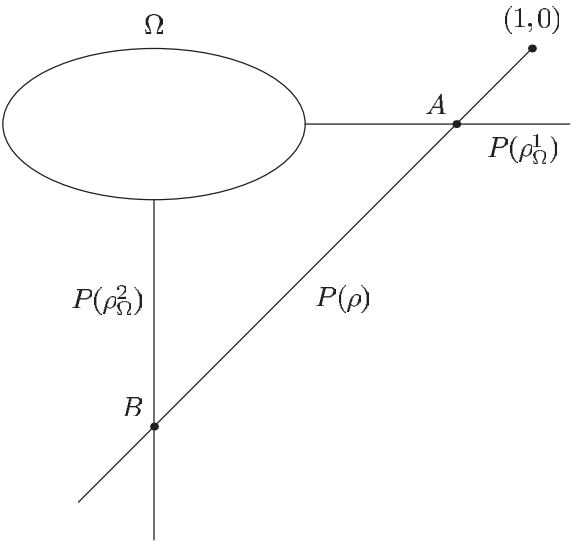

Figure 2.4

rays (e.g. $\rho^1_\Omega, \rho^2_\Omega$) in Γ_Ω. Note that the length of the part $P(\rho^2_\Omega : \rho)$ of $P(\rho^2_\Omega)$ up to the point of intersection with $P(\rho)$ is shorter than length $P(\rho : \rho^2_\Omega)$. Since ρ^2_Ω and ρ rise in t at the same rate the value of t on ρ above the intersection is larger than the corresponding value on ρ^2_Ω. The opposite is the case on $P(\rho) \cap P(\rho^1_\Omega)$.

Now, ρ starts out in the exterior of Γ_Ω. But when it reaches the point above $P(\rho) \cap P(\rho^2_\Omega)$ it is interior to Γ_Ω since it is above a point on Γ_Ω. This means that ρ intersects Γ_Ω at some point between $(0, 1, 0)$ and the point above $P(\rho) \cap P(\rho^2_\Omega)$.

It is geometrically evident that if $\tilde{\rho}$ is a ray for which no ρ^2_Ω exists with $|P(\rho^2_\Omega : \rho)| < |P(\rho : \rho^2_\Omega)|$ then there is a ${}^-\rho^2_\Omega$ with $|P({}^-\rho^2_\Omega : \rho)| < |P(\rho : {}^-\rho^2_\Omega)|$. Thus in all cases ρ or $-\rho$ intersects Γ_Ω.

For general $n > 2$ we can use the same argument in the 3 plane defined by ρ and the origin.

All this implies that the fundamental solution based at any particular point x in $\mathbb{R}^n$ lying outside Ω can be used to compute $F(0, x)$ from the values of F on Γ_Ω. (Standard estimates on solutions of the wave equation imply that for $f \in \mathcal{D}(\mathbb{R}^n)$ there are no convergence problems for $\zeta^\pm_x \cdot F$.) This completes the proof of

Theorem 2.9 *Let Ω be a smooth, strictly convex, compact hypersurface in $\mathbb{R}^n$ (n odd). Any smooth function $f \in \mathcal{D}(\mathbb{R}^n)$ is determined by $\mathbf{R}_\Omega(f)$. Moreover one can explicitly invert the Radon transform $f \to \mathbf{R}_\Omega(f)$ using the wave equation.*

Theorem 2.9 is concerned with spheres tangent to Ω. A natural question arises: what can replace spheres? If we think in terms of the wave equation, spheres arise when we consider the wave equation with CD on the space-like surface $\{t = 0\}$. What happens when we replace $\{t = 0\}$ by a time-like plane such as $\{x_1 = 0\}$?

An immediate difficulty arises from the fact that not all pairs (f, g) of C^∞ functions are admissible CD. One could assume f, g are entire functions and appeal to Cauchy–Kowalewski theory. But this seems to destroy the geometry.

Another method is to assume a spectral condition, meaning a condition on the support of the Fourier transform. Using the ideas of Chapter 4 of FA or of the formalism developed in Chapter 4 below, a natural condition is that $(f, g) \in \mathcal{S}$ and

$$\text{support}\,(\hat{f}, \hat{\mathbf{g}}) \subset \{\hat{t}^2 \geq \hat{x}^2_2 + \cdots + \hat{x}^2_n\}.$$

Under these conditions, for n odd, the value of the solution $F(t^0, x^0_1, x^0_2, \ldots, x^0_n)$ with $\text{CD}(F) = (f, 0)$ on $\{x_1 = 0\}$ and f even in t is given by integration of f over the intersection of the translated light cone $\Gamma + (t^0, x^0)$ with $\{x_1 = 0\}$. This intersection is a hyperboloid of the form $(t - t^0)^2 - (x_2 - x^0_2)^2 - \cdots - (x_n - x^0_n)^2 = (x^0_1)^2$.

In this way the wave equation is related to a Radon transform in which the spheres of Theorem 2.9 are replaced by hyperboloids of the above form. If Ω is a smooth, strictly convex, compact hypersurface in $x_1 = 0$ we denote by $\mathbf{R}_{\mathcal{H}\Omega}$

the Radon transform defined by integration over such hyperboloids which are tangent to Ω.

Problem 2.2 *Is $\mathbf{R}_{\mathcal{H}\Omega}$ injective on functions f satisfying the spectral condition? Can one reconstruct f from $\mathbf{R}_{\mathcal{H}\Omega}(f)$ using the wave equation?*

The relation of $\mathbf{R}_\Omega$ and $\mathbf{R}_{\mathcal{H}\Omega}$ to the wave equation was made possible by the Huygens principle for the wave equation.

Problem 2.3 *Let $P(D)f = 0$ be a hyperbolic system satisfying a Huygens principle. Construct an associated Radon transform and find its inversion.*

In Chapter 6 we relate the usual (linear) Radon transform to partial differential equations (the John equations) in a different manner.

Problem 2.4 *Find a relation of John-like equations to the usual Radon transform in the spirit of the wave equation to $\mathbf{R}_\Omega$.*

2.2 Tensor products and their topology

We have introduced tensor products of function spaces at various points in Chapter 1. In the present section we shall be concerned mainly with the topology of the tensor products.

Let $\mathcal{W}$ be a space of functions which are small at infinity. For each $\mathbf{g} \in \mathbf{G}(n,l)$ we can form the space

$$\mathcal{W}(\mathbf{g}) = \mathcal{W} * \delta_{L(s,\mathbf{g})} = \mathbf{R}^*\mathbf{R}(\mathbf{g})\mathcal{W}.$$

The Fourier transform of $\mathcal{W}(\mathbf{g})$ is

$$\hat{\mathcal{W}}(\mathbf{g}) = \hat{\mathcal{W}}\widehat{\delta_{L(0,\mathbf{g})}} = \hat{\mathcal{W}}(\hat{L}^\perp(0,\mathbf{g})) \times \delta_{\hat{L}^\perp(0,\mathbf{g})} \tag{2.28}$$

where $\hat{L}^\perp(0,\mathbf{g})$ is the normal plane to $L(0,\mathbf{g})$ thought of in the Fourier transform space. $\hat{\mathcal{W}}(\hat{L}^\perp(0,\mathbf{g}))$ is the space of restrictions of functions in $\hat{\mathcal{W}}$ to $\hat{L}^\perp(0,\mathbf{g})$. The discussion of the projection–slice theorem in Section 1.7 shows how to identify $\hat{\mathcal{W}}(\hat{L}^\perp(0,\mathbf{g}))$ with the dual of $\mathcal{W}'(\vec{\partial}L(0,\mathbf{g}))$, where $\vec{\partial}L(0,\mathbf{g})$ are directional derivatives spanning the directions on $L(0,\mathbf{g})$ and $\mathcal{W}'(\vec{\partial}L(0,\mathbf{g}))$ is the kernel of $\vec{\partial}L(0,\mathbf{g})$ on $\mathcal{W}'$.

We want to investigate the possible relation

$$\hat{\mathcal{W}} = \hat{\mathcal{W}}(\hat{S}) \underset{\sim}{\otimes} \mathcal{W}(\mathbf{G}) \tag{2.29}$$

both elementwise and topologically.

In this notation $\hat{S}$ is some fixed $\hat{L}^\perp(0,\mathbf{g})$ and $\hat{\mathcal{W}}(\hat{S})$ is the space of restrictions of functions in $\hat{\mathcal{W}}$ to $\hat{S}$. When $\mathcal{W}$ is LAU (see Section 1.4), as is the case in most of the examples we consider, then $\hat{\mathcal{W}}(\hat{L}^\perp(0,\mathbf{g}))$ is the space of entire functions on the complexification of $\hat{L}^\perp(0,\mathbf{g})$ with the growth conditions and AU structure induced from $\hat{\mathcal{W}}$. The space $\mathcal{W}(\mathbf{G})$ is a suitable space of functions on $\mathbf{G}$ which will usually be $\mathcal{E}(\mathbf{G})$, the space of C^∞ functions on $\mathbf{G}$. The notation $\underset{\sim}{\otimes}$ means the

subspace of the tensor product defined by the compatibility on the intersections $\hat{L}^\perp(0,\mathbf{g})\cap\hat{L}^\perp(0,\mathbf{g}')$ given by conditions (2) or (2*) of Proposition 2.1 for $l=n-1$ and by Proposition 2.5 when $l<n-1$.

The tensor product must be taken in the completed (topological) sense. Since the spaces we deal with are generally Schwartz spaces, all notions of completion coincide [82].

The simplest way of defining the topology on the tensor product is by interpreting it as Hom:

$$\hat{\mathcal{W}}(\hat{S})\otimes\mathcal{W}(\mathbf{G})=\ \text{Hom}\ [\mathcal{W}'(\mathbf{G}),\hat{\mathcal{W}}(\hat{S})]. \tag{2.30}$$

Hom means continuous linear maps. Then the compact–open topology is placed on the tensor product and on the subspace $\otimes$.

The subtensor product $\underset{\sim}{\otimes}$ corresponds to $\underset{\sim}{\text{Hom}}$. This is the space of continuous linear maps α of $\mathcal{W}'(\mathbf{G})\to\hat{\mathcal{W}}(\hat{S})$ which satisfy compatibility. The simplest way of expressing compatibility is by use of the basis $\{\delta_{\mathbf{g}}\}$ for $\mathcal{W}'(\mathbf{G})$. α can be identified with the function

$$F(\mathbf{g},s)=\alpha(\delta_{\mathbf{g}})(\hat{s})\quad\text{on}\ \bigcup_{\mathbf{g}}\hat{L}^\perp(0,\mathbf{g})^{\mathbb{C}}.$$

Compatibility means that this function is well defined on $\bigcup L^\perp(0,\mathbf{g})^{\mathbb{C}}$ and depends smoothly on $\mathbf{g},\hat{s}$. (When $l=n-1$ we also require COMP ORIGIN.)

The results of the previous section show that F is a smooth function on $\mathbb{R}^n$. The Korevaar–Wiegerinck theorem referred to in Section 1.5 shows that F extends to an entire function on $\mathbb{C}^n$. For most of the spaces we deal with this is sufficient to guarantee that F defines an element of $\hat{\mathcal{W}}$. For example in Chapter 3 we show that the Korevaar–Wiegerinck theorem can be modified so that if $\hat{\mathcal{W}}(\hat{S})$ is a space of entire functions of exponential type then the extension of F to $\mathbb{C}^n$ is an entire function of exponential type. (By (2.31) below this is also the case if $l<n-1$.) Moreover the Paley–Wiener theorem then shows that if $\hat{\mathcal{W}}(S)=\hat{\mathcal{D}}(S)$ or $\hat{\mathcal{E}}'(S)$ then $F\in\hat{\mathcal{D}}(\mathbb{R}^n)$ or $\hat{\mathcal{E}}'(\mathbb{R}^n)$ respectively.

We shall make the blanket assumption that $\otimes$ or $\underset{\sim}{\text{Hom}}$ can be identified with $\hat{\mathcal{W}}$. (We can regard this as some version of the Schwartz kernel theorem [46, 82].) Thus elementwise equality holds in (2.29); we are left with the problem of topological equality.

Note that any point $x+iy\in\mathbb{C}^N$ lies in the complex 2 plane $\hat{L}^\perp$ spanned by x and y. Thus for $l<n-1$,[2]

$$\bigcup\hat{L}^\perp(0,\mathbf{g})^{\mathbb{C}}=\mathbb{C}^N. \tag{2.31}$$

However, this is no longer true for $l=n-1$. For then $\hat{L}^\perp(0,\mathbf{g})$ is a line. We call $\hat{L}^\perp(0,\mathbf{g})^{\mathbb{C}}$ a *real complex line* (i.e. a complex line which has real direction numbers) and we write $\mathbb{C}^{\mathbb{R}}$ for the union of the real complex lines. Points of the

[2]This was pointed out to me by my student J. Han [85].

form $(x_1, ix_2, 0, \ldots, 0)$ with x_1, x_2 real and $\neq 0$ certainly do not belong to $\mathbb{C}^{\mathbb{R}}$. On the other hand (2.31) shows that the union of real complex planes of any dimension >1 is all of $\mathbb{C}^n$.

This represents a significant distinction between the tensor products associated to the hyperplane Radon transform and $\mathbf{R}^l$ for $l < n-1$ as we shall see presently.

We can describe $\mathbb{C}^{\mathbb{R}}$ as a real algebraic variety in $\mathbb{R}^{2n}$. For $\hat{x} \in \mathbb{C}^n$ we write $\hat{x} = \Re\hat{x} + i\Im\hat{x}$. $\hat{x} \in \mathbb{C}^{\mathbb{R}}$ means that there is a real point $a = (a_1, \ldots, a_n)$ such that $\hat{x} = \lambda a$ for some $\lambda = \lambda_1 + i\lambda_2$. Thus

$$\Re\hat{x} = \lambda_1 a, \quad \Im\hat{x} = \lambda_2 a \tag{2.32}$$

which means that $\Re\hat{x}$ and $\Im\hat{x}$ are parallel. We have thus arrived at

Remark. $\mathbb{C}^{\mathbb{R}}$ *is the same real algebraic variety as that associated to the parametric Radon transform (see (1.118)).*

There is a difference between the types of functions relevant to the two contexts of this algebraic variety. In the present case of $\mathbb{C}^{\mathbb{R}}$ the functions are holomorphic in λ whereas in the parametric Radon transform they are constant in τ where $\Im\hat{x} = \tau\Re\hat{x}$ so $\tau = \lambda_2/\lambda_1$. In either case the functions are determined by their values in the real space. However, τ is not defined when $\Re\hat{x} = 0$ and this represents one of the main difficulties in the proofs of results concerning the parametric Radon transform (see Chapter 6).

$\mathbf{G}(n,l)$ is compact. Thus the set $\{\delta_{\mathbf{g}}\}$ is a compact set in $\mathcal{E}'(\mathbf{G})$ and, in fact, for any differential operator ∂ on $\mathbf{G}$ the set $\{\partial\delta_{\mathbf{g}}\}$ is compact in $\mathcal{E}'(\mathbf{G})$. Hence if $\mathcal{W}(\mathbf{G}) = \mathcal{E}(\mathbf{G})$, as is usually the case, we can describe the topology in $\underset{\sim}{\mathrm{Hom}}[\mathcal{W}'(\mathbf{G}), \hat{W}(\hat{S})]$ by means of the semi-norms

$$\max_{\substack{\hat{s}\in\hat{L}^{\perp}(0,\mathbf{g})^{\mathbb{C}} \\ \mathbf{g}}} \frac{|\partial F(\hat{s},\mathbf{g})|}{k(\hat{s},\mathbf{g})}. \tag{2.33}$$

Here, for each $\mathbf{g}$, $\{k(\hat{s},\mathbf{g})\}$ is an AU structure (Section 1.4) for $\hat{W}(\hat{S})$. Moreover for any $F \in \underset{\sim}{\mathrm{Hom}}[\mathcal{W}'(\mathbf{G}), \hat{W}(\hat{S})]$ and any differential operator ∂ with constant coefficients in $\hat{s}$, $\partial F(\mathbf{g},\hat{s})$ is an entire function; (2.33) implies that F satisfies the growth conditions of functions in $\hat{\mathcal{W}}(\hat{S})$ uniformly in $\mathbf{g}$.

Let us first treat the case $l < n-1$. From the compactness of $\mathbf{G}$ it follows that, conversely, if $\{k(\hat{s},\mathbf{g})\}$ is an AU structure for a space $\mathcal{W}$ which is invariant under rotations then $\{k^0\}$ defined by

$$k^0(\hat{s}) = \min_{\mathbf{g}} k(\hat{s},\mathbf{g})$$

is an AU structure for $\hat{\mathcal{W}}(\hat{S})$.

This means that the topology of $\hat{\mathcal{W}}$ can be described using the AU structure $\{k^0\}$ thought of as functions on $\mathbb{C}^n$ which are independent of $\mathbf{g}$. One might think

that the presence of the operators $\{\partial\}$ in (2.33) means that the tensor product topology is stronger than that of $\hat{\mathcal{W}}$. We claim that the operators ∂ can be "absorbed" in the norms of $\hat{\mathcal{W}}$.

To verify our claim we use the fact that $\mathbf{G}$ is a homogeneous space of the rotation group. The differential operators on $\mathbf{G}$ are generated by the Lie algebra of the rotation group; generators for this Lie algebra can be chosen as plane rotations.

Let $\partial = \partial/\partial\hat{\theta}$ in the real 2 plane with coordinates $\hat{\xi}, \hat{\eta}$. Thus

$$\partial = \hat{\eta}\frac{\partial}{\partial\hat{\xi}} - \hat{\xi}\frac{\partial}{\partial\hat{\eta}}.$$

Let $\hat{F} \in \hat{\mathcal{W}}$. We want to estimate $\partial\hat{F}$ in terms of $\hat{\mathcal{W}}$ norms of $\hat{F}$. For any point $\hat{x} \in \mathbb{C}^n$ Cauchy's integral formula allows us to express $\partial\hat{F}/\partial\hat{\xi}(\hat{x})$ as an integral of values of $\hat{F}$ at points of the form $\hat{x}+\hat{y}$ where $\hat{y}$ lies in the complex $\hat{\xi}$ plane and, for example, $|\hat{y}| = 1$. For all the spaces $\mathcal{W}$ that we study there is no essential change in norms when we move a finite distance. Moreover multiplication by polynomials has no essential effect on the norms. Thus the term $\hat{\eta}\partial/\partial\hat{\xi}$ does not have any serious effect on the norms. A similar result holds for $\hat{\xi}\partial/\partial\hat{\eta}$ and hence for all ∂.

We have then proven

Theorem 2.10 *For $l < n-1$ equality holds in (2.29) both elementwise and topologically for AU spaces $\mathcal{W}$ which are (real) rotationally invariant.*

For $l = n-1$ the situation is much more complicated. Let us recall the notion of sufficient set for AU spaces as introduced in FA (see also Section 1.4):

Definition. *A closed subset $\sigma \subset \mathbb{C}^n$ is called* sufficient *for the AU space $\mathcal{W}$ if there is an AU structure $\{k\}$ for $\mathcal{W}$ such that the topology of $\hat{\mathcal{W}}$ can be described by the seminorms*

$$||F||_{k,\sigma} = \max_{\hat{x}\in\sigma} \frac{|F(\hat{x})|}{k(\hat{x})}. \tag{2.34}$$

By the maximum modulus theorem we can take $\sigma = \mathbb{C}^n - A$ where A is a compact set. The Phragmén–Lindelöf theorem allows much smaller σ. A striking example due to Taylor [149] is, for $\mathcal{W} = \mathcal{E}'$,

$$\sigma = \{m + in\} \tag{2.35}$$

where m and n are lattice points in $\mathbb{R}^n$.

If σ is a sufficient set then any $T \in \mathcal{W}'$ has a Fourier representation

$$T(x) = \int e^{ix\cdot\hat{x}} \frac{d\mu(\hat{x})}{k(\hat{x})} \tag{2.36}$$

for some k and some measure μ supported on σ. If T is a distribution then

the integral is taken in the distribution sense. The integral converges in the topology of $\mathcal{W}'$.

In particular, Taylor's example (2.35) shows that any C^∞ function f has a convergent "Fourier series" representation

$$f(x) = \sum a_{mn} e^{i(m+in)\cdot x}. \tag{2.37}$$

We can thus rephrase the problem of topological isomorphism as the question of whether $\mathbb{C}^{\mathbb{R}}$ is sufficient for $\mathcal{W}$.

Actually this formulation does not take into account the **G** derivatives ∂ of (2.33). Since the functions in $\hat{\mathcal{W}}(\mathbb{C}^{\mathbb{R}})$ extend to be entire, they extend to be holomorphic in $\hat{\theta}$. As in the proof of Theorem 2.10 we can use Cauchy's formula to estimate any **G** derivative by the maximum on a complex neighborhood.

Call $N\mathbb{C}^{\mathbb{R}}$ the set of points of distance ≤ 1 from $\mathbb{C}^{\mathbb{R}}$. Then our above remarks yield

Proposition 2.11 *If $\mathbb{C}^{\mathbb{R}}$ is sufficient for $\mathcal{W}$ then $\mathcal{W}$ is topologically isomorphic to the (Fourier transform of the) tensor product (2.29). If $N\mathbb{C}^{\mathbb{R}}$ is not sufficient for $\mathcal{W}$ then $\mathcal{W}$ is not isomorphic to the tensor product.*

Remark. *For the spaces that we deal with the functions k can be chosen to be "slowly varying," meaning that the difference between $\mathbb{C}^{\mathbb{R}}$ and $N\mathbb{C}^{\mathbb{R}}$ is insignificant for most of our purposes and we are able to ignore it.*

Let us clarify this remark.

We shall show in Chapter 3 that $\mathbb{C}^{\mathbb{R}}$ is sufficient for the space $\mathcal{E}'$. The space $\mathcal{D}$ is much more difficult to deal with and we shall see below that $\mathbb{C}^{\mathbb{R}}$ is not sufficient for $\mathcal{D}$. Nevertheless the $\mathcal{D}$ topologies on $\mathbb{C}^{\mathbb{R}}$ and $N\mathbb{C}^{\mathbb{R}}$ are essentially the same.

To understand this, we use the fact that $\mathbb{C}^{\mathbb{R}}$ is sufficient for $\mathcal{E}'$. Small in the $\hat{\mathcal{D}}$ norms on $\mathbb{C}^{\mathbb{R}}$ certainly implies small in the $\hat{\mathcal{E}}'$ norms on $\mathbb{C}^{\mathbb{R}}$ and hence on all of $\mathbb{C}$, in particular on $N\mathbb{C}^{\mathbb{R}}$ and, in fact, on $N_2\mathbb{C}^{\mathbb{R}}$ which is the set of points of distance ≤ 2 from $\mathbb{C}^{\mathbb{R}}$. By Cauchy's formula this implies $\hat{\mathcal{E}}'$ bounds on the derivatives on $N\mathbb{C}^{\mathbb{R}}$. Thus the functions which have small $\hat{\mathcal{D}}$ norms on $\mathbb{C}^{\mathbb{R}}$ also have small $\hat{\mathcal{E}}'$ norms for their derivatives on $N\mathbb{C}^{\mathbb{R}}$. Hence they cannot vary much from $\mathbb{C}^{\mathbb{R}}$ to $N\mathbb{C}^{\mathbb{R}}$. By examining the norms in $\hat{\mathcal{E}}'$ and $\hat{\mathcal{D}}$ we see that "$\mathbb{C}^{\mathbb{R}}$ is sufficient for $\mathcal{D}$ on $N\mathbb{C}^{\mathbb{R}}$," meaning that smallness in the $\hat{\mathcal{D}}$ norms on $\mathbb{C}^{\mathbb{R}}$ implies the same on $N\mathbb{C}^{\mathbb{R}}$.

We conclude, as in the remark, that for $\mathcal{D}$ the difference between $\mathbb{C}^{\mathbb{R}}$ and $N\mathbb{C}^{\mathbb{R}}$ is insignificant.

To analyze $\mathcal{W}(\mathbb{C}^{\mathbb{R}})$ it seems natural to use Fourier series on each sphere $\hat{r} = \text{const}$. This is carried out in Chapter 3. For the spaces $\mathcal{E}'$ and $\mathcal{D}_F$ (the space of C^∞ functions of compact support but with a topology which is weaker than that of $\mathcal{D}$ so that $\mathcal{D}'_F$ is the space of distributions of finite order) it is verified

in Chapter 3 that $\mathbb{C}^{\mathbb{R}}$ is sufficient and, in fact, the union of the real and the imaginary spaces is sufficient. Thus

Theorem 2.12 *The Radon transform is a topological isomorphism on $\mathcal{E}'$ and $\mathcal{D}_F$.*

The result for $\mathcal{E}'$ is due to Hertle [96, 97]. His proof differs considerably from ours.

What happens for $\mathcal{D}$? In a slightly different form Hertle proved that $\mathbf{R}$ is not a topological isomorphism.

We have discussed the tensor product from the Fourier transform point of view. One can also examine the question in x space. It is easily seen that $\mathcal{D}_F$ and $\mathcal{E}$ are (sub)tensor products of $\mathcal{E}(\mathbf{G})$ (where $\mathbf{G}$ is the $n-1$ sphere (or $\mathbf{P}^{n-1}$)) with the one-dimensional spaces $\mathcal{D}_F(\mathbb{R}^1)$ or $\mathcal{E}(\mathbb{R}^1)$. This means, interchanging the roles of $\mathcal{W}$ and $\hat{\mathcal{W}}$ in (2.29)ff., that the Radon transform is a topological isomorphism on $\hat{\mathcal{D}}_F$. We cannot, however, define the restriction of $\hat{\mathcal{E}}$ to a line so we do not know how to define the Radon transform on $\hat{\mathcal{E}}$ directly. However, we can define $\mathbf{R}^*\mathbf{R}u(L) = \delta_L * u$ for $u \in \hat{\mathcal{E}}$ by choosing a representative measure for u. The lack of uniqueness is discussed in Chapter 9; it leads to the same lack of uniqueness in $\delta_L * u$ which is thus defined as an element of $\hat{\mathcal{E}}$. (Unlike the usual definition of $\mathbf{R}^*$ we do not map $\mathbf{R}u$ into $\hat{\mathcal{E}}'$ but rather we define $\mathbf{R}^*\mathbf{R}$ by $\{\delta_L *\}$.) Again we obtain a topological isomorphism.

Note that, in terms of x, $\mathcal{D}$ itself is not a topological tensor product

$$\mathcal{D} \neq \mathcal{D}(\mathbb{R}^1) \tilde{\otimes} \mathcal{E}(\mathbf{G}). \tag{2.38}$$

The reason is that although a neighborhood of 0 in the tensor product can be described by inequalities on derivatives, the number of radial (i.e. $\mathbb{R}^1$) derivatives can be unbounded at infinity, whereas the number of $\mathbf{G}$ derivatives is bounded. For $\mathcal{D}$ the number of $\mathbf{G}$ derivatives is also allowed to be unbounded.

One might be tempted to pass from $\mathcal{D}(\mathbb{R}^1)\tilde{\otimes}\mathcal{E}(\mathbf{G})$ to $\hat{\mathcal{D}}(\mathbb{R}^1)\tilde{\otimes}\mathcal{E}(\mathbf{G})$ by Fourier transformation. In order to do this one would have to expand functions in terms of Fourier series on $\mathbf{G}$ and then use suitable Bessel transforms, as the Bessel transform is the Fourier transform on $\mathbb{R}^n$ of functions which are of the form $a(r)\varphi(\theta)$ where φ is a "special function" on the sphere (spherical harmonic).

Problem 2.5 *Carry out this project.*

Nevertheless the nonisomorphism in (2.38) is an indication that $\mathcal{D}$ does not have a tensor product structure. Further evidence of this comes from the result in [46] that, topologically,

$$\mathcal{D}(\mathbb{R}^n) \neq \mathcal{D}(\mathbb{R}^1) \otimes \cdots \otimes \mathcal{D}(\mathbb{R}^1) = \mathcal{D}_\times. \tag{2.39}$$

(Equality holds elementwise.) The right side of (2.39) is called $\mathcal{D}_\times$.

The reason for the topological inequality in (2.39) is that neighborhoods of zero in the tensor product $\mathcal{D}_\times$ allow bounds on only a bounded number of x_2

derivatives near the x_1 axis. Thus, for example,

$$\sum \frac{\partial^m}{\partial x_2^m}\delta_{(m,0,\dots,0)} \notin \mathcal{D}'_\times. \tag{2.40}$$

The advantage of using $\mathcal{D}_\times$ is that it is not difficult to take its n-dimensional Fourier transform. Of course, the elements of $\hat{\mathcal{D}}_\times$ are the same as those of $\hat{\mathcal{D}}$. But it is easily seen that the topology is defined by the semi-norms

$$||F||_k = \max \frac{F(\hat{x})}{k(\hat{x}_1)k(\hat{x}_2)\dots k(\hat{x}_n)} \tag{2.41}$$

where $\{k\}$ is an AU structure for $\mathcal{D}(\mathbb{R}^1)$. We set

$$k^\times(\hat{x}) = k(\hat{x}_1)k(\hat{x}_2)\dots k(\hat{x}_n). \tag{2.42}$$

Note that the topology of $\mathcal{D}_\times$ is strictly stronger than that of $\mathcal{D}(\mathbb{R}^1)\tilde{\otimes}\mathcal{E}(\mathbf{G})$ because, for example, near the diagonal $x_1 = x_2 = \dots = x_n$ a neighborhood of 0 in $\mathcal{D}_\times$ allows an unbounded number of derivatives of any type at infinity whereas in $\mathcal{D}(\mathbb{R}^1)\tilde{\otimes}\mathcal{E}(\mathbf{G})$ unboundedness is only in radial derivatives. Thus if our intuition about the relation between the topologies of $\mathcal{D}(\mathbb{R}^1)\tilde{\otimes}\mathcal{E}(\mathbf{G})$ and $\hat{\mathcal{D}}(\mathbb{R}^1)\tilde{\otimes}\mathcal{E}(\mathbf{G})$ is correct then we could expect that the topology of $\hat{\mathcal{D}}_\times$ is stronger than that of $\hat{\mathcal{D}}(\mathbb{R}^1)\tilde{\otimes}\mathcal{E}(\mathbf{G})$.

Indeed, this is the case, as we shall soon see.

Instead of dealing with $\mathcal{D}$, we can define a space $\mathcal{W}_\times$ for any AU space $\mathcal{W}$ which has restrictions to lines. We write $\hat{\mathcal{W}}_\mathbb{R}$ for the space $\hat{\mathcal{W}}$ with the topology defined using the $\hat{\mathcal{W}}$ norms only on $\mathbb{C}^\mathbb{R}$. $\mathcal{W}_{N(\mathbb{R})}$ is defined in a similar fashion.

Most of the spaces that interest us are invariant under the orthogonal group $\mathcal{O}(n)$. This means that the AU structures $k(\hat{x})$ can be chosen to be $\mathcal{O}(n)$ invariant. There are two interesting ways of constructing $\mathcal{O}(n)$ invariant AU structures from the $\{k_1\}$ of $\mathcal{W}(\mathbb{R}^1)$:

(1) $\{k_1(|\Re\hat{x}|, |\Im\hat{x}|)\}$
(2) $\{k_1(\hat{t})\}$ for $\hat{x} = \omega\hat{t}$, $\omega \in S^1$ (real unit sphere), $\hat{t} \in \mathbb{C}^1$.

It is clear that, in general, (1) induces (2) on $\mathbb{C}^\mathbb{R}$ so that, in general, the topology (1) is stronger than (2). Topology (2) is, except for $\mathbf{G}$ derivatives, the tensor product topology. When $\mathbb{C}^\mathbb{R}$ is sufficient for (1) the two topologies are the same; we have seen that this is so for $\mathcal{W} = \mathcal{D}_F$ or $\mathcal{W} = \mathcal{E}'$.

For the space $\mathcal{D}$ the topology (1) is the topology of $\hat{\mathcal{D}}$. This follows from the fact that the topology of $\mathcal{D}$ is $\mathcal{O}(n)$ invariant, meaning the semi-norms can be chosen to be $\mathcal{O}(n)$ invariant. We are going to show that it is strictly stronger than the topology of $\hat{\mathcal{D}}_\mathbb{R}$. For this we shall use the intermediate space $\mathcal{D}_\times$.

The main difficulty we meet in comparing $\hat{\mathcal{W}}_\times$ and $\hat{\mathcal{W}}_\mathbb{R}$ is that the norms, i.e. the $k_1(\hat{x}_1)$, are not, in general, monotonic under real scalar multiplication. (A positive function $\wp(\hat{x})$ is called monotonically increasing if $\wp(\alpha\hat{x}) \leq \wp(\hat{x})$ for

$0 \leq \alpha \leq 1$.) If $\mathcal{W}$ is a space of distributions, then in the examples of Chapter V of FA the corresponding $k_1(\hat{x}_1)$ can be chosen to be functions of $(|\Re\hat{x}_1|, |\Im\hat{x}_1|)$ which are monotonically increasing in both variables. But if $\mathcal{W}$ is a space of smooth functions, e.g. $\mathcal{W} = \mathcal{D}$, then the functions k_1 are increasing in $|\Im\hat{x}_1|$ but decreasing in $|\Re\,\hat{x}_1|$.

For the space $\mathcal{W} = \mathcal{D}$ we are saved because of the following consideration. For each k_1 in a suitable AU structure K, there exists a majorant k^1 of k_1 which is a function of $|\Im\hat{x}_1|$ only which is monotonically increasing, such that

$$K_1^0 = \{k^1 k_1\} \tag{2.43}$$

forms an AU structure for $\mathcal{D}$.

To understand this point note that the condition that the function $k^1(|\Im\hat{x}_1|)$ is a majorant of some k_1 is that it dominates any exponential $\exp(A|\Im x_1|)$ at infinity. By choosing $\{k_1\}$ and $\{k^1\}$ properly, multiplication by k^1 does not effect k_1 seriously.

Proposition 2.13 *Let $\mathcal{W}$ be an AU space which is $\mathcal{O}(n)$ invariant. Suppose that for any AU structure $K_1 = \{k_1\}$ of $\mathcal{W}_1 = \mathcal{W}(\mathbb{R}^1)$ there is another AU structure $\{k_1''\}$ with the following properties:*

(i) *For any real α_0 with $0 < \alpha_0 \leq 1$ the set $\{k_1' = \min_{\alpha_0 \leq \alpha \leq 1} k_1(\alpha\hat{x}_1)\}$ is an AU structure for $\mathcal{W}_1$.*

(ii) *For any k_1 there is a $k_1'' \leq k_1'$ such that*

$$k_1''(k_1'^*)^{n-1} \leq k_1.$$

Here

$$k_1'^*(\hat{x}_1) = \max_{0 \leq \alpha \leq 1} k_1'(\alpha\hat{x}_1).$$

Then the topology of $\mathcal{W}_\times$ is stronger (perhaps not strictly stronger) than the (tensor product) topology of $\mathcal{W}_{N(\mathbb{R})}$.

When $\mathcal{W} = \mathcal{D}$ (i) can be translated into a simple property of the space $\mathcal{D}$. If $\{N_1\}$ is a fundamental system of neighborhoods of zero in $\mathcal{D}$ then so is $\{N_1'\}$ where $N_1' = \{f | f(\alpha x) \in N_1, \alpha_0^{-1} \geq \alpha \geq 1\}$.

To verify (ii) we use the simple observation that we can assume any k_1 satisfies

$$k_1(\hat{x}_1) \leq cu(|\Im\hat{x}_1|)$$

where u is any fixed function which is monotonically increasing and dominates all linear exponentials. Hence

$$k_1'(\hat{x}_1) \leq cu(\alpha_0|\Im\hat{x}_1|)$$

so that

$$k_1'^*(\hat{x}_1) \leq cu(\alpha_0|\Im\hat{x}_1|).$$

There is very little qualitative difference between $u(\hat{x}_1)$ and $[u(\alpha_0\hat{x}_1)]^{n-1}$. In fact it is easy to see that $[u(\alpha_0\hat{x}_1)]^{n-1}$ could be chosen as an arbitrary function which dominates all linear exponentials. In particular we can make $[u(\alpha_0\hat{x}_1)]^{n-1}$ small enough at infinity so that $k_1(\hat{x})[u(\alpha_0\hat{x}_1)]^{1-n}$ still defines an AU structure on $\hat{\mathcal{D}}$. The point is that multiplication or division of norms in $\hat{\mathcal{D}}$ by u has little effect.

This verifies (ii) for $\mathcal{W} = \mathcal{D}$.

Proof Let $K = \{k\}$ be an AU structure for $\mathcal{W}$. For each k we want to produce a $k^\times$ as above (see (2.42)) so that $k^\times \leq k$ on $\hat{\mathbb{C}}^{\mathbb{R}}$. A real complex line $\hat{L}^\perp$ through the origin is of the form

$$\hat{L}^\perp_\omega = \{\omega\hat{t}\}_{\hat{t}\in\hat{\mathbb{C}}^1} \tag{2.44}$$

for some ω in the real unit sphere in $\hat{\mathbb{R}}^n$. By the $O(n)$ invariance of $\mathcal{W}$, $k(\omega\hat{t})$ can be chosen to be independent of ω, hence a function $k_1(\hat{t})$.

We want to satisfy the inequality

$$\tilde{k}^1_1(\hat{x}_1)\dots\tilde{k}^n_1(\hat{x}_n) \leq k_1(\hat{x}) \quad \text{on } \hat{\mathbb{C}}^{\mathbb{R}} \tag{2.45}$$

for $\{\tilde{k}^j_1\}$ an AU structure of $\mathcal{W}_1$. On $\hat{L}^\perp_\omega = \{(\omega_1,\dots,\omega_n)\hat{t}\}$ this inequality is

$$\tilde{k}^1_1(\omega_1\hat{t})\dots\tilde{k}^n_1(\omega_n\hat{t}) \leq k_1(\hat{t}). \tag{2.46}$$

We should like to set $\tilde{k}^j_1 = (k_1)^{1/n}$. This would be fine if k_1 were monotonically increasing, for then inequality (2.46) would be clear. However, since k_1 is not necessarily monotonically increasing we must hypothesize (i) and (ii). In any case, some $\omega_{j_0} \geq 1/n$; for simplicity of notation, suppose it is ω_1. Then we may assume that $\{k'_1(\hat{x}_1)\} = \{\min_{\frac{1}{n}\leq\alpha\leq 1} k_1(\alpha\hat{x}_1)\}$ is an AU structure for $\mathcal{W}_1$. We now replace $\tilde{k}^1_1$ in (2.46) by k''_1 and the other $\tilde{k}^j_1$ by k'^*_1 according to hypothesis (ii). All this allows us to establish inequality (2.45). The slowly varying hypothesis (see the remark following Proposition 2.11) allows us to replace $\hat{\mathbb{C}}^{\mathbb{R}}$ in (2.45) by $N(\hat{\mathbb{C}}^{\mathbb{R}})$. Hence by our above considerations, the topology of $\hat{\mathcal{W}}_\times$ is stronger than that of $\hat{\mathcal{W}}_{N(\mathbb{R})}$.

This completes the proof of Proposition 2.13.

Since the topology of $\mathcal{D}$ is strictly stronger than that of $\mathcal{D}_\times$ we deduce

Theorem 2.14 *The topology of $\hat{\mathcal{D}}$ is strictly stronger than that of the tensor product $\hat{\mathcal{D}}(\hat{S})\widetilde{\otimes}\mathcal{E}(\mathbf{G})$.*

All this pertains to the Fourier transform space. What does Theorem 2.14 mean for $\mathcal{D}$?

For each $\mathbf{g} \in \mathbf{G}$ the $\mathbf{g}$ component of the tensor product in Theorem 2.14, which is $\hat{\mathcal{D}}\delta_{\mathbf{g}}$ according to the "basis" $\{\delta_{\mathbf{g}}\}$ for $\mathcal{E}(\mathbf{G})$, can be identified via the Fourier transform with $\hat{\delta}_{\mathbf{g}} * \mathcal{D}$ where $\hat{\delta}_{\mathbf{g}} = \delta_{L(0,\mathbf{g})}$, i.e. with $\mathcal{D}(L^{\perp}(0,\mathbf{g})) = \mathcal{D}(S)$. (Recall that, in our notation, $L(0,\mathbf{g})$ is the hyperplane orthogonal to the unit vector defined by $\mathbf{g}$.) The space $\hat{\mathcal{D}}(\hat{S})\underset{\sim}{\otimes}\mathcal{E}(\mathbf{G})$ represents a direct integral with compatibility (using only smooth cross-sections) of the Fourier transforms of $\{\hat{\delta}_{\mathbf{g}} * \mathcal{D}\}$. That is,

$$\hat{\mathcal{D}}(\hat{S})\underset{\sim}{\otimes}\mathcal{E}(\mathbf{G}) = \int_{\sim} \widehat{\hat{\delta}_{\mathbf{g}} * \mathcal{D}} \tag{2.47}$$

so that

$$\widehat{\hat{\mathcal{D}}(\hat{S})\underset{\sim}{\otimes}\mathcal{E}(\mathbf{G})} = \int_{\sim} \hat{\delta}_{\mathbf{g}} * \mathcal{D} = \mathcal{D}(S)\underset{\sim}{\otimes}\mathcal{E}(\mathbf{G}). \tag{2.48}$$

(We are using the symbol $\hat{}$ both for the Fourier transform and its inverse when (hopefully) no confusion is possible.) Although the equality of the first and third terms of (2.48) may seem trivial, it is not because, as we have mentioned several times, the identification of a tensor product such as $\mathcal{D}(S)\underset{\sim}{\otimes}\mathcal{E}(\mathbf{G})$ depends on how we extend functions from S to $\mathbb{R}^n$. In (2.48) the extension consists of making $f(s) \otimes \delta_{\mathbf{g}}$ equal to the function f on $\mathbf{g}$, with the usual identification of $\mathbf{g}$ with S, and then extended to be constant in the direction of $\mathbf{g}^{\perp}$. For this reason we needed the intermediate identity in (2.47). The compatibility is defined by moment conditions.

The Radon transform provides us with a surjective map of $\mathcal{D}$ on this tensor product, namely $h \in \mathcal{D} \to \{\hat{\delta}_{\mathbf{g}} * h\} = \mathbf{R}^*\mathbf{R}h$.

Theorem 2.14 can be rephrased as

Theorem 2.15 $\mathbf{R}$ *is not a topological isomorphism of* $\mathcal{D}$ *onto* $\mathcal{D}(S)\underset{\sim}{\otimes}\mathcal{E}(\mathbf{G})$.

Remark. *Theorem 2.15 is of the same nature as a result of* Hertle [96, 97]. *Hertle uses the space* $\mathcal{D}(S\underset{\sim}{\times}\mathbf{G})$ *in place of* $\mathcal{D}(S)\underset{\sim}{\otimes}\mathcal{E}(\mathbf{G})$ *where* $\mathcal{D}(S\underset{\sim}{\times}\mathbf{G})$ *is the subspace of* $\mathcal{D}(S \times \mathbf{G})$ *defined by the moment conditions. Elementwise* $\mathcal{D}(S \times \mathbf{G})$ *is the tensor product* $\mathcal{D}(S) \otimes \mathcal{E}(\mathbf{G})$ *where, in terms of the basis* $\{\delta_{\mathbf{g}}\}$ *for* $\mathcal{E}(\mathbf{G})$, $\mathcal{D} \otimes \delta_g$ *is identified with* $\mathcal{D}\delta_{\mathbf{g}}$. *This differs from (2.48) but this difference is not very significant; the main difference is the topology.*

Our methods lead to a proof of Hertle's theorem. We need

Theorem 2.16 *The topology of* $\hat{\mathcal{D}}_{N(\mathbb{R})}$ *is stronger (perhaps not strictly stronger) than that of* $\hat{\mathcal{D}}(S\underset{\sim}{\times}\mathbf{G})$.

Proof We have seen that (elementwise) $\mathcal{D}(S\underset{\sim}{\times}\mathbf{G}) = \mathbf{R}\mathcal{D}$. (Actually we showed this for the space $\mathcal{S}$ in place of $\mathcal{D}$ but the same proof (actually easier) works for $\mathcal{D}$. Ideas of this ilk are discussed in detail in the next section.) The Fourier transform of $\mathcal{D}(S\underset{\sim}{\times}\mathbf{G})$ is, elementwise, $\mathcal{D}(S)\underset{\sim}{\otimes}\mathcal{E}(\mathbf{G})$. But the topology of $\mathcal{D}(S\underset{\sim}{\times}\mathbf{G})$ allows

unbounded order of $\mathbf{G}$ derivatives as we go to ∞ in S unlike the tensor product $\mathcal{D}(S)\widehat{\otimes}\mathcal{E}(\mathbf{G})$.

The proof of Theorem 2.16 depends on the explicit description of the topology of $\hat{\mathcal{D}}$ (and its proof) which is quite complicated, so we shall only sketch it.

According to the method of proof in [37] of the description of the Fourier transform space $\hat{\mathcal{D}}$ as an AU space, when we take the inverse Fourier transform of $\hat{f}$ on $\hat{\mathbf{g}}_0^{\perp}$, we shift the contour to large imaginary values to make $\hat{f}|_{\hat{\mathbf{g}}_0^{\perp}}$ and many derivatives small at infinity. Putting all the $\hat{f}|_{\hat{\mathbf{g}}^{\perp}}$ together shows that the topology of $\hat{\mathcal{D}}_{\mathbb{R}}$ is stronger than that of $\hat{\mathcal{D}}(S\widetilde{\times}\mathbf{G})$ if we ignore $\mathbf{G}$ derivatives. We can add as many $\mathbf{G}$ derivatives as we want for large $|x|$ by applying Cauchy's formula on the neighborhood $N(\mathbb{R})$ as in the proof of Theorem 2.8. Going to large $\Im\hat{x}$ on $\hat{\mathbf{g}}_0^{\perp}$ compensates for the $\mathbf{G}$ derivatives.

This completes the proof of Theorem 2.16.

Hertle's theorem follows from Theorems 2.14 and 2.16.

Hertle shows that the distribution

$$\sum \frac{\partial^{m_j}}{\partial x_2^{m_j}}\delta_{(j,0,\dots,0)} = T \tag{2.49}$$

is not in the range of $\mathbf{R}'$ on $\mathcal{D}'(S\times\mathbf{G})$ if the m_j are large enough. Note that, formally,

$$\hat{T} = \sum (i\hat{x}_2)^{m_j} e^{ij\hat{x}_1}. \tag{2.50}$$

Thus in the real complex line $\hat{x}_1 = \hat{x}_3 = \cdots = \hat{x}_n = 0$ we have

$$\hat{T}(0,\hat{x}_2,0,\dots,0) = \sum (i\hat{x}_2)^{m_j}. \tag{2.51}$$

Since the sum is infinite it cannot converge in the topology of $\hat{\mathcal{D}}'(\hat{x}_2)$. It is $\exp(ij\hat{x}_1)$ which acts as a convergence factor in $\hat{\mathcal{D}}$.

This explains why T is not in the range of $\mathbf{R}'$. But for $l < n-1$ it is true that T is in the range of $\mathbf{R}'$. For we can represent T in the form $d\mu/k$ in the 2 plane $(x_1,x_2,0,\dots,0)$. This can be done explicitly as in [43] or FA. To make $\hat{x}_2^{m_j}\exp(ij\hat{x}_1)$ small we represent the distribution $\hat{x}_2^{m_j}\exp(ij\hat{x}_1)\,d\hat{x}_1\,d\hat{x}_2$ on the real $\hat{x}_1,\hat{x}_2$ plane as a measure on the set $\hat{x}_2$ real and $\hat{x}_1$ on a "logarithmic curve"

$$\Im x_1 = a_j + b_j\log(1+|\hat{x}_2|). \tag{2.52}$$

This can be accomplished using Cauchy's theorem to shift contours on integration. For suitable constants a_j, b_j the series converges to $\hat{T}$ in the $\hat{x}_1,\hat{x}_2$ plane.

For $l < n-1$ Theorem 2.10 shows that the topology of $\mathcal{D}(S)\widehat{\otimes}\mathcal{E}(\mathbf{G})$ is the same as that of $\mathcal{D}$. Since the topology of $\mathcal{D}(S\times\mathbf{G})$ lies somewhat between the topology of the tensor product and that of $\mathcal{D}$ (since all standard topologies agree on bounded sets and $\mathcal{D}$ has the strongest such topology), it is also the induced topology from $\mathcal{D}(S\times\mathbf{G})$. Thus all the reasonable topologies are the same.

Finally there remains the question as to how to describe the actual topology of $\mathcal{D}$ in terms of $\mathbf{R} = \mathbf{R}^{n-1}$.

To answer this question we return to our description of $\hat{\mathcal{D}}(S)\widetilde{\otimes}\mathcal{E}(\mathbf{G}) = \hat{\mathcal{D}}(\mathbb{C}^{\mathbb{R}})$. The reason this topology is not that of $\hat{\mathcal{D}}(\mathbb{C}^n)$ is that $\mathbb{C}^{\mathbb{R}}$ is too small—it has too few lines (see e.g. (2.32)). How do we get more lines?

Note that $\mathbb{C}^{\mathbb{R}}$, being the union of real complex lines, is the essential variety for the Korevaar–Wiegerinck theorem (see Section 1.4). The Korevaar–Wiegerinck theorem is proved using power series; these are not good for estimates of decrease at infinity.

On the other hand we can search for an analogous idea using the Bernstein–Hartogs method (see Chapter 5). Instead of using lines through the origin we use complex lines of the form $\hat{x}_1, \ldots, \hat{x}_{j-1}, \hat{x}_{j+1}, \ldots, \hat{x}_n$ fixed real, $\hat{x}_j$ complex. (This set has real dimension $n+1$.)

To understand how things work, we use the method developed in [37] to characterize the topology of $\hat{\mathcal{D}}$. Let f be the inverse Fourier transform of $\hat{f}$. We fix $\hat{x}_2, \ldots, \hat{x}_n$ real and shift the contour in the $\hat{x}_1$ variable to imaginary values. This gives inequalities on x_1 derivatives of f for large x_1. These inequalities are uniform in $x_2, \ldots, x_n$. But this gives us only the topology of $\mathcal{D}_\times$.

To obtain the topology of $\hat{\mathcal{D}}$ itself we need planes of real dimension $n+2$ in $\mathbb{C}^n$, e.g. planes of the form $\hat{x}_3, \ldots, \hat{x}_n$ real, $(\hat{x}_1, \hat{x}_2) \in \mathbb{C}^2$ (and, of course, planes where the indices (1,2) are replaced by an arbitrary pair). Now we obtain inequalities on high-order x_1, x_2 derivatives when $|x_1| + |x_2|$ is large.

If x is large then some coordinate, say x_1, is large. Thus high-order x_1, x_j derivatives are small near x for any j. In particular high-order x_j derivatives are small for each j including $j = 1$. Hence we can make $\Delta^N f$ small for large N, x. It is standard that we can use the powers of the Laplacian to define the topology of $\mathcal{D}$; we do not need $\partial^N / \partial x_1^{N_1} \ldots \partial x_n^{N_n}$ since the latter can be estimated in terms of the former.

We have proven

Theorem 2.17 $\bigcup_{j,j'} \mathbb{R}^n + i\Im\{\hat{x}_j\} + i\Im\{\hat{x}_{j'}\}$ *is sufficient for* $\mathcal{D}$.

Remark. *Theorem 2.17 can be thought of as a "companion" to Theorem 2.10.*

If we insist on using lines through the origin, then we are faced with many difficulties. For each $\mathbf{g} \in \mathbf{G}$ we are given a map $\hat{\rho}_{\mathbf{g}}$ of $\hat{\mathcal{D}}(\mathbb{C}^n) \to \hat{\mathcal{D}}(\mathbb{C}^1)$; namely, restriction to the real complex line $\hat{L}^{\perp}(0, \mathbf{g})^{\mathbb{C}}$. But $\hat{\rho}_{\mathbf{g}}$, when thought of as a function on $\mathbf{G}$ whose range is maps of $\hat{\mathcal{D}}(\mathbb{C}^n) \to \hat{\mathcal{D}}(\mathbb{C}^1)$, has a holomorphic extension to the complexification $\mathbf{G}^{\mathbb{C}}$, not as a map into $\hat{\mathcal{D}}(\mathbb{C}^1)$ but as a map into $\hat{\mathcal{H}}'(\mathbb{C}^1)$ which is the space of entire functions of exponential type.

Looked at from the x viewpoint, $\rho_{\mathbf{g}}$, which is the map $f \to f * \delta_{\mathbf{g}}$, has a holomorphic extension to $\mathbf{G}^{\mathbb{C}}$ as a map into $\mathcal{H}'$ which is the dual of the space $\mathcal{H}$

of entire functions. This means that we should think of f, or more precisely $f\,dx$, as an element of $\mathcal{H}'$ (analytic functional). Then we should like to form its convolution with $\delta_{\mathbf{g}}$ which is the δ function of the complex hyperplane defined by $\mathbf{g} \in \mathbf{G}^{\mathbb{C}}$.

Unfortunately this convolution does not seem to make sense. The reason is that $\delta_{\mathbf{g}}$ does not belong either to $\mathcal{H}$ or to $\mathcal{H}'$.

We can understand this from a slightly different viewpoint. In the case of real $\mathbf{g}$ the fact that $\hat{\delta}_{\mathbf{g}} = \delta_{\hat{\mathbf{g}}^\perp}$, which is the basis for the projection–slice theorem, can be stated in the form

$$f \hat{*} \delta_{\mathbf{g}} = \hat{f}\, \delta_{\hat{\mathbf{g}}^\perp} = \hat{f}|_{\hat{\mathbf{g}}^\perp} \times \delta_{\hat{\mathbf{g}}=0}.$$

But for $\mathbf{g}$ complex the meaning of $\delta_{\mathbf{g}}$ is obscure because it does not belong to any reasonable space. Moreover it is difficult to interpret $\hat{\delta}_{\mathbf{g}}$ because, even if we could find a space $\mathcal{W}$ which contains $\delta_{\mathbf{g}}$, it seems unlikely that the exponentials would be in $\mathcal{W}_1'$ in any natural way since they cannot be integrated over complex planes.

All this means that we cannot interpret the Fourier transform of the restriction $\hat{f} \to \hat{f}|_{\hat{\mathbf{g}}^\perp}$ in the same geometric manner as that given by the projection–slice theorem. In fact, the obvious direct analog of the projection–slice theorem is false. To see this we use

Proposition 2.18 *Let $L(s, \mathbf{g})$ be any nonreal complex affine hyperplane, i.e. $L(s, \mathbf{g})$ is not the translation of the complexification of a real hyperplane. Then, if $n \geq 2$, $L(s, \mathbf{g})$ intersects $\mathbb{R}^n$ in a codimension 2 affine plane. Moreover any affine plane of codimension 2 is of this form.*

Proof Note that $L(s, \mathbf{g})$ is defined by

$$(a + ib) \cdot (x + iy) = s = p + iq \tag{2.53}$$

where $a, b, x, y \in \mathbb{R}^n$, $p, q \in \mathbb{R}^1$. This means

$$\begin{aligned} a \cdot x - b \cdot y &= p \\ b \cdot x + a \cdot y &= q. \end{aligned} \tag{2.54}$$

Thus $L(s, \mathbf{g}) \cap \mathbb{R}^n$ is defined by $y = 0$ and

$$\begin{aligned} a \cdot x &= p \\ b \cdot x &= q \end{aligned} \tag{2.55}$$

which is an arbitrary real affine plane of codimension 2 since a, b are arbitrary linearly independent points in $\mathbb{R}^n$. In fact the linear independence of a, b characterizes nonreal hyperplanes. For, if $L(0, \mathbf{g})$ is the complexification of a real hyperplane then there is a real vector α such that $\alpha \cdot L = 0$. Hence we can choose

$a = b = \alpha$. Conversely if $b = \lambda a$, $\lambda \neq 0$, then we deduce from (2.54) for $s = 0$ that $a \cdot x = a \cdot y = 0$ so $\{x + iy\}$ is the complexification of $\{x\}$ satisfying $a \cdot x = 0$.

Proposition 2.18 shows that the geometrically defined convolution $f * \delta_{\mathbf{g}}$ cannot depend holomorphically on $\mathbf{g}$ since there is a "jump" when $\mathbf{g}$ passes from real to nonreal. The complexification of a real complex affine hyperplane H meets $\mathbb{R}^n$ in a hyperplane while if H is not the complexification of a real hyperplane the intersection has codimension 2. In particular for $n = 2$ the values of $f * \delta_{\mathbf{g}}$ for $\mathbf{g}$ nonreal are the values of f at individual points while for $\mathbf{g}$ real they are integrals of f over real lines.

To determine the topology of the Radon transform of $\mathcal{D}$ intrinsically we must first understand the Fourier transform of the restriction $\hat{\mathcal{D}} \to \hat{\mathcal{D}}|_{\hat{\mathbf{g}}^\perp}$ for both real and nonreal lines $\hat{\mathbf{g}}^\perp$. Calling the Fourier transform space $\mathcal{D}_{\mathbf{g}}$ (spread functions for $\mathbf{g}$) we then have to give a proper determination of a direct integral of the form $\int \mathcal{D}_{\mathbf{g}}$. This direct integral differs from the usual direct integrals because the spaces $\mathcal{D}_{\mathbf{g}}$ are not isomorphic. Moreover it is not only the structures of the individual $\mathcal{D}_{\mathbf{g}}$ that are important, but how they are put together.

As to the individual structures we have

Proposition 2.19 *If* $\mathbf{g}$ *is not real then*

$$\mathcal{D}_{\mathbf{g}} = \mathcal{H}'(\mathbf{g}^\perp). \tag{2.56}$$

(Equality here means that $\mathcal{D}_{\mathbf{g}}$ *is identified with its restriction to* $\mathbf{g}^\perp$*.)*

Proof By Fourier transform Proposition 2.19 takes the form

$$\hat{\mathcal{D}}|_{\hat{\mathbf{g}}^\perp} = \hat{\mathcal{H}}'(\hat{\mathbf{g}}^\perp). \tag{2.57}$$

Now $\hat{\mathbf{g}}^\perp$ is a complex line which is not real; as such it is "elliptic," meaning that the topology induced from $\hat{\mathcal{D}}$ on $\hat{\mathbf{g}}^\perp$ is the same as the topology induced by the space $\hat{\mathcal{H}}'$ of all entire functions of exponential type (Chapter VIII of FA). This establishes our assertion.

Problem 2.6 *Develop a direct integral theory which is applicable to this structure.*

2.3 Support conditions

In this section we study the relation between the support of a function and the support of its (hyperplane) Radon transform. It is clear that for A an open set the support of $\mathbf{R}[\chi(A)]$ is the same as that of $\mathbf{R}[\chi(\text{convex hull } A)]$. ($\chi$ is the characteristic function.) Thus we shall concern ourselves with functions whose supports are given to lie in a convex set K.

For any $\mathbf{g}$ the support of $\mathbf{R}f(s, \mathbf{g})$ lies in the s interval which is the projection of K on the line $\mathbf{g}^\perp$. The union over $\mathbf{g}$ of such projections is called the *1-projection* of K and is denoted by $\mathbf{P}^1(K)$. Since K is the intersection of half-spaces it

follows that $\mathbf{P}^1(K)$ determines K. Note that if K is the positive orthant in $\mathbb{R}^n$ then $\mathbf{P}^1(K)$ is the complement of the negative orthant. This shows that $\mathbf{P}^1(K)$ need not be convex. $\mathbf{P}^1(K)$ is compact if and only if K is compact.

We can place $\mathbf{P}^1(K)$ in a more general setting. Let ϕ be a nonnegative function on $\mathbb{R}^n$. We define the *Legendre transform* (sometimes called *Young conjugate*)

$$\psi(y) = \max_x x \cdot y - \phi(x). \tag{2.58}$$

Note that

$$\max\left[x \cdot \frac{1}{2}(y_1 + y_2) - \phi(x)\right] \leq \frac{1}{2}\max[x \cdot y_1 - \phi(x)] + \frac{1}{2}\max[x \cdot y_2 - \phi(x)]$$

so ψ is convex even if ϕ is not convex. In case ϕ is convex, which we shall assume, it is the Legendre transform of ψ (see Chapter V of FA).

If K is a convex set we define

$$\phi[K](x) = \begin{cases} 0 & \text{if } x \in K \\ \infty & \text{if } x \notin K. \end{cases} \tag{2.59}$$

($\phi[A]$ can be defined in a similar fashion for any set A; $\phi[A]$ is a convex function if and only if the set A is convex.) For any y, $\max x \cdot y - \phi[K](x)$ is taken at the point $x(y) \in K$ for which $x \cdot y$ is maximal. If y is of unit length then $[x(y) \cdot y]y$ represents the extremity of the projection of K on the ray through y. Thus the restriction of $\psi(y)$ to the unit sphere represents $\mathbf{P}^1(K)$ which is the union of lines containing y through the origin with endpoints $y\psi(y)$ and $-y\psi(-y)$.

If K is the interval $[a, b]$ on $\mathbb{R}^1$ then

$$\psi[K](y) = \begin{cases} by & \text{for } y > 0 \\ ay & \text{for } y < 0. \end{cases}$$

For suitable convex functions ϕ we define the space $\mathcal{D}(\phi)$ as consisting of all functions f satisfying

$$\left|x^k f^{(j)}(x)\right| \leq c_{jk} e^{-\phi(x)} \tag{2.60}$$

for all $j, k \geq 0$. (This notation differs somewhat from that of FA which replaces $\phi(x)$ by $\phi(\lambda x)$ for some (or all, depending on the context) $\lambda > 0$.)

In particular $\mathcal{D}(\phi[K])$ is the space of C^∞ functions f which vanish outside K and

$$\left|x^k f^{(j)}(x)\right| \to 0$$

as $x \to \infty$ on K. The point is that on K the right side of (2.60) is constant. But if x is allowed to $\to \infty$ on K then the boundedness of $x^{l+1} f(x)$ implies $x^l f(x) \to 0$.

We now introduce *Paley–Wiener theory* (see Chapter V of FA for details). This asserts that the Fourier transform space $\hat{\mathcal{D}}(\phi)$ consists of all functions $F(\hat{x}) \in \mathcal{S}$ which are holomorphic on the strip

$$\psi(-\Im \hat{x}) < \infty$$

and satisfy

$$\left| x^k F^{(j)}(\hat{x}) \right| \leq c'_{jk} e^{\psi(-\Im \hat{x})}. \tag{2.61}$$

Notation. *If α is a nonnegative function which may take the value ∞ then when we write*

$$|f(x)| \leq \alpha(x)$$

we mean that this inequality holds when $\alpha(x) < \infty$ but we do not require f to be defined where $\alpha = \infty$.

In particular if ϕ is a function of $(|x_1|, \ldots, |x_n|)$ such that $\phi(x)/|x| \to \infty$ as $x \to \infty$ then we define the space $\mathcal{H}(\phi)$ as consisting of all entire functions f satisfying

$$|f(x)| \leq c e^{\phi(|x|)} \tag{2.62}$$

for $x \in \mathbb{C}^n$.

We cannot define the Fourier transform on $\mathcal{H}(\phi)$ but we can define it on the dual space $\mathcal{H}'(\phi)$ since the exponentials belong to $\mathcal{H}(\phi)$. We find as in Chapter V of FA

$$\hat{\mathcal{H}}'(\phi) = \mathcal{H}(\psi).$$

What does the Radon transform tell us? If $f \in \mathcal{D}(\phi)$ the projection–slice theorem tells us that the Fourier transform of $\mathbf{R}f(\mathbf{g})$ is the restriction of $\hat{f}$ to $\mathbf{g}^\perp$. Thus the Radon transform gives information about $\hat{f}$ on $\mathbb{C}^{\mathbb{R}}$. This leads us to the natural

Problem 2.7 *Let $f \in \mathcal{S}$ and suppose that $\hat{f}$ extends on every (real) line $\mathbf{g}^\perp$ to an entire function belonging to $\hat{\mathcal{D}}(\phi)(\mathbf{g}^\perp)$. Under what conditions on ϕ does this imply that $f \in \mathcal{D}(\phi)$?*

Problem 2.7 can be thought of as a weak form of the question of whether $\mathbb{C}^{\mathbb{R}}$ is sufficient for $\mathcal{D}(\phi)$. Problem 2.7 asks if a holomorphic function on $\mathbb{C}^{\mathbb{R}}$ satisfying the growth conditions of $\hat{\mathcal{D}}(\phi)$ extends to a function in $\hat{\mathcal{D}}(\phi)$. Sufficiency of $\mathbb{C}^{\mathbb{R}}$ means that, in addition, the topology of $\hat{\mathcal{D}}(\phi)$ can be described on $\mathbb{C}^{\mathbb{R}}$.

In case $n = 1$ the Phragmén–Lindelöf theorem shows that, for entire functions F of exponential type, exponential bounds on the imaginary axis together with boundedness of $|x^k F(x)|$ on the real axis imply bounds of the form

$$|F(\hat{x})| \leq c_k (1 + |\hat{x}|)^{-k} e^{\alpha^{\pm} \Im \hat{x}} \quad \Im \hat{x} \to \pm\infty \tag{2.63}$$

everywhere. $\alpha^{\pm}$ can be defined by

$$\alpha^{\pm} = \overline{\lim_{\hat{y}\to\pm\infty}} \log|F(\hat{x}_0 + i\hat{y})|/|\hat{y}|. \tag{2.64}$$

The $\overline{\lim}$ is independent of $\hat{x}_0$. In particular $\alpha^{\pm}$ are determined by the growth of F on the imaginary axis.

Similar results are valid in several complex variables (see [111]). This is the heart of the reason that the union of the real and imaginary spaces in $\mathbb{C}^n$ is sufficient for $\mathcal{E}$ or for $\mathcal{D}_F$. (In Chapter 3 we give a different type of proof.)

Let K be a compact convex set which is the closure of its interior. Suppose that $f \in \mathcal{S}$ and support $\mathbf{R}f \subset \mathbf{P}^1(K)$. By our above remarks $\hat{f}$ is holomorphic and satisfies

$$|\hat{f}(\hat{x})| \leq c_k(1+|\hat{x}|)^{-k} e^{\psi[K](\Im\hat{x})} \tag{2.65}$$

on $\mathbb{C}^{\mathbb{R}}$.

In particular $\hat{f}$ defines an entire function on any real line through the origin. The Korevaar–Wiegerinck theorem tells us that $\hat{f}$ extends from $\mathbb{C}^{\mathbb{R}}$ to an entire function on $\mathbb{C}^n$. In Chapter 3 we show that the Korevaar–Wiegerinck theorem can be sharpened so as to apply to entire functions of exponential type. We conclude that $\hat{f}$ is an entire function of exponential type belonging to $\mathcal{S}$ on the real space.

The proof of the Korevaar–Wiegerinck theorem depends on power series so it only gives us bounds of the form

$$|\hat{f}(\hat{x})| \leq ce^{a|\hat{x}|}.$$

However $\mathbb{C}^{\mathbb{R}}$ contains the imaginary space. We have noted above that from (2.65) on the imaginary space together with $f \in \mathcal{S}$ we can deduce (2.65) on all of $\mathbb{C}$ since $\hat{f}$ is of exponential type.

The Paley–Wiener theorem allows us to conclude that

$$\text{support } f \subset K \tag{2.66}$$

which is Helgason's support theorem.

The fact that $\alpha^{\pm}$ in (2.63) can be computed by (2.64) (with the analogous result for entire functions of exponential type in several complex variables) shows that ϕ of the form $\phi[K]$ satisfy the conditions of Problem 2.7. We do not know of any other general class of ϕ.

We want to study the analog of Helgason's theorem for noncompact K. If K is strictly convex with curvature bounded away from 0 then the problem can easily be reduced to that of proper strictly convex cones Γ. We assume that Γ is the closure of its interior and that Γ is smooth except at the origin.

The *generators* of Γ are the lines lying in the boundary of Γ. The set of generators can be identified with the *base* B of Γ which is a suitable cross-section of the set of generators. B is the boundary of a strictly convex compact set K which is the closure of its interior.

By $\hat{\Gamma}$ we denote the dual cone to Γ; that is, $\hat{\Gamma}$ consists of all y such that $y \cdot x \geq 0$ for all $x \in \Gamma$. For example, if Γ is the positive orthant then so is $\hat{\Gamma}$.

We claim that for any line L through the origin

$$\begin{aligned}\text{projection of } \Gamma \text{ on } L &= \mathbf{P}^1(\Gamma) \cap L \\ &= L \cap \text{complement}\,(-\hat{\Gamma}) \cup \{0\}.\end{aligned}$$

To verify this claim, let $x \neq 0$, $x \in L$, and $x' \in \Gamma$. The projection of x' on the line through x is

$$(x' \cdot x)\frac{x}{|x|^2}.$$

For x to be on the projection of the ray through x' we must have $x' \cdot x > 0$. This means that $x \in \text{complement}\,(-\hat{\Gamma}) \cup \{0\}$.

We have shown that

$$\text{closure}\,\mathbf{P}^1(\Gamma) = \text{closure complement}\,(-\hat{\Gamma}). \tag{2.67}$$

In what follows we shall sometimes omit the word "closure."

Note that $\hat{\Gamma}$ is much smaller than complement $-\hat{\Gamma}$. Hence support $f \subset \Gamma$ entails a much greater degree of vanishing than support $\mathbf{R}f \subset \mathbf{P}^1(\Gamma) = \text{complement}(-\hat{\Gamma})$. Thus we should not expect an analog of Helgason's theorem.

Indeed this is the case:

Theorem 2.20 *There is a function $f \in \mathcal{S}$ whose Radon transform vanishes on $-\hat{\Gamma}$ but support $f \not\subset \Gamma$.*

Proof The space of functions in $\mathcal{S}$ supported by Γ was denoted above by $\mathcal{D}(\phi[\Gamma])$. The Legendre transform $\psi[\Gamma]$ is given by

$$\psi[\Gamma](y) = \begin{cases} 0 & \text{if } y \in -\hat{\Gamma} \\ \infty & \text{if } y \in \text{complement } -\hat{\Gamma} \end{cases} \tag{2.68}$$

because, by the above remarks, $y \in -\hat{\Gamma}$ means that $x \cdot y < 0$ for all $x \in \Gamma$, meaning all x for which $\phi[\Gamma](x) = 0$ so $\max x \cdot y - \phi[\Gamma](x) = 0$. On the other hand for $y \in \text{complement}\,(-\hat{\Gamma})$ there is an $x \in \Gamma$ with $x \cdot y > 0$. Going to ∞ on the ray through x shows that $\psi[\Gamma](y) = \infty$.

The space $\hat{\mathcal{D}}(\phi[\Gamma])$ consists of all functions in $\mathcal{S}$ which have analytic continuation to the tube $\tau(\hat{\Gamma})$ over $\hat{\Gamma}$, meaning $\{\hat{x} + i\hat{y}\}$ with $\hat{y} \in \hat{\Gamma}$, and satisfy

$$\left|\hat{f}^{(j)}(\hat{x} + i\hat{y})\right| \leq c_{jk}(1 + |\hat{x}|)^{-k} \tag{2.69}$$

there. (This is (2.61) for $\psi = \psi[\Gamma]$.)

The Fourier transform of functions $h \in \mathcal{S}$ with support $\mathbf{R}h \subset$ complement $(-\hat{\Gamma})$ have analytic continuations to

$$\tau(\hat{\Gamma})^{\mathbb{R}} = \bigcup_{\substack{\hat{L} \text{ real line} \\ \hat{L} \cap -\hat{\Gamma} = \text{half-line}}} \{\hat{L}^{\mathbb{C}} \cap \tau(\hat{\Gamma})\}. \tag{2.70}$$

For, $\hat{h}\big|_{\hat{L}}$ is the Fourier transform of $\mathbf{R}h(L)$ whose support is a line or a half-line. It is a half-line (so $\hat{h}\big|_{\hat{L}}$ has an analytic continuation for all such h) if and only if L meets $-\hat{\Gamma}$ in a half-line, in which case support $\mathbf{R}h(L)$ is contained in the complement of $L \cap -\hat{\Gamma}$, i.e. in $L \cap \hat{\Gamma}$.

The difference between $\hat{\mathcal{D}}(\phi[\Gamma])$ and the space defined by (2.70) is that functions in $\hat{\mathcal{D}}(\phi[\Gamma])$ have analytic extensions from all of $\mathbb{R}^n$ while (2.70) requires analytic continuation only over real lines that meet interior $-\hat{\Gamma}$. We can certainly find functions $\hat{h} \in \mathcal{S}, \hat{h} \not\equiv 0$, which vanish on $\hat{\Gamma} \cup -\hat{\Gamma}$, which is the union of all such lines, but have no holomorphic extension from all of $\mathbb{R}^n$. (Clearly such an $\hat{h}$ extends holomorphically from $\hat{\Gamma} \cap -\hat{\Gamma}$, namely to $\equiv 0$.)

Theorem 2.20 is thereby proven.

The above construction of h is possible because the functions in $\mathcal{S}$ can vanish in open sets. In particular if we replace $\hat{\mathcal{D}}(\phi[\Gamma])$ by a space of functions which are real analytic or quasianalytic (see Chapter IX or Chapter XIII of FA) on all of $\mathbb{R}^n$ then our construction of $\hat{h}$ would fail. To achieve more regularity we introduce a new space $\mathcal{D}(\phi_1[K])$. $\phi_1[K]$ is convex; it is still $\equiv \infty$ on the complement of K, but ϕ_1 increases rapidly (at least exponentially) on K. The functions in $\hat{\mathcal{D}}(\phi_1[K])$ have holomorphic extensions from all of $\mathbb{R}^n$.

Problem 2.8 *For which ϕ_1 does the support property hold: if $f \in \mathcal{D}(\phi_1[\Gamma])$ and $\mathbf{R}f = 0$ on $-\hat{\Gamma}$ then support $f \subset \Gamma$?*[3]

In particular if $\phi_1 \equiv \infty$ outside a compact set then Helgason's support theorem applies.

We can analyze this problem using a variation of Lemma 1.1 combined with Korevaar–Wiegerinck theory. Let K be any strictly convex closed set with curvature bounded away from 0. If $x \notin K$ then there is an open set $\{L\}$ of hyperplanes through x which do not meet K. The spread functions corresponding to these hyperplanes span $\mathcal{D}'(\phi_1[K])$ if ϕ_1 is large enough by Lemma 1.1. The difficulty in applying this result is that we do not know that a function which has support in a small neighborhood of x can be approximated by linear combinations of spread functions for L such that these spread functions vanish on the points $y \in L^{\perp}$ for which $y + L$ meets K.

We can prove a positive result. The Korevaar–Wiegerinck idea, as made somewhat more precise in Chapter 3, shows that for spaces $\mathcal{H}(\phi)$ of entire functions

[3]As this book was about to go to press, this problem was solved in the affirmative by T. Banh in the quasianalytic case. He gave a simple example for Theorem 2.20.

defined as in (2.62) with $x \in \mathbb{C}^n$, we can pass from $\mathbb{C}^{\mathbb{R}}$ to $\mathbb{C}$ if $\phi(x) = \phi(|x|)$ and $\phi(x)/|x| \to \infty$ as $x \to \infty$ and ϕ has suitable regularity. This yields

Theorem 2.21 *If $\hat{f}$ belongs to $\hat{\mathcal{H}}(\psi)$ on $\mathbb{C}^{\mathbb{R}}$ then $\hat{f}$ extends to a function in $\hat{\mathcal{H}}(\psi)$.*

Remark. *If $f \in \mathcal{H}'(\phi)$ then we cannot define $\mathbf{R}f$ directly because $\delta_L \notin \mathcal{H}(\phi)$. However, the restriction of $\hat{f}$ to complex lines does make sense and it is reasonable to think of these restrictions as the Fourier transform of $\mathbf{R}f$.*

All the above considerations were concerned with the hyperplane Radon transform $\mathbf{R}^{n-1}$, which has been denoted by $\mathbf{R}$. We now make our notation precise and write $\mathbf{R}^l$ where $l \leq n-1$.

To understand what we can obtain using $\mathbf{R}^l$, suppose $n = 3$. Let

$$\text{support } f \subset \{y^2 - x_1^2 - x_2^2 \geq 0\} = \Gamma^+ \cup \Gamma^-. \tag{2.71}$$

Since the convex hull of $\Gamma^+ \cup \Gamma^-$ is all of $\mathbb{R}^3$ there is no support condition on $\mathbf{R}^2 f$. But if $\mathbf{g}^\perp$ represents any plane (dimension 2) containing the y axis

$$\text{support } \mathbf{R}^1 f(\mathbf{g}) \subset \Gamma^+_{\mathbf{g}^\perp} \cup \Gamma^-_{\mathbf{g}^\perp} \tag{2.72}$$

where $\Gamma^\pm_{\mathbf{g}^\perp} = \Gamma^\pm \cap \mathbf{g}^\perp$. (For an intuitive example let $\mathbf{g} = y, x_1$ plane so $\mathbf{g}^\perp = x_2$ axis.)

Conversely, (2.72) for all such $\mathbf{g}$ implies that support $f \subset \Gamma^+ \cup \Gamma^-$ if $f \in \mathcal{S}$. For, consider the restriction f_{y_0} of f to the plane $\{y = y_0\}$. Any affine line L in this plane is orthogonal to some $\mathbf{g}^\perp$, namely to the $\mathbf{g}^\perp$ spanned by the y axis and the line through the origin in $\{y = y_0\}$ orthogonal to L. Clearly

$$\mathbf{R}^1 f(y_0, s, L) = \mathbf{R}^1 f_{y_0}(s, L)$$

if s is in the orthogonal to L in $\{y = y_0\}$. Our hypothesis (2.72) allows us to apply Helgason's support theorem; we deduce support $f_{y_0} \subset x_1^2 + x_2^2 \leq y_0^2$. Since this holds for every y_0 our assertion is verified.

The same proof leads to

Proposition 2.22 *Let $f \in \mathcal{S}$. A necessary and sufficient condition for*

$$\textit{support } f \subset \{y^2 \geq x_1^2 + \cdots + x_n^2\} = \Gamma^+ \cup \Gamma^- \tag{2.73}$$

is

$$\textit{support } (\mathbf{R}^{n-1} f)(\mathbf{g}) \subset \Gamma^+_{\mathbf{g}^\perp} \cup \Gamma^-_{\mathbf{g}^\perp}$$

for any $n-1$ plane $\mathbf{g}$ orthogonal to the y axis.

Naturally there are other results for $l \leq n-1$ and for other nonconvex sets. Also the space $\mathcal{S}$ can be replaced by other spaces of functions. There is no problem at infinity because the $L(s, \mathbf{g})$ meets support f in a compact set.

Proposition 2.22 has an interesting relation to the CP for certain hyperbolic or ultrahyperbolic equations. Suppose $f \in \mathcal{S}'$ is a solution of

$$\frac{\partial^2}{\partial u^2} + \sum_1^p \frac{\partial^2 f}{\partial x_j^2} - \sum_1^q \frac{\partial^2 f}{\partial y_j^2} = 0. \tag{2.74}$$

We assume that neither p nor q is 0. By a modified version of the fundamental principle (see Section 1.4; the fundamental principle applies to AU spaces but for the ultrahyperbolic operator one can prove this version for the space $\mathcal{S}$) we can write f as a Fourier integral

$$f(u,x,y) = \int_{\hat{u}^2+\hat{x}^2=\hat{y}^2} e^{iu\hat{u}+ix\cdot\hat{x}+iy\cdot\hat{y}}\, d\hat{\mu}(\hat{u},\hat{x},\hat{y}). \tag{2.75}$$

$\hat{\mu}$ is a suitable distribution in $\mathcal{S}'$ supported on the real points of $\hat{u}^2 + \hat{x}^2 = \hat{y}^2$. The CD of f on $u = 0$ is given by

$$\begin{aligned} \mathrm{CD}(f) &= \left[f|_{u=0}, \frac{\partial f}{\partial u}|_{u=0} \right] \\ &= \left\{ \int e^{ix\cdot\hat{x}+iy\cdot\hat{y}}\, d\hat{\mu}^0(\hat{x},\hat{y}), \int e^{ix\cdot\hat{x}+iy\cdot\hat{y}}\, d\hat{\mu}^1(\hat{x},\hat{y}) \right\}. \end{aligned} \tag{2.76}$$

$\hat{\mu}^0$ is the sum of $\hat{\mu}$ at $\hat{u} = \pm(\hat{y}^2 - \hat{x}^2)^{1/2}$ and $\hat{\mu}^1$ is the sum of $i\hat{u}\hat{\mu}$ at the same points. (There is a slight modification at $\hat{u} = 0$.)

Since $\hat{u}, \hat{x}, \hat{y}$ are real

$$\text{support } \hat{\mu}^j \subset \{\hat{y}^2 \geq \hat{x}^2\}. \tag{2.77}$$

Conversely, if $\hat{\mu}^j$ satisfy (2.77) then they define a solution f of (2.74) for which $(\hat{\mu}^1, \hat{\mu}^2)$ is the Fourier transform of $\mathrm{CD}(f)$.

We can now apply Proposition 2.22 if $q = 1$ to $n-2$ planes orthogonal to the y axis.

If $q \geq 1$ then, in fact, we can use planes $\mathbf{g}$ of any fixed dimension $0 < p' < p$ orthogonal to the y axis. (We assume $p > 1$ for this to be of interest.) The proof of Proposition 2.22 shows that the analogous result is valid.

The projection–slice theorem asserts that the Fourier transform of the Radon transform $\mathbf{R}^{p'}\hat{\mu}(\hat{\mathbf{g}})$ is the restriction of μ to $\mathbf{g}^\perp$. Assuming $p' = p-1$ the support condition on $\mathbf{R}^{p'}f$ applied to $\hat{f} = (\hat{\mu}^0, \hat{\mu}^1)$ (see (2.77)) means that $(\mu^0, \mu^1)\big|_{\mathbf{g}^\perp}$ is the CD on $\mathbf{g}^\perp$ of a solution of

$$\frac{\partial^2 f}{\partial u^2} + \frac{\partial^2 f}{\partial x_{\mathbf{g}}^2} - \sum \frac{\partial^2 f}{\partial y_j^2} = 0 \tag{2.78}$$

for every such $\mathbf{g}^\perp$ containing the y axis. We have written $x_{\mathbf{g}}$ for the orthogonal to y in $\mathbf{g}^\perp$.

Using Proposition 2.22 we deduce another instance of the Radon ansatz:

Theorem 2.23 *Let (h^0, h^1) be functions of x, y $(q \geq 1)$ belonging to the Schwartz space $\hat{\mathcal{O}}'_C = \mathcal{O}_M$. A necessary and sufficient condition that (h^0, h^1) be the CD of a solution $f \in \mathcal{S}'$ of (2.74) on $u = 0$ is that for any plane $\mathbf{g}$ orthogonal to the y axis of dimension $p-1$ the restriction of (h^0, h^1) to $\mathbf{g}^\perp$ is the CD on $u = 0$ of a solution of (2.78).*

Theorem 2.23 deals with functions in $\mathcal{O}_M$. What is the analogous result for functions in $\mathcal{E}$? By the fundamental principle we do have a Fourier representation of the form (2.75) for f where μ is a suitable measure on the whole complex variety $\hat{u}^2 + \hat{x}^2 = \hat{y}^2$. We can then derive (2.76). It is easily seen that μ^j are exponentially decreasing in the real $\hat{x}, \hat{y}$ space outside a conical neighborhood of the cone (2.77) (instead of vanishing as was the situation when $f \in \mathcal{S}'$). Moreover, using the ideas of sufficient sets one can prove that the μ^j can be chosen to be exponentially decreasing except in a proper conical neighborhood of (2.77).

Subtracting the part of μ^j lying outside the real set $\{y^2 \geq (1-\varepsilon)x^2\}$ is equivalent to subtracting an entire function from $\mathrm{CD}(f)$. We are left with the part of μ^j lying in $\mathbb{R}^n \cap \{y^2 \geq (1-\varepsilon)x^2\}$. The fact that $\epsilon \neq 0$ causes difficulty. For this reason we see no method of proving an analog of Theorem 2.23.

3

HARMONIC FUNCTIONS IN $\mathbb{R}^n$

Given a family $\{\mathcal{O}_\alpha\}$ of algebraic varieties on $\mathbb{R}^n$, a harmonious way of relating functions h_α on the various $\mathcal{O}_\alpha$ is by means of a system $\vec{\partial}$ of linear partial differential equations such that each $\mathcal{O}_\alpha$ is a parametrization surface for solutions f of $\vec{\partial} f = 0$. We study analogs of the Dirichlet problem, pioneered by Chevalley, for which the parametrization problem involves only the value of f on $\mathcal{O}_\alpha$, i.e. no normal derivatives, even though the order of $\vec{\partial}$ may be high.

In Section 3.1 we show that this Dirichlet problem is well posed for $\vec{\partial}$ on the space of polynomials. This leads to a tensor product decomposition of the space of polynomials. It also leads to a (dual) Radon decomposition of $f = \int u_\alpha \, d\mu(\alpha)$ as the integral of solutions u_α of $\vec{\partial} u_\alpha = 0$ where support $\hat{u}_\alpha$ is a line $\hat{L}_\alpha$. The decomposition of harmonic functions into $\{u_\alpha\}$ is closely related to the Penrose transformation (Section 1.5).

In Section 3.2 we show that the same results persist for entire solutions of $(\partial^4/\partial x^4 + \partial^4/\partial y^4) f = 0$. Precisely, such solutions are parametrized by their values on $x^4 + y^4 = 1$ in contrast to the usual Dirichlet problem for smooth solutions, in which case the normal derivative is relevant. Other extensions of our results to entire functions are given. These are the first such results which do not require bounds on f.

In Section 3.3 we change our viewpoint and study the same question when $\vec{\partial}$ is group invariant. In the simplest case $\vec{\partial}$ is the Laplacian and the $\mathcal{O}_\alpha$ are spheres. To accomplish this we characterize various spaces of entire functions by means of their spherical harmonic expansions. Such expansions play a central role in our extension of Penrose theory and our treatment of the Watergate problem in Chapter 4.

Section 3.4 deals with analogous ideas when spheres are replaced by hyperboloids.

Up to now all the results depend on the homogeneity of the operators ∂_j which are the components of $\vec{\partial}$. In Section 3.5 certain ameliorations are presented.

In Section 3.6 we examine to what extent the coefficients in the harmonic expansion of polynomials can be expressed as integrals.

3.1 Algebraic theory

The Radon transform leads to a representation of a function $f(x)$ lying in a space $\mathcal{W}$ in terms of a linear combination (integral) of $\{\mathbf{R}^* \mathbf{R} f(\mathbf{g})\}$ for varying $\mathbf{g}$.

Each $\mathbf{g}$ corresponds to a spread: that is, a decomposition of the underlying space $\{x\}$ into fine grains called *leaves*. $\mathbf{R}^*\mathbf{R}f(\mathbf{g})$ is constructed from contributions of f to these fine grains; it is a solution of a system of linear partial differential equations $\partial(\mathbf{g})\mathbf{R}^*\mathbf{R}f(\mathbf{g}) = 0$ which depends on $\mathbf{g}$, i.e. on the coarse grain defined by $\mathbf{g}$. ($\mathbf{R}^*\mathbf{R}f$ actually belongs to the dual $\mathcal{W}'$ of $\mathcal{W}$ so in writing f as an integral of the $\mathbf{R}^*\mathbf{R}f(\mathbf{g})$ we have provided $\mathcal{W}$ with an injection into $\mathcal{W}'$.) For the usual Radon transform the leaves of the spread are parallel planes and $\partial(\mathbf{g})$ is the system of differential equations defined by differentiation in the directions of these planes.

Solutions of $\partial(\mathbf{g})h = 0$ comprise the spread functions $\{h(\mathbf{g})\}$. They are determined by their values on any cross-section S of the leaves of the spread. Moreover $\partial(\mathbf{g})$ is "hyperbolic," meaning that the restriction $h \to h|_S$ is an isomorphism of solutions of $\partial(\mathbf{g})h = 0$ onto functions on S. We call the system $\partial(\mathbf{g})h = 0$ an *analytic spread*.

The Fourier transform converts spread functions into slice functions. A slice is a subset $\hat{A}$ of the dual space $\hat{\mathbb{R}}^k$; slice functions are functions supported by $\hat{A}$. The relation between slice functions and spread functions is provided by the fundamental principle (Section 1.4) which expresses solutions of $\partial(\mathbf{g})h = 0$ as Fourier transforms of functions (measures) supported by an $\hat{A} = \hat{A}(\mathbf{g})$.

Slice functions have the advantage over spread functions in that if we are given a covering of $\hat{\mathbb{R}}^k$ by slices $\hat{A}_\alpha$ then the decomposition of an arbitrary f on $\hat{\mathbb{R}}^k$ into slice components for the $\hat{A}_\alpha$ is simple. It is for this reason that Fourier analysis plays such an important role in the standard Radon theory as, for example, in Chapter 2.

The system $\partial(\mathbf{g})h = 0$ is first order corresponding to the fact that $\hat{A}(\mathbf{g})$ is a linear variety. If we replace $\partial(\mathbf{g})$ by a higher order system $\vec{P}(D)h = 0$ where $\vec{P} = (P_1, \ldots, P_r)$ and $\vec{P}h = 0$ means $P_j h = 0$ for all j, then the analytic spread $\vec{P}(D)h = 0$ may not fit into a family of spreads (in contrast to $\{\partial(\mathbf{g})h = 0\}_{\mathbf{g}}$).

This suggests that we must accomplish with the single spread $\vec{P}h = 0$ what was accomplished with the whole family $\{\partial(\mathbf{g})h = 0\}$. To understand how this can be done let us examine the simplest case when $r = 1$ and $P = \partial/\partial x_1$. The corresponding slice is $\hat{A} = \{\hat{x}|\hat{x}_1 = 0\}$. Certainly not all functions are determined by their restrictions to $\hat{A}$. But if $\hat{f}$ is an entire function then $\hat{f}$ is determined by

$$\left\{ \left. \frac{\partial^j \hat{f}}{\partial \hat{x}_1^j} \right|_{\hat{x}_1 = 0} \right\} \tag{3.1}$$

since $\partial/\partial\hat{x}_1$ is the normal derivative to $\hat{A}$.

This means that we have to replace the variety $\hat{A}$ by the multiplicity variety (see Section 1.4)

$$\underline{\hat{A}} = \left\{ \hat{A}, \text{identity}; \hat{A}, \frac{\partial}{\partial \hat{x}_1}; \ldots : \hat{A}, \frac{\partial}{\partial \hat{x}_1^j}; \ldots \right\}. \tag{3.2}$$

By Fourier transformation we replace the spread functions $h(\mathbf{g})$ by the set of spread functions $h(\mathbf{g}_j) = \{x_1^j h(\mathbf{g})\}$. Our heuristics indicate that the spread functions from these spreads should span suitable function spaces on $\mathbb{R}^n$.

For general $\vec{P}$ the fundamental principle associates an algebraic multiplicity variety $\underline{V}$ which describes the kernel of $\vec{P}$. What plays the role of "normal derivatives" to $\underline{V}$? We shall see that when the P_j are homogeneous then it is the $P_j(\partial/\partial\hat{x})$ which serve this function; this seems to be the geometric content of Fischer's theorem (Lemma 3.1).

Although it is difficult to ascribe geometry to differential operators of order greater than 1 we can get some insight into the geometry in some simple cases (as the case of $\partial/\partial x_1$ treated above). Suppose $r = 1$, $n = 2$,

$$P = \frac{\partial^3}{\partial x_1 \partial x_2 \partial(x_1 - x_2)}. \tag{3.3}$$

The algebraic variety $\hat{V}$ associated to P is given by

$$\hat{V} = \{\hat{x}_1 = 0\} \cup \{\hat{x}_2 = 0\} \cup \{\hat{x}_1 = \hat{x}_2\}. \tag{3.4}$$

Our theory is restricted to polynomial solutions of $Ph = 0$. The Fourier transform of a polynomial can be expressed as

$$\hat{q} = \sum a_{j_1 j_2} \frac{\partial^{j_1 + j_2}}{\partial \hat{x}_1^{j_1} \partial \hat{x}_2^{j_2}} \delta_0. \tag{3.5}$$

It is not hard to verify that if h is a polynomial satisfying $P(\hat{x})\hat{h} = 0$ then $\hat{h}$ is of the form

$$\hat{h} = \sum a_j^1 \frac{\partial^j}{\partial \hat{x}_1^j} \delta_0 + \sum a_j^2 \frac{\partial^j}{\partial \hat{x}_2^j} \delta_0 + \sum a_j^3 \frac{\partial^j}{\partial(\hat{x}_1 + \hat{x}_2)^j} \delta_0. \tag{3.6}$$

If we examine the argument of Lemma 1.1, by duality (or else by a simple argument), we find that any homogeneous $\hat{q}$ of order 0, 1, 2 is equal to an $\hat{h}$ satisfying $P(\hat{x})\hat{h} = 0$ of the same order. But for order 3 the dimension of the space of $\{\hat{q}\}$ is four while that of $\{\hat{h}\}$ is three.

The operator $P(\partial/\partial\hat{x})$ is the product of the normals to the three lines that constitute $\hat{A}$. It provides the extra fourth dimension. This is the geometric content of the "normal derivative" quality of $P(\partial/\partial\hat{x})$.

For more complicated $\vec{P}$ the geometry is unclear. Perhaps the reader can supply more insight.

We shall deal with homogeneous P_j. Suppose that the V corresponding to $\vec{P}$ is a variety (not a multiplicity variety) so V is homogeneous. This means that V can be decomposed into slices which are complex lines L_α. If $\vec{P}h = 0$ then by the fundamental principle h can be expressed as the Fourier transform of a suitable

measure $\hat{h}$ on V. We shall show later that $\hat{h}$ can be chosen to be smooth. Thus $\hat{h}$ can be decomposed as

$$\hat{h} = \int \hat{h}_\alpha \, d\nu(\alpha) \tag{3.7}$$

where $\hat{h}_\alpha$ is a measure supported on L_α. Since L_α is a line it is defined by complex linear equations

$$L_\alpha = \{\hat{x} | \lambda_\alpha(\hat{x}) = 0\}.$$

Again by the fundamental principle, for α fixed, $\{\hat{h}_\alpha\}$ describes the space of solutions of the equations

$$\lambda_\alpha \left(\frac{\partial}{\partial x} \right) h_\alpha(x) = 0. \tag{3.8}$$

Thus we have decomposed a solution of $\vec{P}h = 0$ into an integral of simpler first-order systems $\lambda_\alpha(D)h_\alpha = 0$. We can regard this on the one hand as a general form of the Whittaker–Bateman decomposition of harmonic functions described in Section 1.5, and on the other hand, as a decomposition of the spread $\{\vec{P}h = 0\}$ into the semi-coarse grains $\{\lambda_\alpha(D)h_\alpha = 0\}$.

Precise forms of this decomposition are given in Theorems 3.5ff.

The theory presented here had its inception in the work of Chevalley on finite reflection groups; it was developed in detail by Todd and Shephard and many subsequent mathematicians (see [74, 75]). Let W be a finite group of real linear transformations of R^k with $t = (t_1, \ldots, t_k)$ as variable. Call $\vec{i} = (i_1, \ldots, i_r)$ where the i_j are homogeneous generators of the ring I of polynomial W invariants without a constant term. We say that a polynomial h is *harmonic* if $\partial(I)h = 0$; that is, h satisfies the equations

$$(\partial i_j)h \equiv i_j(\partial/\partial t)h = 0 \quad \text{all } j. \tag{3.9}$$

Remark. *The homogeneity of the i_j and the positivity of degree i_j play a crucial role in Chevalley's theory. The true significance of these properties is clarified at the end of Section 3.2 and in Section 3.3.*

As in the case of Laplace's equation (which corresponds to taking for W the rotation group) we pose the

Dirichlet problem (DP). *Given any function f on an orbit $\mathcal{O}$ of W, is there a unique harmonic h such that $h|_{\mathcal{O}} = f$?*

If this is the case we say that the DP is *well posed*. The DP is well posed for polynomials if and only if there is a *harmonious* way of comparing functions on orbits, namely f on $\mathcal{O}$ is compared to f' on $\mathcal{O}'$ if there is an harmonic h which equals f on $\mathcal{O}$ and f' on $\mathcal{O}'$.

We must distinguish two cases: *generic (nondegenerate) orbits* (i.e. having a maximum number of points) and *degenerate orbits*. In the latter case we must

interpret $\mathcal{O}$ as carrying with it a multiplicity; that is, $\mathcal{O}$ is a multiplicity variety in the sense of FA (see Section 1.4).

Let us return to the theory of Chevalley. Each orbit $\mathcal{O}$ is defined (set-theoretically) by equations of the form $i_j = c_j$ for suitable constants c_j. We write $\vec{c} = \vec{\imath}(\mathcal{O})$. We think of $\underline{\mathcal{O}}$, more precisely, as the multiplicity variety associated to the polynomials $\vec{\imath} - \vec{c}$. $\mathcal{O}$ is generic if there are no multiplicities.

Chevalley and Todd–Shephard proved the equivalence of the following three properties:

(A) *The DP is well posed.*
(B) *The group W is a reflection group (i.e. W is generated by reflections in hyperplanes).*
(C) *I is a free ring (on k generators).*

As an example of a nonreflection group we can choose $k = 2$, W consisting of the identity and $(t_1, t_2) \to (-t_1, -t_2)$. The ring I is generated by t_1^2, t_2^2, $t_1 t_2$. The ring is not free since

$$(t_1 t_2)^2 = t_1^2 t_2^2. \tag{3.10}$$

The harmonic functions have a (linear) basis

$$1,\ t_1,\ t_2.$$

Since the space of harmonic functions is three dimensional, there is no uniqueness possible for the DP.

Our point of departure is that the condition that W be a reflection group is equivalent to the fact that the polynomials i_j are *strongly independent*, meaning that sets of the form

$$i_{j_1} = c_{j_1}, \ldots, i_{j_p} = c_{j_p} \tag{3.11}$$

are complete intersections in the algebraic sense; that is, they have codimension p.

We drop the group W and deal with an arbitrary system of strongly independent[1] homogeneous polynomials $i_1, \ldots, i_r$ (so $r \leq k$). The "orbits" $\vec{\imath} = \vec{c}$ may now be of positive dimension. Harmonic is still defined by (3.9). We shall prove that the DP is well posed on such orbits in the framework of polynomials.

In general the DP is well posed for strongly free systems; it is also well posed for some other systems. The Chevalley theory deals only with finite groups; in the finite group case the well-posedness of the DP is equivalent to strong independence of the generators of the ring of invariants. But this is not true in general.

The well-posedness of the DP is capable of another interpretation: the polynomial ring $\mathcal{P}$ is a tensor product (see Proposition 3.5 below)

$$\mathcal{P} = H \otimes \tilde{I} \tag{3.12}$$

[1]We sometimes use the expression *strongly free*.

where H is the space of harmonic polynomials and $\tilde{I}$ is the ring generated by the i_j with 1 adjoined. The relation between the tensor product and the DP is clarified below.

H is a linear space. Of course, the product of two harmonics is not harmonic. However, the tensor product shows (more precisely, see Proposition 3.5 below) that H can be regarded as a sort of filtered algebra over the ring $\tilde{I}$. (This structure is called a filtered *cross-product* or *smash product*.)

When $r = 1$ we can pass from one $\mathcal{O}$ to another by scalar multiplication. Scalar multiplication leads to a harmonious geometric identification of the orbits. It also leads to a "polar coordinate" system (ρ, θ) in R^k. We first choose coordinates in some way on some fixed nondegenerate orbit $\mathcal{O}_1$ which does not contain any line. These are the θ coordinates. We also set $\rho = 1$ for $\mathcal{O}_1$. Then the θ coordinates of a point on any other nondegenerate orbit $\mathcal{O}$ are the same as the ones on $\mathcal{O}_1$ which correspond under scalar multiplication. The ρ coordinate is defined by scalar multiplication.

There is a slight problem in defining the ρ coordinate because there are d scalar multiplications corresponding to the $d =$ (degree i) roots of unity which preserve the orbits. (We allow the c_j to be complex.) But this is not serious if we require that the passage from $\mathcal{O}_1$ to any $\mathcal{O}$ depends continuously on $\mathcal{O}$.

Scalar multiplication leads in an obvious way to an isomorphism between function spaces on $\mathcal{O}$ and on $\mathcal{O}_1$ when $\mathcal{O}$ is nondegenerate and contains no lines.

Scalar multiplication represents one harmonious geometric way of identifying all the nondegenerate orbits $\mathcal{O}$. In the "group" case there is another method. Thus, suppose each nondegenerate $\mathcal{O}$ is the orbit of a group W. If W^0 represents the isotropy group of some fixed point $t_0^0 \in \mathcal{O}_0$ then we can explicitly identify $\mathcal{O}_0$ with W/W^0. To find coordinates on another orbit $\mathcal{O}$ we start with $t^0 \in \mathcal{O}$ whose isotropy group is W^0. We assume that there are only a finite number of W^0 fixed points in $\mathcal{O}$ so t^0 is chosen by moving continuously from some initial t_0^0. Then coordinates in W/W^0 give coordinates in $\mathcal{O}$ if we identify t^0 with the base point, i.e. the chosen W^0 fixed point. We call this *isotropy identification*.

For the orthogonal group acting on R^n the assumption that there are only a finite number of fixed points of W^0 holds when $n \geq 3$. For $n = 2$ it does not hold for the orthogonal group but it does hold if we include improper rotations (reflections).

Thus we have defined the θ coordinates; for the ρ coordinate we can use the values $i(\mathcal{O})$ or a suitable function of these values.

Note that this construction works even if $r > 1$.

If we are not in the group case and if $r > 1$ then it is rare that we can do anything in a "harmonious" way.

Problem 3.1 *Find a harmonious way of introducing "polar coordinates" when* $r > 1$ *and the orbits are not defined by a group.*

To each geometric object (orbit) $\mathcal{O}$ we can associate a unique analytic object which plays the role of a measure on $\mathcal{O}$. This is the "projection on the constant."

When the DP is well posed, every function f on $\mathcal{O}$ has a unique harmonic extension h. Since the equation (3.9) preserves homogeneity, we can write $h = \sum h^n$ where the h^n are the homogeneous components of h. Then the constant h^0 plays the role of the integral of f over $\mathcal{O}$ and we shall often write

$$h^0 = \int_{\mathcal{O}} f. \tag{3.13}$$

We shall discuss this orbital integral in Section 3.6.

The theory we present can be further generalized. We still require the i_j to be strongly independent and homogeneous of positive degree and harmonicity to be defined by (3.9). But we generalize the notion of orbit to include complex affine algebraic varieties X_Y of the form

$$X_Y : \vec{\imath} - \vec{Y} = 0. \tag{3.14}$$

Here Y_j is a polynomial of degree $<$ degree i_j. $\vec{\imath}$ is regarded as fixed and $\vec{Y}$ as variable. "Strongly independent" implies that all X_Y are of codimension r; for this it suffices to check the case $\vec{Y} = 0$. More precisely, we consider X_Y as a multiplicity variety; that is, as the multiplicity variety related to the ideal generated by the $i_j - Y_j$. In particular, X_0 (or, more precisely, $\underline{X}_0$) is the multiplicity variety associated to the ideal generated by the i_j. For convenience we require the i_j to be real; without this reality condition our formulas would become more complicated.

We shall sometimes refer to X_Y as an "orbit." We regard $\vec{\imath} - \vec{Y}$ as a deformation of $\vec{\imath}$. In some future work we shall consider deformations in which we allow Y_j to have the same degree as i_j as long as the homogeneous part of Y_j of highest degree is suitably restricted.

In the classical theory of the DP for the Laplacian, great significance is attached to the Dirichlet norm, which is an inner product for functions on the domain. The Dirichlet norm has the (Weyl) property that the harmonic functions and the functions vanishing on the boundary (i.e. functions with vanishing Dirichlet data) are orthogonal complements of each other. This means that the kernel of $\partial(\vec{\imath})$ and the image of (ideal generated by) $\vec{\imath} - \vec{c}$ are orthogonal complements. A general algebraic counterpart which was developed by Fischer (see [74, 75]) is the Fischer inner product of polynomials P, Q defined by

$$\langle P, Q \rangle = [\bar{P}(\partial/\partial t)Q]\big|_{t=0}. \tag{3.15}$$

It is readily verified that

Lemma 3.1 (Fischer's theorem) *The harmonic polynomials and the ideal $I_{\vec{\imath}}$ generated by i_j are orthogonal complements.*

Although this corresponds to the case $\vec{c} = 0$ in the usual Dirichlet norm, we shall use it to derive corresponding results for any $\vec{c}$ and, in fact, for any $\vec{Y}$.

Remark. *The space H of harmonics can be constructed either by using the orthogonal complement of $I_{\vec{\imath}}$ in the Fischer inner product or by means of differential equations. Conversely, the Fischer norm allows us to construct the ideal $I_{\vec{\imath}}$ from H.*

If, for $r = k$, we consider the DP for X_0, which is the multiplicity variety associated to $\vec{\imath}$, it consists of the origin with multiplicity (in which case it should be denoted by $\underline{X}_0$ and the DP can also be thought of as the Cauchy problem). Lemma 3.1 can be regarded as saying that the harmonics and the polynomials with vanishing Dirichlet data (at the origin) are orthogonal complements, since $I_{\vec{\imath}}$ is exactly the set of polynomials with vanishing Dirichlet data at the origin. Thus the Weyl property holds.

Problem 3.2 *Find an analog of Fischer's norm for more general X_Y, i.e. for $Y \neq 0$, which is related to the orthogonality of harmonic functions and functions with vanishing Dirichlet data (DD).*

In case $r \leq k$ we now consider the DP in the category of polynomials. If X_Y is an orbit, we call a function f on X_Y a polynomial on X_Y if it is the restriction to X_Y of a polynomial. This means (see FA) that f is holomorphic on X_Y and is of polynomial growth at infinity if X_Y is not compact. Later we shall consider the DP for other classes of entire functions.

Theorem 3.2 *If the i_j are strongly independent, the DP is well posed for any X_Y.*

Remark. *Theorem 3.2 can be thought of within the framework of Fourier transformation on algebraic (multiplicity) varieties. Thus, by the fundamental principle (Theorem 1.2 of Chapter 1), harmonic functions are Fourier transforms of measures on $\underline{X}_0$. The well-posedness of the DP is the assertion that the Fourier transform establishes an isomorphism between suitable measures (modulo an equivalence relation which is discussed in Chapter 9) on $\underline{X}_0$ and polynomials on X_Y.*

Proof We shall consider generic Y as the case of degenerate Y involves only some technical modifications. The proof is divided into two parts:

(1) Existence for the DP.
(2) Uniqueness for the DP.

In case the number of points in X_Y is finite, i.e. $r = k$, we have also

(3) dim H = number of points (counting multiplicity) in X_Y

Note that any two of (1), (2), (3) imply the third when the number of points on X_Y is finite. However, we allow the situation in which (3) is meaningless since both sides of the equation are infinite.

Proof of (3) (This proof uses some facts from algebraic geometry. The reader who is unfamiliar with such concepts can ignore this proof as we are most concerned with (1) and (2).) It is a consequence of intersection theory and the fact

that the i_j are strongly independent that the number of points in V_Y including infinity, i.e. in the projective space, is independent of Y. In fact, by Bezout's theorem this number is the product of the degrees of the i_j. When $\vec{c} = 0$ all the zeros are at the origin. It follows by continuity (or directly) that for any $\vec{c}$ there are no points of X_Y at infinity.

On the other hand the dimension of H is equal to the number of points in $\underset{\sim}{X}_0$. This can be seen in many ways: for example, it is an immediate consequence of the fundamental principle (see Section 1.4); (3) is thereby proven.

Proof of (1) We rely on the Fischer inner product. By definition, the functions on X_Y that we consider can be interpolated by (are restrictions of) polynomials so we must prove that for any polynomial P we can find an harmonic polynomial h so that $P = h$ on X_Y. Given P we use Fischer's theorem (Lemma 3.1) to write the orthogonal decomposition

$$P = h_0 + \sum a_l i_l \tag{3.16}$$

where h_0 is harmonic and the a_l are polynomials. (Although the Fischer orthonormal property is stated in terms of all polynomials—which do not form a Hilbert space—we can restrict our considerations to polynomials of degree $\leq$ degree P which form a Hilbert space under the Fischer inner product.) Since i_l are homogeneous, it is clear that each homogeneous part of any harmonic polynomial is also harmonic. Thus we can assume that each term on the right side of (3.16) is of degree $\leq$ degree P so that

$$\begin{aligned} \deg h_0 &\leq \deg P \\ \deg a_l &\leq \deg P - \deg i_l. \end{aligned} \tag{3.17}$$

Inequalities (3.17) allows us to iterate (3.16); that is, to apply (3.16) to a_l in place of P. Proceeding in this way we arrive at

$$P = h_0 + \sum \beta_l h_l \tag{3.18}$$

where the h_l are harmonic and β_l are polynomials with no constant terms belonging to the ring generated by the i_l; deg $h_0 \leq$ deg P, and deg $\beta_l h_l \leq$ deg P.

We now observe what happens on X_Y. We can replace each i_j in β_l by Y_j without changing the value of the right side of (3.7) on X_Y. This lowers the degree of β_l. Thus,

$$P = h_0 + Q \quad \text{on } X_Y \tag{3.19}$$

where h_0 is harmonic and deg $Q <$ deg P.

We can thus iterate the process with Q replacing P thereby lowering the degree of Q to 0. Since the constants are harmonic we have obtained the desired result: $P = h$ on X_Y.

Proof of (2) We shall give a complete proof for general r below. That proof, though simple, is not "elementary" as it uses the theory of multiplicity varieties. Therefore we start by giving elementary (combinatorial), albeit complicated, proofs in cases $r = 2, 3$, as these proofs are instructive and apply to some subtle situations. (For $r = 1$ the proof is essentially trivial.) Though the purely combinatorial method works for all r, we shall stop at $r = 3$. We begin with $r = 2$.

Suppose uniqueness does not hold. Then there exists a nontrivial harmonic function h that can be written in the form

$$h = (i_1 - Y_1)P_1 + (i_2 - Y_2)P_2. \tag{3.20}$$

Now, i_1 and i_2 are homogeneous. We write $d_j = \deg i_j$. We set

$$Y_j = \sum_{l<d_j} Y_j^{(l)} \tag{3.21}$$

where $Y_j^{(l)}$ is the homogeneous part of Y_j of degree l. Let c_j be integers satisfying

$$c_1 \geq \deg P_1, \quad c_2 \geq \deg P_2$$
$$\deg h \leq c_j + d_j = a \quad \text{independent of } j.$$

We want to prove that $h = 0$. We write

$$h = \sum_{l\leq a} h^{(l)}. \tag{3.22}$$

We shall prove that each $h^{(l)} = 0$. Note that each $h^{(l)}$ is harmonic.

We begin with $h^{(a)}$. We have, by (3.20),

$$h^{(a)} = i_1 {P_1}^{(c_1)} + i_2 {P_2}^{(c_2)}. \tag{3.23}$$

Thus $h^{(a)}$ belongs to $\underline{I}_{\vec{i}}$ so by Fischer's theorem, $h^{(a)} = 0$. Moreover, since i_1 and i_2 are strongly independent, they have no common factor, so we have

$${P_1}^{(c_1)} = i_2 Q^{(c_1-d_2)}, \quad {P_2}^{(c_2)} = -i_1 Q^{(c_2-d_1)}. \tag{3.24}$$

(Note that $c_1 - d_2 = c_2 - d_1$.)

In order to understand the general induction assumption, let us continue the process for two more steps: We compute the term of degree $a - 1$ in (3.20):

$$\begin{aligned} h^{(a-1)} &= i_1 P_1^{(c_1-1)} - Y_1^{(d_1-1)} P_1^{(c_1)} + i_2 P_2^{(c_2-1)} - Y_2^{(d_2-1)} P_2^{(c_2)} \\ &= i_1 P_1^{(c_1-1)} - Y_1^{(d_1-1)} i_2 Q^{(c_1-d_2)} + i_2 P_2^{(c_2-1)} + Y_2^{(d_2-1)} i_1 Q^{(c_2-d_1)} \end{aligned} \tag{3.25}$$

by (3.24). The right side belongs to $\underline{I}_{\vec{i}}$ so, by Fischer's theorem $h^{(a-1)} = 0$. This means that

$$i_1\big[P_1^{(c_1-1)} + Y_2^{(d_2-1)}Q^{(c_2-d_1)}\big] + i_2\big[P_2^{(c_2-1)} - Y_1^{(d_1-1)}Q^{(c_1-d_2)}\big] = 0. \tag{3.26}$$

Again, using the fact that i_1 and i_2 have no common factor we obtain

$$\begin{aligned} P_1^{(c_1-1)} + Y_2^{(d_2-1)}Q^{(c_1-d_2)} &= i_2 Q^{(c_1-d_2-1)} \\ P_2^{(c_2-1)} - Y_1^{(d_1-1)}Q^{(c_2-d_1)} &= -i_1 Q^{(c_2-d_1-1)}. \end{aligned} \tag{3.27}$$

We now compute the term of degree $(a-2)$ in (3.24):

$$\begin{aligned} h^{(a-2)} &= i_1 P_1^{(c_1-2)} - Y_1^{(d_1-1)}P_1^{(c_1-1)} - Y_1^{(d_1-2)}P_1^{(c_1)} \\ &\quad + i_2 P_2^{(c_2-2)} - Y_2^{(d_2-1)}P_2^{(c_2-1)} - Y_2^{(d_2-2)}P_2^{(c_2)} \\ &= i_1 P_1^{(c_1-2)} + Y_1^{(d_1-1)}Y_2^{(d_2-1)}Q^{(c_1-d_2)} \\ &\quad - Y_1^{(d_1-1)}i_2 Q^{(c_1-d_2-1)} - Y_1^{(d_1-2)}i_2 Q^{(c_1-d_2)} \\ &\quad + i_2 P_2^{(c_2-2)} - Y_2^{(d_2-1)}Y_1^{(d_1-1)}Q^{(c_2-d_1)} \\ &\quad + Y_2^{(d_2-1)}i_1 Q^{(c_2-d_1-1)} + Y_2^{(d_2-2)}i_1 Q^{(c_2-d_1)} \end{aligned} \tag{3.28}$$

by (3.24) and (3.27),

$$\begin{aligned} &= i_1\big[P_1^{(c_1-2)} + Y_2^{(d_2-2)}Q^{(c_1-d_2)} + Y_2^{(d_2-1)}Q^{(c_1-d_2-1)} \\ &\quad + i_2\big[P_2^{(c_2-2)} - Y_1^{(d_1-2)}Q^{(c_2-d_1)} - Y_1^{(d_1-1)}Q^{(c_2-d_1-1)}\big]. \end{aligned}$$

Thus $h^{(a-2)} \in \underline{I}_{\vec{i}}$ so $h^{(a-2)} = 0$. Again, using the fact that i_1 and i_2 have no common factor,

$$\begin{aligned} P_1^{(c_1-2)} + Y_2^{(d_2-2)}Q^{(c_1-d_2)} + Y_2^{(d_2-1)}Q^{(c_1-d_2-1)} &= i_2 Q^{(c_1-d_2-2)} \\ P_2^{(c_2-2)} - Y_1^{(d_1-2)}Q^{(c_2-d_1)} - Y_1^{(d_1-1)}Q^{(c_2-d_1-1)} &= -i_1 Q^{(c_2-d_1-2)}. \end{aligned} \tag{3.29}$$

We can now see the general induction pattern. We assume $h^{(a-j)} = 0$ for $j = 0, 1, \ldots, j_0$. Moreover, for the same range of j,

$$\begin{aligned} &P_1^{(c_1-j)} + Y_2^{(d_2-j)}Q^{(c_1-d_2)} + Y_2^{(d_2-j+1)}Q^{(c_1-d_2-1)} \\ &\quad + \cdots + Y_2^{(d_2-1)}Q^{(c_1-d_2+1-j)} = i_2 Q^{(c_1-d_2-j)} \\ &P_2^{(c_2-j)} - Y_1^{(d_1-j)}Q^{(c_2-d_1)} - Y_1^{(d_1-j+1)}Q^{(c_2-d_1-1)} \\ &\quad - \cdots - Y_1^{(d_1-1)}Q^{(c_2-d_1+1-j)} = -i_1 Q^{(c_2-d_1-j)} \end{aligned} \tag{3.30}$$

for some $Q^{(c_1-d_2)}, Q^{(c_1-d_2-1)}, \ldots, Q^{(c_1-d_2-j)}$. (Again recall that $c_1 - d_2 = c_2 - d_1$.) Of course, if $c_2 - d_1 - j \leq 0$ we define $Q^{(c_2-d_1-j)} = 0$.

We can now prove the result for $j_0 + 1$. Let us compute the term of order $a - j_0 - 1$ in (3.20):

$$\begin{aligned}
h^{(a-j_0-1)} &= i_1 P_1^{(c_1-j_0-1)} - \sum_{j=1}^{j_0+1} Y_1^{(d_1-j)} P_1^{(c_1+j-j_0-1)} \\
&\quad + i_2 P_2^{(c_2-j_0-1)} - \sum_{j=1}^{j_0+1} Y_2^{(d_2-j)} P_2^{(c_2+j-j_0-1)} \\
&= i_1 P_1^{(c_1-j_0-1)} + \sum_{j=1}^{j_0+1} \sum_{l=0}^{j_0-j} Y_1^{(d_1-j)} Y_2^{(d_2+j-j_0-1+l)} Q^{(c_1-d_2-l)} \\
&\quad - \sum_{j=1}^{j_0+1} Y_1^{(d_1-j)} i_2 Q^{(c_1-d_2+j-j_0-1)} \\
&\quad + i_2 P_2^{(c_2-j_0-1)} - \sum_{j=1}^{j_0+1} \sum_{l=0}^{j_0-j} Y_2^{(d_2-j)} Y_1^{(d_1+j-j_0-1+l)} Q^{(c_2-d_1-l)} \\
&\quad + \sum_{j=1}^{j_0+1} Y_2^{(d_2-j)} i_1 Q^{(c_2-d_1+j-j_0-1)} \\
&= i_1 \left(P_1^{(c_1-j_0-1)} + \sum_{j=1}^{(j_0+1)} Y_2^{(d_2-j)} Q^{(c_1-d_2+j-j_0-1)} \right) \\
&\quad + i_2 \left(P_2^{(c_2-j_0-1)} - \sum_{j=1}^{(j_0+1)} Y_1^{(d_1-j)} Q^{(c_2-d_1+j-j_0-1)} \right). \qquad (3.31)
\end{aligned}$$

(Note that when $j = j_0 + 1$ the double sums above vanish since they do not contain any terms. Thus the double sums are actually

$$\sum_{j=1}^{j_0} \sum_{l=0}^{j_0-1}$$

from which the cancellation of double sums is clear.)

Again using Fischer's theorem we deduce that $h^{(a-j_0-1)} = 0$ and using the fact that i_1 and i_2 have no common factor we see that (3.30) holds for $j = j_0+1$. Thus our assertion is verified for $r = 2$.

What are the modifications necessary to pass to the case of arbitrary r? We used the assumption $r = 2$ in several places, e.g. (3.24) to know that the solution

of $i_1R_1 + i_2R_2 = 0$ is $R_1 = i_2Q, R_2 = -i_1Q$. In case $r > 2$ we should replace this equation by

$$\sum i_j R_j = 0. \tag{3.32}$$

For example, (3.23) is to be replaced by

$$h^{(a)} = \sum i_j P_j^{(c_j)}. \tag{3.23*}$$

We again assume that $d_j + c_j$ is independent of j. By Fischer's theorem this leads to (3.32) with R_j replaced by $P_j^{(c_j)}$. To continue we must know the solutions of (3.32).

Let us now examine the case $r = 3$. The fact that the i_j are strongly independent means that anything in the module of relations of i_j, i.e. any $\vec{R}$ satisfying (3.32), is of the form

$$\vec{R} = \vec{\imath} \times \vec{Q} \tag{3.33}$$

in the usual notation of cross- and dot products. Here $\vec{Q} = (Q_1^{c_2-d_3}, Q_2^{c_3-d_1}, Q_3^{c_1-d_2})$. This assertion as well as the generalization to arbitrary r will be proven below.

We need the readily verifiable properties

$$\begin{aligned} \vec{Y} \cdot (\vec{\imath} \times \vec{Q}) &= -\vec{\imath} \cdot (\vec{Y} \times \vec{Q}) \\ \vec{Y} \cdot (\vec{Y} \times \vec{Q}) &= 0. \end{aligned} \tag{3.34}$$

Our induction assumption is $h^{(a-j)} = 0$ for $j = 0, 1, \ldots, j_0$ and

$$\vec{P}^{(c-j)} + \sum \vec{Y}^{(d-j+l)} \times \vec{Q}^{(c-d-l)} = \vec{\imath} \times \vec{Q}^{(c-d-j)}. \tag{3.30*}$$

Here $\vec{Q}^{(c-d-l)} = (Q_1^{(c_2-d_3-l)}, Q_2^{(c_3-d_1-l)}, Q_3^{(c_1-d_2-l)})$ so that all vectors in (3.30*) have the same homogeneous degrees $(c_1 - j, c_2 - j, c_3 - j)$. To prove the result for $j = j_0 + 1$, we compute the term of order $(a - j_0 - 1)$ in the analog of (3.20) for $r = 3$:

$$\begin{aligned} h^{(a-j_0-1)} &= \vec{\imath} \cdot \vec{P}^{(c-j_0-1)} - \sum_{j=1}^{j_0+1} \vec{Y}^{(d-j)} \cdot \vec{P}^{(c+j-j_0-1)} \\ &= \vec{\imath} \cdot \vec{P}^{(c-j_0-1)} + \sum_j \sum_l \vec{Y}^{(d-j)} \cdot \big[\vec{Y}^{(d+j-j_0-1+l)} \times \vec{Q}^{(c-d-l)}\big] \\ &\quad - \sum_j \vec{Y}^{(d-j)} \cdot \big[\vec{\imath} \times \vec{Q}^{(c-d+j-j_0-1)}\big] \\ &= \vec{\imath} \cdot \vec{P}^{(c-j_0-1)} + \vec{\imath} \cdot \big[\sum \vec{Y}^{(d-j)} \times \vec{Q}^{(c-d+j-j_0-1)}\big]. \end{aligned} \tag{3.31*}$$

The double sum in (3.31*) vanishes by (3.34) since each term with distinct subscripts occurs twice with opposite signs. The induction is now completed as before.

We have thus completed the proof for $r = 3$. Let us pass to the case of general r.

The simplest way of proceeding is by introduction of the exterior algebra generated by the i_j with polynomial coefficients. We could prove that the exterior algebra defines an acyclic complex, meaning that its cohomology vanishes. To avoid the abstraction we shall deal only with the first two stages.

In simple terms what we need is

Proposition 3.3 *If $\vec{R}$ belongs to the module of relations of $\vec{\imath}$, meaning (3.32) holds, then $\vec{R}$ is of the form*

$$R_j = \sum i_l (Q_{jl} - Q_{lj}). \tag{3.35}$$

Remark 1 *Equation (3.35) reduces to (3.33) for $r = 3$ where the vector $\vec{Q}$ of (3.33) is replaced by $\{Q_{ij}\}$ with $Q_1 \leftrightarrow Q_{23}$, $Q_2 \leftrightarrow Q_{31}$, $Q_3 \leftrightarrow Q_{12}$ and other components of $\boxed{Q}$ vanish.*

Remark 2 *In terms of the exterior algebra (3.35) states that if $\delta\vec{R} = 0$ then $\vec{R} = \delta\boxed{Q}$.*

Assuming Proposition 3.3 we can complete the proof of Theorem 3.2 as follows. We replace $\vec{\imath} \times \vec{Q}$ in (3.33), (3.34), etc., by $\delta\boxed{Q}$ (the right side of (3.35)). The cross-product $\vec{Y} \times \vec{Q}$ becomes the interior product which takes the same form as (3.35) with $\vec{Y}$ replacing $\vec{\imath}$ (so $\vec{Y}$ is a 1 form). The dot product is again the interior product. $\vec{Q}$ is regarded as a 2 form while $\vec{Y}$ and $\vec{Y} \times \vec{Q}$ and $\vec{\imath} \times \vec{Q}$ are regarded as 1 forms (the components depend on a single index).

With this in mind formulas (3.34) are true for general r. Thus the induction argument of (3.30*) and (3.31*) proceeds as before.

The proof of Theorem 3.2 is completed by

Proof of Proposition 3.3 We want to use induction but this does not work unless we prove the following strengthened form of Proposition 3.3. Let $J = (j_1, \ldots, j_p)$ be a subset of $(1, \ldots, r)$. Call $\underline{V}_J$ the multiplicity variety associated to the ideal generated by $\{i_j\}_{j \in J}$. If the indices of $\vec{R}$ do not belong to J and

$$\sum i_j R_j = 0 \quad \text{on } \underline{V}_J$$

then $\vec{R} = \delta\boxed{Q}$ on $\underline{V}_J$ where the indices of $\boxed{Q}$ do not belong to J. (For precision we sometimes write $\delta^{(J)}$ for δ.)

For simplicity of notation we assume that $J = \{(1, 2, \ldots, p)\}$. The proof goes by induction on $r - p$. The case $p = r$ is trivial. For $r - p = 1$ we have $i_r R_r = 0$ on

$\underline{V}_J$. Since the $\{i_l\}$ are strongly independent, the dimension of each irreducible component of the varieties in $\underline{V}_J$ is positive and i_r does not vanish on them. Thus $R_r = 0 = \delta^{(J)}(0)$ on $\underline{V}_J$, which is the desired result.

Now assume that the result is proven for $p = r-1, r-2, \ldots, r-q$. We want to prove the result for $p = r-q-1$. (We assume $q < r$ otherwise there is nothing to prove.) Our hypothesis is

$$i_{p+1}R_{p+1} + \cdots + i_r R_r = 0 \quad \text{on } \underline{V}_J$$

which is equivalent to

$$i_{p+1}R_{p+1} + \cdots + i_r R_r \equiv 0 \mod I_J \tag{3.36}$$

where I_J is the ideal generated by $(i_1, \ldots, i_p)$. We denote $(1, \ldots, p, p+1)$ by J'. Thus the above relation implies

$$i_{p+2}R_{p+2} + \cdots + i_r R_r = 0 \quad \text{on } \underline{V}_{J'}.$$

By our induction assumption this gives, with $\vec{R'} = (R_{p+2}, \ldots, R_r)$,

$$\vec{R'} \equiv \delta^{(J')}Q', \qquad \mod I_{J'}$$

or, what is the same thing,

$$\vec{R'} \equiv \delta^{(J')}Q' + i_{p+1}\vec{T'} \qquad \mod I_J. \tag{3.37}$$

We insert (3.37) in (3.36). Note that $\delta\delta = 0$ means that

$$i_{p+2}(\delta^{(J')}Q')_{p+2} + \cdots + i_r(\delta^{(J')}Q')_r = 0$$

so we obtain

$$i_{p+1}(R_{p+1} + i_{p+2}T'_{p+2} + \cdots + i_r T'_r) = 0 \quad \text{on } \underline{V}_J. \tag{3.38}$$

Since $i_1, \ldots, i_r$ are strongly independent, it follows that i_{p+1} does not vanish on any irreducible component of any variety in $\underline{V}_J$. Thus

$$R_{p+1} + i_{p+2}T'_{p+2} + \cdots + i_r T'_r = 0 \quad \text{on } \underline{V}_J.$$

We have thus shown that

$$\begin{aligned}\vec{R} &= (R_{p+1}, \vec{R'}) \\ &= (-i_{p+2}T'_{p+2} - \cdots - i_r T'_r, \delta^{(J')}Q' + i_{p+1}\vec{T'}) \quad \text{on } \underline{V}_J.\end{aligned} \tag{3.39}$$

We now define Q as follows. Q' is a function on pairs of indices taken from $p+2, \ldots, r$. We define Q to be Q' for these indices and we also define

$$Q_{p+1,l} = -T'_l.$$

It is clear that

$$\vec{R} = \delta^{(J)} Q \quad \text{on } \underline{V}_J \tag{3.40}$$

which is the desired result (the homogeneity properties of Q being clear).

This completes the proof of Proposition 3.3 and hence of Theorem 3.2.

Remark 1 *Our proof of Proposition 3.3 does not use the homogeneity of the* i_j.

Remark 2 *The complex made up of* $\vec{R}$ *and* δ *is the Koszul complex associated to the* i_j. *The equivalence of vanishing of the cohomology of the Koszul complex, the strong independence of* $\{i_j\}$, *and the well posed nature of the DP seem to be well understood (see [109, 110]), but our proof using multiplicity varieties is novel.*

Remark 3 *In Section 9.1 we shall prove a result which is equivalent to the vanishing of the first cohomology group when codim* $V > 1$.

Problem 3.3 *Verify or disprove.*

Conjecture *If codim* $V > l$ *then the first* l *cohomology groups vanish.*

Some special cases of the conjecture have been verified by Karen Taylor when $l = 2, 3$.

We have noted that the varieties V_Y are generally complex whereas Chevalley's original theory applies to groups acting on real space. Actually the theory of reflection groups was extended by Steinberg [146] to cover the complex case.

Consider the simplest example: $k = 1$ and W is the group generated by

$$t \to \zeta t \quad \text{where} \quad \zeta = e^{2\pi i/l}.$$

The invariants are generated by t^l. The harmonics are polynomials of degree less than l. The solution of the DP of Chevalley is given by Lagrange's interpolation formula for sets of the form

$$t, \zeta t, \zeta^2 t, \ldots, \zeta^{l-1} t.$$

On the other hand, Theorem 3.2 for $r = 1$ and $i = t^l$ is equivalent to the general Lagrange interpolation formula for l points.

For another classical example, let $r = 1$ and $i = \sum x_i^2$. The sets $i = c$ are the orbits of the rotation group. Harmonic in our sense is ordinary harmonicity. The restriction of an homogeneous harmonic to the unit sphere is called a spherical harmonic. Thus Theorem 3.2 asserts that every function f on $i = 1$ which extends to a polynomial on R^k coincides on $i = 1$ with a finite linear combination of spherical harmonics.

Our theory applies to much more general classes of f. This will be discussed in the next section.

Proposition 3.4 *If the DP is well posed for X_c for all constants c, then the ring of polynomials* $\mathbf{P}$ *is a tensor product*

$$\mathbf{P} = H \otimes \tilde{I}. \tag{3.41}$$

Conversely, if $\mathbf{P}$ *is such tensor product then the DP is well posed on any X_c which is nonempty.*

Recall that $\tilde{I}$ is the ring I with the constants adjoined.

Proof Suppose the DP is well posed. In "Proof of (1)" in Theorem 3.2 we showed by means of Fischer's theorem that for any $\vec{\imath}$ we can write any polynomial P in the form $P = \sum h_j a_j$ where h_j are harmonics and $a_j \in \tilde{I}$. Thus we have to verify that there is no nontrivial relation of the form

$$\sum h_j a_j = 0$$

where h_j are linearly independent harmonics and $a_j \in \tilde{I}$. We are now dealing with the case $\vec{Y}$ is a constant $\vec{c}$, so the a_j are constants on X_c, say $a_j = b_j$. If the a_j are not identically zero then we can find $\vec{c}$ so that not all $b_j = 0$. Thus

$$\sum h_j b_j = 0 \quad \text{on } X_{\vec{c}}$$

contradicting the uniqueness of the DP.

We claim that, conversely, the tensor decomposition $\mathbf{P} = H \otimes \tilde{I}$ implies the uniqueness of DP. Again it was shown in the "Proof of (1)" in Theorem 3.2 that for any $\vec{\imath}$ there is existence for DP. Thus our claim means that tensor decomposition implies well-posedness.

Suppose we had a relation $\sum b_j h_j = 0$ on $X_{\vec{c}}$; that is, since $X_{\vec{c}}$ is defined as $\vec{\imath} = \vec{c}$,

$$\sum b_j h_j = \sum (i_l - c_l) v_l \tag{3.42}$$

for some polynomials v_l and some constants b_j not all zero. Since $\mathbf{P} = H \otimes \tilde{I}$ we can write

$$v_l = \sum u_l^j h_j$$

with $u_l^j \in \tilde{I}$; this yields

$$\sum b_j h_j = \sum_j h_j \sum_l (i_l - c_l) u_l^j. \tag{3.43}$$

By the uniqueness of the tensor product

$$b_j = \sum (i_l - c_l) u_l^j$$

which means that all b_j vanish on $X_{\vec{c}}$. Since the b_j are constant and $X_{\vec{c}}$ is not empty this means that all b_j vanish.

Our assertion is thereby established.

Our approach to the tensor product goes via the DP. One might wonder if there is a direct approach using the vanishing of the cohomology of the Koszul complex. Indeed, such is the case. For the vanishing cohomology shows that the relations amongst the generators $i_1, \ldots, i_r$ are generated by the trivial relations. In particular, the relations are generated by elements in I.

For the next step, one uses the same ideas to show that the relations amongst the quadratics $i_j i_k$ are again generated by I. Proceeding in this manner we derive an analogous result for monomials of any degree and then of polynomials.

We conclude that all relations among elements of $\tilde{I}$ are generated by (vectors of) elements of $\tilde{I}$. This is the same as saying that $\mathbf{P} = H \otimes \tilde{I}$; that is, there are no nontrivial relations that require the use of H.

This method of proof can be applied to certain situations in which the generators are not strongly free. We can determine all relations in I from the relations amongst the generators. However, the procedure is very complicated so we shall restrict ourselves to the relevant examples as they appear.

An interesting situation occurs when the multiplicity variety $\underline{\mathrm{X}}_0$ defined by $\vec{i} = 0$ is an actual variety X_0. Since the i_j are homogeneous X_0 is homogeneous. By the fundamental principle (see Section 1.4) (or else by Fischer's lemma) the harmonics are the Fourier transforms of distributions[2] of the form $\hat{h}\delta$ for those differential operators $\hat{h}$ which annihilate the ideal generated by $\vec{i}$. In the case X_0 is nice, i.e. a cone over a nonsingular variety (or a variety with singularities which are not bad), then such $\hat{h}$ are linear combinations of powers of the directional derivatives in the directions along the generators of the cone X_0.

For example, in the case of the rotation group, X_0 is defined by $\sum t_j^2 = 0$. Let a be any complex vector on X_0, i.e. $\sum a_j^2 = 0$. Then $\sum a_j \partial/\partial t_j$ is a directional derivative along the generator of X_0 through a. Thus the harmonics are linear combinations of $(\sum a_j t_j)^m$.

In the particular case $k = 2$ this says that harmonic polynomials of degree m are linear combinations of $(t_1 \pm it_2)^m$.

We can state these results as (see the remark following Theorem 7.2 for a detailed analysis of this structure)

Theorem 3.5 *Suppose $\vec{i} = 0$ defines a multiplicity variety $\underline{X}_0$ for which the underlying irreducible varieties are cones over smooth varieties and the multiplicities can be described by constant coefficient differential operators; then the harmonics are linear combinations of polynomials of the form $Q(t)(\sum a_j t_j)^m$*

[2]We shall denote the Fourier transform of the operator "multiplication by the polynomial $P(x)$" by $\partial(P)$, or $P(D)$ or $\hat{P}$.

where $\sum a_j \partial/\partial t$ is a differentiation along a generator of one of the irreducible components of X_0, and Q is a polynomial corresponding to the multiplicity.

If there are singularities then the description of harmonics is somewhat more complicated.

The above described results for the rotation group are well known. Some generalizations are discussed in Section 3.3.

Suppose for simplicity that there are no multiplicities. We can think of X_0 in Theorem 3.5 as being decomposed into lines $\{\mathbf{g}\}$ through the origin. As mentioned in Section 1.1 a decomposition of the underlying variety yields a semi-coarse grain decomposition of function spaces. Theorem 3.5 gives a decomposition of the space H of harmonic polynomials for $\vec{\imath}$ (spread) into $\{H(\mathbf{g})\}$. We can describe $H(\mathbf{g})$ as the space of $\vec{\imath}$ harmonic polynomials which are constant in the directions orthogonal to $\mathbf{g}$; it can also be described as the space of spread functions for $\mathbf{g}$ which are harmonic.

In this way we can think of Theorem 3.5 as giving a Radon-like slice decomposition of H.

Thus far all our considerations have involved polynomials. In order to treat the space $\mathcal{E}$ of C^∞ functions we make the assumption that $\vec{P}(D)$ is *Cauchy hyperbolic*.[3] This means that the cone X_0 (which we assume has no multiplicity) is a branched s sheeted covering of a linear variety L which is the complexification of its real part. We write $t = (x, y)$ where x is a coordinate on L and y is the orthogonal coordinate. Cauchy hyperbolic means that for x real the points $(x, y_j(x)) \in X_0$ are real. Moreover functions $f(t) \in \hat{\mathcal{E}}(X_0)$ can be written in the form

$$f(t) = \sum_{l=1}^{s} q_l(y) f_l(x) \quad \text{on } X_0$$

where q_l are fixed polynomials and $f_l \in \hat{\mathcal{E}}'(L)$ depend on f. The values $f_l(x)$ can be expressed by Lagrange-like interpolation formulas as described in the beginning of Section 1.4. (This is discussed in detail in Chapter IX of FA.)

This relation $f \longleftrightarrow \{f_l\}$ is easily seen to be a topological isomorphism of $\hat{\mathcal{E}}'(X_0)$ and $\hat{\mathcal{E}}'(L)^s$ which is the direct sum of $\hat{\mathcal{E}}'(L)$ with itself s times. Note that each generator Λ of the cone X_0 lies over a line through the origin $\lambda \subset L$; in fact each line $\lambda \subset L$ is the projection of s generators Λ_l of X_0. (Recall that a generator of a cone is a line through the origin lying in this cone.)

It is shown in Theorem 3.21ff. below that $\mathbb{C}^{\Re}(L)$, which is the union of the complexifications of real lines[4] in L, is sufficient for $\hat{\mathcal{E}}'(L)$. (see Section 1.4 for the definition of "sufficient".) This means that every $h \in \mathcal{E}(\hat{L})$ has a Fourier

[3]We shall usually use the notation P instead of i in what follows as P is generally used when we are concerned with differential equations while i is used for invariants.

[4]Complex lines which are complexifications of real lines, i.e. have real direction numbers, are called *real complex lines*.

representation

$$\hat{h}(\hat{x}) = \int e^{ix\cdot\hat{x}} \frac{d\mu(x)}{k(x)}$$

where μ is a bounded measure and k belongs to an AU structure for $\mathcal{E}(\hat{L})$ with

$$\text{support } \mu \subset \mathbb{C}^{\Re}(L).$$

($\hat{L}$ is the same as L but in the dual space.)

We take a polar decomposition of μ, meaning

$$\mu = \int \mu_\theta \, d\theta$$

where μ_θ is a bounded measure on the complexification of the real line θ and the μ_θ are uniformly bounded measures. The Fourier transform of $d\mu_\theta/k$ is clearly a spread function for the spread of functions which are constant in the directions orthogonal to θ.

Remark. *One might expect that there is a problem taking the polar decomposition of a measure μ. However, in our case μ is a measure representing an analytic functional, meaning μ is in the dual of a space of holomorphic functions. Since, for holomorphic functions, uniform convergence is the same, locally, as distribution convergence, we can choose a representative measure μ to be in the dual of a distribution space which makes μ smooth so there is no difficulty in defining the polar decomposition.*

The isomorphism $f \longleftrightarrow \{f_l\}$ of $\hat{\mathcal{E}}'(X_0)$, with $\hat{\mathcal{E}}'(L)^s$ shows that $\mathbb{C}^{\Re}(X_0)$, which is the union of the complexifications of real generators of X_0, is sufficient for $\hat{\mathcal{E}}'(X_0)$. By our above remarks this leads to

Theorem 3.6 *Suppose $\vec{P}(D)$ is Cauchy hyperbolic. Then any solution $f \in \mathcal{E}$ of $\vec{P}(D)f = 0$ admits the Radon decomposition*

$$f(t) = \int f_\gamma(t) \, d\mu(\gamma) \tag{3.44}$$

where f_γ is a spread solution in $\mathcal{E}_{\vec{P}}$ corresponding to the real generator γ of X_0.

In general we cannot express f_γ in terms of f because the f_γ are not uniquely defined by f (direct integral with amalgamation). Explicit forms of this decomposition in special circumstances are discussed below. In Chapter 9 there is a discussion of the lack of uniqueness in the representation (3.44).

Replacing $\mathcal{E}$ by $\mathbf{P}$, the space of polynomials, means replacing $\hat{\mathcal{E}}'$ by $\hat{\mathbf{P}}'$ which is the space of formal power series (Chapter V of FA). Of course $\hat{\mathbf{P}}'$ is not a space of entire functions so $\mathbf{P}$ is certainly not AU. In fact, $\hat{\mathbf{P}}'$ is simpler to deal with than $\hat{\mathcal{E}}'$ so that, under the hypotheses of Theorem 3.5, we can show that (3.41)

is valid for polynomials $f \in \mathbf{P}_{\vec{P}}$, the kernel of $\vec{P}(D)$ in $\mathbf{P}$, with $f_\gamma \in \mathbf{P}_{\vec{P}}$. This means that it suffices to use real directions in Theorem 3.4.

Combining these ideas with Theorem 3.2 yields a *Radon transform on* $\mathcal{O}$:

Theorem 3.7 *Suppose* $\vec{P}(D)$ *is Cauchy hyperbolic. Then for any orbit* $\mathcal{O}$ *and any polynomial* f *on* $\mathcal{O}$ *we can write*

$$f(t) = \int f_\gamma(t)\, d\mu(\gamma) \tag{3.45}$$

where f_γ *is the restriction to* $\mathcal{O}$ *of a spread polynomial in* $\mathbf{P}_{\vec{P}}$ *corresponding to a real generator.*

Now suppose that the orbit $\mathcal{O}$ is nonsingular and irreducible and is *essentially real*, meaning the real part $\mathcal{O}^{\mathbb{R}}$ is a nonsingular real algebraic variety whose real dimension equals the complex dimension of $\mathcal{O}$. Since the real part $\mathcal{O}^{\mathbb{R}}$ of $\mathcal{O}$ is a nonsingular variety we can define the space $\mathcal{D}(\mathcal{O}^{\mathbb{R}})$ of C^∞ functions on $\mathcal{O}^{\mathbb{R}}$ of compact support. Any convenient positive smooth measure $d\nu$ on $\mathcal{O}^{\mathbb{R}}$ gives a means of identifying functions on $\mathcal{O}^{\mathbb{R}}$ with distributions. In particular the polynomials are dense in $\mathcal{D}'(\mathcal{O}^{\mathbb{R}})$.

Let γ be a real generator of X_0. On the intersection of each translate $\hat{\gamma}_s^\perp$ of $\hat{\gamma}^\perp$ with $\mathcal{O}^{\mathbb{R}}$ we have a measure, namely the measure defined by ν. If f_γ is a real spread polynomial corresponding to γ then for any $u \in \mathcal{D}(\mathcal{O}^{\mathbb{R}})$

$$\int f_\gamma u\, d\nu = \int ds \int_{\hat{\gamma}_s^\perp \cap \mathcal{O}^{\mathbb{R}}} f_\gamma u \tag{3.46}$$

for a suitable measure ds (which may depend on γ). Since f_γ is constant on $\hat{\gamma}_s^\perp$ for each s, the inner integral can be thought of as $f_\gamma(s)(\mathbf{R}u)(s,\gamma)$ where the Radon transform is defined by

$$\mathbf{R}u(s,\gamma) = \int_{\hat{\gamma}_s^\perp \cap \mathcal{O}^{\mathbb{R}}} u. \tag{3.47}$$

Theorem 3.7 and the density of the polynomials yield

Theorem 3.7* *If the base of* X_0 *and* $\mathcal{O}$ *are essentially real and smooth then the Radon transform on* $\mathcal{D}(\mathcal{O}^{\mathbb{R}})$ *is injective.*

Problem 3.4 *Describe the range of the Radon transform on* $\mathcal{O}^{\mathbb{R}}$.

We have noted that Theorem 3.7 can be considered as the decomposition of functions on an orbit into spread functions. In general it is difficult to give an explicit formula for the decomposition. But when we are in a group situation then we can (sometimes) give explicit formulas.

Suppose W is an orthogonal group, meaning that it leaves invariant a nondegenerate definite quadratic form. It is this quadratic form that we use for

Fourier transformation and to make explicit the relation between polynomials and constant coefficient differential operators. We assume that the ideal generated by $\vec{P}$ is W invariant and that W preserves a measure $d\theta$ on each orbit.

We denote by $\hat{W}$ the action of W on the Fourier transform space. $\hat{W}$ acts on X_0. Moreover $\hat{W}$ commutes with scalar multiplication and hence acts on the set of generators of X_0; we assume that the action on generators is transitive. Also $\hat{W}$ acts on the space of differential operators which are linear combinations of powers of derivation along the generators. Again, since $\hat{W}$ is linear it preserves the degree of derivation.

Assumption *We can find a countable set of rays* $\{\hat{\lambda}_k\} \subset X_0$ *such that:*

(a) *For each* l *we can choose a minimal set* $\{\hat{\lambda}_{lj}\} \subset \{\hat{\lambda}_k\}$ *such that the* l*-th powers of derivation along the* $\hat{\lambda}_{lj}$ *span the* l*-th powers of derivation along all rays on* X_0.
(b) *If* $l' > l$ *then* $\hat{\lambda}_{lj} = \hat{\lambda}_{l'j'}$ *for some* j'.
(c) *Every* $\hat{\lambda}_k$ *is of the form* $\hat{\lambda}_{lj}$ *for some* l, j.

We denote by $\partial/\partial\hat{\lambda}_k$ the directional derivation along $\hat{\lambda}_k$ and call h_{lj} the Fourier transform of the l-th order derivative of δ_0 along $\hat{\lambda}_{lj}$. We can write

$$h_{lj}(x) = \lambda_{lj}^l(x) = (x \cdot \hat{\lambda}_{lj}^0)^l \tag{3.48}$$

where λ_k is the Fourier transform of $(\partial/\partial\hat{\lambda}_k)\delta_0$ and $\hat{\lambda}_k^0$ is a unit vector along the ray $\hat{\lambda}_k$.

Clearly the h_{lj} are harmonic for $\vec{P}$; they are homogeneous of degree l.

We set

$$\mathbf{P}_l = \text{ linear span of } \{h_{lj}\}_j. \tag{3.49}$$

Thus $\mathbf{P}_l$ are representation spaces for W.

The linear span of $\{h_{lj}\}_{l \text{ fixed}}$ might be reducible as a representation space for W, in which case $\mathbf{P}_l$ decomposes into a direct sum of irreducible representation spaces:

$$\mathbf{P}_l = \sum_{\oplus} \mathbf{P}_l^m.$$

By (3.48) the h_{lj} are spread functions for $\hat{\lambda}_{lj}^{0\perp}$, meaning they are constant in directions orthogonal to $\hat{\lambda}_{lj}^0$.

It may happen that $\hat{\lambda}_{lj}^0 = \hat{\lambda}_{l'j'}^0$. We collect all such $l'j'$ and define the (analytic) spread related to $\hat{\lambda}_{lj}^0$ as the linear span of $\{h_{l'j'}\}$ for such $l'j'$. It seems that the "basis" $\{h_{l'j'}\}$ for the spread of $\vec{P}$ harmonic functions which are constant on $\lambda_{lj}^{0\perp}$ (orthogonal to $\hat{\lambda}_{lj}^0$) has more group theoretical significance than the usual basis which consists of δ functions on the affine hyperplanes orthogonal to $\hat{\lambda}_{lj}^0$.

Let $\mathcal{O}$ be a smooth, essentially real orbit. We write

$$\mathbf{R}f(l,j) = \int_{\mathcal{O}^{\mathbb{R}}} f h_{lj}$$

for $f \in \mathcal{D}(\mathcal{O}^{\mathbb{R}})$, a smooth function on $\mathcal{O}^{\mathbb{R}}$. Thus by our general construction

$$\mathbf{R}^*\mathbf{R}f = \sum [\mathbf{R}f(l,j)] h_{lj}.$$

The problem of explicit inversion of the Radon transform reduces to that of explicit orthonormalization of the h_{lj}.

The functions h_{lj} are related to various types of hypergeometric functions (see e.g. [58]) but these relations are beyond the scope of the present work.

Sometimes we can use the representation theory of W to express δ_{x_0}, for $x_0 \in \mathcal{O}^{\mathbb{R}}$, as an infinite series

$$\delta_{x_0} = \sum a_{lj} h_{lj} \tag{3.50}$$

on $\mathcal{O}^{\mathbb{R}}$. Generally the coefficients a_{lj} grow polynomially in l so the series converges in suitable distribution spaces.

For example, if $n = 2$ and $\vec{P} = P \equiv x_1^2 + x_2^2$ and W is the rotation group then (3.50) becomes

$$\delta_{z=1} = \sum_{m \geq 0} z^m + \sum_{m>0} \bar{z}^m. \tag{3.51}$$

This equation holds on $\mathcal{O} = \{|z| = 1\}$; $\delta_{z=1}$ means the δ function of $\theta = 0$ on this circle.

Problems for further study

A. All the above applies to generic orbits or to corresponding multiplicity varieties on degenerate orbits. Is there an analog that applies to degenerate orbits themselves?

Consider the example $W = S_3$, the symmetric group on three letters (which is a reflection group) acting on R^3. The orbits

$$(a,a,b),\ (a,b,a),\ (b,a,a) \tag{3.52}$$

are generally *not* complete intersections. For, being three points, if it were a complete intersection the degrees of the polynomials would be 1, 1, 3. But (3.52) does not generally lie in a line so this is impossible.

We want a harmonic theory that produces a basis for functions on such orbits. As the usual harmonics for S_3 form a six-dimensional space we need a mechanism to "split" the space of harmonics. For this we pass from *invariants* to *covariants*.

The general theory we envisage is the following. We start with an abstract representation ρ of the finite group W which has two realizations ρ_1 and ρ_2 in

the ring of polynomials. We choose bases for ρ_1 and ρ_2 so that the matrices for ρ_1 and ρ_2 are identical. If $\vec{\imath}^1, \vec{\imath}^2$ denote these bases then we are interested in varieties of the form

$$\vec{\imath}^2 = \alpha \vec{\imath}^1 \tag{3.53}$$

where α is a constant, with differential equations corresponding to $\alpha = 0$. A somewhat more general form of this occurs when ρ_1 and ρ_2 are direct sums of representations and α (which is now called $\vec{\alpha}$) is an operator which is a scalar on each of the subrepresentation spaces. By varying α we obtain varying W invariant sets.

To make things precise, we assume that the degree of the polynomials in $\vec{\imath}^2$ corresponding to each representation in ρ_2 is greater than the corresponding degree in ρ_1. We can form an analog of harmonicity by the equation

$$\partial(\vec{\imath}^2)h = 0.$$

The Chevalley theory deals with the case in which ρ is the direct sum of r copies of the trivial representation. Thus both ρ_1 and ρ_2 are the sum of the trivial representation with itself r times. ρ_1 is the representation on the constant repeated r times and ρ_2 is the representation on the generators of invariant polynomials with no constant term. The bases are

$$\begin{aligned} \vec{\imath}_0{}^1 &= (1, \ldots, 1) \\ \vec{\imath}_0{}^2 &= (i_1, \ldots, i_r) \end{aligned} \tag{3.54}$$

where $i_1, \ldots, i_r$ form a basis for the ring of invariants with no constant term and instead of α being a constant in (3.53), $\alpha = (\alpha_1, \ldots, \alpha_r)$ where the α_j are constants. Equation (3.53) thus defines the generic orbits of W and the differential equations define the usual (Chevalley) harmonicity.

Let us return to our example of S_3. By the well-posedness of the DP the representation of S_3 on the harmonic polynomials is the same as the representation on functions on a generic orbit, which is the regular representation. The decomposition of the representation on harmonic polynomials into irreducible components has the bases

$$\begin{aligned} 1 \text{ (trivial rep.)} \\ \rho_1 : t_1 - t_2,\, t_2 - t_3 \text{ (dim. 2)} \\ \rho_2 : t_1^2 - t_2^2 - 2t_3(t_1 - t_2), t_2^2 - t_3^2 - 2t_1(t_2 - t_3) \text{ (dim. 2)} \\ \rho_3 : (t_1 - t_2)(t_2 - t_3)(t_1 - t_3) \text{ (dim. 1)} \end{aligned} \tag{3.55}$$

ρ_2 is the two-dimensional representation on harmonic polynomials of degree 2 and ρ_1 is the two-dimensional representation on harmonic polynomials of

degree 1. We denote the respective bases of ρ_1, ρ_2 by $\vec{\imath}_1{}^1$, $\vec{\imath}_1{}^2$. The two representations of dimension 2 are equivalent; we have adjusted the bases so that the matrices for $w \in S_3$ are the same. Thus we can apply the idea discussed in (3.53) and the following.

Note that

$$t_1^2 - t_2^2 - 2t_3(t_1 - t_2) = (t_1 - t_2)(t_1 + t_2 - 2t_3)$$
$$t_2^2 - t_3^2 - 2t_1(t_2 - t_3) = (t_2 - t_3)(t_2 + t_3 - 2t_1).$$

For (3.53) to be satisfied we must have

$$t_1 + t_2 - 2t_3 = \alpha \qquad \text{if } t_1 \neq t_2$$
$$t_2 + t_3 - 2t_1 = \alpha \qquad \text{if } t_2 \neq t_3.$$

In particular not all t_1, t_2, t_3 can be distinct, for if they were then $t_1 + t_2 - 2t_3 = t_2 + t_3 - 2t_1$, which implies $t_1 = t_3$.

In case of the orbit (3.52), $\alpha = b - a$. But this does not fix a, b. To fix them we need the ordinary invariants.

Now, the orbit (3.52), when defined by the equations involving ordinary invariants, i.e. according to (3.54),

$$\vec{\imath}_0^{\,2} = (t_1 + t_2 + t_3, t_1t_2 + t_2t_3 + t_1t_3, t_1t_2t_3)$$

with $\vec{\alpha} = (2a + b, a^2 + 2ab, a^2b)$, has multiplicity, i.e. is a multiplicity variety. A little "experimentation" shows that the multiplicity is defined by the differential operator

$$\delta = (t_2 - t_3)\frac{\partial}{\partial t_1} + (t_3 - t_1)\frac{\partial}{\partial t_2} + (t_1 - t_2)\frac{\partial}{\partial t_3}.$$

This means that the polynomials Q and δQ vanish on (3.52) if and only if Q belongs to the ideal generated by $\vec{\imath}_0{}^2 - \alpha\vec{\imath}_0{}^1$. For this $\vec{\alpha}$ and for $\alpha = b - a$ in (3.53) we obtain the orbit (3.52) with no multiplicity.

It is readily verified that the usual harmonic functions (3.55) for S_3 which also satisfy

$$\partial(\vec{\imath}_1{}^2)h = 0 \tag{3.56}$$

are spanned by

$$1,\ t_1 - t_2,\ t_2 - t_3. \tag{3.57}$$

These form a basis for functions on the orbit (3.52). Thus we have succeeded in constructing a basis for the variety (3.52) with no multiplicity by adding the covariant equation (3.56) to the usual (invariant) equations for harmonicity, i.e. we have "split" the space of harmonics.

We hope that the reader will expand these ideas.

A different approach is given in Chapter 4.

B. Up to now we have been dealing with equations for scalar functions. We want to use matrix equations to split harmonicity. The best known example is the Cauchy–Riemann system

$$\begin{pmatrix} \dfrac{\partial}{\partial x} & -\dfrac{\partial}{\partial y} \\ \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial x} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = 0 \tag{3.58}$$

which splits the Laplace equation $\Delta h = 0$ in two variables. We shall present this splitting within the framework of our ideas.

It is important to note that Δ is the determinant of the matrix in (3.58).

By Fourier transform the equation becomes (formally)

$$\begin{pmatrix} \hat{x} & -\hat{y} \\ \hat{y} & \hat{x} \end{pmatrix} \begin{pmatrix} \hat{u} \\ \hat{v} \end{pmatrix} = 0.$$

Since $\Delta u = \Delta v = 0$ the fundamental principle asserts that we can assume that the supports of $\hat{u}, \hat{v}$ are on $\hat{y} = \omega\hat{x}$ where $\omega = \pm i$.

Looked at from another point of view, the rank of the matrix is 2 unless $\hat{x}^2 + \hat{y}^2 = 0$, in which case it is 1 unless $\hat{x} = \hat{y} = 0$. We concentrate on the case of rank 1. $\hat{x}^2 + \hat{y}^2 = 0$ means $\hat{y} = \omega\hat{x}$ where $\omega = \pm i$. We can write the equation as

$$\hat{x} \begin{pmatrix} 1 & -\omega \\ \omega & 1 \end{pmatrix} \begin{pmatrix} \hat{u}(\hat{x}, \omega\hat{x}) \\ \hat{v}(\hat{x}, \omega\hat{x}) \end{pmatrix} = 0$$

where $\hat{x} \neq 0$. The solution of

$$\begin{pmatrix} 1 & -\omega \\ \omega & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 0$$

is

$$\begin{pmatrix} a \\ b \end{pmatrix} = c \begin{pmatrix} 1 \\ -\omega \end{pmatrix}.$$

Thus

$$\begin{pmatrix} \hat{u}(\hat{x}, \omega\hat{x}) \\ \hat{v}(\hat{x}, \omega\hat{x}) \end{pmatrix} = \hat{\alpha}_\omega(\hat{x}, \omega\hat{x}) \begin{pmatrix} 1 \\ -\omega \end{pmatrix}$$

for some suitable function (measure) $\hat{\alpha}_\omega$.

We find

$$\alpha_\omega(x, y) = \int e^{ix\hat{x} + i\omega y\hat{x}} \hat{\alpha}_\omega(\hat{x}, \omega\hat{x})\, d\hat{x} = \beta_\omega(x + \omega y) \tag{3.59}$$

so that

$$\begin{aligned} u + iv &= \beta_\omega(x + \omega y)(1 - i\omega) \\ &= \begin{cases} 2\beta_i(x + iy), & \text{if } \omega = i \\ 0, & \text{if } \omega = -i, \end{cases} \end{aligned}$$

while

$$u - iv = \begin{cases} 0, & \text{if } \omega = i \\ 2\beta_{-i}(x - iy), & \text{if } \omega = -i. \end{cases}$$

We can now understand how the Cauchy–Riemann system splits the Laplace equation. Solutions of the Laplace equation are Fourier transforms of measures on $\hat{y} = \pm i\hat{x}$. Hence u and v also have similar Fourier representations. Forming $u \pm iv$ restricts the support to one of the two lines. This is the splitting of $\Delta h = 0$ in two variables.

Remark. *This splitting is possible because*

$$(1 \quad \pm i) \cdot \begin{pmatrix} 1 \\ -\omega \end{pmatrix} = \begin{cases} 0, & \text{if } \pm i\omega = 1 \\ 2, & \text{if } \pm i\omega = -1. \end{cases}$$

Thus multiplying $\binom{u}{v}$ *by* $(1 \quad \pm i)$ *"projects"* $\binom{u}{v}$ *onto (twice) the contribution from the complex line* $\pm i\omega = -1$.

We want to examine to what extent this splitting can be carried out in higher dimensions. The two-dimensional splitting depended on the factorization of the Laplacian into linear factors. In higher dimensions there is a standard splitting of the Laplacian by use of spinors. Rather than deal with spinors we shall discuss the case of dimension $n = 4$, in which case the resulting split equation, the (elliptic) Dirac equation, can be thought of formally in terms of quaternions as the differential system

$$\left(\frac{\partial}{\partial x_1} + i\frac{\partial}{\partial x_2} + j\frac{\partial}{\partial x_3} + k\frac{\partial}{\partial x_4} \right) (f_1 + if_2 + jf_3 + kf_4) = 0. \tag{3.60}$$

Here i, j, k are the usual quaternionic units. In matrix form this is

$$\begin{pmatrix} \frac{\partial}{\partial x_1} & -\frac{\partial}{\partial x_2} & -\frac{\partial}{\partial x_3} & -\frac{\partial}{\partial x_4} \\ \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_1} & -\frac{\partial}{\partial x_4} & \frac{\partial}{\partial x_3} \\ \frac{\partial}{\partial x_3} & \frac{\partial}{\partial x_4} & \frac{\partial}{\partial x_1} & -\frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_4} & -\frac{\partial}{\partial x_3} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_1} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix} = 0. \tag{3.61}$$

The determinant of the matrix is Δ^2. By applying $(\partial/\partial x_1 - i(\partial/\partial x_2) - j(\partial/\partial x_3) - k(\partial/\partial x_4))$ to the quaternionic equation we see that all f_l are harmonic.

The Fourier transform of the matrix in (3.61) is (up to a factor of i)

$$\boxed{\hat{D}} = \begin{pmatrix} \hat{x}_1 & -\hat{x}_2 & -\hat{x}_3 & -\hat{x}_4 \\ \hat{x}_2 & \hat{x}_1 & -\hat{x}_4 & \hat{x}_3 \\ \hat{x}_3 & \hat{x}_4 & \hat{x}_1 & -\hat{x}_2 \\ \hat{x}_4 & -\hat{x}_3 & \hat{x}_2 & \hat{x}_1 \end{pmatrix}.$$

Note that the rows (columns) of $\boxed{\hat{D}}$ are orthogonal. Also $\det \boxed{\hat{D}} = (\hat{x}_1^2 + \hat{x}_2^2 + \hat{x}_3^2 + \hat{x}_4^2)^2$. Thus $\operatorname{rank} \boxed{\hat{D}} < 4$ when $\hat{x}_1^2 + \hat{x}_2^2 + \hat{x}_3^2 + \hat{x}_4^2 = 0$, in which case each row or column has length 0.

We can easily verify that when $\det \boxed{\hat{D}} = 0$ the rank of $\boxed{\hat{D}}$ is 2 or 0. As in the case of the Cauchy–Riemann equations the interesting case is that of rank 2.

Since all f_l are harmonic the fundamental principle allows us to think of $\hat{f}_l$ as being supported on $\sum \hat{x}_l^2 = 0$ which is a cone. We write the generators of the cone in the form $\hat{t}(\omega_1, \omega_2, \omega_3, \omega_4)$ where we normalize "almost all" rays by $\omega_1 = i$ so $\omega_2^2 + \omega_3^2 + \omega_4^2 = 1$. On this ray the functions $\hat{f}_l$ satisfy

$$\begin{pmatrix} \omega_1 & -\omega_2 & -\omega_3 & -\omega_4 \\ \omega_2 & \omega_1 & -\omega_4 & \omega_3 \\ \omega_3 & \omega_4 & \omega_1 & -\omega_2 \\ \omega_4 & -\omega_3 & \omega_2 & \omega_1 \end{pmatrix} \begin{pmatrix} \hat{f}_1 \\ \hat{f}_2 \\ \hat{f}_3 \\ \hat{f}_4 \end{pmatrix} = 0. \tag{3.62}$$

It is important to note that the rank of the matrix is 2. Thus on each ray there is a two-dimensional space of $\vec{\hat{f}}_\omega$ (two-dimensional over the space of scalar functions on the ray) which satisfy (3.62). Note that the rows (columns) of the matrix (3.62) are orthogonal. Thus we can choose basis vectors $\vec{\lambda}^1(\omega), \vec{\lambda}^2(\omega)$ to be, for example, the first two rows of the matrix (generically). For each $\vec{\hat{f}}_\omega$ we can find functions $\hat{v}_\omega^i$ so that

$$\vec{\hat{f}}_\omega = \vec{\lambda}^1(\omega)\hat{v}_\omega^1 + \vec{\lambda}^2(\omega)\hat{v}_\omega^2.$$

By Fourier transformation we conclude

$$\vec{f}(x) = \int \sum \vec{\lambda}^j(\omega) v_\omega^j(x \cdot \omega)\, d\omega. \tag{3.63}$$

Of course, this result is only formal. It is the analog of the Whittaker representation of harmonic functions discussed in Section 1.4. We shall show how to prove this result below.

On each ray we need two functions $\hat{v}_\omega^j$ for solutions of the Dirac equation whereas for the Whittaker representation of harmonic functions there is only one. This is a further insight into the fact that the Dirac equation has twice as many solutions as Laplace's equation—a fact which was already pointed out by Dirac [31].

Another phenomenon of the Dirac equation is that we can find vectors in the cone $\sum \hat{x}_j^2 = 0$ whose annihilator (orthogonal complement) in the cone is a 2 plane. Thus the annihilator of $(1, -i, 0, 0)$ is the 2 plane defined by

$$\hat{x}_1 = i\hat{x}_2 \qquad \hat{x}_4 = i\hat{x}_3. \tag{3.64}$$

(Actually there are two 2 planes which annihilate $(1, -i, 0, 0)$, namely $x_1 = ix_2$, $x_4 = \pm ix_3$.) The Fourier transform of (3.62) on this plane is

$$\begin{pmatrix} i\hat{x}_2 & -\hat{x}_2 & -\hat{x}_3 & -i\hat{x}_3 \\ \hat{x}_2 & i\hat{x}_2 & -i\hat{x}_3 & \hat{x}_3 \\ \hat{x}_3 & i\hat{x}_3 & i\hat{x}_2 & -\hat{x}_2 \\ i\hat{x}_3 & -\hat{x}_3 & \hat{x}_2 & i\hat{x}_2 \end{pmatrix} \begin{pmatrix} \hat{f}_1 \\ \hat{f}_2 \\ \hat{f}_3 \\ \hat{f}_2 \end{pmatrix} = 0.$$

A basis (over multiplication by scalar functions on the plane) for solutions is $(1, i, 0, 0)$ and $(0, 0, 1, i)$. Thus, by comparison with Bateman's representation of harmonic functions we again see that there are twice as many solutions of Dirac's equation as of Laplace's equation.

As in the case of the Cauchy–Riemann equations we can "split" the Dirac equations. In the case of (3.63) we can take the scalar product with $(\hat{x}_2, \hat{x}_1, \hat{x}_4, -\hat{x}_3)$; this eliminates the $\vec{\lambda}^1$ term. In the case of (3.64) we can take the inner product with $(0, 0, 1, -i)$ to eliminate the contribution from $(1, i, 0, 0)$. However, there do not appear to be any "natural" further splittings.

These examples deal with harmonic functions and Dirac operators or spinors. Let us give an example of a higher degree equation related to a system. We use the simplest example, namely

$$P(\hat{x}, \hat{y}) = \hat{x}^4 + \hat{y}^4.$$

There is a natural system that "splits" P, namely

$$\boxed{P} = \begin{pmatrix} \hat{x} & -\hat{y} & 0 & 0 \\ 0 & \hat{x} & -\hat{y} & 0 \\ 0 & 0 & \hat{x} & -\hat{y} \\ \hat{y} & 0 & 0 & \hat{x} \end{pmatrix}. \tag{3.65}$$

Since $\det \boxed{P} = P$, any solution of $\boxed{P}\,\vec{\hat{f}} = 0$ has support on $\{\hat{x} = \alpha_j \hat{y}\}$ where α_j are the fourth roots of -1.

The rank of the matrix $\boxed{P}$ on $P = 0$ is 3 except at $\hat{x} = \hat{y} = 0$. The point $\hat{x} = \hat{y} = 0$ is unimportant since its complement is sufficient. Thus for each α_j there is an "essentially" unique 4 vector $\vec{\lambda}_j$ such that, on $\hat{x} = \alpha_j \hat{y}$, we can write

$$\vec{\hat{f}} = \vec{\lambda}_j \hat{\mu}_j$$

where $\hat{\mu}_j$ is a function (measure). A simple calculation shows that we can choose

$$\vec{\lambda}_j = (\alpha_j, \alpha_j^2, \alpha_j^3, -1).$$

This leads to the representation

$$\vec{f}(x, y) = \sum \vec{\lambda}_j \mu_j(\alpha_j x + y) \tag{3.66}$$

for solutions of $\partial\left(\boxed{P}\right)\vec{f} = 0$.

If $\vec{\nu}_j$ is a vector (of constants) such that

$$\vec{\nu}_j \cdot \vec{\lambda}_{j'} = \begin{cases} 0 & \text{if } j' \neq j \\ 1 & \text{if } j' = 0, \end{cases}$$

then $\vec{\nu}_j \cdot f$ cuts down the solutions of $\boxed{P}\vec{f} = 0$ to the contribution from the single line corresponding to α_j. Thus the "splitting" is as in the Cauchy–Riemann equations.

The same idea works for $\hat{x}^m \pm \hat{y}^m$ for any m (even or odd).

It would be very interesting to put these ideas in a general setting, including a treatment in higher dimensions.

C. We mentioned in Chapter 1 that one of the cornerstones of analysis is the ability to multiply functions. How about multiplication of (scalar) harmonic functions?

Let h_1 and h_2 be homogeneous harmonics of degrees d_1, d_2. In general, $h_1 h_2$ is not harmonic. However, degree $(h_1 h_2) = d_1 + d_2$. If we apply Fischer's method as in the existence part of Theorem 3.2 or if we use Proposition 3.5 we see that

$$h_1 h_2 = h^{d_1+d_2} + \sum j_l h^{d_1+d_2-\deg j_l} \tag{3.67}$$

where j_l belong to the ring generated by the i_k.

In particular if $r = 1$ and $i = \sum t_p^2$ then we derive the classical Clebsch–Gordon result that the symmetric Kronecker product of two irreducible representations of the rotation group is a sum of irreducible components with degrees starting with the sum of the degrees of the factors and diminishing degrees in steps of 2. However, we do not know how to prove that the sum stops at $|d_1 - d_2|$. This lower bound is crucial for a true understanding of the Kronecker product.

Problem 3.5 *Use harmonic theory to deduce this lower bound. Is there a lower bound in the nongroup case, e.g. for* $x_1^4 + x_2^4$*?*

For the rotation group much more precise information is available. This involves choosing a special basis for the harmonics and hence for the spherical harmonics which are their restrictions to the unit sphere. In this basis the spherical harmonics are hypergeometric functions. Moreover the coefficients in the expression (3.67) can be computed in this basis. They are the $3 - J$ coefficients of Raccah (see [17]); they can also be expressed in terms of hypergeometric series.

It would be of great interest to understand the algebra behind this special basis and the algebraic significance of the corresponding $3 - J$ coefficients.

One important application of the Kronecker product is that it can often be used to replace group integration (when there is a group). In this way it gives an algebraic setting for group integration. Thus one computes the integral $\int_{\vec{i}=\vec{c}} f_1 f_2$

by expanding f_1 and f_2 in harmonics and then computing the constant term in the product.

An application of this idea to the famous Rogers–Ramanujan identities is given in [57].

The Kronecker product can sometimes be used to obtain an explicit basis for the harmonics.

To understand how things work, let us start the investigation by studying a simple situation. Let $i^0(x_1, x_2)$ be a homogeneous polynomial of degree $d > 1$ in two variables with no multiple factors. Thus, by Proposition 3.4, a basis for harmonics is $\{(x_1 + \alpha_j x_2)^n\}$ where α_j are the roots of $i^0(1, \alpha) = 0$.

Next we add another variable t and consider

$$i(t, x_1, x_2) = t^d - i^0(x_1, x_2). \tag{3.68}$$

The variety $V : i = 0$ is a cone whose base B is $\{i^0 = 1\}$ which we identify with $\{i = 0\} \cap \{t = 1\}$. For each $b \in B$ there are d points in V whose (x_1, x_2) coordinate is b, namely those whose t coordinate is $\zeta^j, j = 0, 1, \ldots, d-1$, where ζ is a primitive d-th root of unity. Using Proposition 3.5, a spanning set for harmonics (for i) is given by

$$\{(t + b_1 x_1 + b_2 x_2)^n\}_{b \in B, n = 0, 1, \ldots}. \tag{3.69}$$

Actually we should consider similar expressions in which t is replaced by $\zeta^j t$. But this gives nothing new since we can factor out the constant ζ^j because B is invariant under scalar multiplication by ζ^j.

For an intuitive understanding of this construction think of the simple case when $i^0(x_1, x_2) = x_1^2 + x_2^2$. Then V is the light cone and the monomials (3.69) can be thought of as spread functions corresponding to the generators $\{\lambda(1, b_1, b_2)\}$ of the light cone; that is, for fixed $(1, b_1, b_2)$ they are constant on the leaves of the spread defined by $(1, b_1, b_2)$ which are the orthogonal planes.

Expression (3.69) provides us with a spanning set for the i harmonics; this spanning set is parametrized by $\{\delta_b\}_{b \in B}$. In view of our theory of harmonic functions we should be able to replace the δ functions on B by the restrictions of i^0 harmonics to B as the latter provides a different basis for "functions" on B.

This suggests that we consider

$$(t + b_1 x_1 + b_2 x_2)^n \tag{3.69*}$$

as a function of (t, x_1, x_2) and of $(b_1, b_2) \in B$. We expand it in terms of $i^0(b)$ harmonics and restrict the harmonics to $i^0(b) = 1$, i.e. to $b \in B$.

We obtain

$$(t + b_1 x_1 + b_2 x_2)^n = \sum g_\beta^n(t, x_1, x_2) h_\beta^0(b).$$

Here h_β^0 are harmonics for i^0 and g_β^n are "interesting" functions. The g_β^n are harmonic for i because the $h_\beta^0(b)$ are linearly independent over polynomials in

(t, x_1, x_2). They can generally be used to produce an explicit basis for the i harmonics. For the 3 sphere (in which case $i^0(x_1, x_2) = x_1^2 + x_2^2$ as before but the minus sign in (3.68) is replaced by plus) this is equivalent to the classical way of writing spherical harmonics.

We can iterate the process and pass to higher dimensional varieties.

Let us return to the orbital integral (3.13). In the (compact) group case the orbital integral is essentially the integral on $\mathcal{O}^{\mathbb{R}}$ (which is the group W or a quotient space) using Haar measure. $\mathcal{O}^{\mathbb{R}}$ is not all of $\mathcal{O}$; rather it is the real part of $\mathcal{O}$. The fact that this integral is the constant term in the group case is just Schur's orthogonality relation if we identify representation functions of the group W on $\mathcal{O}^{\mathbb{R}}$ with the restrictions of harmonic functions. This identification can be made because by Theorem 3.2 the restriction of harmonics to $\mathcal{O}^{\mathbb{R}}$ consists of all "nice" functions on $\mathcal{O}^{\mathbb{R}}$ and W acts on "nice" functions on $\mathcal{O}^{\mathbb{R}}$ and hence on the harmonics, which are uniquely determined by their restriction to $\mathcal{O}^{\mathbb{R}}$. Thus the representation on harmonics is the regular representation when $W = \mathcal{O}^{\mathbb{R}}$. Since W is compact (actually more general W are possible) any representation can be decomposed into irreducible representations. Hence the space of harmonics has a basis consisting of representation functions so Schur orthogonality is valid.

The delta functions $\{\delta_x\}_{x \in \mathcal{O}}$ and the harmonics represent two bases for functions on $\mathcal{O}^{\mathbb{R}}$. If f is a polynomial and we express

$$f = \sum \alpha_n h^n \quad \text{on } \mathcal{O}^{\mathbb{R}}$$

where $\{h^n\}$ are homogeneous harmonics forming a basis for all harmonic polynomials, then we can evaluate

$$\delta_x \cdot f = f(x) = \sum \alpha_n h^n(x).$$

Thus the passage from $\{h^n\}$ to $\{\delta_x\}$ is simple.

When the orbital integral (3.13) has an interpretation in terms of Riemann-like sums, then it tells us how to compute α_0 in terms of $\{\delta_x \cdot f\}$. (h^0 is the constant function.) If there were a good Kronecker product theory then, for example, we could multiply f by h^n and study the harmonic expansion of fh^n. The constant terms of the various fh^n should, eventually, give a method for computing all α_j.

We conclude:

When the orbital integral (3.13) has a Riemann-like definition and there is a "good" Kronecker product theory, then we have an explicit way of computing $\{\alpha_n\}$ *in terms of* $\{\delta_x \cdot f\}$.

The above theory cannot be as well behaved in general as it is for the group case. Thus if things worked perfectly then, at least for elliptic i, meaning the real part $\mathcal{O}^R$ of $\mathcal{O}$ is compact, we expect that harmonic expansions could be defined for a more or less arbitrary function f on $\mathcal{O}^{\mathbb{R}}$. Suppose that we can expand f in terms of $\{h^n\}$ as before. (Existence of expansion for nonpolynomial f is

generally difficult to establish, even if f has an extension from $\mathcal{O}^{\mathbb{R}}$ to a function which is holomorphic (see Section 3.2).) If the orbital integral and Kronecker products behave well then the expansion coefficients α_n could be computed in terms of the values $f(x), x \in \mathcal{O}^{\mathbb{R}}$. But, say for $i = x_1^4 + x_2^4$, solutions of $\partial(i)h = 0$ inside $x_1^4 + x_2^4 = 1$ are *not* determined by their values on $\mathcal{O}^{\mathbb{R}}$. (For, the usual DP involves the values of f and its normal derivative.) Taking f to be a solution of $\partial(i)f = 0$ which vanishes on $i = 1$ would then lead to a contradiction since the above indicates that all expansion coefficients in terms of harmonics of $f|_{\mathcal{O}}$ vanish. Thus we should expect $f \equiv 0$ which is not necessarily the case.

It might be possible that a Riemann-like definition of the orbital integral exists, but its support would have to be on some "enlargement" of $\mathcal{O}$ rather than $\mathcal{O}^R$ or else it might apply only to a class of solutions of $\partial(i)f = 0$ which are determined by their values on $\mathcal{O}^R$ (like polynomials by Theorem 3.2). It would be of great interest to clarify this situation for at least one nontrivial example.

3.2 Analytic theory

In accordance with the above discussion there is a certain way in which Theorem 3.2 seems to contradict the classical analytic theory of the DP. For example, for the operator

$$P(D) = \frac{\partial^4}{\partial x^4} + \frac{\partial^4}{\partial y^4} \tag{3.70}$$

there is a standard (analytic) DP for the region $A = \{x^4 + y^4 \leq 1\}$ which prescribes the function f and its normal derivative on the real points

$$\tilde{A} = \text{ bd } (A) = \{x^4 + y^4 = 1\}.$$

On the other hand, in the Chevalley (polynomial) DP, i.e. Theorem 3.2, we prescribe only the function on $\tilde{A}$ and hence, by analytic continuation, on $\tilde{A}^{\mathbb{C}} =$ complexification of $\tilde{A}$.

Wherein lies the rational for the difference in DD?

To clarify the issues, let us first examine the simpler case

$$P_1(D) = \frac{d^4}{dx^4} \tag{3.71}$$

on the interval $B = [-1, 1]$. The standard DP with data on bd (B) prescribes f and f' on bd (B). This is four conditions as it should be. The Chevalley DP prescribes f on all of the complex points of $V_1 = \{x^4 = 1\}$ which constitutes four conditions again. Thus it is the fact that $\{x^4 = 1\}$ is "twice as large as bd (B)" which implies that we need twice as much data on bd (B).

We can now clarify what happens when $n = 2$. There is a new phenomenon, namely that $V_1 = \{x^4 + y^4 = 1\}$ is now irreducible. Moreover, since $P(D)$ is elliptic, the solutions of $P(D)f = 0$ on A (more precisely, solutions on interior

A which are smooth on A) extend to some a priori neighborhood A^P of A in C^2 by the Cauchy–Kowalewski theorem. (For the purpose of this discussion we can consider the situations in which the DD of f is either smooth or is real analytic. The difference is insignificant in this case. *However*, for other examples discussed below the difference between smooth and real analytic has a profound effect.)

Our first task is to show that, in a suitable sense, V is twice as large as the part of $\tilde{A}^{\mathbb{C}}$ which lies in A^P.

For any x with $|x| < 1$ there are two real values of y for which $x^4 + y^4 = 1$ while there are four complex values of y.

Claim *The same persists on* A^P*; namely, for any* $\tilde{x}$ *there are at most two complex values of* $\tilde{y}$ *on* $\tilde{x}^4 + \tilde{y}^4 = 1$ *for which* $(\tilde{x}, \tilde{y}) \in A^P$.

In order to understand this result, we must first explain how to compute A^P explicitly. The method depends on the fundamental principle of FA (see Section 1.4). Note that A is a strictly convex set. It is proven in Chapter V of FA that the space $\mathcal{E}(A)$ of C^∞ functions on A is LAU (localizable analytically uniform). This means that the solutions of $\partial(P)f = 0$ on A are represented by absolutely convergent Fourier integrals on the complex algebraic variety

$$\begin{aligned} V &= \{P = 0\} \\ &= \{\hat{y} = e^{ij\pi/4}\hat{x}\}_{j=1,3,5,7} \\ &= \{\hat{y} = \alpha_j \hat{x}\} \end{aligned} \tag{3.72}$$

in the form

$$f(x,y) = \int_{(\tilde{x},\tilde{y})\in V} e^{ix\hat{x}+iy\hat{y}} \frac{d\mu(\hat{x},\hat{y})}{k(\hat{x},\hat{y})} \tag{3.73}$$

where μ is a measure of bounded variation and k belongs to an AU structure for $\mathcal{E}(A)$.

The methods of Chapter V of FA indicate that the growth of k can be computed "up to unimportant factors" by the condition

$$|e^{ix\hat{x}+iy\hat{y}}| \le k(\hat{x},\hat{y}) \quad \text{for all } x, y \in A, \tag{3.74}$$

since this is essentially the condition that the integral in (3.73) converges absolutely on A. We can study (3.74) on the rays $\hat{x} = \beta t, \hat{y} = \alpha_j \beta t$ where β is a complex number of absolute value 1 and $t \in [0, \infty)$. It is seen that

$$k(\beta t, \alpha_j \beta t) \sim \max_{x,y\in A} \left| e^{-\Im[(x+\alpha_j y)\beta]t} \right| \tag{3.75}$$

where we use the symbol $\sim$ to denote equality up to unimportant terms.

The domain A^P consists of all $\tilde{x}, \tilde{y}$ for which the inequality (3.74) holds for $\hat{x}, \hat{y} \in V$ (with $\tilde{x}, \tilde{y}$ replacing x, y) because the integral (3.73) converges absolutely there. Thus k is determined by A, i.e. by (3.74), (3.75), and once k is determined

it determines A^P. We claim that for any $\tilde{x}$ except $\tilde{x} = 1$ there are at least two values of $\tilde{y}$ in the complexification of bd (A) for which $(\tilde{x}, \tilde{y}) \notin A^P$.

The simplest computation comes by taking $\beta = \pm\alpha_j^{-1}$ in (3.75). Since x, y are real,

$$\max_{x,y\in A} -\Im\left[(x \pm \alpha_j y)\alpha_j^{-1}\right] t = \max |\Im\alpha_j^{-1} xt| = \frac{\sqrt{2}}{2} t. \tag{3.76}$$

(We do not have to be concerned about the $\pm$ sign because we can reverse the signs of x or y.) By (3.75) this is the size of k at the points corresponding to $\beta = \pm\alpha_j^{-1}$.

To verify our claim we must show that, given $\tilde{x}$, for at least two solutions $\tilde{y}_l$ of $\tilde{x}^4 + \tilde{y}^4 = 1$ there is a $j = j(l)$ such that

$$|\Im(\alpha_j^{-1}\tilde{x} + \tilde{y})| > \frac{\sqrt{2}}{2}. \tag{3.77}$$

Such points $\tilde{x}, \tilde{y}$ cannot be in A^P since (3.74) (with $\tilde{x}, \tilde{y}$ replacing x, y) is not satisfied.

Note that for any $\tilde{x}$ we can find a j so that $\alpha_j^{-1}\tilde{x}$ lies in any given quadrant of the complex plane. Moreover if $\tilde{y}$ is a root of $\tilde{x}^4 + \tilde{y}^4 = 1$ so are $-\tilde{y}$, $i\tilde{y}$, $-i\tilde{y}$. In particular there are two choices, say $\tilde{y}_1, \tilde{y}_2$, amongst $\tilde{y}, -\tilde{y}, i\tilde{y}, -i\tilde{y}$ so that

$$\begin{aligned} \frac{\pi}{4} &\le \arg\, \tilde{y}_1 \le \frac{3\pi}{4} \\ -\frac{\pi}{4} &\ge \arg\, \tilde{y}_2 \ge -\frac{3\pi}{4}. \end{aligned} \tag{3.78}$$

Corresponding to these choices we define $j(1), j(2)$ so that $\alpha_{j(l)}^{-1}\tilde{x}$ lies in the same quadrant. With this choice

$$\begin{aligned} |\Im(\alpha_{j(l)}^{-1}\tilde{x} + \tilde{y}_l)| &\ge \frac{\sqrt{2}}{2}[|\tilde{x}| + |\tilde{y}_l|] \\ &\ge \frac{\sqrt{2}}{2}[|\tilde{x}| + |(1 - |\tilde{x}|^4)|^{1/4}]. \end{aligned} \tag{3.79}$$

Note that for any real number x

$$|(1 - x^4)^{1/4}| \ge \begin{cases} (1 - |x|^4)^{1/4}, & |x| \le 1 \\ (|x|^4 - 1)^{1/4} & |x| \ge 1. \end{cases}$$

To show that the right side of (3.79) is larger than $\sqrt{2}/2$ we are left to verify

$$\begin{aligned} u + (1 - u^4)^{1/4} &> 1 \quad 0 < u < 1 \\ u + (u^4 - 1)^{1/4} &> 1 \quad u > 1 \end{aligned} \tag{3.80}$$

and

$$|\tilde{x}| + |(1-\tilde{x}^4)|^{1/4} > 1 \quad |\tilde{x}| = 1, \quad \tilde{x} \neq 1. \tag{3.81}$$

We must also examine the case $\tilde{x} = 0$.

The second and third inequalities are trivial. For the first inequality we observe that the derivative of $u + (1-u^4)^{1/4}$ is

$$1 - u^3(1-u^4)^{-3/4}.$$

The derivative is positive at $u = 0$, has a simple zero at $u = 2^{-1/4}$, and then is negative, decreasing to $-\infty$ at $u = 1$. (It is readily verified that the second derivative is negative.) Since $u + (1-u^4)^{1/4}$ takes the values 1 at $u = 0$ and at $u = 1$ it is >1 in $(0,1)$. This establishes the first inequality in (3.80).

When $\tilde{x} = 0$ the values of $\tilde{y}$ are $\pm 1, \pm i$. In this case clearly, for $\tilde{y} = \pm i$,

$$|\Im(\alpha_j^{-1}\tilde{x} + \tilde{y})| = 1 > \frac{\sqrt{2}}{2}.$$

Our assertion that there are at most two $\tilde{y}$ for each $\tilde{x}$ except $\tilde{x} = 1$ with $(\tilde{x}, \tilde{y}) \in A^P \cap V_1$ is completely verified.

The fundamental principle implies that the kernel of $\partial^4/\partial x^4 + \partial^4/\partial y^4$ is parametrized by points on V which is a four sheeted covering of $\{\hat{x}\}$.

On the other hand we have shown that A^P forms at most a two sheeted covering of $\{x\}$ (except at $x = 1$). Since the values of a solution f of $P(D)f = 0$ on A with data given on $V_1^{\mathbb{R}}$ = real points of $x^4 + y^4 = 1$ are determined by this data (assuming enough regularity) on A^P, the data given by the value of f on $V_1^{\mathbb{R}}$ is only half the information needed to determine f. For this reason we need 2 data on $V_1^{\mathbb{R}}$ in accordance with the usual DP.

Remark. *This argument is only heuristic because V is a four sheeted covering in $\hat{x}$ space while A^P is (at most) a two sheeted covering in x space.*

A simple modification of the above calculation of k (see (3.74)ff.) shows that if f is a solution of $\partial^4 f/\partial x^4 + \partial^4 f/\partial y^4$ on all of $\mathbb{R}^2$ then f extends to a holomorphic solution $f^{\mathbb{C}}$ on all of $\mathbb{C}^2$. Thus if f vanishes on the real set $V_1^{\mathbb{R}}$ then f vanishes on the whole complex variety V_1. For each x there are four points in V_1 above x. This means that global solutions that vanish on $V_1^{\mathbb{R}}$ have twice as much vanishing as do solutions on A. This explains why one needs normal derivatives to restore uniqueness to solutions on A while for polynomial solutions the vanishing on $V_1^{\mathbb{R}}$ suffices by the harmonic function ideas of Section 3.1. (Recall that there are four points in V above each $\hat{x}$.)

We explained that $(\mathbb{R}^2)^P = \mathbb{C}^2$ so for each $\tilde{x} \in \mathbb{R}^2$ there are four points $(\tilde{x}, \tilde{y}) \in (\mathbb{R}^n)^P \cap V_1$. By modifying our above computation of A^P we can find compact convex domains $\tilde{A} \supset A$, e.g.

$$\tilde{A} = \{x^4 + y^4 \leq 2\},$$

such that for all real x with $|x| \leq 1$ there are four points $(x, \tilde{y}) \in \tilde{A}^P \cap V_1$ which "connect" to $V_1^{\mathbb{R}}$ in the sense described below. Since the branches of V_1 meet at $x = \pm 1$ this implies a four-fold vanishing above $\tilde{A}$ for functions which vanish on $V_1^{\mathbb{R}}$ and are solutions on all of $\tilde{A}$. One might expect that there is uniqueness for the DP for solutions on $\tilde{A}$ with DD being the function value alone on $V_1^{\mathbb{R}}$. Although we shall prove this (even for nonpolynomial solutions) below for $\tilde{A} = \mathbb{R}^2$ we do not know how to prove it for other $\tilde{A}$.

Problem 3.6 *Find compact convex $\tilde{A}$ so that f, being a solution on $\tilde{A}$ which vanishes on $V_1^{\mathbb{R}}$, implies $f = 0$.*

The above discussion concerns $\partial^4/\partial x^4 + \partial^4/\partial y^4$ with data given on $x^4 + y^4 = 1$. Is there a general result which explains the relation between the analytic and Chevalley DPs? Presumably there is one, but the algebra seems highly impenetrable.

We shall study a problem which is not so general in one sense yet more general in another. Let $P \equiv x^m + y^m$, and $Q = x^l + y^l$ where m and l are both even integers. We use the notation $A = \text{real } \{Q \leq 1\}$, $V_Q^{\mathbb{R}} = \text{ bd } A$, and $V_Q = \{x^l + y^l = 1, x, y \text{ complex}\}$ = complexification of $V_Q^{\mathbb{R}}$, and A^P is an a priori set in $\mathbb{C}^2$ into which solutions of $P(D)f = 0$ on A continue.

We define $(A^P \cap V_Q)_0$ as the union of all branches of $A^P \cap V_Q$ which meet $V_Q^{\mathbb{R}}$ in an open set of $V_Q^{\mathbb{R}}$. Put in more practical terms, $(A^P \cap V_Q)_0$ is the subset of $A^P \cap V_Q$ on which holomorphic functions are uniquely determined by their values on $V_Q^{\mathbb{R}}$.

As in case $l = m = 4$ we want to study the size of $A^P \cap V_Q$ since it is on this set that a solution f of $P(D)f = 0$ which vanishes on bd A must vanish.

Theorem 3.8 *If $l \leq m$ then for any real $\tilde{x}$ there are at most two points $(\tilde{x}, \tilde{y}) \in A^P \cap V_Q$. If m divides l then there are at most $2l/m$ such points. However, there are at most two $\tilde{y}$ for which $(\tilde{x}, \tilde{y}) \in (A^P \cap V_Q)_0$ even for $\tilde{x}$ complex.*

Proof We proceed along the lines of our proof in case $m = l = 4$. We may suppose that $l \geq 4$; otherwise $l = 2$ so there is nothing to prove.

We explained above (see (3.75)) that the AU structure $\{k\}$ is determined on $\hat{x} = \beta t, \hat{y} = \pm\alpha_j \beta t$ by

$$\max_{x,y \in A} -\Im[(x \pm \alpha_j y)\beta]t.$$

Here α_j is an m-th root of -1 and $|\beta| = 1$. Choosing $\beta = \alpha_j^{-1}$ as before leads to

$$\max |\Im x \alpha_j^{-1}| = |x| \sin \frac{\pi}{m} \tag{3.82}$$

if we choose j so that $|\arg \alpha_j|$ or $|\arg(\alpha_j - \pi)|$ is minimal, i.e. $\alpha_j = \pm \exp(\pm i\pi/m)$. This is the analog of (3.76) and determines k when $\beta = \alpha_j^{-1}$ for this α_j.

If $(\tilde{x}, \tilde{y}_p) \in V_Q$ then all $\tilde{y}_p$ are of the form $\omega_p \tilde{y}_1$ where ω_p is some l-th root of unity.

Suppose first that $l \leq m$. Given $(\tilde{x}, \tilde{y}_p) \in V_Q$, with the possible exception of one value of p, $|\arg \tilde{y}_p| \geq \pi/l$, and also except for possibly one value of p, $|\arg \tilde{y}_p - \pi| \geq \pi/l$, since multiplication by ω_p changes the argument by at least $2\pi/l$.

Let us suppose that

$$\frac{\pi}{l} \leq \arg \tilde{y}_p \leq \frac{(l-1)\pi}{l}.$$

By (3.82) $\arg \alpha_j \tilde{x}$ satisfies the same inequality with m replacing l. Hence

$$|\Im(\tilde{x}\alpha_j^{-1} + \tilde{y}_p)| \geq (|\tilde{x}| + |\tilde{y}_p|) \sin \frac{\pi}{m}. \tag{3.83}$$

Note that

$$|\tilde{y}_p| = |1 - \tilde{x}^l|^{1/l} \geq \begin{cases} (1 - |\tilde{x}|^l)^{1/l} & |\tilde{x}| < 1 \\ (|\tilde{x}|^l - 1)^{1/l} & |\tilde{x}| > 1. \end{cases}$$

In order to contradict (3.82) and hence have $(\tilde{x}, \tilde{y}_p) \notin A^P$ we are left to prove that

$$\begin{aligned} u + (1 - u^l)^{1/l} > 1 \quad & 0 < u < 1 \\ u + (u^l - 1)^{1/l} > 1 \quad & u > 1 \\ |\tilde{x}| + |(1 - \tilde{x}^l)|^{1/l} > 1 \quad & |\tilde{x}| = 1, \tilde{x} \neq 1. \end{aligned}$$

The previous proof that we gave for $l = 4$ carries over *mutatis mutandis*.

Let us consider the case $l > m$. If m divides l then there might be $2l/m$ values of p for which $|\arg \tilde{y}_p| < \pi/m$ or $|\arg \tilde{y}_p - \pi| < \pi/m$. For the other values of p we proceed as before.

Finally we must show that there are not more than two values of p for which $(\tilde{x}, \tilde{y}_p) \in (A^P \cap V_Q)_0$. To verify this we examine the geometry of V_Q. Regard V_Q as the branched covering of the complex x plane defined by $y^l = 1 - x^l$. Note that for any real point x with $|x| < 1$ there are only two real values of y in V_Q lying above x. These define two branches which are the only two branches that can meet $V_Q^{\mathbb{R}}$ in an open set as the branches defined by nonreal roots $y = (1 - x^l)^{1/l}$ when x is real and $|x| < 1$ have the point $x = 1,\ y = 0$ as their only real point in $V_Q^{\mathbb{R}}$.

This concludes the proof of Theorem 3.8.

As we have mentioned, the variety $\hat{x}^m + \hat{y}^m = 0$ which, by the fundamental principle, defines the solutions of $\partial(P)f = 0$ is an m sheeted covering of $\{\hat{x}\}$. On the other hand the last part of Theorem 3.8 indicates that boundary conditions on bd A relate to a two sheeted covering of $\{x\}$ since the boundary data can only impinge on $(A^P \cap V_Q)_0$. For this reason the usual DP for $\partial(P)$ (when m is even) prescribes $m/2$ data.

On the other hand V_Q is an irreducible l sheeted covering of $\{x\}$. If we give entire data on bd A (i.e. functions on bd A which are the restrictions of entire functions) then they promulgate by analytic continuation to all of V_Q. Thus each datum corresponds (heuristically) to l points above $\{x\}$. This suggests that if $l|m$ then we should give m/l data on bd A when dealing with solutions on all of $\mathbb{R}^n$. We now present some ideas as to how this can be accomplished.

Of course when $l = m$ then this is exactly the theory treated in Section 3.1, except that we only know how to deal with the general problem within the framework of polynomials, rather than entire functions. (We shall discuss entire solutions presently.)

Our results are restricted to dimension 2. Let P, Q be homogeneous polynomials on $\mathbb{R}^2$ of respective degrees m, l. We write

$$Q(x,y) = \prod_1^l (y + \beta_p x)$$

$$P(x,y) = \prod_1^m (y + \alpha_j x)$$

where the $\{\alpha_j\}$ and the $\{\beta_j\}$ are distinct. (We have ignored unimportant constants.) We need

Assumption *For $N \geq m$*

$$\det(\beta_p - \alpha_j)^{N-k} \neq 0. \tag{3.84}$$

(Here $k = 0, 1, \ldots, m/l - 1$. The rows of the above matrix are indexed by j and the columns by p, k.)

It is easy to find polynomials for which our assumption is satisfied, e.g. $P = Q = x^m + y^m$. Thus the condition on pairs of polynomials is "generic." But we do not know any general conditions on pairs (P, Q) which guarantee that the assumption is satisfied.

Theorem 3.9 *Suppose the assumption is valid for (P, Q). The DP for polynomial solutions of $P(D)f = 0$ is well posed if we prescribe*

$$f|_{Q=1}, \quad \frac{\partial f}{\partial \eta}|_{Q=1}, \ldots \frac{\partial^{m/l-1}}{\partial \eta^{m/l-1}} f|_{Q=1}.$$

Here $\partial/\partial\eta$ is the normal derivative

$$\frac{\partial}{\partial \eta} = Q_x \frac{\partial}{\partial y} - Q_y \frac{\partial}{\partial x}.$$

Proof We begin with a proof of uniqueness.

Our first observation is that the vanishing of the DD of f is equivalent to f being divisible by $(Q-1)^{m/l}$. This follows from the fact that $Q-1$ is irreducible.

Next we apply the fundamental principle for solutions f of $P(D)f = 0$ to write

$$f(x,y) = \sum w_j(x - \alpha_j y) \tag{3.85}$$

where w_j are polynomials. Suppose f is of degree $N \geq m$; at least one of the w_j has degree N. Call $w_j^N = \gamma_j^N (x - \alpha_j y)^N$ the homogeneous part of w_j of degree N; the (γ_j^N) are constants, not all 0. If DD(f) vanishes then the equation

$$f(x,y) = (Q-1)^{m/l} g(x,y)$$

implies degree $g = N - m$. It leads to

$$\sum \gamma_j^N (x - \alpha_j y)^N = Q^{m/l} g^{N-m} \tag{3.86}$$

where g^{N-m} is the homogeneous part of g of degree $N - m$.

Now

$$Q(x,y) = \prod (x - \beta_p y).$$

The right side of (3.86) has a zero of order m/l at $x = \beta_p y$. Thus

$$\sum \gamma_j^N (\beta_p - \alpha_j)^{N-k} = 0$$

for all p and for $0 \leq k \leq m/l - 1$. This is a system of m equations for the m unknowns γ_j^N. By the assumption it follows that $\gamma_j^N = 0$.

We conclude that degree $f < m$. But then $f \equiv 0$ since such an f cannot be divisible by $(Q-1)^{m/l}$.

As for the existence, suppose we are given m/l polynomials $u_0, \ldots, u_{m/l-1}$ which we want to be the DD of a polynomial solution f of $P(D)f = 0$. Let v be a polynomial (not necessarily a solution of $P(D)v = 0$) whose DD is $(u_0, \ldots, u_{m/l-1})$. (It is clear that such v exist.) We want to write v in the form (recall that, by the fundamental principle, f is given by (3.85))

$$v = f + (Q-1)^{m/l} v_1$$

where $P(D)f = 0$. This is the analog of the Fischer orthogonal decomposition.

Suppose v is of degree $N \geq m$. Then we may suppose that f is of degree N and v_1 is of degree $N - m$. The homogeneous part of our relation is

$$v^N(x,y) = \sum \gamma_j^N (x - \alpha_j y)^N + Q^{m/l}(x,y) v_1^{N-m}(x,y).$$

(Our notation is such that $(x - \alpha_j y)^N$ is the N-th power of $x - \alpha_j y$ and $Q^{m/l}$ is the m/l power of Q but v^N, v_j^{N-m} represent homogeneous terms.) $Q^{m/l}$ has an

m/l-fold zero along $x = \beta_p y$. As in the proof of uniqueness we can solve for γ_j^N by setting $x = \beta_p y$ and using the assumption.

We have thus reduced our search for f from a polynomial of degree N to degree $N - m$. The result follows by iteration as in Section 3.1.

Let us now analyze the case m divides l. The simplest example (which, strictly speaking, does not fall into the above theory) is $m = 1, l = 2$ where $m = 1$ refers to the Cauchy–Riemann equations and $Q(x, y) = x^2 + y^2$. A polynomial f on $\{Q = 1\}$ is the DD of a holomorphic function if and only if its negative Fourier series coefficients vanish.

But when $l > 2$ the situation is not clear. However, we can reformulate the case $m = 1, l = 2$ in such a way as to have some meaning in general.

First, there is a unique (ordinary) harmonic polynomial h which agrees with f on $\{Q = 1\}$. The above ideas tell us how to compute h explicitly, namely let f^N be the highest homogeneous part of f. Then we can write

$$f^N(x, y) = h_+^N(x + iy)^N + h_-^N(x - iy)^N + (x^2 + y^2)f_1$$

where degree $f_1 = N - 2 < N$. Our task is to compute $h_+(x + iy)$. Writing as before $h - f = (Q-1)g$ and examining the terms of degree N allows us to calculate the constants $h_\pm^N$ from the values of f^N on $x = \pm iy$ as before. We have thus avoided the nonexplicit projection theorem which we employed in connection with the Fischer theory and we replaced it by an explicit computation using the fundamental principle and the nonvanishing of the determinant $\{(\lambda_p - \alpha_j)^N\}$ where $\lambda_p = \pm i$ and $\alpha_j = \pm i$.

Continuing in this way we can determine if f is the DD of a holomorphic function—namely, all the $h_-^{N'}$ for $N' \geq 1$ vanish.

Problem 3.7 *What is the general setting for elliptic polynomials P, Q for the above theory? In particular, when does the analog of Theorem 3.9 hold?*

The algebraic theory presented in the previous section depended on Fischer's Lemma 3.1 and on the projection theorem in Hilbert space. As such the results were purely qualitative and so applied only to polynomials. To go further we need to quantify Fischer's lemma.

To understand some of the difficulties we face, consider the case $n = 2$, $i = x_1 x_2$. The harmonics are spanned by the powers of x_1 and of x_2. If f is any entire function, we write

$$f(x) = \sum a_{kl} x_1^k x_2^l. \tag{3.87}$$

Then

$$f(x) = \sum_{k \geq l} a_{kl}(x_1 x_2)^l x_1^{k-l} + \sum_{k<l} a_{kl}(x_1 x_2)^k x_2^{l-k}. \tag{3.88}$$

Thus

$$f(x) = \sum_{k \geq l} a_{kl} x_1^{k-l} + \sum_{k<l} a_{kl} x_2^{l-k} \tag{3.89}$$

on $i = 1$. If f is entire all series converge to entire functions so f agrees with an entire i harmonic function h on $i = 1$. (The right side of (3.89) is harmonic.)

h is unique modulo constants. For if $h_1(x_1) + h_2(x_2)$ vanishes on $i = 1$ then

$$h_1(x_1) = -h_2(x_1^{-1})$$

which is impossible for nonconstant entire functions (let $x_1 \to 0$).

The difficulty in extending this example to a general class of polynomials is obvious. We shall discuss this method in detail below.

As a sideline observe that uniqueness modulo constants holds for smooth functions of the real variables x_1, x_2. For if $\partial(i)h = 0$ then $h(x_1, x_2)$ is of the form $h_1(x_1) + h_2(x_2)$. If $h_1(x_1) + h_2(x_2) = 0$ on $\{x_1 = 0\} \cup \{x_2 = 0\}$ then

$$h_1(x_1) = -h_2(0)$$
$$h_2(x_2) = -h_1(0)$$

so h_1 and h_2 are constants.

We shall now present three approaches to a quantified theory that can go beyond polynomials. Unfortunately none of these methods is very successful and we have only scattered results.

If we work within the framework of power series then quantification requires a study of how the Fischer inner product works on homogeneous polynomials of fixed degree. Because of the manifold combinatorial complications we shall restrict our considerations to the case $r = 1$; that is, a single i which is homogeneous of degree d.

Let $\mathcal{P}^n$ denote the Hilbert space of homogeneous polynomials of degree n with the Fischer norm. Thus the harmonics and the image $i\mathcal{P}^{n-d}$ are orthogonal complements. We need two quantitative properties:

Quant A. *Compare the norm* $\|Q\|_n$ *of a polynomial in* $\mathcal{P}^n$ *to the maximum of its coefficients. Precisely, for any homogeneous polynomial* Q *of degree* n

$$a_n\|Q\|_n \leq \max|\text{coeff. } Q| \leq b_n\|Q\|_n. \tag{3.90}$$

Quant B. *Estimate the norm of the inverse of multiplication by* i. *Thus, for* $P \in \mathcal{P}^{n-d}$

$$\|P\|_{n-d} \leq c_n\|iP\|_n. \tag{3.91}$$

We shall, presently, give precise estimates for the constants in Quant A and Quant B. Before doing this, let us examine what is needed to extend Theorem 3.2 to entire functions. By this we mean that any entire function on $\mathcal{O} = \{i = 1\}$ has a unique entire harmonic extension to all of C^n.

We know how to determine whether $f = \sum f^n$ (f^n is the homogeneous part of f of degree n) is an entire function in terms of its maximum coefficient norm $|||f|||$; clearly the condition is

$$|||f^n||| \leq \alpha_\varepsilon \varepsilon^n \tag{3.92}$$

for any $\varepsilon > 0$. Our first task is to estimate a_n and b_n so as to characterize the entire nature of f in terms of the Fischer norms of $\{f^n\}$.

Before making the estimates, let us see how they can be applied. Suppose there is an entire harmonic function h which vanishes on $i = 1$. Thus there is a relation $h = (i-1)u$ for some entire function u, i.e.

$$\sum h^n = (i-1)\sum u^n. \tag{3.93}$$

This gives

$$h^n = iu^{n-d} - u^n. \tag{3.94}$$

Equation (3.94) breaks up the $\{n\}$ in (3.93) into residue classes mod d. Consider $\{n\} = \{kd\}$ as the other residue classes are similar. Also, assume for convenience that $h^0 = 1$ so that $u^0 = -1$.

Let us examine (3.94). Since $h^n \perp iu^{n-d}$ it follows that

$$\|u^n\| \geq \|iu^{n-d}\|. \tag{3.95}$$

But $\|u^0\| = 1$ from which we conclude by (3.91) that

$$\|u^{jd}\| \geq (c_{jd}c_{(j-1)d}\dots c_d)^{-1}. \tag{3.96}$$

This means that if we have a good lower estimate on the c_n and if a_n in (3.90) is not small then by (3.92) the series $\sum u^n$ could not define an entire functions. This obviates (3.93) and so proves uniqueness of the DP for entire functions. (We shall discuss existence presently.)

Now for the real work of estimating $a_n, b_n,$ and c_n. We use multi-index notation $t^m = t_1^{m_1}\dots t_k^{m_k}$. Let $u^{|m|}$ be a homogeneous polynomial of degree $|m|$,

$$u^{|m|} = \sum \alpha_m t^m. \tag{3.97}$$

Then the Fischer norm of $u^{|m|}$ is given by

$$\begin{aligned}\|u^{|m|}\|^2 &= \partial(\overline{u^{|m|}})\cdot u^{|m|}\\ &= \sum m!|\alpha_m|^2.\end{aligned} \tag{3.98}$$

We have written $m! = m_1!\dots m_k!$. The number of terms in the sum in (3.98) is

$$\binom{|m|+k-1}{k-1} \sim |m|^k \tag{3.99}$$

which is an unimportant factor for most estimates. (We are ignoring terms which are obviously of smaller magnitude than $|m|^{c|m|}$ such as $|m|^c, c^{|m|}$.) Thus in (3.90) we may take

$$a_{|m|} = \frac{1}{[|m|^k \max m!]^{1/2}} \sim |m|^{-k/2}|m|^{-|m|/2}e^{|m|/2}. \tag{3.100}$$

Moreover since the sum in (3.98) majorizes any of its terms

$$b_{|m|} = (\min m!)^{-1/2} \sim |m|^{-|m|/2} k^{|m|/2} e^{|m|/2}. \tag{3.101}$$

(The max and min are over all m with fixed $|m|$.)

It is clear that max $m!$ occurs when some $m_j = |m|$ and all others are 0. Thus

$$\max m! = |m|! \sim e^{-|m|} |m|^{|m|}$$

by Stirling's formula. This is (3.100).

On the other hand, min $m!$ occurs when each $m_j \sim |m|/k$ so that

$$\begin{aligned} \min m! &\sim \Gamma(|m|/k)^k \\ &\sim \left(\frac{|m|}{k}\right)^{|m|} e^{-|m|} \end{aligned}$$

which is (3.101).

It is important to note that the difference between max $m!$ and min $m!$ is the factor $k^{-|m|}$ which is insignificant for most of our purposes.

Combining these estimates with (3.92) yields the inequality

$$\|f^n\| \leq \alpha_\varepsilon \varepsilon^n n^{n/2} \tag{3.102}$$

as the characterization of entire functions in terms of the Fischer norm.

In order to apply (3.102) to (3.96) to obtain uniqueness, we need the estimate

$$c_{jd} c_{(j-1)d} \dots c_d \leq \alpha c^{jd} (jd)^{-jd/2} \tag{3.103}$$

for some constants $\alpha, c \neq 0$. For then, by (3.96),

$$\|u^{jd}\| \geq \alpha^{-1} c^{-jd} (jd)^{jd}$$

which contradicts (3.102) so $u = \sum u^n$ would not be entire.

Actually, (3.103) is a consequence of

$$c_n \leq c n^{-d/2}. \tag{3.104}$$

For then we have the estimate

$$\begin{aligned} c_{jd} c_{(j-1)d} \dots c_d &\leq c^j \{(jd)[(j-1)d] \dots d\}^{-d/2} \\ &= c^j d^{-jd/2} (j!)^{-d/2} \sim c^j (jd)^{-jd/2} \end{aligned}$$

(Stirling) as required.

Thus (3.104) implies uniqueness. (We emphasize that we have not proven (3.104).)

For existence we write f^n in its Fischer decomposition

$$f^n = h_0^n + ig_d^n. \tag{3.105}$$

(g_d^n has degree $n-d$.) Applying the same decomposition to g_d^n in place of f^n and iterating (still assuming $n = jd$) leads to

$$f^n = h_0^n + ih_d^n + i^2 h_{2d}^n + \cdots + i^{n/d} h_{dn/d}^n. \tag{3.106}$$

We can estimate $\|h_{jd}^n\|$ as follows. Since (3.105) represents orthogonal projection the Fischer norms of both h_0^n and ig_d^n are majorized by $\|f^n\|$. Use of (3.91) allows us to estimate

$$\|g_d^n\| \le c_n \|ig_d^n\| \le c_n \|f^n\|.$$

We have noted that $\|h_0^n\| \le \|f^n\|$. Moreover the relation of h_d^n to g_d^n is the same as the relation of h_0^n to f^n. We can thus iterate the process to obtain

$$\begin{aligned} \|h_{jd}^n\| &\le c_n c_{n-d} \dots c_{n-(j-1)d} \|f^n\| \\ &\le \alpha_\varepsilon \varepsilon^n c_n c_{n-d} \dots c_{n-(j-1)d} n^{n/2}. \end{aligned} \tag{3.107}$$

The last inequality arises from (3.102).

If f is equal to an harmonic function $h = \sum h^m$ on $i = 1$ then, assuming the uniqueness of the DP, we derive from (3.106)

$$h^n = \sum h_{jd}^{n+jd}$$

on $i = 1$. For h to be entire means that the h^n satisfy the inequality (3.102). It is sufficient to have the individual summands h_{jd}^{n+jd} satisfy a similar inequality in which the right side of (3.102) is replaced by $\alpha_\epsilon \epsilon^{n+jd} n^{n/2}$. By (3.107) this means that we want

$$c_{n+jd} c_{n+(j-1)d} \dots c_{n+d} \le c^n (n+jd)^{-(n+jd)/2} n^{n/2}.$$

The natural inequality we might expect for c_n is (3.104) again. However, the best possible result we could obtain from inequalities of this form is that the product is

$$\le c^j (n+jd)^{-jd/2}.$$

For large n and j this inequality is too weak.

More work is needed to obtain proper estimates. In any case it seems easier to prove uniqueness than existence.

It remains to verify (3.104). The most reasonable way to obtain estimates for c_n is given by estimating i^{-1} using the maximum coefficient norm. Such an estimate could be obtained by relating the maximum coefficient norm to the sup norm on the polydisk and using the standard division method as set forth, e.g. in FA.

However, this method is doomed to failure! The reason is that the difference between a_n and b_n in (3.90) is a multiple of $k^{n/2}$ by (3.100) and (3.101). This is too large to be absorbed in any estimate for c_n (compare (3.104)).

Thus we must make a careful investigation of the inequalities (3.100) and (3.101). The main point is that the weights $m!$ where $|m| = n$ are largest near the "outside," i.e. when one of the components m_j is close to n and the others are small. It is smallest "inside," i.e. where all components are close to n/k. We would be in trouble if g_d^n had large coefficients outside and small coefficients inside and the situation were reversed for ig_d^n.

"Experimentation" seems to indicate that such a condition is impossible and, in fact, large coefficients in g_d^n "flow out" upon multiplication by i. But a detailed proof is lacking.

We now present a second approach to a nonpolynomial Fischer theory. At present this method works only for $k = 2$. We do obtain results (Theorem 3.6) for $k = 2$ on the uniqueness of the DP. Let us illustrate this first for $i = x_1^2 + x_2^2$. Any solution h of $\partial(i)h = 0$ on all of $\mathbb{R}^2$ is of the form $h = h^+ + h^-$ where h^+ is a holomorphic function of $x_1 + ix_2$ and h^- is a holomorphic function of $x_1 - ix_2$.

Suppose h vanishes on $x_1^2 + x_2^2 = 1$. Then it vanishes on the complexification of this circle. The lines $x_1 = \pm ix_2$ are asymptotes of the complexification. Now, as we go along the complexification of the circle to a point at infinity on the complex line $x_1 = ix_2$, $x_1 + ix_2$ tends to a point at infinity on the complex circle, which can be chosen arbitrarily, while $x_1 - ix_2$ stays bounded; hence h^- stays bounded. Since $h^+ + h^- = 0$ this means that h^+ is bounded at infinity and so is a constant. Similarly h^- is constant. Since h vanishes on $i = 1$, it follows that $h \equiv 0$. (In fact the same argument shows that a nonidentically vanishing h which vanishes on $x_1^2 + x_2^2 = 1$ must have singularities both in the unit circle and outside the unit circle (including infinity). The example $h = z - \overline{z}^{-1}$ illustrates what can happen.)

The above argument applies, in principle, to any homogeneous polynomial i in two variables which has distinct factors. Let

$$i = \prod (x_1 + \alpha_j x_2).$$

Any entire i harmonic function h can be expressed in the form

$$h = \sum h^j(\alpha_j x_1 - x_2)$$

where the h^j are arbitrary entire functions. This can be proven in many ways. For example, it is an immediate consequence of the Fourier representation of h using the fundamental principle.

We claim that h cannot vanish on $i = 1$. To this end we examine the asymptotes $x_1 = -\alpha_j x_2$ of $i = 1$. We cannot completely "separate" the h^j along these asymptotes as we did for $x_1^2 + x_2^2$. We can, nevertheless, prove that h cannot vanish on $i = 1$ by a slight modification of that argument.

We need the following:

Assumption *For each p there is an $l = l(p)$ so that*

$$|\alpha_p \alpha_l + 1| > |\alpha_j \alpha_l + 1| \quad \textit{for } j \neq p. \tag{3.108}$$

Our first claim is that no h^p can have a larger modulus at infinity than the others. (We do not give a precise meaning to "larger" as there are many possibilities.) For, if there were such an h^p then the sum $\sum h^j(\alpha_j x_1 - x_2)$ could not vanish as we approach the branch of $i = 1$ at infinity where $x_1 \sim -\alpha_{l(p)} x_2$ because h^p would dominate the sum. In fact the strict inequality in (3.108) shows that no h^p can even grow "at least as fast as" all the others. Thus no h^p can grow at all so all h^p are constants. Finally, if h vanishes on $i = 1$ then $h \equiv 0$. We have thus proven

Theorem 3.10 *When the assumption (3.108) holds no entire i harmonic function can vanish on $i =$ const.*

This argument seems to apply only to entire h. I do not know how to modify it to apply to h which are holomorphic on large sets.

The first surprising case of Theorem 3.10 (or Theorem 3.2) occurs for $k = 2$, $r = 1$ and $i = x_1^4 + x_2^4$. We have discussed the difference between the classical DP and the polynomial (Chevalley) DP. The classical DP is well posed for $i \leq 1$, meaning that we can uniquely determine a solution h of $\partial(i)h = 0$ on $i \leq 1$ by the values (h_0, h_1) of h and its normal derivative on $i = 1$. In particular we can set $h_0 = 0$ and obtain h which are real analytic on $i < 1$ and which vanish on $i = 0$. By Theorem 3.2 such an h cannot be a polynomial and by Theorem 3.10 h cannot be entire.

A fundamental problem in this connection is

Problem 3.8 *Determine the maximal domain (domains) on which there exists a nontrivial solution h of $(\partial^4/\partial x_1^4 + \partial^4/\partial x_2^4)h = 0$ which vanishes on $x_1^4 + x_2^4 = 1$.*

Instead of going to ∞ as we did in the proof of Theorem 3.10 we shall examine a possible extension of the power series method used in the above example for which $i = xy$ (see (3.87)ff.). Now let i be any homogeneous polynomial of degree m in two variables. We write

$$i(x, y) = \prod_{j=1}^{m} (x + a_j y). \tag{3.109}$$

We assume the α_j are distinct. (If they are not distinct the same ideas apply but with considerable technical complication.) The harmonics for i are spanned by

$$h_j^N(x, y) = (\alpha_j x - y)^N. \tag{3.110}$$

Suppose we want to write the explicit Fischer decomposition

$$x^k y^{N-k} = \sum_j a_{jN}^k (\alpha_j x - y)^N + R_{jN}^k(x, y) i(x, y). \tag{3.111}$$

To solve for a^k_{jN} we set $x = -\alpha_l y$. Then

$$\pm\alpha_l^k = \sum_j a^k_{jN}(\alpha_j\alpha_l - 1)^N. \tag{3.112}$$

This is a system of m equations for the m unknowns a^k_{jN}.

Since the dimension of the space of homogeneous polynomials of degree N is $N+1$, their restrictions to $V = \{P = 0\}$ form a linear space of dimension $N + 1 - (N - m + 1) = m$. (The kernel of the restriction map is the set of multiples by P of homogeneous polynomials of degree $N-m$.) This is exactly the dimension of the restriction to V of the space of homogeneous harmonics of degree N as follows from the Fischer decomposition. For if we write a homogeneous polynomial of degree N

$$f = h + iR$$

then h is homogeneous of degree N and R is an arbitrary homogeneous polynomial of degree $N - m$. h determines f on V so dim $\{h\} \geq m$. By Fischer's theorem h is unique so dim $\{h\} = m$.

We conclude from (3.112) that

$$\det[\alpha_j\alpha_l - 1]^N \neq 0$$

for any N. For example, for $m = 2$ the determinant is

$$[(\alpha_1^2 - 1)(\alpha_2^2 - 1)]^N - [(\alpha_1\alpha_2 - 1)^2]^N.$$

The nonvanishing is the statement that $\alpha_1^2 + \alpha_2^2 \neq 2\alpha_1\alpha_2$ unless $\alpha_1 = \alpha_2$ (which we have excluded).

We can go somewhat further when the α_j are real (which we now assume); in fact, the determinant is $\geq C^N$ for some C. To see this, note that

$$|\det(\alpha_j\alpha_l - 1)^N| = (\alpha_1 \dots \alpha_m)^N |\det(\alpha_j - \alpha_l^{-1})^N|. \tag{3.113}$$

Lemma 3.11 *Let $(\gamma_1, \dots, \gamma_m)$ and $(\delta_1, \dots, \delta_m)$ be sequences of distinct real numbers such that $\gamma_i > \delta_j$ for all i, j. There is a permutation p_0 of $1, 2, \dots, m$ for which*

$$|(\gamma_1 - \delta_{p_0(1)}) \dots (\gamma_m - \delta_{p_0(m)})| \tag{3.114}$$

is strictly larger than the value of the product for any other permutation p.

Proof Let us begin with $m = 2$. We may assume that $\gamma_1 > \gamma_2 > \dots > \gamma_m$. We shall show that p_0 orders the δ_j by $\delta_1 > \delta_2 > \dots > \delta_m$.

Let us begin with $m = 2$. We have

$$\begin{aligned}(\gamma_1 - \delta_2)(\gamma_2 - \delta_1) &= [(\gamma_1 - \delta_1) + (\delta_1 - \delta_2)][(\gamma_2 - \delta_2) - (\delta_1 - \delta_2)] \\ &= (\gamma_1 - \delta_1)(\gamma_2 - \delta_2) + (\delta_1 - \delta_2)[(\gamma_2 - \delta_2) - (\delta_1 - \delta_2) - (\gamma_1 - \delta_1)] \\ &= (\gamma_1 - \delta_1)(\gamma_2 - \delta_2) - (\delta_1 - \delta_2)(\gamma_1 - \gamma_2) \\ &< (\gamma_1 - \delta_1)(\gamma_2 - \delta_2).\end{aligned}$$

We can now iterate this argument. If $\delta_1, \ldots, \delta_m$ is not in decreasing order, we use the result for $m = 2$ to "slide" the largest δ_j to the left step by step using the result for $m = 2$. We have assumed that each $\gamma_i - \delta_j$ is positive so this process can continue. Since all products are positive this sliding increases the absolute value.

Continuing in this way with the next largest δ_j, etc., we complete the proof of the lemma.

The lemma implies that, in fact, the same result is true generically for complex γ_j, δ_k.

Now, assume the conclusion of the lemma. (The hypothesis is valid if, for example, α_j are real, distinct, and > 1 with $\gamma_i = \alpha_i$ and $\delta_j = \alpha_j^{-1}$.) Then for N large enough

$$|\det(\alpha_j - \alpha_l^{-1})^N| \geq (1 - \varepsilon) \prod |(\alpha_j - \alpha_{p_0(j)}^{-1})^N| \tag{3.115}$$

because when we expand the determinant the N-th power of the maximum term dominates all the rest.

It follows from (3.113), (3.112), and (3.115) that

$$|a_{jN}^k| \leq c_1^N. \tag{3.116}$$

Suppose now that $f = \sum_{kN} u_{kN} x^k y^{N-k}$ is entire. This means that $|u_{kN}| \leq b\varepsilon^N$. Thus when we express

$$f = h + Ri$$

using (3.111), our estimates (3.116) show that h and hence R are entire.

Iterating this process, i.e. replacing f by R and continuing in this manner, we arrive at the tensor product expansion

$$f = h_0 + h_1 i + h_2 i^2 + \cdots$$

where each h_j is entire. Since the power series for $h_k P^k$ has no homogeneous terms of degree $< km$, the series converges to f coefficientwise. But the estimate $|a_{jN}^k| \leq c_1^N$ may *not* be good enough to establish convergence of the series in the space of entire functions.

To understand this point note that the estimate for a_{jN}^k given in (3.116) depends on the coefficients b_N^k of $x^k y^{N-k}$ in the power series for f and the inverse of the determinant which by (3.113)ff. is (in modulus) $\geq C^N$ where (essentially)

$$C = |\alpha_1 \ldots \alpha_m| \prod |\alpha_j - \alpha_{p_0(j)}^{-1}|.$$

In the next step of iteration we estimate the coefficients of h_1 with f replaced by R. Since $Ri = f - h$ the estimate is of a similar nature except that the bounds for the coefficients of Ri depend on the previous estimate for those of h. The coefficients of homogeneous terms of order N in h have been estimated by C^{-N}

times the corresponding estimate for f. If C is large then we are all right so long as division by i does not upset the estimates.

One must also take into account the size of α_j^k in (3.112).

Problem 3.9 *For which i do we have convergence of the tensor product series for f in the space of entire functions?*

Actually the variety $i = 1$ is not the simplest case. In fact $\vec{i} = 0$ can be treated simply and completely:

Proposition 3.12 *The ring of formal power series $\mathcal{F}$ is a (topological) tensor product $\mathcal{F} = H \otimes \tilde{I}$, meaning that any formal power series f has a unique expansion*

$$f = \sum \vec{i}^m h_m.$$

Here $\vec{i} = (i_1, \ldots, i_r)$, m is a multi-index, and h_m is a harmonic formal power series

$$h_m = \sum h_m^j$$

where h_m^j is a homogeneous harmonic of degree j.

Proof Note that for fixed N there are only a finite number of $\vec{i}^m h_m^j$ of degree N so that the sum is a well defined formal power series.

We write $f = \sum f^n$. Then by Theorem 3.2 we have

$$f^n = \sum_{j+m\cdot\deg\vec{i}=n} \vec{i}^m h_m^j. \tag{3.117}$$

h_m^j are uniquely determined by (3.117). By summing over n we have the result.

Corollary 3.13 *If h is a formal power series harmonic function which vanishes on the multiplicity variety $\underline{X}_0 : \{\vec{i} = 0\}$ then $h = 0$.*

Proof If h vanishes on $\underline{X}_0$ then $h = \sum i_\alpha g_\alpha$. Apply Proposition 3.12 to write the g_α in accordance with the tensor product decomposition. This expresses h in the form $\sum \vec{i}^m \cdot h_m$ where $h_0 = 0$. By the Fischer orthogonality and the uniqueness of tensor product expansion it follows that $h \equiv 0$.

There is a third approach to the analytic theory which has the promise of being applicable even to functions which are not entire. We have not yet developed this technique to its fullest extent. We hope that some readers will find it enticing.

Again we consider only the case $r = 1$. We rewrite the relation

$$f = h + ig.$$

as

$$f \cdot \frac{1}{i} = h \cdot \frac{1}{i} + g.$$

Although this change seems trivial, it is not. For it asserts that there is a unique harmonic h for which h/i has the same residues as f/i. Actually the real power of the division is seen by the Fourier transform:

$$\hat{f} * e = \hat{h} * e + \hat{g}$$

where e is a fundamental solution for $\partial(i)$.

Suppose that f, hence also h and g, is a polynomial. Since h is harmonic $\hat{h}$ can be expressed in terms of powers of directional derivatives along the generators of the cone $i = 0$, if we assume that the cone is nice. Moreover the support of $\hat{g}$ is the origin.

Suppose that we can construct e fairly explicitly, or, at least, asymptotically. Then, since

$$\hat{f} * e = \hat{h} * e$$

away from the origin, this gives an explicit method for constructing $\hat{h}$ and hence h. Suitable norms of $\hat{h}$ should be computable in terms of asymptotics of $\hat{f} * e$. In this way we can bound $\hat{h}$ and hence $\hat{g}$ in terms of bounds on $\hat{f}$, therefore obtaining an explicit Fischer decomposition for entire functions.

The reader should find it interesting to examine the above idea for the simplest nontrivial case $i = \sum t_j^2$.

3.3 Fourier series expansions on spheres

The three ideas just presented refer to general $\vec{\imath}$. We can obtain more complete results when $\vec{\imath}$ represents the invariants of a Lie group W. In this case we can often give complete proofs to results which go beyond the polynomial (algebraic) theory.

We remarked at the beginning of this chapter that harmonics give a harmonious analysis to the orbits. In particular they define "the same" basis on all orbits. If we have some well-defined class $\mathcal{T}$ of functions on R^k we can study how restrictions to orbits of functions in this class vary with the orbits $\mathcal{O}$. We can measure this variation by the variation with $\mathcal{O}$ of the expansion coefficients of these functions in terms of the harmonic basis. We consider only the case when $\mathcal{T}$ is the space of restrictions to R^k of a well-defined space of entire functions.

Let W be the rotation group acting on $\mathbb{R}^k$. Let S be the real unit sphere so S is a generic orbit for W; θ is a convenient parameter on S. We want to study the harmonic expansion of functions in the space $\mathcal{T}$ on the spheres rS.

Since W acts on S it acts on functions on S so there is a representation of W on $L_2(S)$. The usual DP for the Laplacian Δ on R^k allows us to identify $L_2(S)$ with a space of functions which are harmonic inside S. Since, when acting on $\mathbb{R}^k$, W commutes with Δ, it preserves harmonicity. Moreover W commutes with scalar multiplication so it preserves the space of harmonics of any fixed degree p.

By decomposing the space of harmonics into homogeneous harmonics we have a decomposition of the representation of W on $L_2(S)$ into subrepresentations.

W clearly preserves the norm in $L_2(S)$ so all the representations we encounter in this way are unitary.

Each irreducible representation of W occurs once in $L_2(S)$. Moreover, the representation of W on homogeneous harmonics is irreducible when $k > 2$. For $k = 2$ the representation of W on homogeneous harmonics of degree $p > 0$ decomposes into two one-dimensional irreducible spaces spanned by

$$r^p e^{\pm ip\theta} = (t_1 \pm it_2)^p.$$

We are thus in the favorable situation where we can compute the harmonic expansion of a function on rS in terms of its decomposition into W irreducible components.

The classical theory of representations of W asserts that for $k > 2$ a basis for homogeneous harmonics is

$$r^p h_1^p(\theta), \ldots, r^p h_{d_p}^p(\theta) \tag{3.118}$$

where h_j^p are the spherical harmonics of degree p and d_p is the degree of the representation. In particular

$$(t_1, \ldots, t_k) = (rh_1^1(\theta), \ldots, rh_k^1(\theta)).$$

The spherical harmonics are normalized *à là* Schur, meaning

$$\|h_j^p\|^2 = \int_S |h_j^p(\theta)|^2 \, d\theta = \frac{1}{d_p}. \tag{3.119}$$

The difference between the Schur normalization and orthonormalization is insignificant for us because d_p is a polynomial in p, and our estimates are insensitive to polynomial factors.

We begin with $\mathcal{T}$ = the space $\hat{\mathcal{H}}'$ of entire functions of exponential type on $\mathbb{C}^k$. $\hat{\mathcal{H}}'$ is the Fourier transform of the dual of $\mathcal{H}$ which is the space of entire functions (see Chapter V of FA).

Remark. *We could approach this problem via tensor products as described in the remark following Theorem 3.19. That would entail using a quantitative version of the Korevaar–Wiegerinck theorem and Paley–Wiener theory. We prefer the present method as it seems to be more revealing.*

Theorem 3.14 *The Fourier series coefficients on each real sphere of radius r*

$$A_l^p(r) = \int A[rh_1^1(\theta), \ldots, rh_k^1(\theta)] \, \bar{h}_l^p(\theta) \, d\theta \tag{3.120}$$

of $A \in \hat{\mathcal{H}}'$ are characterized as being extendable to entire functions of exponential type of the complex variable r which form a bounded set in $\hat{\mathcal{H}}'[k = 1]$. Moreover the power series for $A_l^p(r)$ contains only terms of the form r^{p+2j}.

Remark. *The invariants for W are generated by*

$$r^2 = \sum t_j^2.$$

It is crucial for Theorem 3.14 that we express the functions A_l^p in terms of r rather than r^2 as the description in terms of r^2 is very complicated. In order to generalize theorems like Theorem 3.14 to other groups we must understand the meaning of r.

A cross-section of the orbits is provided, for example, by the t_1 axis with the proviso that t_1 and $-t_1$ correspond to the same orbit. Moreover the correspondence between t_1 and $-t_1$ comes from the group

$$W' = \{\pm \text{ identity}\}$$

which is a group of linear transformations on all of $\mathbb{R}^n$, acts on the t_1 axis, and commutes with W.

Thus we have a euclidean space (the t_1 axis) $\mathbb{R} \subset \mathbb{R}^n$ which is invariant under the action of the group W' which commutes with W. The orbits of W' on $\mathbb{R}$ form a cross-section for the action of W. r is defined as the euclidean coordinate on $\mathbb{R}$; it transforms under the nontrivial representation of W'. The Fourier series coefficients (3.120) are defined using the W orbit of r.

At the end of this section we shall discuss this setting in more detail and also how it extends to other groups.

Proof of Theorem 3.14 For each θ the integrand is an entire function of r and the convergence of the integral is uniform on compact sets of r in $\mathbb{C}^1$. Thus $A_l^p(r)$ are entire functions. Since the h_l^p are matrix coefficients of unitary representations, $|h_l^p| \leq 1$. Thus

$$|A_l^p(r)| \leq \max_{|t|=|r|} |A(t)| \leq ce^{c|r|} \tag{3.121}$$

which means that $\{A_l^p\}$ is a bounded set in $\hat{\mathcal{H}}'$.

To prove the statement on the vanishing power series coefficients observe that the power r^m occurring in $A_l^p(r)$ arises from monomials of the form $t_1^{m_1} \dots t_k^{m_k}$ with $m_1 + \cdots + m_k = m$ in the power series of the integrand $A(t)$ in (3.120). Now for $t_j = rh_j^1(\theta)$

$$t_1^{m_1} \dots t_k^{m_k} = r^m (h_1^1)^{m_1} \dots (h_k^1)^{m_k}. \tag{3.122}$$

The Clebsch–Gordon theory (see [159]) asserts that the product of the harmonics on the right side of (3.122) is a linear combination of harmonics of degrees m, $m-2, \dots$ on S (see Section 3.1). Such harmonics are orthogonal to h_l^p unless p is of the form $m - 2j$. Reading things backwards, $m = p + 2j$, which is our assertion.

This completes the proof of the direct part of Theorem 3.14.

For the converse we need the following observation. If an entire function $A_l^p(r)$ satisfies (3.121) and has vanishing power series coefficients of r^m for $m < p$ then actually

$$|A_l^p(r)| \leq c_0 \frac{(c_1|r|)^p}{p!} e^{c_1|r|} \tag{3.123}$$

where c_0, c_1 depend only on c.

Remark. *We can regard (3.123) as an analog for the space* $\hat{\mathcal{H}}'$ *of the classical Schwarz lemma which asserts that if* $f(z)$ *is holomorphic in* $|z| < 1$ *and* $|f(z)| \leq M$*, and* $f(0) = 0$*, then* $|f(z)| \leq M|z|$.

To verify this assertion we write

$$A_l^p(r) = \sum_{m \geq p} \frac{a_m r^m}{m!}.$$

A standard inequality for entire functions of exponential type is

$$|a_m| \leq c_0 c_1^m.$$

Thus

$$\begin{aligned}
|A_l^p(r)| &\leq \sum_{m \geq p} \frac{|a_m|\,|r|^m}{m!} \\
&\leq c_0 \sum_{m \geq p} \frac{(c_1|r|)^m}{m!} \\
&= c_0 \frac{(c_1|r|)^p}{p!} \sum (c_1|r|)^j \frac{p!}{(p+j)!} \\
&\leq c_0 \frac{(c_1|r|)^p}{p!} \sum \frac{(c_1|r|)^j}{j!} \\
&= c_0 \frac{(c_1|r|)^p}{p!} e^{c_1|r|}
\end{aligned}$$

which is our assertion.

We have used the trivial fact that $(p+j)! \geq p!j!$.

The $h_j^p(\theta)$ are matrix coefficients of unitary representations so are bounded by 1. Using (3.123) we deduce, for $t = [rh_1^1(\theta) \ldots, rh_k^1(\theta)]$,

$$\begin{aligned}
|A(t)| &= \left| \sum_{pl} d_p A_l^p(r) h_l^p(\theta) \right| \\
&\leq c_0 e^{c_1|r|} \sum_p d_p^2 \frac{(c_1|r|)^p}{p!} \\
&\leq c_0' e^{c_1'|r|}
\end{aligned} \tag{3.124}$$

where $c_1' = 2c_1 + \varepsilon$. d_p is the number of l for a given p; since d_p is polynomially bounded we can absorb it in c^p.

Actually we have verified (3.124) only for the points $t = [rh_1^1(\theta), \dots, rh_k^1(\theta)]$ with r complex and $\theta \in S$ (real unit sphere). These form $\mathbb{C}^{\mathbb{R}}$ which is the union of the complexifications of real lines through the origin; such lines are called *real complex lines*. A is holomorphic on $\mathbb{C}^{\mathbb{R}}$ (meaning on every complex line) and we want to extend A to an entire function on $\mathbb{C}^k$ which should be of exponential type. In our discussion of the Korevaar–Wiegerinck theorem in Section 1.5 we introduced the condition COMP ORIGIN, which in the present case is equivalent to the vanishing power series conditions on the $A_l^p(r)$. Thus A extends to an entire function on $\mathbb{C}^k$ and it remains to verify that it is of exponential type.

In what follows we shall use a somewhat different approach which does not appeal to the Korevaar–Wiegerinck theorem.

Another way of extending A to all of $\mathbb{C}^k$ is to employ the fact that $r^p h_l^p(\theta)$ are the restrictions to $\mathbb{C}^{\mathbb{R}}$ of homogeneous polynomials of degree p (the harmonic polynomials for the Laplacian) which we denote by $h_l^p(t)$. In particular $t = \{h_l^1(t)\}$. It follows that $r^{p+2j} h_l^p(\theta) = r^p h_l^p(\theta) r^{2j}$ are also polynomials in a natural way. (We hope there is no confusion in the notation: the p in h_l^p is a superscript while in r^p it is a power.) In this way we extend A to all of $\mathbb{C}^k$. It remains to make the proper estimates on the convergence of the series and the fact that the extended A is of exponential type.

Since $|h_l^p(\theta)| \le 1$ for θ real

$$|h_l^p(t)| = |r^p h_l^p(\theta)| \le r^p \quad r, t \text{ real}, \;\; r^2 = t^2. \tag{3.125}$$

We claim that this implies

$$|h_l^p(t)| \le (1 + |t|)^p \quad \text{for all } t. \tag{3.126}$$

Consider points of the form $t = (t_1, t_2^0, \dots, t_k^0)$ where $t_2^0, \dots, t_k^0$ are real. Such points form the complex line $t_2 = t_2^0, \dots, t_k = t_k^0$. Expression (3.125) gives the bounds (3.126) for h_l^p on the real points of this complex line. In the t_1 upper half-plane (meaning $\Im t_1 \ge 0$) we know that

$$\frac{h_l^p(t)}{|(t_1 + i|t_2^0| + \cdots + i|t_k^0|)|^p}$$

is bounded because h_l^p is a polynomial of degree p and

$$|(t_1 + i|t_2^0| + \cdots + i|t_k^0|)| \ge \left(\left|t_1\right| + \sum |t_j^0|\right)/\sqrt{2}.$$

Thus by the maximum modulus theorem (Phragmén–Lindelöf), (3.126) holds in the upper half-plane. Similarly it holds in the lower half-plane.

By the same argument the inequality is valid when all but one of the variables are real. Next we fix $t_1, \dots, t_k$ with all but two variables (say t_1, t_2) real

and let t_2 be arbitrary (complex). Then (3.126) holds for t_1 real and by the same Phragmén–Lindelöf argument for all t_1. (Our use of the Phragmén–Lindelöf argument does not require t_2^0 to be real.)

Proceeding in this manner we have established (3.126).

We can now complete the proof that $A \in \hat{\mathcal{H}}'$. The inversion of "Fourier series" on S gives

$$\begin{aligned}
|A(t)| &= d_p \Big| \sum A_l^p(r) h_l^p(\theta) \Big| \\
&= d_p \Big| \sum a_{lj}^p r^{p+2j} h_l^p(\theta) \Big| \\
&= d_p \Big| \sum a_{lj}^p r^{2j} h_l^p(t) \Big| \\
&\le c_2 d_p \sum \frac{c_3^{p+2j}}{(p+2j)!} |r|^{2j} |h_l^p(t)| \\
&\le c_2 d_p \sum \frac{(c_3|r|)^{2j}}{(2j)!} \frac{[c_3(1+|t|)]^p}{p!} \quad \text{by (3.126)} \\
&\le c_5 e^{c_6|t|} \qquad (3.127)
\end{aligned}$$

for all t since r^2 is the restriction of t^2 to $\mathbb{C}^{\mathbb{R}}$. (The fourth line follows from the fact that a_{lj}^p is the coefficient of r^{p+2j} in the power series expansion of $A_l^p(r)$.)

This completes the proof of Theorem 3.14.

For $k = 2$ we could prove the direct part of Theorem 3.14 by observing that $h_1^1(\theta) = \cos(\theta), h_2^1(\theta) = \sin\theta$ so

$$\begin{aligned}
r h_1^1(\theta) &= \frac{z + z^{-1}}{2} \\
r h_2^1(\theta) &= \frac{z - z^{-1}}{2i}
\end{aligned}$$

on $|z| = 1$, where z is a complex variable. We then shift the contour from $|z| = 1$ to $|z| = \lambda_p$ where λ_p is suitably chosen. In this way we derive (3.123) which, as we have seen, is the crucial inequality for establishing the direct part of Theorem 3.14.

This is the method that was employed by Helgason in [94], [95] in proving Theorem 3.14 for $k = 2$. Unfortunately this method does not seem to be extendable to $k > 2$.

Another method that works for $k = 2$ starts with

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \qquad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}.$$

We can expand $t_1^{m_1} t_2^{m_2}$ in terms of $r^{m_1+m_2} e^{i(m_1+m_2-2j)\theta}$ using the binomial theorem. Using estimates on binomial coefficients leads again to (3.123).

It is possible to extend this method to $k > 2$. The main point is the combinatorial decomposition of powers $t_1^{m_1} \dots t_k^{m_k}$ into $r^{m_1+\dots+m_k} h_{m-2j}(\theta)$ where h_{m-2j}

are the spherical harmonics. This amounts to an explicit Clebsch–Gordon decomposition of symmetric products of representations of the rotation group. Such technique started with the work of Rakkah (see [21]). (The binomial coefficients become Rakkah $3-j$ coefficients.)

Remark. *There is a certain aspect of Theorem 3.14 and the analogous results Theorems 3.15 and 3.16 which seems puzzling at first sight. On each real sphere $r = \text{const.}$ the functions A are real analytic so we should expect an exponential decrease in p in the Fourier coefficients A_l^p. Yet Theorem 3.14 seems to impose only a boundedness condition on these coefficients. The point is that (3.123) shows that the vanishing conditions imply exponential decrease in p. (Actually they imply decrease of order $(p!)^{-1}$ which is the expected order for entire functions of exponential type.)*

An allied problem concerns expansions in various coordinate systems. For example, if we use spherical coordinates in $\mathbb{R}^3$ then we could search for expansions in terms of

$$r^p \cos n\theta$$
$$r^p \sin n\theta \cos m\varphi$$
$$r^p \sin n\theta \sin m\varphi$$

instead of harmonics.

The most interesting problem of this type involves characterizing expansions in confocal coordinates in $\mathbb{R}^3$. This leads to deep questions concerning Mathieu and Lamé functions which we shall not pursue.

Actually, our main interest is in the space $\mathcal{E}$ of C^∞ functions on $\mathbb{R}^k$ rather than $\mathcal{H}$.

Theorem 3.15 *The Fourier coefficients B_l^p of $B \in \hat{\mathcal{E}}'$ are characterized by being extendable to entire functions of exponential type such that $\{B_l^p(r)\}$ is bounded in $\hat{\mathcal{E}}'[k=1]$. Moreover the power series for $B_l^p(r)$ contains only terms of the form r^{p+2j}.*

Proof Since $\hat{\mathcal{E}}' \subset \hat{\mathcal{H}}'$ Theorem 3.14 shows that the B_l^p are entire functions of exponential type and the condition on power series coefficients holds. Since $B \in \hat{\mathcal{E}}'$ we have

$$|B(\hat{t})| \leq c(1+|\hat{t}|)^c e^{c|\Im \hat{t}|}$$

so we derive from (3.120) the uniform inequality

$$|B_l^p(r)| \leq c_0(1+|r|)^c e^{c|\Im r|}.$$

This says that $\{B_l^p\}$ is a bounded set in $\hat{\mathcal{E}}'[k=1]$.

For the converse, we know from Theorem 3.14 that

$$B(r,\theta) = B[rh_1^1(\theta), \ldots, rh_k^1(\theta)] = \sum d_p B_l^p(r) h_l^p(\theta) \tag{3.128}$$

on $\mathbb{C}^{\mathbb{R}}$ extends to an entire function $B(t)$ of exponential type on $\mathbb{C}^k$. Moreover for any fixed real r the series (3.128) converges in the space of distributions $\mathcal{S}'(S)$ on S because the Fourier coefficients $B_l^p(r)$ are bounded in p and d_p is polynomially bounded. (It is standard that Fourier series coefficients in $\mathcal{S}'(S)$ are characterized by being bounded by p^c.)

Call $B_r(\theta)$ the restriction of (3.128) to the real sphere S_r of radius r. By (3.128) and the above inequality on $B_l^p(r)$ we see that $\{B_r(\theta)(1+|r|)^{-c}\}_r$ is a bounded set in $\mathcal{S}'(S)$. This means that $B(r,\theta)$ defines a distribution $\tilde{B}$ in $\mathcal{S}'(\mathbb{R}^k)$.

We conclude that $B(t)$ is an entire function of exponential type whose restriction to $\mathbb{R}^k$ belongs to $\mathcal{S}'(\mathbb{R}^k)$. The Paley–Wiener–Schwartz theorem allows us to conclude that $B(t) \in \hat{\mathcal{E}}'$, which completes the proof of Theorem 3.15.

Let us clarify the last point. If we form the convolution $\tilde{B} * f$ on $\mathbb{R}^k$ with $f \in \mathcal{D}$ then $\tilde{B} * f$ extends to an entire function of exponential type which is bounded by $c'(1+|t|)^{c'}$ on $\{t \text{ real}\}$. Thus for any $\hat{g} \in \hat{\mathcal{D}}$ the function $\hat{g}(\tilde{B} * f)$ is rapidly decreasing on $\mathbb{R}^k$ and is an entire function of exponential type and hence belongs to $\hat{\mathcal{D}}$. We conclude that multiplication by $(\tilde{B} * f)$ maps $\hat{\mathcal{D}} \mapsto \hat{\mathcal{D}}$. It follows that $\tilde{B} * f \in \hat{\mathcal{E}}'$ by the results of Chapter V of FA. Thus $\tilde{B}$ maps $\mathcal{D} \mapsto \hat{\mathcal{E}}'$ by convolution. Again this implies that $\tilde{B} \in \hat{\mathcal{E}}'$.

It should be pointed out that the fact that an entire function of exponential type which is rapidly decreasing on $\mathbb{R}^k$ is in $\hat{\mathcal{D}}$ is proven by the same method we used to derive (3.126) from (3.125).

Remark. *It is interesting to observe that the condition* $\{B_l^p(r)\}$ *is bounded in* $\hat{\mathcal{E}}'[k=1]$ *is equivalent to the boundedness of*

$$\{(1+p)^q B_l^p(r)\}$$

for any fixed q. *This is easily seen from our proof of the converse part of Theorem 3.15, for the convergence of the series (3.128) in the space* $\mathcal{S}'$ *is insensitive to the factor* $(1+p)^q$.

It is somewhat more interesting to understand the irrelevance of $(1+p)^q$ from the viewpoint of the function $B(t)$. Let us first illustrate what happens for $n=2$. Observe that, for any q (we call $B_{\pm}^p(r)$ the Fourier series coefficients of B),

$$(\pm ip)^q B_{\pm}^p(r) = \int \left[\frac{\partial^q}{\partial\theta^q} B(r\cos\theta, r\sin\theta)\right] e^{\mp ip\theta} d\theta. \tag{3.129}$$

Note that

$$\frac{\partial}{\partial\theta} = t_2\frac{\partial}{\partial t_1} - t_1\frac{\partial}{\partial t_2}. \tag{3.130}$$

To estimate $\partial^q B/\partial\theta^q$ we use the fact that $\hat{B}$ is a distribution in $\hat{t}$ of order $\leq c$ whose support is in the square of center origin, side $2c$. By (3.130) and the fact that the Fourier transform interchanges multiplication by coordinate functions

with differentiation the Fourier transform of $\partial^q B/\partial\theta^q$ has order $\leq c+q$. There are also factors of powers of $\hat{t}_1, \hat{t}_2$ of order $\leq q$. Any such power can be estimated by $1+c^q$ on the support of $\hat{B}$ and the derivatives of order $\leq c+q$ of such powers can be estimated by $c_{B_q}(1+q^q)$ where c_{B_q} is a constant depending only on B, q.

Taking the inverse Fourier transform yields the estimate

$$\left|\frac{\partial^q B}{\partial\theta^q}(t_1,t_2)\right| \leq c'_{B_q}(1+q^q)(1+|t_1|+|t_2|)^{c+q}e^{c|\Im t|}.$$

Combining this with (3.129) shows that factors of order $(1+p)^q$ for $q \geq 0$ in estimates on the $B^p_\pm$ introduce, at worst, polynomial factors and hence do not play any essential role, which verifies our assertion for $k=2$.

The proof for $k>2$ goes along similar lines except that we replace $\partial/\partial\theta$ by the Laplacian Δ_θ on the sphere and use the observation that the $h^p_l(\theta)$ are eigenfunctions of Δ_θ whose eigenvalue is quadratic in p.

All the above indicates that we can regard

$$\hat{\mathcal{E}}' = \mathcal{S}'(\theta)\underset{\sim}{\otimes}\mathcal{E}'[k=1]. \tag{3.131}$$

The identification is made using Fourier series in terms of spherical harmonics on S. The subnature of the tensor product is the vanishing of the power series coefficients. We shall see below that the isomorphism of $\hat{\mathcal{E}}'$ with this (sub) tensor product is valid topologically.

The discussion following Theorem 2.12 shows that

$$\mathcal{E} = \mathcal{S}(\theta)\underset{\sim}{\otimes}\mathcal{E}[k=1].$$

The subnature of this tensor product is the compatibility at the origin, i.e. COMP ORIGIN. Thus

$$\mathcal{E}' = \mathcal{S}'(\theta)\underset{\sim}{\otimes}\mathcal{E}'[k=1].$$

$\mathcal{E}'$ is a supra tensor product; the relations in this supra tensor product come from identities in powers of directional derivatives at the origin, e.g. for $n=2$

$$\partial_{x+y}\delta_0 = \frac{\partial}{\partial x}\delta_0 + \frac{\partial}{\partial y}\delta_0$$

where ∂_{x+y} is the directional derivative in the direction with direction numbers (1,1). In terms of tensor products this is the identity

$$\delta_{\theta=\pi/4}\otimes\frac{\partial}{\partial x}\delta_0 = \delta_{\theta=0}\otimes\frac{\partial}{\partial x}\delta_0 + \delta_{\theta=\pi/2}\frac{\partial}{\partial x}\delta_0.$$

Nevertheless the Fourier transform of $\mathcal{E}'$ is a sub tensor product, which seems strange. The point is that the Fourier transform kills elements of the actual tensor product which are zero in the supra tensor product because the kernel of the Fourier transform is a smooth function. Moreover the Fourier transform of

functions of r times spherical harmonics involves Bessel functions, which vanish at $r = 0$. This vanishing, which is a consequence of Schur orthogonality, i.e. the orthogonality of spherical harmonics of different degrees, is the basis for the sub tensor nature of $\hat{\mathcal{E}}'$. It is exploited in Chapter 7 in our study of discrete series.

We want to study the topology of $\hat{\mathcal{E}}'$ in terms of $\{B_l^p(r)\}$. The original derivation of the AU topology of $\hat{\mathcal{E}}'$ as given in Chapter 5 of FA uses the fact that the topology of $\hat{\mathcal{E}}'$ can be described as the compact–open topology of maps

$$G \to BG$$

of $\hat{\mathcal{D}}_F \to \hat{\mathcal{D}}_F$ for $B \in \hat{\mathcal{E}}'$. ($\mathcal{D}_F$ is the same as $\mathcal{D}$ elementwise but topologically $\mathcal{D}_F$ is the space whose dual consists of distributions of finite order.) In order to translate the map topology into AU topology we need to have direct knowledge of the topology of $\hat{\mathcal{D}}_F$. This space is much easier to deal with than $\hat{\mathcal{E}}$ because the inverse Fourier transform on $\hat{\mathcal{D}}_F$ can be defined directly.

Theorem 3.16 *The Fourier coefficients $G_l^p(r)$ of $G \in \hat{\mathcal{D}}_F$ are characterized as being entire functions of exponential type belonging to $\hat{\mathcal{D}}_F[k = 1]$ such that for any q*

$$\{(1 + p)^q G_l^p\}$$

is bounded in the space $\hat{\mathcal{D}}_F[k = 1]$. Moreover the only nonvanishing power series terms in G_l^p occur for r^{p+2j}. The topology of $\hat{\mathcal{D}}_F$ is given by the semi-norms

$$\|\{G_l^p\}\|_{k_1 N} = \max \frac{|G_l^p(r)|(1 + p)^N}{k_1(r)}$$

where $\{k_1\}$ is an AU structure for $\hat{\mathcal{D}}_F[k = 1]$.

(We have, inadvertently, used the letter k for dim $\{t\}$ and for the AU semi-norms.)

The proof of the characterization of $\{G_l^p\}$ is essentially the same as for $\hat{\mathcal{E}}'$. Instead of convergence of (3.128) in $\mathcal{S}'$ we now have convergence in $\mathcal{S}$ and the estimate $\mathcal{O}[(1+|r|)^{-m} \exp(c|\Im r|)]$ for all m replaces $\mathcal{O}[(1+|r|)^c \exp(c|\Im r|)]$; this suffices for the characterization of $\{G_l^p\}$.

To analyze the topology we study the Fourier series decomposition of $\mathcal{D}_F$ (which is easier than $\hat{\mathcal{D}}_F$). It is important to observe that $\mathcal{D}_F$ is a *convergent direct sum*

$$\mathcal{D}_F = \sum \mathcal{D}_F(p, l). \tag{3.132}$$

Here $\mathcal{D}_F(p, l)$ is the space of functions in $\mathcal{D}_F[k = 1]$ whose Taylor series coefficients at zero satisfy the usual vanishing conditions, i.e. only r^{p+2j} appear. ($\mathcal{D}_F(p, l)$ does not depend on l.) We used the term "convergent direct sum"

to indicate that only sequences $\{g_l^p\} \in \{\mathcal{D}_F(p,l)\}$ satisfying the boundedness condition

$$\{(1+p)^q g_l^p\} \text{ is bounded in } \mathcal{D}_F[k=1] \quad \text{for any } q$$

appear as coefficients. The identity (3.132) is defined by associating to any $g \in \mathcal{D}_F$ the series

$$g = \sum g_l^p(r) h_l^p(\theta)$$

where $g_l^p(r)$ are the Fourier series coefficients of g on rS. Moreover the topology of the space of $\{g_l^p(r)\}$ which makes $g \longleftrightarrow \{g_l^p(r)\}$ a topological isomorphism is given by semi-norms

$$\|\{g_l^p(r)\}\|_{\kappa,N} = \max(1+p)^N \|g_l^p\|_\kappa$$

where $||\ ||_\kappa$ are semi-norms for $\mathcal{D}_F[k=1]$. This characterization of the topology of $\mathcal{D}_F$ and its topology are easily verified.

Put in other terms, $\mathcal{D}_F$ is the space of continuous maps of $\mathcal{S}(S)$ into $\mathcal{D}_F[k=1]$ with compatibility at the origin; we can regard this as the sub tensor product $\mathcal{D}_F[k=1] \underset{\sim}{\otimes} \mathcal{S}(S)$. The topology of $\mathcal{D}_F$ is the compact–open topology of this space of maps.

We now take the Fourier transform of (3.132). There are two issues that must be addressed:

(1) The topology of $\mathcal{D}_F(p,l)$ for fixed p,l.
(2) The uniformity in p,l of the norms.

To deal with these issues we use the fact that the k-dimensional Fourier transform of $\mathcal{D}_F(p,l)h_l^p$ is a Bessel transform of $\mathcal{D}_F(p,l)$ [11, vol. II, p. 73]. We then have to establish the Paley–Wiener theory for the Bessel transforms, with uniformity in p,l.

The only major difference in the Paley–Wiener theory for the Bessel transforms and for the Fourier transform is that the Bessel function satisfies a functional relation involving r and $-r$. This means that there is no vanishing at $i\infty$ as there is for $\exp(ir)$ or at $-i\infty$ as there is for $\exp(-ir)$. It is this vanishing which is crucial for the usual proofs of the Paley–Wiener theorem.

For the simpler case of the cosine transform we can establish Paley–Wiener theory by use Euler's identity

$$\cos x = \tfrac{1}{2}(e^{ix} + e^{-ix})$$

thereby reducing to Fourier transform. In [68] we met similar problems for the Legendre transform. This was overcome by using an Euler-like formula relating the Legendre function P_u to the associated Legendre functions Q_u and Q_{1-u}. For the usual J_ν Bessel functions the role of Q_u is played by Bessel functions of the third kind $H_\nu^{(1)}$ and $H_\nu^{(2)}$ [11, vol. II, p. 4]. These have the required asymptotic vanishing at infinity. The relation (7) on p. 4 of [11, vol. II] serves as

the analog of Euler's identity. (This is for Bessel functions corresponding to $k = 2$; there are analogous functions for $k > 2$.) Using this Euler-like decomposition the calculations assume the same character as in the case of the ordinary Paley–Wiener theory. We leave the details of this argument to the reader. (See the remark following Theorem 3.19.)

We can now proceed, as in FA, using the fact that the topology of $\hat{\mathcal{E}}'$ is the compact–open topology of multiplication maps of $\hat{\mathcal{D}}_F \to \hat{\mathcal{D}}_F$ to determine the topology of $\{B_l^p(r)\}$ for $B \in \hat{\mathcal{E}}'$ which makes $B \longleftrightarrow \{B_l^p(r)\}$ a topological isomorphism. The vanishing Taylor coefficients and the fact that we have a sequence instead of a single function present no difficulties.

We find

Theorem 3.17 *The topology of* $\hat{\mathcal{E}}' = \{B_l^p(r)\}$ *is given by the semi-norms*

$$\|\{B_l^p(r)\}\| = \max \frac{|B_l^p(r)|}{k_2(r)\tilde{k}(p)}$$

where $\{k_2\}$ *is an AU structure for* $\hat{\mathcal{E}}'[k = 1]$ *and* $\tilde{k}(p)$ *is larger at infinity than* $|p|^s$ *for any* s.

Our constructions have profound consequences for harmonic expansions and for Radon transforms.

Concerning harmonic expansions, observe that $[A_l^p(r)]r^{-p}$ is a function of r^2, i.e. is an invariant for the orthogonal group W, and $r^p h_l^p(\theta)$ is (the restriction to $\mathbb{C}^{\mathbb{R}}$ of) a homogeneous harmonic.

Theorem 3.18 *For* $A \in \hat{\mathcal{H}}'$ *the harmonic expansion*

$$A(t) = \sum d_p[A_l^p(r)r^{-p}]r^p h_l^p(\theta)$$

converges in the topology of $\hat{\mathcal{H}}'$. *Similar results hold for the spaces* $\hat{\mathcal{E}}', \hat{\mathcal{D}}_F$.

Proof We have noted in the remark preceding (3.129) that the boundedness of $\{B_1^p(r)\}$ in $\hat{\mathcal{E}}'[k = 1]$ is the same as the boundedness of $\{(1 + p)^q B_1^p(r)\}$ for any q. Since $|h_l^p(\theta)| \leq 1$ it follows that the boundedness of $\{B_1^p(r)\}$ implies the convergence in the topology of $\hat{\mathcal{E}}'$ for the harmonic expansion given in Theorem 3.18 (for B in place of A). The same is true for the spaces $\hat{\mathcal{H}}'$, $\hat{\mathcal{D}}_F$ by Theorems 3.14, 3.15, and 3.16.

The validity of Theorem 3.18 is established.

We have seen in Chapter 2 that the Fourier transform of the hyperplane Radon transform involves, directly, only $\mathbb{C}^{\mathbb{R}}$ rather than all of $\mathbb{C}^k$. Note that the construction of, and bounds on, A_l^p, B_l^p, G_l^p depend only on the respective functions A, B, G on $\mathbb{C}^{\mathbb{R}} = \cup_{z \in \mathbb{C}} zS$. Theorems 3.16 and 3.17 show that the AU topologies of $\{B_l^p\}, \{G_l^p\}$ define the respective topologies of $\hat{\mathcal{E}}', \hat{\mathcal{D}}_F$. The same is true for the space $\hat{\mathcal{H}}'$; in fact the proof is easier. This establishes the important result.

Theorem 3.19 $\mathbb{C}^{\mathbb{R}}$ *is sufficient for the spaces* $\mathcal{H}, \mathcal{E}, \mathcal{D}_F$.

Remark. *Theorem 3.19 for the spaces* $\mathcal{D}_F$ *and* $\mathcal{E}$ *can be understood from the following viewpoint. These spaces are sub tensor products* $\mathcal{D}_F[k=1]\underset{\sim}{\otimes}\mathcal{S}[S]$ *and* $\mathcal{E}[k=1]\underset{\sim}{\otimes}\mathcal{S}[S]$ *respectively. Here* S *is the unit sphere and the subnature of the tensor product refers to the compatibility at the origin. It is "reasonable," therefore, that the topologies of the duals of these spaces should exhibit this tensor product structure, and this is exactly Theorem 3.19.*

The argument for $\mathcal{H}$ *is similar in view of the Korevaar–Wiegerinck theorem (made slightly more precise so as to include the topology).*

In Chapter 5 of FA we proved that the union of the real and imaginary axes is sufficient for $\mathcal{E}(\mathbb{R}^1)$ and for $\mathcal{D}_F(\mathbb{R}^1)$. This means that we can replace "real complex lines" in Theorem 3.19 by "union of real and imaginary axes in each real complex line." Such an axis is of the form $\lambda(\alpha_1, \ldots, \alpha_m)$ where α_j are real and λ is either real or imaginary. The axis is thus either pure real or pure imaginary. $\mathbb{C}^{\mathbb{R}}$ is replaced by the union of the pure real and pure imaginary "axes" in $\mathbb{C}^k$.

We have established[5]

Theorem 3.20 *The union of the real and imaginary subspaces of* $\mathbb{C}^k$ *is sufficient for* $\mathcal{E}$ *and for* $\mathcal{D}_F$.

This answers Problem 5.2 of FA.

The corresponding result for $\mathcal{D}$ is false because the semi-norms for $\hat{\mathcal{D}}$ and $\hat{\mathcal{D}}_F$ agree there; that is, they dominate exponentials on the imaginary subspace and they are of the form $(1+|\hat{x}|)^{-N}$ on the real subspace but the topologies of $\mathcal{D}$ and $\mathcal{D}_F$ are different.

It is possible to sharpen Theorem 3.19 somewhat. By duality Theorem 3.19 implies (see Section 1.4) that any f in any of the spaces $\mathcal{H}, \mathcal{E}, \mathcal{D}_f$ has a Fourier representation

$$f(t) = \int e^{it\cdot\hat{t}} \frac{d\mu(\hat{t})}{k(\hat{t})}$$

where μ is a bounded measure on $\mathbb{C}^{\mathbb{R}}$ and k belongs to an AU structure for the appropriate space.

Let us concentrate on the space $\mathcal{E}$. For certain applications in Chapter 4 it is important to be able to choose μ as a smooth measure; that is, $\mu(\hat{t}) = \hat{f}(\hat{t})d\hat{t}$ where $\hat{f} \in \mathbb{C}^{\mathbb{N}}$ for suitable N (actually we can take $N = \infty$) and

$$d\hat{t} = d\hat{\theta}d\hat{r}.$$

$\hat{\theta}$ is the coordinate on the real unit sphere and $d\hat{r}$ is the area measure on the complex line. (Actually only the smoothness in $\hat{\theta}$ is important for our applications.)

[5]This theorem was proved for the first time by B. Alan Taylor by a different method (personal communication).

The improvement in the structure of μ depends on the remarks following the proof of Theorem 3.15. We explained that polynomial factors in p are irrelevant for the description on the Fourier series of functions in $\hat{\mathcal{E}}'$. This has the consequence (3.131). The topology is the topological tensor product topology; that is, we think of $\mathcal{S}'(\hat{\theta}) \otimes \hat{\mathcal{E}}'[k=1]$ as the space of maps of $\mathcal{S}(\hat{\theta})$ into $\hat{\mathcal{E}}'[k=1]$ and give this space the compact–open topology.

The dual of $\mathcal{S}'(\hat{\theta}) \otimes \hat{\mathcal{E}}'[k=1]$ is $\mathcal{S}(\hat{\theta}) \otimes \hat{\mathcal{E}}[k=1]$ so the dual $\hat{\mathcal{E}}$ of $\mathcal{S}'(\hat{\theta}) \underset{\sim}{\otimes} \mathcal{E}'[k=1]$ is $\mathcal{S}(\hat{\theta}) \underset{\sim}{\otimes} \mathcal{E}[k=1]$. In particular any $\hat{f} \in \hat{\mathcal{E}}$ can be represented in the form

$$\hat{f}(\hat{r}, \hat{\theta}) = \int \hat{f}_{\hat{\theta}}(\hat{r})\, d\hat{\theta}.$$

$\hat{f}_{\hat{\theta}}$ depends smoothly on $\hat{\theta}$ and the integral is over the real unit sphere.

For each $\hat{\theta}$, $\hat{f}_{\hat{\theta}}$ is an element of $\hat{\mathcal{E}}[k=1]$. In FA and Section 1.5 we described elements of $\hat{\mathcal{E}}$ by measures because we described $\hat{\mathcal{E}}'$ as an AU space, meaning a space defined by uniform norms. It is a standard consequence of Cauchy's integral formula that, for holomorphic functions, uniform norms can be replaced by norms in the dual of the space C^N of N times differentiable functions.

This suggests that (by duality) the $\hat{f}_{\hat{\theta}}(\hat{r})$ can be chosen as C^N functions in $\hat{r}$ which are uniformly bounded by $k(\hat{r})^{-1}$ for some k in an AU structure for $\mathcal{E}'(k=1]$.

Actually this last remark needs some verification; one has to check that application of Cauchy's integral formula does not distort the AU growth conditions. This is easy to verify because the growth conditions do not vary quickly. (In FA we showed how the slowly varying notion of AU norms is at the basis of division theorems.)

Theorem 3.21 *Any $f \in \mathcal{E}$ has a Fourier representation*

$$f(t) = \iint \hat{f}(\hat{r}, \hat{\theta}) e^{it\cdot\hat{t}}\, d\hat{r}\, d\hat{\theta}$$

where $\hat{f}$ can be chosen in C^N for any N. We can also choose $\hat{f}$ to vanish in a fixed neighborhood of 0 in $\hat{r}$.

The space $\mathcal{E}$ can be replaced by $\mathcal{H}$ or $\mathcal{D}_f$.

Everything is proven except for the vanishing of $\hat{f}$ in a neighborhood of $\hat{r} = 0$. This follows from the fact that the complement of a compact neighborhood of the origin is sufficient for $\mathcal{E}, \mathcal{H}$, and $\mathcal{D}_F$.

We write

$$\hat{t} = \hat{r}\hat{t}_{\hat{\theta}}$$

where $\hat{t}_{\hat{\theta}}$ is the unit vector in the (real) direction of $\hat{t}$. We can rewrite the Fourier integral in Theorem 3.21 as

$$f(t) = \int d\hat{\theta} \int \hat{f}(\hat{r}, \hat{\theta}) e^{i\hat{r}t\cdot\hat{t}_{\hat{\theta}}}\, d\hat{r}.$$

For each $\hat{\theta}$, $t \cdot \hat{t}_{\hat{\theta}}$ is constant in the t directions orthogonal to $\hat{\theta}$; that is, it is a spread function for $\hat{\theta}$, and hence the inner integral is a spread function $f_{\hat{\theta}}$ for $\hat{\theta}$.

We have proven

Theorem 3.22 *Any $f \in \mathcal{E}$ has a Radon decomposition*

$$f(t) = \int f_{\hat{\theta}}\, d\hat{\theta}$$

where $f_{\hat{\theta}}$ is a spread function corresponding to the point $\hat{\theta}$ in the real unit sphere.

The same result is true for the spaces $\mathcal{H}, \mathcal{D}_F$.

In Chapter 2 we explained the relation of $\mathbb{C}^{\mathbb{R}}$ to the hyperplane Radon transform. What happens for $\mathbf{R}^{k-l}$?

For the hyperplane Radon transform we used coordinates r, θ. According to the ideas of Chapter 2 we can think of r as a coordinate on $\mathbb{R}^1$ which is the natural Fourier transform of a spread parameter of a hyperplane spread; θ is the coordinate on the unit sphere in $\mathbb{R}^k$. We can think of the unit sphere in $\mathbb{R}^k$ as the unit sphere in the orthogonal complement of the origin, the origin being a hyperplane in $\mathbb{R}^1$.

This suggests introducing analogous coordinates r, θ for $\mathbb{R}^k$ as follows.

Pick a fixed real plane L of dimension l with coordinate r. Let L_0 be a hyperplane of L (both L, L_0 containing the origin). θ is a coordinate on the unit sphere in $L_0^{\perp}$ which is the orthogonal complement of L_0 in $\mathbb{R}^k$.

r, θ are a form of cylindrical coordinates; they express the rotation of a point $r \in L$ by the amount θ in the $L_0^{\perp}$ plane.

For example, let $k = 3, l = 1$; L is the t_2, t_3 plane and L_0 is the t_3 axis. We can think of the θ coordinate as defining a rotation of the t_2, t_3 plane about the t_3 axis. Thus if $r = (\tilde{r}, \varphi)$ represents polar coordinates in the t_2, t_3 plane then $(r, \theta) = (\tilde{r}, \varphi, \theta)$ represents spherical coordinates in $\mathbb{R}^3$ with φ being the latitude and θ the longitude.

For the analog of $\mathbb{C}^R$ we form the complex l plane $L^{\mathbb{C}}$ with complex coordinate r. θ is a real coordinate on the real unit sphere in $\mathbb{L}_0^{\perp}$. Note that the real dimension of $\{(L^{\mathbb{C}}, \theta)\}_{\theta}$ is

$$2l + (k - l) = l + k < 2k$$

so we are in a somewhat similar situation to that described in case $l = 1$ except that we are not using all real l planes but only those obtained from L by rotation in $L_0^{\perp}$. (We showed in Chapter 2 that if we use all l planes, then we cover all of $\mathbb{C}^k$.)

If we choose an L_0 of codimension >1 then there are more rotations and the exact description of the compatibility (vanishing) conditions (analog of our above result asserting the vanishing of coefficients of r^j unless $j = p + 2a$) seems complicated. If L_0 is the origin then we obtain all l planes by rotation of L. We have not computed the compatibility conditions.

We leave the development of these ideas to the reader.

For the space $\mathcal{H}$ we can prove a stronger sufficiency result than those given by Theorems 3.19 or 3.20. Theorem 3.19 does not use all complex lines but only those with pure real direction numbers. We can, in fact, restrict ourselves to any open set of such lines. To prove this, we want to verify (3.124) using only an open set of θ. We write the proof for two dimensions to avoid complications in notation; the general result follows exactly along the same lines.

We start with

$$\sum A^p(r)e^{ip\theta} = A(r,\theta). \tag{3.133}$$

We want to express $A^p(r)$ in terms of $A(r,\theta)$, say for $\theta \in [-\pi/4, \pi/4]$. We cannot use the usual Fourier formula because that involves integration over $\theta \in [-\pi, \pi]$. In fact we cannot use (3.133) directly to compute $A^p(r)$ in terms of $A(r,\theta)$ for $\theta \in [-\pi/4, \pi/4]$ because the $\{\exp(ip\theta)\}$ are not linearly independent in $[-\pi/4, \pi/4]$.

What saves us is the condition that the power series for $A^p(r)$ involves only $r^{|p|+2j}$. Using the idea of the calculations following (3.123) we write

$$\begin{aligned} A(r,\theta) = \sum A^p(r)e^{ip\theta} &= \sum_{j,p} \frac{\alpha_p^{|p|+2j}}{(|p|+2j)!} r^{|p|+2j} e^{ip\theta} \\ &= \sum_n \sum_j \frac{\alpha_p^n}{n!} r^n e^{\pm i(n-2j)\theta} \end{aligned} \tag{3.134}$$

where $p = \pm(n-2j)$. For any given n the sum is finite since $|p| + 2j = n$ and $j \geq 0$ so $j \leq n/2$. (The superscripts $|p| + 2j$ and n of α_p are not powers.)

Since the sum is finite, the exponentials are linearly independent over any interval. We can, therefore, find linear functions $\lambda_{j_0}^n$ which are measures supported in $[-\pi/4, \pi/4]$ such that

$$\int \lambda_{j_0}^n(\theta) e^{\pm i(n-2j)\theta} = \begin{cases} 1, & \text{if } j = j_0 \text{ and } + \\ 0, & \text{otherwise.} \end{cases} \tag{3.135}$$

Any "reasonable" norm of $\lambda_{j_0}^n$ will provide an estimate for α_p^n in terms of the values of $A(r,\theta)$ for $\theta \in [-\pi/4, \pi/4]$. (Actually $\lambda_{j_0}^n$ depends on whether $p > 0$ or $p < 0$, so we should write ${}_{\pm}\lambda_{j_0}^n$. But we consider only the case $p \geq 0$ as $p < 0$ is treated similarly.)

How is $\lambda_{j_0}^n$ to be constructed? If we regard $\lambda_{j_0}^n$ as a function on $\theta \in (-\infty, \infty)$ with support in $[-\pi/4, \pi/4]$, then its Fourier transform $\hat{\lambda}_{j_0}^n$ is an entire function of exponential type $\pi/4$. Equation (3.135) is equivalent to

$$\hat{\lambda}_{j_0}^n((n-2j)) = \begin{cases} 1, & \text{if } j = j_0 \text{ and } + \\ 0, & \text{otherwise.} \end{cases} \tag{3.136}$$

The usual construction of such functions starts with $\sin \pi\hat{\theta}/4$. We must modify $\sin \pi\hat{\theta}/4$ so that it satisfies (3.136). Now, $\sin \pi\hat{\theta}/4$ vanishes at $\{4k\}_{k\in\mathbb{Z}}$.

Thus we multiply it by a polynomial to make it vanish at the points in (3.136). The polynomial depends on the parity of p; suppose for definiteness that p (hence n) is even and >0. Then we form

$$[(\hat{\theta}^2 - 2^2)(\hat{\theta}^2 - 6^2) \dots (\hat{\theta}^2 - n^2) \sin \pi\hat{\theta}/4]/[\hat{\theta} - (n - 2j_0)]. \tag{3.137}$$

(The last factor $(\hat{\theta}^2 - n^2)$ is not present if n is divisible by 4.)

This function is not good enough for our purposes because it grows too fast at infinity. We divide out about $n/2$ zeros of $\sin \pi\hat{\theta}/4$ so as to make it bounded. Thus we form

$$\hat{\lambda}^n_{j_0}(\hat{\theta}) = c \frac{(\hat{\theta}^2 - 2^2) \dots (\hat{\theta}^2 - n^2)}{[\hat{\theta}^2 - (n+2)^2] \dots [\hat{\theta}^2 - (2n)^2][\hat{\theta} - (n - 2j_0)]} \sin \pi\hat{\theta}/4. \tag{3.138}$$

$c = c_{p_0}$ is chosen so as to make $\hat{\lambda}^n_{j_0}(n - 2j_0) = 1$. (This formula assumes n is not divisible by 4; it is clear how to write an analogous formula if n is divisible by 4.)

We have to estimate c from above. Suppose that $n - 2j_0 \equiv 0 \bmod 4$; the case $n - 2j_0 \equiv 2 \bmod 4$ requires simple modifications.

We have

$$\frac{\sin \pi\hat{\theta}/4}{\hat{\theta} - (n - 2j_0)}\Big|_{\hat{\theta}=n-2j_0} = \frac{1}{4}.$$

The polynomial factors of the numerator and denominator in (3.138) can be factored into $2([n/4] + 1)$ linear factors. When $\hat{\theta} = n - 2j_0$ it may happen that some pairs of factors in the numerator or in the denominator have the same absolute value. The other absolute values change by at least 2 and at most 4; they are all ≥ 2. Thus the numerator is bounded from below by

$$c_0^n \left[\frac{n}{2}\right]!$$

and the denominator is bounded from above by

$$c_1^n \left[\frac{n}{2}\right]!.$$

We conclude

$$c \leq c_2^n. \tag{3.139}$$

The same type of argument shows that, in fact,

$$\max_{\hat{\theta} \text{ real}} |\hat{\lambda}^n_{j_0}(\hat{\theta})| \leq c_3^n. \tag{3.140}$$

By a similar argument we could modify $\hat{\lambda}^n_{j_0}$ and obtain an analogous function for which the max norm is replaced by the L_2 or L_1 norm. (Just add more factors to the denominator in (3.138).)

Suppose $\{A\}$ lies in a bounded set B in $\hat{\mathcal{H}}'(r, |\theta| \leq \frac{1}{4})$ so $|A(r,\theta)| \leq c \exp(cr)$ uniformly for $A \in B$ and $|\theta| \leq \pi/4$. Hence

$$|\lambda_{j_0}^n \cdot A| \leq c_4^n e^{cr}. \tag{3.141}$$

Now, $\lambda_{j_0}^n \cdot A \in \hat{\mathcal{H}}'(r)$. Therefore, by standard use of Cauchy's formula we can estimate the p-th coefficient a_p^n of r^n in $\lambda_{j_0}^n \cdot A$. We find $|a_p^n| \leq c_5^n$ for all $p \leq n$ by (3.135) and (3.136). (Recall $p = \pm(n - 2j_0)$.)

We have proven that $A \in B$ implies $\{A^p\}$ is bounded. It is easy to modify this argument so as to apply to neighborhoods of 0 in place of bounded sets. By duality and our above remarks we have

Theorem 3.23 *Let U be any open set on the unit sphere in $\mathbb{R}^n$. The union of the complexifications of lines in $\mathbb{R}^n$ passing through U is sufficient for $\mathcal{H}$.*

Remark 1 *We do not need a full open set U. What is really needed is a set U large enough so that if a finite Fourier series*

$$\left| \sum_{p \leq N} \alpha_l^p h_l^p(\theta) \right| \leq 1 \quad \text{on } U \tag{3.142}$$

then

$$|\alpha_l^p| \leq c_0 c^N$$

for constants c_0, c depending only on U.

Such sets U have been studied extensively in approximation theory (see [72]).

Remark 2 *We cannot prove the analogous result for $\mathcal{E}$ using only an open set of lines. Theorem 3.20 states that a sufficient set for $\mathcal{E}$ is the union of $\mathbb{R}^k$ with a sufficient set for $\mathcal{H}$. By Theorem 3.23 the latter can be taken as the union of complexifications of lines in $\mathbb{R}^n$ whose direction cosines lie in an open set U of the unit sphere. But for estimates of $B_l^p(r)$ on r real it seems we need all real directions.*

We introduced a general setting for Theorems 3.14–3.22 in the remark following the statement of Theorem 3.14. We now expound on that idea.

We start with a compact reductive Lie group W acting linearly and transitively on $\mathbb{R}^n$. (The reader unfamiliar with reductive groups can think of W as a product of rotation groups and scalar multiplication acting on different variables.) For simplicity we assume that W is connected.

If W_0 is any Lie subgroup of W the set of points fixed by W_0 forms a euclidian space since W is linear. In particular we assume that there is a closed connected subgroup W_R of W whose fixed point set R "almost" parametrizes the orbits of W. By this we mean that there is a finite group W' of linear transformations of $\mathbb{R}^n$ which commutes with W, which acts as a reflection group W_R' on R, and such that each of the fundamental domains of W_R' in R forms a cross-section of the orbits of W on $\mathbb{R}^n$.

Let r be a euclidean coordinate on R. We pick (smoothly) a set of representatives $\{w_\theta\} \subset W$ for W/W_R. Any point $t \in \mathbb{R}^n$ is of the form $t = \omega_\theta r$. We can think of (r, θ) as *polar coordinates* in $\mathbb{R}^n$. (Actually (r^+, θ) where r^+ is the restriction of r to some fixed fundamental domain for W'_R is a more precise analog of polar coordinates.) Call N the order of W'. Generically each point $x \in \mathbb{R}^n$ has N polar coordinates because (generically) we can choose one r coordinate from each fundamental domain for W'_R and then θ is determined. (If we use r^+ there is a simple polar coordinate for each generic x.)

In the example of $W =$ rotation group, we can set $R = x_1$ axis, $W_R =$ subgroup fixing R, $W/W_R =$ unit sphere S, $W' = \{x \to \pm x\}$.

W' commutes with W. Thus for $w' \in W'$

$$\begin{aligned} w'(r, \theta) &= w' w_\theta r \\ &= w_\theta w' r \\ &= (w' r, \theta). \end{aligned}$$

Since $w'r$ lies on the same W orbit as r it follows that W' preserves the orbits of W. Thus we can write

$$\begin{aligned} w'(r, \theta) &= (w' r, \theta) \\ &= (r, w'\theta) \end{aligned}$$

where $w'\theta$ is defined by this equation. In the case of Theorem 3.14,

$$\begin{aligned} w' r &= -r \\ w'\theta &= \text{ antipode of } \theta. \end{aligned}$$

Call $\vec{\imath} = (i_1^0, \ldots, i_r^0)$ generators for the ring of invariants of W. We assume degree $i_q > 1$ for all q. Since $i_q(t)$ is constant on any orbit, i_q is a function of r. We assume that t is a linear function of r (this will be made more precise below). Thus $i_q(t) = i'_q(r)$ is a polynomial of the same degree as i_q. Since $i_q(t)$ depends only on the orbit of t it follows that $i'_q(r)$ is invariant under W'_R. Conversely any W'_R invariant defines a W invariant. This implies that $\vec{\imath}'(r)$, which generates the invariants of W'_R, is the same as $\vec{\imath}(t)$.

Since W'_R is a reflection group $\vec{\imath}'$ and $\vec{\imath}$ are strongly free so the harmonic function theory of Section 3.1 applies.

Let $\mathcal{O}^0$ be a fixed generic orbit of W. Since W is compact so is $\mathcal{O}^0$. By Weierstrass' theorem the restrictions of polynomials to $\mathcal{O}^0$ are dense in any "reasonable" function space on $\mathcal{O}^0$.

W commutes with scalar multiplication and hence preserves the degrees of polynomials. In particular each p defines a representation of W on the polynomials which are homogeneous of degree p. Since W preserves $\vec{\imath}$ it preserves $\vec{\imath}$ harmonicity so we obtain a representation ρ^p on the homogeneous harmonics of

degree p. ρ^p can be decomposed into a finite number of irreducible representations ρ^p_m on harmonic spaces H^p_m. Call d^p_m the dimension of H^p_m.

Restricting harmonics to $\mathcal{O}^0$ is an injective map by Theorem 3.2. Let $\{h^p_{jl}\}$ be a basis for H^p_j. We can normalize the basis *à la* Schur, meaning

$$\int_{\mathcal{O}^0} h^p_{ml}\,\bar{h}^{p'}_{m'l'} = \frac{1}{d^p_m}\delta^{p=p'}_{m=m',l=l'}.$$

(We could also orthonormalize; we have already noted that the difference between the two normalizations is insignificant.)

Let us use the notation $\mathcal{O}(r)$ for the orbit corresponding to r. The Schur normalization is given on an orbit $\mathcal{O}^0 = \mathcal{O}(r_0)$ for a convenient r_0.

Since the restrictions of harmonics to any generic $\mathcal{O}(r)$ are the same as the restrictions of polynomials by Theorem 3.2, they are dense in $L_2(\mathcal{O}(r))$. They provide a basis for $L_2(\mathcal{O}(r))$, meaning that we can expand any nice function in a Fourier series in the restrictions of h^p_{jl} to $\mathcal{O}(r)$.

Let $A(t)$ be a function belonging to some space of entire functions such as $\hat{\mathcal{H}}', \hat{\mathcal{E}}', \hat{\mathcal{D}}_F$. We want to study the expansion coefficients

$$A^p_{ml}(r) = \int_{\mathcal{O}^0} A(r,\theta)\bar{h}^p_{ml}(\theta)\,d\theta$$

where $d\theta$ is the W invariant measure suitably normalized.

To develop results as in Theorems 3.14, 3.15, and 3.16 we need to assume that t is a linear function of r. Precisely we write

$$t = r \cdot \vec{h}^1(\theta).$$

This "inner product" is some particular linear combination of products of components of r with the $h^1_{ml}(\theta)$. (r, θ are the polar coordinates of t.)

This inner product allows us to replace the powers t^q of t (q is a multi-index) in the power series expansion of $A(t)$ by powers of r times products of the h^1_{ml}. These products can be expressed as linear combinations of the h^p_{ml} in accordance with Kronecker product (Clebsch–Gordon) decomposition (Section 3.2). In this way we can find some "vanishing theorems" as in Theorem 3.14. Hopefully these can lead to analogs of Theorem 3.14.

Problem 3.10 *Give a precise description (analog of Theorem 3.14) of the functions* A^p_{ml} *for an arbitrary compact semi-simple Lie group* W.

3.4 Fourier expansions on hyperbolas

Thus far we have been concerned with (spherical) polar coordinates and the orthogonal group, which is compact. The noncompact theory is much more complicated. Let us begin with the group W of hyperbolic motions of the plane; that

is, $W = SO(1,1) =$ group leaving $t_1^2 - t_2^2$ invariant. We define hyperbolic polar coordinates by

$$\begin{aligned} t_1 &= \rho \cosh \zeta \\ t_2 &= \rho \sinh \zeta \end{aligned} \tag{3.143}$$

in the region where $|t_1| > |t_2|$. Thus ρ has the same sign as t_1. For $|t_2| > |t_1|$ we use

$$\begin{aligned} t_1 &= \rho \sinh \zeta \\ t_2 &= \rho \cosh \zeta \end{aligned} \tag{3.144}$$

with the sign of ρ the same as the sign of t_2.

Since W is noncompact (W is isomorphic to the additive group $\mathbb{R}^1$) we replace the Fourier series of the previous section by the Fourier integral. Suppose, for simplicity, that we are in the region (3.143) with $\rho > 0$. We define the Fourier transform by

$$A_s(\rho) = \int A(\rho \cosh \zeta, \rho \sinh \zeta) e^{is\zeta} \, d\zeta. \tag{3.145}$$

For the precise analog of Theorem 3.15 we seek to characterize the ζ Fourier integral coefficients $A_s(\rho)$ of $A \in \hat{\mathcal{E}}'$ on the hyperbolas. However, the integral is not defined for $A \in \hat{\mathcal{E}}'$. To avoid this difficulty we assume that $A \in \hat{\mathcal{D}}$. Thus $A(t_1, t_2)$ decreases at (real) infinity faster than $|t|^{-N}$ for any N so (3.145) is well defined. In fact, the polynomial decrease in t implies, via (3.143), that for ρ fixed $A(\rho \cosh \zeta, \rho \sinh \zeta)$ falls off exponentially in ζ so $A_s(\rho)$ is an entire function of s which is bounded in any strip $|\Im\, s| \leq c$.

We search for an analog of the results of the previous section, namely a characterization of $\{A_s(\rho)\}$ for $A \in \hat{\mathcal{D}}$. We are unable to accomplish this completely so we content ourselves with some results on the structure of $A_s(\rho)$. They involve:

(1) The relation of the Fourier transform $\hat{A}(\hat{t})$ of $A(t)$ to the Mellin transform of $A_s(\rho)$.
(2) The relation of $A_s(t)$ to the Fourier series coefficients $A_{\pm}^p(r)$ discussed in the previous section.
(3) The behavior of $A_s(\rho)$ as $\rho \to 0$.

(1) We define

$$\mathbb{A}(\lambda_\alpha, s_\alpha; \hat{t}_1, \hat{t}_2) = \iint A(\rho_\alpha \cosh \zeta_\alpha, \rho_\alpha \sinh \zeta_\alpha) \rho_\alpha^{i\lambda_\alpha} e^{is_\alpha \zeta_\alpha} e^{i(t_1 \hat{t}_1 + t_2 \hat{t}_2)} \, dt_1 \, dt_2. \tag{3.146}$$

Our notation is designed so that $(\rho_\alpha, \zeta_\alpha)$ represent the (ρ, ζ) coordinates in the four quadrants, which are denoted by α. The integral is taken over the quadrant.

The function $\mathbb{A}$ satisfies the differential–difference equations (and so differs from the examples of Chapter 5 which all satisfy differential equations):

$$\begin{aligned} \frac{1}{i}\frac{\partial \mathbb{A}_\alpha}{\partial \hat{t}_1} &= \tfrac{1}{2}\left[\tau_{-i}^{\lambda}\tau_{-i}^{s} + \tau_{-i}^{\lambda}\tau_{i}^{s}\right]\mathbb{A}_\alpha \\ \frac{1}{i}\frac{\partial \mathbb{A}_\alpha}{\partial \hat{t}_2} &= \tfrac{1}{2}\left[\tau_{-i}^{\lambda}\tau_{-i}^{s} - \tau_{-i}^{\lambda}\tau_{i}^{s}\right]\mathbb{A}_\alpha. \end{aligned} \tag{3.147}$$

Here τ is the translation operator and we have written τ_a^s (resp. τ_b^λ) to mean translation in s_α (resp. λ_α) by a (resp. b). Equations (3.147) are merely restatements of the fact that

$$\begin{aligned} t_1 &= \frac{\rho}{2}(e^{\zeta} + e^{-\zeta}) \\ t_2 &= \frac{\rho}{2}(e^{\zeta} - e^{-\zeta}). \end{aligned}$$

In addition there are "initial conditions"

$$\begin{aligned} \mathbb{A}_\alpha(\lambda_\alpha, s_\alpha; 0, 0) &= \int A(\rho_\alpha \cosh\zeta_\alpha, \rho_\alpha \sinh\zeta_\alpha)\rho_\alpha^{i\lambda_\alpha} e^{is_\alpha\zeta_\alpha}\rho_\alpha \, d\rho_\alpha \, d\zeta_\alpha \\ &= \left[\underset{\vee}{A}_{s_\alpha}(\lambda_\alpha - 2i)\right]. \end{aligned} \tag{3.148}$$

We have written $\underset{\vee}{B}$ for the Mellin transform

$$\underset{\vee}{B}(\lambda) = \int B(\rho)\rho^{i\lambda}\frac{d\rho}{\rho}.$$

Also

$$\sum_\alpha \mathbb{A}_\alpha(0, 0; \hat{t}_1, \hat{t}_2) = \hat{A}(\hat{t}). \tag{3.149}$$

The nonlinear Fourier transform thus relates $\hat{A}(\hat{t})$ to the Mellin transform of $A_s(\rho)$ via the equations (3.147). Information about either $\underset{\vee}{A}_\rho$ or $\hat{A}(\hat{t})$ can be used to study the other.

Of course, $\hat{A}$ is an arbitrary function in $\mathcal{D}$.

It can be shown that (3.147), (3.148), and (3.149) determine $\mathbb{A}_\alpha$. The proof depends on a "transcendental" version of the fundamental principle (Section 1.5) which describes all solutions of (3.147). This is the analog of the fundamental principle for differential difference equations like (3.147). We omit the details.

(2) We relate $A_s(\rho)$ to $A_\pm^p(r)$ which are the Fourier series coefficients as defined in the previous section. We start with the function

$$\underset{\approx}{\mathbb{A}}_\alpha(\lambda_\alpha, s_\alpha; \beta, p, \pm) = \iint A(\rho_\alpha \cosh\zeta_\alpha, \rho_\alpha \sinh\zeta_\alpha)\rho_\alpha^{i\lambda_\alpha} e^{is_\alpha\zeta_\alpha} r^{i\beta} e^{\pm ip\theta}\, dt_1\, dt_2. \tag{3.150}$$

As in equations (3.147) we find

$$\tfrac{1}{2}\left\{[\tau_{-i}^{\lambda}\tau_{-i}^{s}+\tau_{-i}^{\lambda}\tau_{i}^{s}]\pm i[\tau_{-i}^{\lambda}\tau_{-i}^{s}-\tau_{-i}^{\lambda}\tau_{i}^{s}]\right\}\underset{\sim}{\mathbb{A}}_{\alpha}=\tau_{-i}^{i\beta}\tau_{1/p}^{\theta}\underset{\vee}{\mathbb{A}}_{\alpha}. \tag{3.151}$$

This equation is a restatement of $t_1 \pm it_2 = r\exp(\pm i\theta)$.

The "initial conditions" are

$$\begin{aligned}\underset{\sim}{\mathbb{A}}_{\alpha}(\lambda_{\alpha},s_{\alpha};0,0,\pm)&=\underset{\sim}{A}_{s_{\alpha}}(\lambda_{\alpha}-2i)\\ \underset{\sim}{\mathbb{A}}_{\alpha}(0,0;\beta,p,\pm)&=\underset{\vee}{A}{}^{p}_{\pm}(\beta-2i).\end{aligned} \tag{3.152}$$

In the first equation of (3.152) the Mellin transform is taken in the variable ρ_α while in the second equation it is in the variable r.

We shall discuss below the relations of the Mellin transform to the power series coefficients (in r) of $A_{\pm}^{p}(r)$.

It is possible to show that $\underset{\sim}{\mathbb{A}}$ is determined by equations (3.151) and (3.152); we shall omit the proof.

In the above discussion of $A_s(\rho)$ we studied the relation of $A_s(\rho)$ to $\hat{A}$ and to $A_{\pm}^{p}$. These relations do not seem to give much insight into the structure of A_s so we give a more direct approach. As we shall see, our ideas do not lead to a completely satisfactory characterization of $A_s(\rho)$.

(3) Let us begin by making precise estimates. It is generally more convenient to write

$$\begin{aligned}u&=\rho e^{\zeta}=t_1+t_2\\ v&=\rho e^{-\zeta}=t_1-t_2.\end{aligned} \tag{3.153}$$

Since $A(t)\in\hat{\mathcal{D}}$ decreases at infinity faster than $|t|^{-N}$ for any N

$$|A(u,v)|\le c\rho^{-N}e^{-N|\zeta|} \tag{3.154}$$

for t large. In particular for ρ fixed A is exponentially decreasing in ζ. This means that $A_\rho(s)$ is holomorphic and bounded in any strip $|\Im s|\le N$. Since the same is true for the ζ derivatives of A it follows that $A_\rho(s)$ decreases polynomially in any such strip, i.e.

$$|A_s(\rho)|\le c_{\rho MN}(1+|s|)^{-M}\quad |\Im s|\le N. \tag{3.155}$$

It is clear that we can bound $c_{\rho MN}$ by $c'_{\rho_0 MN}$ for $\rho\ge\rho_0>0$. We now embark on the important study of the growth of $c_{\rho MN}$ as $\rho\to 0$.

We write

$$\begin{aligned}A_s(\rho)&=\int_0^{\infty}A(\rho e^{\zeta},\rho e^{-\zeta})e^{is\zeta}\,d\zeta+\int_{-\infty}^{0}A(\rho e^{\zeta},\rho e^{-\zeta})e^{is\zeta}\,d\zeta\\ &=A_s^{+}(\rho)+A_s^{-}(\rho)\end{aligned} \tag{3.156}$$

say. We have

$$A_s^+(\rho) = \rho^{-is} \int_\rho^\infty A(u, \rho^2/u) u^{is} \frac{du}{u} \tag{3.157+}$$

$$A_s^-(\rho) = \rho^{is} \int_\rho^\infty A(\rho^2/v, v) v^{-is} \frac{dv}{v}. \tag{3.157-}$$

In (3.157+) $u \geq \rho$ so $\rho^2/u \leq \rho$. It follows that as $\rho \to 0$ the integrand in (3.157+) $\to A(u,0)u^{is}$. When $\Re is > 0$ we can apply Lebesgue's theorem to conclude that the integral

$$\to \int_0^\infty A(u,0) u^{is} \frac{du}{u} = \underset{\vee 1}{A}(s)$$

where $\underset{\vee 1}{A}$ is the Mellin transform in the first variable when the second $= 0$.

Similarly if $\Re is < 0$ the integral in (3.157−) $\to \underset{\vee 2}{A}(s)$.

On the other hand if $\Re is < 0$ and $\rho < 1$

$$\left| \int_\rho^\infty A(u, \rho^2/u) u^{is} \frac{du}{u} \right| \leq c_N \int_\rho^\infty (1+u)^{-N} \left| u^{is} \right| \frac{du}{u}$$

$$\leq \frac{c}{|\Re is|} \rho^{\Re is} \tag{3.158+}$$

while for $\Re is > 0$

$$\left| \int_\rho^\infty A(\rho^2/v, v) v^{-is} \frac{du}{u} \right| \leq \frac{c}{|\Re is|} \rho^{-\Re is}. \tag{3.158-}$$

These inequalities follow from the fact that the integrals over $(1, \infty)$ are bounded while

$$\left| \int_\rho^1 u^{is} \frac{du}{u} \right| \leq \frac{\rho^{\Re is}}{|\Re is|} + c$$

with a similar inequality for (3.158−).

Application of these results to (3.156) yields

$$\lim_{\substack{\rho \to 0 \\ \Re is > 0}} \rho^{is} A_s(\rho) = \underset{\vee 1}{A}(s) \tag{3.159+}$$

$$\lim_{\substack{\rho \to 0 \\ \Re is < 0}} \rho^{-is} A_s(\rho) = \underset{\vee 2}{A}(s). \tag{3.159-}$$

Let us apply the same idea to the ζ derivative of A. For any l

$$(is)^l A_s(\rho) = \int A(\rho e^\zeta, \rho e^{-\zeta}) \frac{\partial^l}{\partial \zeta^l} e^{is\zeta} \, d\zeta$$

$$= \rho^l \sum_{j=0}^{l} \int e^{(2j-l)\zeta} A^{(j)}(\rho e^\zeta, \rho e^{-\zeta}) e^{is\zeta} \, d\zeta \tag{3.160}$$

where $A^{(j)}$ is of the form

$$A^{(j)}(\rho e^{\zeta}, \rho e^{-\zeta}) = c_{lj} \frac{\partial^l}{\partial u^j \partial v^{l-j}} A(\rho e^{\zeta}, \rho e^{-\zeta}). \tag{3.161}$$

Equation (3.160) can be written in the form

$$(is)^l A_s(\rho) = \sum \int u^j v^{l-j} A^{(j)}(u,v) e^{is\zeta}\, d\zeta. \tag{3.162}$$

When $\rho \to 0$ for u fixed, $v \to 0$. Thus when $\Re\, is > 0$

$$\begin{aligned}(is)^l \rho^{is} A_s(\rho) &= \sum \int u^{j+is} v^{l-j} A^{(j)}(u,v)\, \frac{du}{u} \\ &\to \int u^{l+is} A^{(l)}(u,0)\, \frac{du}{u} \\ &= \underset{\vee 1}{A}{}^{(l)}(-il+s)\end{aligned} \tag{3.163+}$$

while for $\Re\, is < 0$

$$(is)^l \rho^{-is} A_s(\rho) \to \underset{\vee 2}{A}{}^{(l)}(-il-s). \tag{3.163-}$$

These limits involve the Mellin transform of $A^{(l)}(u,0)$ or of $A^{(l)}(0,v)$. We shall show, below, how to use the Mellin transform to study the power series coefficients of A.

Before doing that we digress to give a heuristic treatment of what results we should expect.

We write

$$\begin{aligned}A(u,v) &= \sum a_{mj} u^m v^j \\ &= \sum a_{mj} \rho^{m+j} e^{(m-j)\zeta} \\ &= \sum \rho^k e^{k\zeta} \sum a_{k+j,j} \rho^{2j} \\ &= \sum u^k \sum a_{k+j,j} \rho^{2j}.\end{aligned} \tag{3.164}$$

Think of A as a function of u and ρ; we write $A(u,v) = B(u,\rho)$. By the last line of (3.164)

$$B(u,0) = \sum a_{k,0} u^k \tag{3.165}$$

so we can recover the Laurent series coefficients $a_{k,0}$ by the values of B at $\rho = 0$. We can recover the other Laurent series coefficients as

$$\left[\left(\frac{1}{2\rho}\frac{\partial}{\partial \rho}\right)^l B\right](u,0) = l! \sum a_{k+l,l} u^k. \tag{3.166}$$

Recall that $k = m - j$ so $k + j = m > 0$. Thus (3.166) is the power series of an entire function.

This analysis shows the significance of studying the behavior of A at $\rho = 0$. It remains to understand how to determine the power series coefficients of $B(u, 0)$ in terms of its Mellin transform.

The following argument is often useful in this connection. If

$$f(x) = \sum c_n x^n$$

is an entire function which is small, say $\mathcal{O}(x^{-N})$ for all N, at $x = +\infty$, then we write the Mellin transform

$$\begin{aligned} \underset{\vee}{f}(s) &= \int_0^\infty f(x) x^{is} \frac{dx}{x} \\ &= \int_0^\alpha + \int_a^\infty . \end{aligned} \tag{3.167}$$

The Mellin transform is defined for $\Re is > 0$. It is clear that the integral over $[\alpha, \infty)$ is an entire function. For the integral over $[0, \alpha]$ we use Taylor's formula

$$\begin{aligned} \int_0^\alpha f(x) x^{is} \frac{dx}{x} &= \sum_{n=0}^N c_n \int_0^\alpha x^{n+is} + \int_0^\alpha x^{is} \left[\mathcal{O}(x^{N+1})\right] \frac{dx}{x} \\ &= \sum_{n=0}^N c_n \frac{\alpha^{i(s+in)}}{i(s-in)} + \beta(s). \end{aligned} \tag{3.168}$$

$\beta(s)$ is holomorphic for $\Re is > -(N+1)$.

We conclude that

$\underset{\vee}{f}(s)$ has a meromorphic extension to the whole plane with poles at $\{in\}_{n \geq 0}$ with residues $\{-ic_n\}$.

Note that $\Re is > 0$ means $\Im s < 0$. The integral (3.167) certainly converges and hence is regular when $\Im s < 0$. The above locates the poles in the meromorphic continuation at the nonnegative imaginary integers.

We have seen in (3.163±) how to express the Mellin transform of derivatives of A on $\rho = 0$ as suitable limits of $(is)^l \rho^{\pm is} A_s(\rho)$. Unfortunately we are expressing the u derivatives on $v = 0$ and the v derivatives on $u = 0$. To determine A from A_s we need to examine u derivatives on $u = 0$ or v derivatives on $v = 0$. Such are provided in (3.166) in our heuristic argument by means of the operators $[(1/2\rho)(\partial/\partial\rho)]^l$. We now show that this heuristic argument points to the correct results.

We have constructed the Mellin transform of A on the line $\rho = 0$ from $A_s(\rho)$. Since $A(u, 0)$ and $A(0, v)$ are entire functions we can use the argument of (3.167)ff. to find the power series coefficients a_{0k} and a_{k0} of A.

To determine the other power series coefficients of A we have to perform the same process for the u, v derivatives of A as in our heuristic argument (3.164)f.

Since $u = \rho e^{\zeta}$, $v = \rho e^{-\zeta}$

$$\begin{aligned} \frac{\partial}{\partial u} &= \frac{e^{-\zeta}}{2}\left(\frac{\partial}{\partial \rho} + \frac{1}{\rho}\frac{\partial}{\partial \zeta}\right) \\ \frac{\partial}{\partial v} &= \frac{e^{\zeta}}{2}\left(\frac{\partial}{\partial \rho} - \frac{1}{\rho}\frac{\partial}{\partial \zeta}\right). \end{aligned} \tag{3.169}$$

Call A_{us}, A_{vs} the analogs of A_s for $\partial A/\partial u = A_u$, $\partial A/\partial v = A_v$. We find

$$\begin{aligned} A_{us}(\rho) &= \frac{1}{2}\int e^{-\zeta + is\zeta}\left[\frac{\partial}{\partial \rho} + \frac{1}{\rho}\frac{\partial}{\partial \zeta}\right] A(\rho e^{\zeta}, \rho e^{-\zeta})\, d\zeta \\ &= \frac{1}{2}\frac{\partial}{\partial \rho}\int e^{i(s+i)\zeta} A(\rho e^{\zeta}, \rho e^{-\zeta})\, d\zeta - i\frac{s+i}{2\rho}\int e^{i(s+i)\zeta} A(\rho e^{\zeta}, \rho e^{-\zeta})\, d\zeta \\ &= \frac{1}{2}\left(\frac{\partial}{\partial \rho} - i\frac{s+i}{\rho}\right) A_{s+i}(\rho) \\ &= \frac{1}{2}\rho^{i(s+i)}\frac{\partial}{\partial \rho}\rho^{-i(s+i)} A_{s+i}. \end{aligned} \tag{3.170}$$

Similarly

$$A_{vs}(\rho) = \frac{1}{2}\rho^{-i(s-i)}\frac{\partial}{\partial \rho}\rho^{i(s-i)} A_{s-i}. \tag{3.171}$$

As before we multiply by $\rho^{\pm is}$ and examine the limits as $\rho \to 0$. We are interested in the v derivatives of A on $v = 0$ and the u derivatives on $u = 0$. We have

$$\begin{aligned} \rho^{is} A_{vs}(\rho) &= \frac{1}{2\rho}\frac{\partial}{\partial \rho}\rho^{i(s-i)} A_{s-i} \\ &\to \frac{1}{2\rho}\frac{\partial}{\partial \rho}\int A(u,0) u^{i(s-i)}\,\frac{du}{u} \end{aligned} \tag{3.172}$$

for $\Re is > 0$. This agrees with our heuristic result (3.166). The last expression in (3.172) is defined by writing the integrand in terms of ρ, ζ, then applying the operator $(1/2\rho)(\partial/\partial\rho)$, and then setting $u = \rho e^{\zeta}, v = \rho e^{-\zeta} = 0$. The fact that the power of u is $i(s-i)$ instead of is is a reflection of the appearance of $a_{k+j,j}$ (for $j = 1$) in (3.166).

We can summarize our results by

Theorem 3.24 *Let $A \in \hat{\mathcal{D}}$. Suppose we are in the region $t_1 \geq |t_2|$; we set $u = t_1 + t_2, v = t_1 - t_2$. The Fourier transform $A_{v^l s}(\rho)$ of $\partial^l A/\partial v^l$ on the (half-)hyperbola defined by $\rho > 0$ is an entire function of s. For $\rho \to 0$, $\Re is > 0$*

$$\rho^{is} A_{v^l s}(\rho) \to \left(\frac{1}{2\rho}\frac{\partial}{\partial \rho}\right)^l \int A(u,0) u^{l+is}\,\frac{du}{u}. \tag{3.173}$$

The right side of (3.173) is meromorphic in the entire s plane with poles at $\{s = i(l+n)\}_{n\geq 0}$. *The residues at these poles are, up to a constant, the power series coefficients* a_{jl} *of* $A(u,v)$.

The same result holds if the roles of u, v *are interchanged and* ρ^{is} *is replaced by* ρ^{-is} *in (3.173) with* $\Re is < 0$.

The homogeneity of the operator $(1/2\rho)(\partial/\partial\rho)$ *of degree* -2 *is the analog of the statement in Theorem 3.14 that the power series for* A_l^p *contains only the terms* r^{p+2j}.

Let us illustrate our ideas (actually in a somewhat modified form) for the classical case

$$A = e^{-t_1} = e^{-1/2(u+v)}.$$

Although $A \notin \hat{\mathcal{D}}$ it has a sufficiently nice behavior in $t_1 > |t_2|$ for our methods to apply

$$\begin{aligned} A_s(\rho) &= \int e^{-\zeta\cosh\zeta} e^{is\zeta}\, d\zeta \\ &= K_{is}(\rho). \end{aligned} \tag{3.174}$$

$K_{is}(\rho)$ is a standard Bessel function.

According to (3.157±) we write

$$\begin{aligned} K_{is}(\rho) &= K_{is}^+(\rho) + K_{is}^-(\rho) \\ &= \rho^{-is}\int_\rho^\infty e^{-1/2(u+\rho^2/u)} u^{is}\,\frac{du}{u} + \rho^{is}\int_\rho^\infty e^{-1/2(v+\rho^2/v)} v^{-is}\,\frac{dv}{v}. \end{aligned} \tag{3.175}$$

Since $\rho^2 = uv$, if u is fixed and $\rho \to 0$ then $\rho^2/u = v \to 0$, as does ρ^2/v for fixed v as $\rho \to 0$. On the other hand the integrals in (3.175) over $(1,\infty)$ are clearly entire functions of s. Thus the singularities at $\rho = 0$ can be determined by expanding $\exp[-\frac{1}{2}(u+\rho^2/u)]$ and $\exp[-\frac{1}{2}(v+\rho^2/v)]$ in power series in $\rho^2/u, \rho^2/v$ respectively.

We find

$$\begin{aligned} K_{is}^+(\rho) &= \rho^{-is}\int_\rho^\infty e^{-(u+\rho^2/u)/2} u^{is}\,\frac{du}{u} \\ &= \sum (-1)^n \frac{\rho^{2n-is} 2^{-2n+is}}{n!}\int_{\rho/2}^\infty e^{-u} u^{is-n}\,\frac{du}{u}. \end{aligned} \tag{3.176}$$

We have thus expressed K_{is} in terms of the incomplete Γ functions $\Gamma(is-n, \rho/2)$ which is the standard notation for the integral.

$\Gamma(is-n,\rho/2)$ is an entire function of s. When $\Re\,(is-n)>0$ we can write

$$\begin{aligned}\int_{\rho/2}^{\infty} e^{-u}u^{is-n}\frac{du}{u} &= \left[\int_0^{\infty}-\int_0^{\rho/2}\right] e^{-u}u^{is-n}\frac{du}{u}\\ &= \Gamma(is-n)-\int_0^{\rho/2}\sum\frac{(-1)^j}{j!}u^{is-n+j}\frac{du}{u}\\ &= \Gamma(is-n)-\sum\frac{(-1)^j(\rho/2)^{is-n+j}}{j!(is-n+j)}.\end{aligned}\tag{3.177}$$

Note in passing that the poles of Γ are canceled by the residues of the series, confirming the entire nature of the incomplete Γ function.

We have arrived at the expression

$$K_{is}^{+}(\rho)=\sum(-1)^n\frac{(\rho/2)^{2n-is}}{n!}\Gamma(is-n)-\sum(-1)^{n+j}\frac{(\rho/2)^{n+j}}{n!j!(is-n+j)}.\tag{3.178}$$

Since

$$\rho^{-is}e^{i\zeta s}=(t_1-t_2)^{-is}$$

we can repeat the same argument to deduce

$$K_{is}^{-}(\rho)=K_{-is}^{+}(\rho).$$

Note that the terms in K_{is}^{+} which come from $\int_0^{\rho/2}$ are symmetric in n and j except for the denominator $(is-n+j)$. Changing s to $-s$ changes the sum into its negative. Thus these terms cancel in $K^{+}+K^{-}$ and we arrive at the classical formula for K:

$$K_{is}(\rho)=\sum\frac{(\rho/2)^{2j+is}}{j!}(-1)^j\Gamma(-is-j)-\sum\frac{(\rho/2)^{2j-is}}{j!}(-1)^j\Gamma(is-j).\tag{3.179}$$

Such types of splittings are crucial for the deeper study of special functions and eigenfunction expansions. In the present case the splitting is related to the classical expression [11, Vol. II, p. 5]

$$K_{is}(\rho)=(\pi/2)[\sin i\pi s]^{-1}[I_{-is}(\rho)-I_{is}(\rho)]$$

where I_{is} is the standard Bessel function

$$\begin{aligned}I_{is}(\rho) &= e^{s\pi/2}J_{is}(\rho e^{i\pi/2})\\ &= \sum\frac{(\rho/2)^{2j+is}}{j!\Gamma(j+is+1)}.\end{aligned}\tag{3.180}$$

The identity of this expression for K_{is} with (3.179) is a consequence of

$$\Gamma(s)\Gamma(1-s)=\pi\csc\pi s$$

(see [11, Vol. II, p. 3]).

Let us now pass from $\exp(-t_1)$ to a general function $A(t_1, t_2) \in \hat{\mathcal{D}}$. We set

$$\begin{aligned} A_s(\rho) &= \int A(u,v) e^{is\zeta}\, d\zeta \\ &= \rho^{-is} \int_\rho^\infty A(u, \rho^2/u) u^{is} \frac{du}{u} + \rho^{is} \int_\rho^\infty A(\rho^2/v, v) v^{-is} \frac{dv}{v} \\ &= A_s^+(\rho) + A_s^-(\rho). \end{aligned} \tag{3.181}$$

We evaluate $A^\pm$ as in the special case $A_s = K_{is}$. We expand $A(u,v)$ as a power series in v and set $v = \rho^2/u$:

$$A(u, \rho^2/u) = \sum a_m^+(u)(\rho^2/u)^m.$$

Thus

$$A_s^+(\rho) = \sum \rho^{2m-is} \int_\rho^\infty a_m^+(u) u^{is-m} \frac{du}{u}. \tag{3.182+}$$

In analogy to the incomplete Γ function we have, for $\Re(is) > m$,

$$\int_\rho^\infty a_m^+(u) u^{is-m} \frac{du}{u} = \int_0^\infty a_m^+(u) u^{is-m} \frac{du}{u} - \sum a_{m_j}^+ \frac{\rho^{is-m+j}}{is-m+j} \tag{3.183+}$$

where $a_{m_j}^+$ are the power series coefficients of $a_m^+(u)$. The integral on $[0,\infty)$ is $\underset{\vee}{a}_m^+(s+im)$.

This leads to the expression

$$A_s^+(\rho) = \sum_m \left[\rho^{2m-is} \underset{\vee}{a}_m^+(s+im) - \sum_j \rho^{m+j} \frac{a_{m_j}^+}{is-m+j} \right]. \tag{3.184+}$$

(This agrees with (3.167) when $A(t_1, t_2) = e^{-t_1}$.)

Observe that

$$a_m^+(u) = \frac{1}{m!} \frac{\partial^m}{\partial v^m} A(u,0) \tag{3.185}$$

$$a_{m_j}^+ = \frac{1}{j!} \frac{d^j}{du^j} a_m^+(0) = \frac{1}{m!\,j!} \frac{\partial^{j+m}}{\partial u^j \partial v^m} A(0,0). \tag{3.186}$$

We carry out the same expansion for $A_s^-(\rho)$. Equation (3.182+) becomes

$$A_s^-(\rho) = \sum \rho^{2m+is} \int_\rho^\infty a_m^-(v) v^{-is-m} \frac{dv}{v}. \tag{3.182−}$$

We write

$$\int_\rho^\infty a_m^-(v) v^{-is-m} \frac{dv}{v} = \int_0^\infty a_m^-(v) v^{-is-m} \frac{dv}{v} - \sum a_{m_j}^- \frac{\rho^{-is-m+j}}{-is-m+j} \tag{3.183−}$$

so that

$$A_s^-(\rho) = \sum_m \left[\rho^{2m+is} \underset{\vee}{a}{}_m^-(-s+im) - \sum \rho^{m+j} \frac{a_{m_j}^-}{-is-m+j} \right]. \qquad (3.184-)$$

By (3.186) $a_{m_j}^+$ is symmetric in j, m; also it is equal to $a_{m_j}^-$. Since the sums of the terms involving $a_{m_j}^\pm$ in (3.184±) are over all m, j, we can replace $a_{m_j}^-/(-is-m+j)$ by

$$\frac{a_{m_j}^-}{-is+m-j} = -\frac{a_{m_j}^-}{is-m+j} = -\frac{a_{m_j}^+}{is-m+j}.$$

Thus these two sums cancel in $A_s^+(\rho)+A_s^-(\rho)$ and we obtain the general splitting

$$A_s(\rho) = \sum \rho^{2m-is} \underset{\vee}{a}{}_m^+(s+im) + \sum \rho^{2m+is} \underset{\vee}{a}{}_m^-(-s+im). \qquad (3.187)$$

The functions $\underset{\vee}{a}{}_m^\pm$ are defined by meromorphic continuation from $\Re(\pm is-m) > 0$. Equation (3.187) is the general form of (3.179).

For many applications it is important to study functions which are periodic in ζ, say of period 1. Let Γ be the group

$$\Gamma = \{\gamma_l\} = \left\{ \begin{pmatrix} \cosh l & \sinh l \\ \sinh l & \cosh l \end{pmatrix} \right\}$$

where l is an integer. Γ is a discrete subgroup of the group of hyperbolic motions of the plane (see Section 10.1). Note that

$$\begin{aligned} \gamma_1 \cdot t &= \begin{pmatrix} \cosh 1 & \sinh 1 \\ \sinh 1 & \cosh 1 \end{pmatrix} \cdot \begin{pmatrix} \rho\cosh\zeta \\ \rho\sinh\zeta \end{pmatrix} \\ &= \begin{pmatrix} \rho\cosh(\zeta+1) \\ \rho\sinh(\zeta+1) \end{pmatrix}. \end{aligned} \qquad (3.188)$$

Thus periodicity is the same as invariance under the discrete group Γ.

Γ acts on the light cone $\rho = 0$. In fact,

$$\gamma_1 \begin{pmatrix} u \\ \pm u \end{pmatrix} = \begin{pmatrix} ue^{\pm 1} \\ \pm ue^{\pm 1} \end{pmatrix}. \qquad (3.189)$$

This suggests that the meaning of "periodicity of period 1 in ζ" on the light cone is multiplicative periodicity with period $\exp(\pm 1)$.

We can understand this from the hyperbolic geometry of the plane. Suppose we are in the quadrant $t_1 \geq |t_2|$ and we let $t_1 \to t_2$ which is positive and fixed. Let Δ be the fundamental domain for Γ on the hyperbola containing (t_1, t_2) with Δ centered at (t_1, t_2). If we write $t_1 = \rho\cosh\zeta, t_2 = \rho\sinh\zeta$ then a change of 1 in ζ corresponds to a change of about $e(t_1, t_2)$ in (t_1, t_2). This is best seen by writing $t_1 + t_2 = \rho e^\zeta, t_1 - t_2 = \rho e^{-\zeta}$ and observing that $\rho \to 0$, $\zeta \to \infty$. Thus

the fundamental domains have a limit Δ^0 which is a fundamental domain for the multiplicative group generated by multiplication by e.

These fundamental domains differ from the usual fundamental domains which are defined by angular regions based at the origin (see Figures 3.1 and 3.2).

We want to find an analog of Theorem 3.24 for periodic functions. There is a major difficulty because the orbit of any point on the light cone under Γ has the origin as a limit point. Thus a smooth periodic function is constant on the light cone.

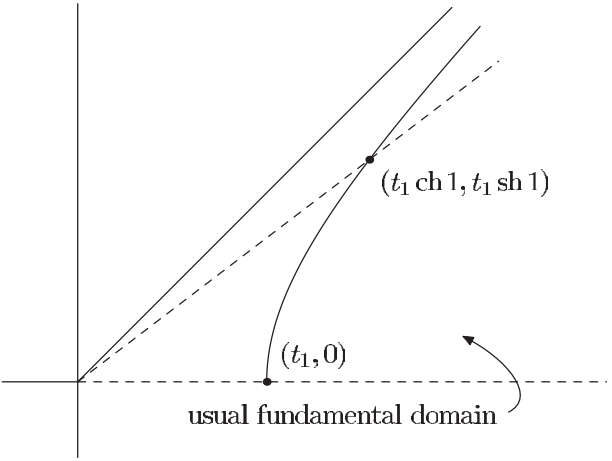

Figure 3.1

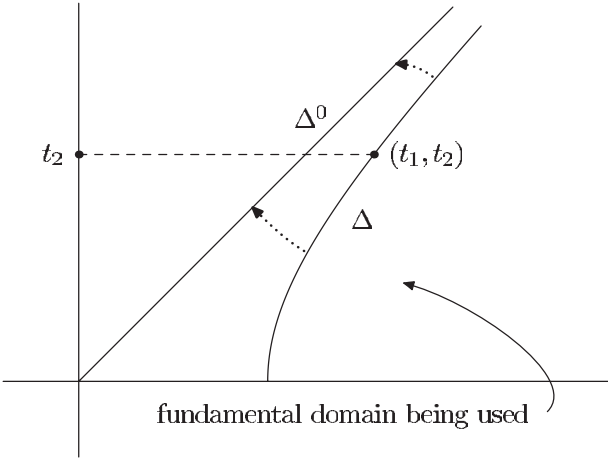

Figure 3.2

Theorem 3.25 *Any entire function A which is Γ invariant is an entire function of ρ^2.*

Proof Since $A = c_0$ is constant on the light cone, $A - c_0$ is divisible by ρ^2, i.e.

$$A = c_0 + \rho^2 A_1.$$

Clearly A_1 is also Γ invariant and hence is a constant c_1 on the light cone. Proceeding in this manner we find

$$A = c_0 + c_1\rho^2 + c_2\rho^4 + \cdots + c_m\rho^{2m} + \rho^{2(m+2)}A_{m+1}. \tag{3.190}$$

It is not difficult to show that the series converges so A is an entire function of ρ^2.

Problem 3.11 *Does there exist a function f which is real analytic except at the origin and is Γ invariant but is not a function of ρ^2?*

The nonexistence of f would be a sharpening of Theorem 3.25.

What is the Fourier series characterization of periodic functions which are smooth in the closed quadrant except at the origin? We should expect the Fourier series coefficients

$$ {}_pA(\rho) = \int_\Delta A(u,v)e^{2\pi ip\zeta}\, d\zeta \tag{3.191}$$

to converge to the multiplicative Fourier coefficients on the light cone.

To study the limit as $\rho \to 0$ we multiply by $\rho^{2\pi ip}$. Thus

$$\rho^{2\pi ip}\, {}_pA(\rho) = \int_\Delta A(u,v)u^{2\pi ip}\, d\zeta. \tag{3.192}$$

The periodicity of A tells us that it does not matter which fundamental domain Δ is used. Hence we use the fundamental domain depicted in Figure 3.2. We have seen that $\Delta = \Delta(\rho)$ converges to a fundamental domain $\Delta(0)$ for Γ on the light cone, so we obtain the multiplicative (Mellin) Fourier series coefficients ${}_p\overset{\vee}{A}$ of A on the positive light cone (i.e. $t_2 > 0$).

If we were in the part of $\rho =$ const. where $\zeta < 0$ then the fundamental domains would approach a fundamental domain Δ^- on the negative light cone. The limit of

$$\rho^{-2\pi ip}\, {}_pA(\rho) = \int_{\Delta^-} A(u,v)v^{-2\pi ip}\, d\zeta \tag{3.193}$$

as $\rho \to 0$ would again be the multiplicative (Mellin) coefficients of A but this time on the negative light cone.

The fact that we can obtain the Fourier coefficients on $\rho = 0$ from either the positive or negative light cones shows that the values of A on the two half-lines which are the respective limits of $\rho \to 0$, $\zeta > 0$ ($\zeta < 0$) are closely related. (The Fourier coefficients are not equal because of the factors $\exp(\pm 2\pi ip)$.) Certainly this makes sense in terms of the geometry of the hyperbolas.

Of course one's immediate reaction is: what happens in higher dimensions?

One generalization of the above method is to study expansions related to hyperboloids defined by

$$t_1^2 - t_2^2 - \cdots - t_k^2 = \rho^2.$$

Although the passage from $k = 2$ to $k > 2$ involves several difficulties, they are all essentially present when $k = 2$ so we shall restrict our considerations to that case.

Let us start with the region $\rho^2 > 0, t_1 > 0$ (so we may assume $\rho > 0$). If A is a nice function we can decompose A under the group K of rotations around the t_1 axis, meaning, for each fixed t_1^0, ρ, we write A in terms of its Fourier series coefficients (decomposition under K) on the circle which is the intersection of the hyperboloid defined by ρ and the plane defined by $t_1 = t_1^0$. For simplicity we shall restrict our considerations to the case when A is invariant under rotations about the t axis so A is a function of $t_1, r = \sqrt{t_2^2 + t_3^2}$.

Call W the orthogonal group of the quadratic form $t_1^2 - t_2^2 - t_3^2$. W is the natural analog of the group $SO(1,1)$ which played an essential role when $k = 2$. In the present case $W = SO(1,2)$ which is essentially $SL(2,\mathbb{R})$. This group is studied extensively in Chapter 7.

The polar coordinates for $k = 2$ are replaced by hyperbolic spherical coordinates. (For $\rho > 0$ they describe the interior of the forward light cone.)

$$\begin{aligned} t_1 &= \rho \cosh \zeta \\ t_2 &= \rho \sinh \zeta \cos \theta \\ t_3 &= \rho \sinh \zeta \sin \theta. \end{aligned} \tag{3.194}$$

The natural group theoretical analogs of $e^{is\zeta}$ are (see Chapter 7)

$$P_{is}^m(\cosh \zeta) = \frac{\Gamma(is + m + 1)}{\pi \Gamma(is + 1)} \int (\cosh \zeta + \sinh \zeta \cos \theta)^{is} \cos m\theta \, d\theta. \tag{3.195}$$

P_{is}^m is the Legendre function [11, vol. I, p. 157]. Its significance in the representation theory of W was promulgated by Bargman [9] and by Ehrenpreis and Mautner [68].

Since A is rotationally invariant only $m = 0$ is relevant; we write P_{is} for P_{is}^0.

We should like to extend Theorem 3.24 to $k > 2$. We set

$$\begin{aligned} u &= \rho e^{\zeta} \\ &= t_1 + t_2 \cos \theta + t_3 \sin \theta, \\ v &= \rho e^{-\zeta} \\ &= t_1 - t_2 \cos \theta - t_3 \sin \theta. \end{aligned} \tag{3.196}$$

Since A is rotationally invariant it is a function of ρ and ζ, hence of u and v; in fact, Weyl invariance (Chapter 7) says $A(u, v) = A(v, u)$. On the hyperboloid

defined by ρ we have

$$uv = \rho^2$$

so

$$v = \rho^2/u$$

We are led to define

$$A_s(\rho) = \int A(u, v) P_{is}((u + v)/2\rho) \sinh^2 \zeta \, d\zeta. \tag{3.197}$$

(The measure $\sinh^2 \zeta \, d\zeta$ is the W invariant measure on the half-hyperboloids $\rho = \text{const.} > 0$.)

We can expand $A(u, \rho^2/u)$ in a power series for small ρ or large u. Now we notice the first instance of the difficulty in passing from $k = 2$ to $k = 3$. The function u^{is-m} is to be replaced by $P_{is}u^{-m}$. It seems that the simplicity of the passage from (3.182) to (3.187) is lost.

However, the situation becomes simpler when $\rho \to 0$. For we show in Chapter 7 that, in a suitable limit sense,

$$\rho^{is} P_{is} \to \alpha(s) r^{is}$$

as $\rho \to 0$. Here r^{is} is the euclidean distance from a point on the light cone to the origin and $\alpha(s)$ can be expressed in terms of Γ functions.

Thus some analog of the limit part of Theorem 3.24 is valid.

The exterior of the light cone, meaning $\rho^2 < 0$, is more difficult to deal with because of the presence of the discrete series in the analog of (3.187) (see Chapter 7). This causes difficulty in comparing the expansions in the interior and in the exterior of the light cone. We hope to return to these questions in a future publication.

When $k > 2$ the expansion of "periodic" functions is more complicated. Consider the case $W = SO(1, 2)$ as before. W is essentially the same as $SL(2, \mathbb{R})$ which is the group of 2×2 real matrices of determinant 1, and so contains the modular group $\Gamma = SL(2, \mathbb{Z})$ of 2×2 integral matrices of determinant 1. Unlike the situation in case $n = 2$, Γ does not have any sort of fundamental domain on the light cone, for it is shown in [62] that Γ has dense orbits on the light cone.

When dealing with functions A on the hyperboloids $t_1^2 - t_2^2 - t_3^2 = \rho^2 > 0$ which are Γ invariant we expand A in terms of the "harmonic analysis" on the hyperboloid modulo Γ. This consists of the Eisenstein series which depend on a continuous parameter and of L_2 eigenfunctions of the group Laplacian (see Chapter 10 and [69].

Much work remains to understand how the expansion coefficients behave as $\rho \to 0$.

3.5 Deformation theory

Let us take stock of the theory of harmonic functions as developed thus far. The ingredients are

(1) *Grading.* This allows us to define homogeneity.
(2) *Homogeneous polynomials* $\vec{\imath}$ of positive degree.
(3) *Homogeneous differential operators.* $\vec{\partial} = \partial(\vec{\imath})$ of positive degree.

The abstract properties we use are

(a) Multiplication of homogeneous polynomials adds degrees.
(b) Differentiation by a homogeneous operator subtracts degrees.
(c) The kernel of $\vec{\partial}$ and the image of $\vec{\imath}$ are orthogonal complements (Fischer property).

Actually, our main (algebraic) uniqueness result can be paraphrased as saying that the orthogonal complement H of the image of $\vec{\imath}$ is a complement (though not orthogonal) not only for the image of $\vec{\imath}$ but also for the deformation which is the image of $\vec{\imath} - \vec{X}$ when $\vec{X}$ is of lower degree than $\vec{\imath}$. As such there is a glaring lack of symmetry; for we allow deformations of the image of $\vec{\imath}$ but not of its orthogonal complement.

One natural way of deforming the harmonics H is to form the space $H_{\vec{a}}$ defined by

$$(\vec{\partial} - \vec{a})h = 0. \tag{3.198}$$

We allow $\vec{a}$ to be a vector of constants or of lower order differential operators. The problem with which we are dealing in (3.198) is that not only have we destroyed the grading of H (as was also the case for the complement of H in replacing $\vec{\imath}$ by $\vec{\imath} - \vec{X}$) but, in general, there are no polynomials in the kernel of $\vec{\partial} - \vec{a}$.

We can take things one step further. We can simultaneously deform $\vec{\partial}$ according to (3.198) and $\vec{\imath}$ into $\vec{\imath} - \vec{X}$. An immediate problem arises even in the simplest case of $r = k = 1$ and $i = x^2$. For $\sin \pi x$ is a solution of $\partial(i) + \pi^2$ and also $\sin \pi x$ is in the image of $i - 1$.

This suggests that if we deform $\vec{\imath}$ to $\vec{\imath} - \vec{X}$ and $\vec{\partial}$ to $\vec{\partial} - \vec{a}$ then we should require, in some suitable sense,

$$\|\vec{a}\| \, \|\vec{X}\| < \varepsilon. \tag{3.199}$$

We have not made a general analysis of (3.199). Sections 3.1 and 3.2 dealt with the case $\vec{a} = 0$. The corresponding case $\vec{X} = 0$ can be considered as the CP rather than the DP. In order to avoid the complications that were dealt with in Theorem 3.2 we shall restrict our considerations to $r = 1$ and $a = \text{const}$.

Theorem 3.26 *There are no formal power series which are in the image of* i *and in the kernel of* $\partial(i) - a$.

Proof We want to show that it is impossible to have

$$(\partial(i) - a) \sum ig^n = 0. \tag{3.200}$$

Comparing homogeneous terms (using (b)) gives

$$\partial(i)(ig^n) = aig^{n-d} \tag{3.201}$$

where $d =$ degree i. In (3.221) only the residue class of n mod d is important; we assume $n = jd$ as the other residue classes are treated in the same fashion. We write

$$\partial(i)(ig^{jd}) = aig^{(j-1)d}. \tag{3.202}$$

For $j = 0$ this says $\partial(i)(ig^0) = 0$. Since $\partial(i)i$ is a nonzero constant it follows that $g^0 = 0$. We could iterate this process if we knew that $\partial(i)(ig^n) = 0$ implies $ig^n = 0$. But this is just Fischer's Lemma 3.1 which implies that no harmonic polynomial is in the image of i.

Theorem 3.26 is thereby proven.

To go further and deal with simultaneous deformations of i and $\partial(i)$ requires a study of the differential equation

$$(\partial(i) - 1)(i - \varepsilon)g = 0. \tag{3.203}$$

There is an essential difference between equations (3.200) and (3.203). For the operator $(\partial(i) - a)i$ is a sum of terms having two degrees of homogeneity, namely $\partial(i)i$ of degree 0, and ai of degree d. But the operator in (3.203) has three degrees of homogeneity: $\partial(i)i + \varepsilon$ of degree 0, $\varepsilon\partial(i)$ of degree $-d$, and i of degree d.

Operators which have two degrees of homogeneity are termed "hypergeometric" in [58]. Their solutions generalize the classical notion of hypergeometric series. Operators having three degrees of homogeneity generalize the more mysterious class which includes Mathieu and Lamé functions.

As usual we write $g = \sum g^{jd}$. The recursion equation corresponding to (3.203) is

$$\varepsilon\partial(i)g^{(j+1)d} = (\partial(i)i + \varepsilon)g^{jd} - ig^{(j-1)d}. \tag{3.204}$$

Instead of formal power series which formed the essence of Theorem 3.26, we examine entire solutions. We want to use (3.204) to show that $g^{(j+1)d}$ is large. This amounts to showing that, in a suitable sense, the inverse of the operator $\partial(i)$ acting on homogeneous functions of order $(j + 1)d$ is about the same size as $\partial(i)i$ acting on functions which are homogeneous order jd. We also need to know that there is no cancelation on the right side.

A little consideration shows the difficulties in carrying out such a program. In the first place, $g^{(j+1)d}$ is not completely determined by (3.204). For we can add to $g^{(j+1)d}$ any harmonic which is homogeneous of degree $(j + 1)d$. Thus we need auxiliary conditions to establish an analog of Fischer's theorem.

Problem 3.12 *Find an analog of Fischer's theorem in the framework of (3.199).*

There is one other type of deformation which is of great interest. $f(x) = \sin x$ is a solution (for $i = x^2$) of

$$\begin{aligned} [\partial(i) + 1]f &= 0 \\ f &= 0 \quad \text{on } i = \pi^2. \end{aligned} \tag{3.205}$$

It is a classical result of Hermite that π is transcendental. One might wonder if there is a deformation theory in which rational or algebraic deformations (meaning adding lower order terms with rational or algebraic coefficients) are "small" (in a sense different from (3.199)).

Let us recall the famous Lindemann–Weierstrass theorem: if $\alpha_1, \ldots, \alpha_n$ are distinct algebraic numbers then $\exp(\alpha_1), \ldots, \exp(\alpha_n)$ are linearly independent over the ring of algebraic numbers.

This result can be shown to be more or less equivalent to

Proposition 3.27 *If ∂ is an ordinary linear differential operator of order n with constant coefficients, the coefficients being algebraic numbers, then any solution of $\partial f = 0$ cannot vanish at n algebraic points unless $f \equiv 0$.*

The example of $f = \sin x$ shows that n is best possible.

For partial differential equations we can prove the following result:

Theorem 3.28 *Let $n = 2$ and let i be real and elliptic; that is, 0 is the only real zero of i. Suppose i has real rational coefficients and a is a rational number. Then no solution of $[\partial(i) - 1]f = 0$ can vanish on $i = a$ unless $f \equiv 0$.*

In case $i = \sum t_j^2$ then, by decomposition in spherical harmonics, this result states that the zeros of Bessel functions are irrational. This result is due to C. L. Siegel [141] whose proof depends on the ordinary differential equations satisfied by Bessel functions. (Siegel actually proved the deeper result that the roots are transcendental.) Our proof is complicated and will be presented elsewhere.

Another interesting example of deformation occurs for certain elliptic operators. For example, solutions of the equation

$$\Delta + \lambda = 0$$

(Δ = Laplacian) cannot vanish on $t_1^2 + t_2^2 = \alpha$ if $\alpha > 0$ and λ is not positive. This is a consequence of the positivity of the operator $-\Delta$.

In the simplest case of $k = 1$ our assertion is that solutions of $d^2/dt^2 + \lambda = 0$ cannot vanish on $\{t = \pm\sqrt{\alpha}\}$ for $\alpha > 0$ unless $\lambda > 0$. Note that the variety

$$\hat{Y}_\lambda = \{\hat{t}^2 - \lambda = 0\}$$

is an orbit of the reflection group S_2 as is $\{t = \pm\sqrt{\alpha}\} = Y_\alpha$.

For $\alpha < 0$ the variety $\hat{Y}_\alpha$ is pure imaginary. This leads us to pose

Problem 3.13 *Let $\vec{\imath} = (i_1, \dots, i_k)$ be a strongly free system with real i_j. Let $\vec{c}, \vec{c}'$ be chosen such that*

$$\hat{X}_{\vec{c}} = \{\vec{\hat{\imath}} = \vec{c}\}$$

(which consists of a finite number of points) has no real points, and $X_{\vec{c}'}$ is real. Can the Fourier transform of any function supported on $\hat{X}_{\vec{c}}$ vanish on $X_{\vec{c}'}$?

We suspect that the solution to Problem 3.13 is in the affirmative. In particular it should be possible to prove the result when $X_{\vec{c}}$, $X_{\vec{c}'}$ are orbits of a reflection group, $X_{\vec{c}'}$ being the orbit of a real point and $\hat{X}_{\vec{c}}$ of a pure imaginary point.[6]

Theorem 3.28 and Problem 3.13 belong to the realm of *Fourier transform on varieties* (see Section 1.4): if $V_1 \subset \{\hat{t}\}$ and $V_2 \subset \{t\}$ are algebraic varieties (real or complex) then the Fourier transform defines a map from a suitable space of functions or measures μ on V_1 to functions or measures on V_2, namely

$$\mu \to \hat{\mu}|_{V_2}. \tag{3.206}$$

We call this map the *Fourier transform from V_1 to V_2*. This map makes sense when it is injective, as is the case in Theorem 3.28 and those cases of Problem 3.13 which have a positive solution.

Another example is given by the CP or any other boundary value problem. If V_2 is a Cauchy (or parametrization) surface for $\partial(V_1)$, meaning the differential system whose solutions (see the fundamental principle in Section 1.4) are Fourier integrals over V_1, then the Fourier transform from V_1 to V_2 is an isomorphism on suitable function spaces.

The theory of harmonic functions, as we have presented in Section 3.3 proscribes that the extension of $\exp(ip\theta)$ from the unit circle in $\mathbb{R}^2$ to all of $\mathbb{R}^2$ or $\mathbb{C}^2$ should be given by $r^{|p|}e^{(ip\theta)}$ as that is the harmonic function which agrees with $\exp(ip\theta)$ on that circle. Similarly in Section 3.4 we used $(x+y)^s$ to extend $\exp(s\xi)$ from the half-hyperbola $x^2 - y^2 = 1, x > 0$ to the quadrant $x \geq |y|$.

One major difference between $(x \pm iy)^p$ and $(x \pm y)^s$ is that the Fourier transform of $(x \pm iy)^p$ is a differential operator and so has support at a point. But it seems difficult to define the Fourier transform $\widehat{(x \pm y)^s}$ of $(x \pm y)^s$ directly; rather $\widehat{(x \pm y)^s}$ is defined via the Fourier transform.

It makes sense, therefore, to replace $(x \pm y)^s$ by different extensions of $\exp(\pm s\xi)$. We can also replace $(x \pm iy)^p$ by other functions which extend $\exp(ip\theta)$. We shall find some constructions of this type in Chapter 4.

3.6 Orbital integrals and Fourier transform

Whenever a linear space has a specific basis $\{v^j\}$ with a special element v^0, the "projection" on v^0 can be thought of as an integral. Thus in (3.13) the basis

[6]Some work of C. Berenstein and the author points to an affirmative solution when the reflection group is the symmetric group.

is $\{h^n\}$ and $v^0 = h^0 = 1$. We shall denote the integral in (3.13) more precisely by $\int_H \cdot$

Using the basis $\{h^n\}$ and the special function $h^0 = 1$ we can define the *harmonic Fourier transform* of a function f on an orbit $\mathcal{O}$ formally by

$$\hat{f_H}(\hat{t}) = \int_H f(t) \exp(\sqrt{-1} t \cdot \hat{t}). \tag{3.207}$$

(We have written $\sqrt{-1}$ because i is used in author context.) To make (3.207) meaningful we expand

$$\exp(\sqrt{-1} t \cdot \hat{t}) = \sum \frac{(\sqrt{-1} t \cdot \hat{t})^n}{n!}. \tag{3.208}$$

By the tensor product decomposition for polynomials in t

$$(t \cdot \hat{t})^n = \sum \alpha_{jn}(t, \hat{t}) h^j(t). \tag{3.209}$$

Here α_{jn} are invariants in t which depend on $\hat{t}$.

Combining (3.209) and (3.208) we obtain

$$\exp(\sqrt{-1} t \cdot \hat{t}) = \sum_j h^j(t) \sum_n \frac{\alpha_{jn}(t, \hat{t}) \sqrt{-1}^n}{n!}. \tag{3.210}$$

Since $(t \cdot \hat{t})^n$ is homogeneous of degree n only those j for which degree $h^j \leq n$ occur in (3.209). Thus the sum over j for fixed n is finite so α_{jn} is well defined. But $n \geq$ degree h^j so the sum over n in (3.210) could involve infinitely many terms and its convergence is unclear. In the group case described in Section 3.2 there is no problem of convergence.

We set (formally)

$$J_j(t, \hat{t}) = \sum \frac{\alpha_{jn}(t, \hat{t}) \sqrt{-1}^n}{n!}. \tag{3.211}$$

For the rotation group J_j is, up to a simple factor, the usual Bessel function. Since the $\alpha_{jn}(t, \hat{t})$ are invariants in t the function $J_j(t, \hat{t})$ depends only on the $\mathcal{O}$ orbit containing t, so we can write $J_j(\mathcal{O}, \hat{t})$.

To define (3.207) precisely, we replace f on $\mathcal{O}$ by the harmonic h equal to f. Then we define the harmonic Fourier transform by

$$\hat{f_H}(\hat{t}) = \sum J_j(\mathcal{O}, \hat{t}) \int_H h h^j. \tag{3.212}$$

The integral is certainly meaningful if h is a polynomial or in group cases.

How does (3.212) compare with the usual Fourier transform? In order to give meaning to this question we must define a measure on $\mathcal{O}$. In the group case we

would use the group invariant measure on $\mathcal{O}$. In this case it is clear that the usual Fourier transform agrees with the harmonic Fourier transform.

In general we want to associate a measure $df_{\mathcal{O}}$ on $\mathcal{O}$ with f so that the usual Fourier transform $\hat{df_{\mathcal{O}}}$ of $df_{\mathcal{O}}$ agrees with $\hat{f_H}$. By (3.210) and (3.212) we are forced to define $df_{\mathcal{O}}$ by

$$df_{\mathcal{O}} \cdot h^l = \int_H f h^l. \tag{3.213}$$

In particular $d1_{\mathcal{O}}$ would play the role of the Haar measure in the group case because, by (3.213), $d1_{\mathcal{O}}$ is orthogonal to all harmonics except the constant.

Since the h^l are linearly independent this defines $df_{\mathcal{O}}$ on polynomials h^l when f is a polynomial. In order to have an actual measure we need some conditions on $\vec{\imath}$.

Fourier conditions on $\vec{\imath}$

(**a**) $\vec{\imath}$ has real coefficients and is elliptic, meaning that the origin is the only real point on $\vec{\imath} = 0$.

Condition (**a**) implies that for $\vec{c}$ real the set of real points $\mathcal{O}^{\mathbb{R}}$ on $\mathcal{O} : \{\vec{\imath} = \vec{c}\}$ is compact. Moreover since $\vec{\imath}$ is real and homogeneous, for an open set of real $\vec{c}$ each complex component X of $\mathcal{O}$ has real part $X^{\mathbb{R}}$ of the same dimension as the complex dimension of X. For such $\vec{c}$ no holomorphic function on $\mathcal{O}$ can vanish on $\mathcal{O}^{\mathbb{R}}$.

(**b**) 1 is not the uniform limit on $\mathcal{O}^{\mathbb{R}}$ of harmonic polynomials with no constant term.

Condition (**b**) means that $df_{\mathcal{O}}$ defined by (3.213) is a continuous linear function on the polynomials on $\mathcal{O}$ with the uniform norm. For, let g_j be polynomials and $g_j \to g_0$ uniformly on $\mathcal{O}^{\mathbb{R}}$ so $f_j = fg_j \to f_0 = fg_0$. We write the harmonic expansions $f_j = \sum a_{jl} h^l$ where $h^0 = 1$. Then

$$\sum_{l \neq 0} (a_{jl} - a_{0l}) h^l + (a_{j0} - a_{00}) \to 0$$

uniformly. If $a_{j0} - a_{00}$ does not approach 0 on $\mathcal{O}^{\mathbb{R}}$ then we can divide this relation by $a_{j'0} - a_{00}$ on a subset $\{j'\}$ where $a_{j'0} - a_{00}$ is bounded away from 0. (Note that $a_{j'0}$ and a_{00} are constant on $\mathcal{O}$ so if $a_{j'0} - a_{00}$ is not small at some point of $\mathcal{O}$ it is uniformly large on $\mathcal{O}$.) This would contradict (**b**). We conclude that

$$a_{j0} = \int_H f g_j \to a_{00} = \int_H f g_0$$

which is our assertion.

The closure of the polynomials is the space of continuous functions. Thus $df_{\mathcal{O}}$ is an actual measure on $\mathcal{O}^{\mathbb{R}}$.

Proposition 3.29 *If Conditions* (**a**) *and* (**b**) *hold then* $\widehat{df}_{\mathcal{O}} = \hat{f}_H$ *whenever the series (3.211) converges uniformly on* $\mathcal{O}^R$.

Proposition 3.29 is just an interchange of summation and integration. Thus when Conditions (**a**) and (**b**) hold, the harmonic Fourier transform is injective.

Problem 3.14 *Find conditions on* $\vec{\imath}$ *which are sufficient for Condition* (**b**).

Let us note that the harmonic Fourier transform is far from injective, and hence uninteresting, in some cases. For example, if $k = 2$ and $r = 1$ and $i = t_2$ then the harmonics are the functions of t_1 and the invariants are the functions of t_2. A harmonic basis is $\{t_1^m\}$. From (3.209), (3.210), by an easy calculation, the Bessel function is

$$J_m(t, \hat{t}) = \frac{(\sqrt{-1}\hat{t}_1)^m}{m!} e^{\sqrt{-1}t_2\hat{t}_2}.$$

Thus the harmonic Fourier transform on $\mathcal{O} : \{t_2 = 1\}$ is determined by

$$(\hat{t_1^k})_H = \begin{cases} e^{\sqrt{-1}t_2\hat{t}_2} & \text{if } k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Problem 3.15 *Find conditions on* $\vec{\imath}$ *to guarantee the injectivity of the harmonic Fourier transform.*

The ordinary Fourier transform can be defined in a way which is analogous to the harmonic Fourier transform: the ordinary Fourier transform uses the basis $\exp(it \cdot \hat{t})$ and the special basis element 1.

A few words about the comparison between the ordinary Fourier transform and the harmonic Fourier transform are in order. Let $\chi_{\vec{c}}$ (thought of as a measure) be the characteristic function of $\vec{\imath} = \vec{c}$ or, more precisely, some minor modification of the euclidean measure on $\vec{\imath} = \vec{c}$. One might be tempted to think that $\chi_{\vec{c}}$, or some other nicely defined positive measure on $\vec{\imath} = \vec{c}$, is the $1_{\mathcal{O}}$ of (3.213). This is certainly true in the case of groups because $\chi_{\vec{c}}$ is group invariant and hence is orthogonal to all harmonics except the constant.

But this may no longer be true in general even in very "favorable" situations. For example, in the elliptic case, meaning $\{\vec{\imath} = \vec{c}\}$ is compact, we have

$$\chi_{\vec{c}} = \int e^{i\alpha_1(i_1 - c_1) + \cdots + i\alpha_r(i_r - c_r)} \, d\alpha_1 \ldots d\alpha_r. \tag{3.214}$$

($\chi_{\vec{c}}$ is not the euclidean measure but is the euclidean measure divided by a suitable Jacobian formed from the i_l. In case $r = 1$ we divide by $|\text{grad} i|$.) From this expression, or from many other viewpoints, $\chi_{\vec{c}}$ is a limit of polynomials in $\vec{\imath}$ and $\vec{c}$. Such polynomials are orthogonal to the nonconstant harmonics *in the Fischer norm* but possibly not in the usual sense of integration. (Of course we cannot integrate the product of polynomials over R^n directly, but we could modify (3.214) by adding a convergence factor which is a function of $\vec{i}$, e.g. $\exp[-\varepsilon \sum (i_l - c_l)^2]$.) This leads us to suspect that $\chi_{\vec{c}}$ is not orthogonal to the harmonics in the usual sense of integration except in the group case.

Problem 3.16 *For* $r = 1, i = t_1^4 + t_2^4$, *is* χ_1 *orthogonal to the harmonics?*

4

HARMONIC FUNCTIONS AND RADON TRANSFORM ON ALGEBRAIC VARIETIES

The harmonic function theory of Chapter 3 is applied to functions on algebraic varieties. Harmonicity gives rise to Cauchy problems (CPs) and other parametrization problems for which we have developed the Watergate method (Section 1.4). Section 4.1 treats the "finite CP," meaning that a finite number of Cauchy data are prescribed; it is dual to a Radon transform on algebraic varieties.

Infinitely many Cauchy data have to be prescribed when the Cauchy surface is of lower dimension than the natural dimension associated to solutions of the equation. In this case the CP is called the Watergate problem (WP); the simplest example of the WP is the wave equation with data given on the time axis. In Section 4.2 we treat the compact WP for which the data is naturally associated to a compact manifold, while Section 4.3 deals with the noncompact WP. The solutions of the WP depend on the harmonic function theory developed in Chapter 3.

4.1 Algebraic theory and finite Cauchy problem

In Chapter 3 we were concerned with decomposing $\hat{T} = \mathbb{R}^k$ into orbits $\{\vec{\imath} = \vec{c}\}$ and finding a consistent (harmonious) way of parametrizing functions on the orbits. (In this chapter we change notation using $\hat{T}$ in place of T.) We now introduce the *pulling apart* process on orbits, meaning that we introduce another space $\hat{Y}$ and place the orbits over the points $\hat{y} \in \hat{Y}$ in an "algebraic" manner which we shall make precise presently. Each $\hat{y}$ corresponds to a $\vec{c}$ but the maps $\hat{y} \mapsto \vec{c}$ may not be one-to-one.

From an abstract viewpoint we start with a covering of $\mathbb{R}^k$ by sets $\hat{L}_s$. We can think of $\{\hat{L}_s\}$ as the leaves of a spread but more generally we allow the $\hat{L}_s$ to have "small" intersections. We suppose that the parameter space $S = \{s\}$ is covered[1] by a euclidean space $\hat{Y}$. We *pull apart* the sets $\hat{L}_s$ by the mapping

$$\hat{L}_s \to \{[\hat{L}_s, \hat{y}(s)]\} \subset \mathbb{R}^k \times \hat{Y}$$

where $\{\hat{y}(s)\}$ are the points in $\hat{Y}$ which cover the point s.

[1]The word "covering" has two connotations: $\mathbb{R}^k$ being covered by $\{\hat{L}_s\}$ means $\mathbb{R}^k$ is the union of the $\hat{L}_s$, while $\hat{Y}$ is a covering space of the parameter space S. Thus the $\hat{L}_s$ are injected into $\mathbb{R}^k$ while $\hat{Y}$ is projected on S.

Remark 1 *The pulling apart transforms the orbits into a variety in* $\mathbb{R}^k \times \hat{Y}$ *to which we can apply the Watergate method (Section 1.4).*

To clarify this concept we give two examples.

Example 1 $\{\hat{L}_s\} = \{O_c\}$ *where* O_c *is the sphere* $\hat{t}^2 = c^2$.

$\hat{Y}$ *is one dimensional; we set* $\hat{y}^2 = c^2$ *so each* $s = c$ *has two points* $\hat{y} = \hat{y}(c)$ *"above" it. The sphere* $\hat{t}^2 = c^2$ *in* $\mathbb{R}^k$ *is mapped into*

$$\{\hat{t}, \pm c\}_{\hat{t}^2 = c^2}.$$

The union of the images forms the light cone $\hat{t}^2 = \hat{y}^2$ *in* $\mathbb{R}^k \times \mathbb{R}$. *(*$\hat{t}$ *is the space variable and* $\hat{y}$ *is the time variable.)*

Example 2 $\{\hat{L}_s\}$ *is the set of lines through the origin on* $\mathbb{R}^k$. *We call* $\hat{Y} = \mathbb{R}^k$ *and we set* $\{\hat{y}(s)\}$ *= the set of direction numbers of* $\hat{L}_s$, *so* $\{\hat{y}(s)\}$ *is a line.*

In fact $\{\hat{y}(s)\}$ *is the same line in* $\hat{Y}$ *as* $\hat{L}_s$ *is in* $\mathbb{R}^k(\hat{t})$.

Our pulling apart is

$$\hat{L}_s \to \{(\hat{L}_s, \hat{y}(s))\}.$$

The image of $\mathbb{R}^k$ *is the set* $\{(\hat{t}, \hat{y})\}$ *where* $\hat{t} \parallel \hat{y}$. *Thus the pulling apart transfers the set of lines through the origin, which is the fundamental construct of the Fourier transform of the nonparametric Radon transform (Chapter 2), into the set of parallel lines which is the fundamental construct of the parametric Radon transform (Chapter 6).*

In this chapter we shall concentrate on $\hat{L}_s$ of the form $O_{\vec{c}} = \{\vec{i} = \vec{c}\}$.

Let us begin with a formal introduction to our methods. The simplest way of pulling apart is to introduce r (=number of components of $\vec{i}$) strongly free polynomials $\vec{j}(\hat{y}) = j_1(\hat{y}), \ldots, j_r(\hat{y})$ and construct the algebraic variety

$$\begin{aligned} V &= \{\vec{\imath}(\hat{t}) - \vec{j}(\hat{y}) = 0\} \\ &= \{\vec{P}(\hat{t}, \hat{y}) = 0\}. \end{aligned} \tag{4.1}$$

We assume for the present that V is a variety, not a multiplicity variety (Section 1.4). For each $\hat{y}^0$ we have an orbit

$$\begin{aligned} \mathcal{O}_{\hat{y}^0} : \vec{\imath}(\hat{t}) &= \vec{j}(\hat{y}^0) \\ &= V \cap \{\hat{y} = \hat{y}^0\}. \end{aligned}$$

Since the $\hat{t}$ orbits are defined by the constancy of $\vec{\imath}(\hat{t})$ it follows that for each $\hat{y}^0$ the orbit $\mathcal{O}_{\hat{y}^0}$ is an orbit in $\hat{t}$ space.

These orbits are the same for all $\hat{y}$ for which $\vec{j}(\hat{y}) = \vec{j}(\hat{y}^0)$. Moreover since $\vec{j}$ is strongly free, for every $\vec{c} \in \mathbb{C}^r$ the variety $\vec{j}(\hat{y}) = \vec{c}$ is nonempty, so the map $\hat{y} \mapsto \vec{j}(\hat{y})$ is surjective on $\mathbb{C}^r$. Thus every orbit

$$\mathcal{O}(\vec{\imath}) = \{\vec{\imath}(\vec{t}) = \vec{c}\}$$

is lifted to orbits $\mathcal{O}_{\hat{y}}$ in V.

In this way we have pulled apart the orbits $\mathcal{O}(\vec{\imath})$ in $\hat{T}$ and placed them in V (in a possibly nonsingle-valued manner).

We now apply the Watergate method.

If $U(\hat{t},\hat{y})$ is a polynomial then for each $\hat{y}^0$ we can write the harmonic expansion of $\hat{U}$ on $\mathcal{O}_{\hat{y}^0}$ thought of as an orbit in $\hat{t}$ for this fixed $\hat{y}^0$:

$$\hat{U}(\hat{t},\hat{y}^0) = \sum \hat{S}_l(\vec{\imath}(\hat{t}),\hat{y}^0)\, h_l(\hat{t}) \quad \text{on } \mathcal{O}_{\hat{y}^0}. \tag{4.2}$$

$\{h_l\}$ are the harmonics for $\vec{\imath}$ which are extended to functions of $(\hat{t},\hat{y})$ by making them constant in $\hat{y}$.

Remark. *We must be careful to distinguish two types of harmonics for $\vec{\imath}$. Since $\vec{\imath} = \vec{\imath}(\hat{t})$ there is the theory of harmonics for $\vec{\imath}$ as discussed in Chapter 3; these are functions h of $\hat{t}$ only, satisfying $\vec{\imath}(\partial/\partial\hat{t})h = 0$.*

We can also think of harmonicity in $\hat{t},\hat{y}$. Since $\vec{\imath}$ does not involve $\hat{y}$ any function $\widetilde{h}(\hat{y})$ satisfies $\vec{\imath}(\partial/\partial\hat{t})\widetilde{h} = 0$. The tensor product $\{h\}\otimes\{\widetilde{h}\}$ is the kernel of $\vec{\imath}(\partial/\partial\hat{t})$.

We shall use the word "harmonic" to refer to $\{h(t)\}$. The harmonics which are functions of $(\hat{t},\hat{y})$ will be called $(\hat{t},\hat{y})$ *harmonics.*

Using (4.1) we can rewrite (4.2) as

$$\hat{U}(\hat{t},\hat{y}) = \sum \hat{S}_l(\vec{j}(\hat{y}),\hat{y})h_l(\hat{t}) \quad \text{on } V. \tag{4.3}$$

We shall obtain similar expansions for U on varieties which are somewhat more general than (4.1).

The algebraic theory works nicely for polynomials U but if, for example, $\hat{U}\in\hat{\mathcal{E}}'(\hat{t},\hat{y})$ then we would want $\hat{S}_l(\vec{j}(\hat{y}),\hat{y})\in\hat{\mathcal{E}}'(\hat{y})$. This is true only for special varieties V called *hyperbolic*.

This construction can be viewed within the framework of the Radon transform. The orbits $\mathcal{O}_{\hat{y}}$ are the leaves of a single spread on V. We regard this spread as a multiplicity projection spread with multiplicity $\{h_l(\hat{t})\}$. We can think of l as a Grassmann parameter and $\hat{y}$ as a spread parameter.

Suppose that $r = k = \dim \hat{t}$. Then the number of points ${}_p\hat{t}(\hat{y})$ in a generic orbit is $N < \infty$. If $\hat{U}(\hat{t},\hat{x})$ is a holomorphic function on V then we define its multiplicity Radon transform $\mathbf{R}_V\hat{U}$ relative to this spread by

$$\mathbf{R}_V\hat{U}(\hat{y},l) = \sum_p \hat{U}({}_p\hat{t}(\hat{y}),\hat{y})h_l({}_p\hat{t}(\hat{y})) \tag{4.4}$$

for generic $\hat{y}$ (meaning $\mathcal{O}_{\hat{y}}$ is generic). We shall see that $\mathbf{R}_V U$ extends to be entire in $\hat{y}$.

Let us compute the adjoint $\mathbf{R}_V'$ of $\mathbf{R}_V$. $\mathbf{R}_V$ maps functions on V into N-tuples of functions on $\hat{Y}$. Let $\vec{\hat{w}} = \{\hat{w}_l\}$ be suitable functions of $\hat{y}$. We use a convenient

measure $d\hat{y}$ on $\hat{Y}$ to define a measure on V by

$$\int_V \alpha(\hat{t},\hat{y}) = \int \sum \alpha({}_p\hat{t}(\hat{y}),\hat{y})\, d\hat{y}.$$

We have

$$\begin{aligned}\mathbf{R}'_V\vec{\hat{w}}\cdot\hat{U} &= \sum_l \int \hat{w}_l(\hat{y}) \sum_p \hat{U}({}_p\hat{t}(\hat{y}),\hat{y}) h_l({}_p\hat{t}(\hat{y}))\, d\hat{y} \\ &= \int \sum_p \left[\sum_l \hat{w}_l(\hat{y}) h_l({}_p\hat{t}(\hat{y}),\hat{y})\right] \hat{U}({}_p\hat{t}(\hat{y}),\hat{y})\, d\hat{y} \\ &= \int_V \left[\sum_l h_l(\hat{t})\hat{w}_l(\hat{y})\right] \hat{U}(\hat{t},\hat{y}).\end{aligned}$$

This means that, if we define $\mathbf{R}^*_V$ as $\mathbf{R}'_V$ using the identification of functions α on V with measures $\alpha d\hat{y}$ (see Section 1.1),

$$\mathbf{R}^*_V\vec{\hat{w}} = \sum h_l(\hat{t})\hat{w}_l(\hat{y}). \tag{4.5}$$

In particular if $\vec{\hat{w}} = \mathbf{R}_V\hat{U}$ then by (4.4)

$$\mathbf{R}^*_V\mathbf{R}_V\hat{U}(\hat{t},\hat{y}) = \sum_l h_l(\hat{t}) \sum_p \hat{U}({}_p\hat{t}(\hat{y}),\hat{y}) h_l({}_p\hat{t}(\hat{y})). \tag{4.6}$$

According to the general principle of Radon transform as expounded in Chapter 1, $\mathbf{R}^*_V\mathbf{R}_V\hat{U}$ should be the harmonic function in $\hat{t}$ whose "restriction to the parametrization (Cauchy) surface $\hat{Y}$" is $\mathbf{R}_V\hat{U}$. "Restriction to $\hat{Y}$" is to be interpreted as the set of coefficients of $\{h_l(\hat{t})\}$; a comparison of (4.4) and (4.6) shows the agreement with the general principle. (Recall that for the classical Radon transform $\mathbf{R}^*_\mathbf{g}\mathbf{R}_\mathbf{g}f$ is the integral over the parametrization surface S (which in the present case is Y) of the basis elements $\delta_{L(s,\mathbf{g})}$ with coefficients $\mathbf{R}f(s,\mathbf{g})$. The role of $\{\delta_{L(s,\mathbf{g})}\}$ is played by the basis $\{h_l\}$ or, more precisely, by $\{h_l\delta_{\hat{y}}\}$ with $s \leftrightarrow \hat{y}$ and $l \leftrightarrow \mathbf{g}$.) Moreover we see from (4.6) that the value of $\mathbf{R}^*\mathbf{R}\hat{U}$ at the point $({}_q\hat{t}(\hat{y}),\hat{y}) \in V$ depends only on the values of $\hat{U}$ on the leaf $\mathcal{O}_{\hat{y}}$ containing the given point, as in the usual Radon transform.

Call $\boxed{h(\hat{y})}$ the matrix whose $p\,l$ entry is $\{h_l({}_p\hat{t}(\hat{y}))\}$. Since the $h_l(\hat{t})$ form a basis on each orbit, det $\boxed{h}$ is generically $\neq 0$.

We think of $\mathbf{R}_V\hat{U}$ as a vector whose l component is $\sum_p \hat{U}({}_p\hat{t}(\hat{y}),\hat{y})h_l({}_p\hat{t}(\hat{y}))$. Thus we can write

$$\begin{aligned}\mathbf{R}_V\hat{U}(\hat{y},l)\boxed{h}^{-1}(\hat{y}) &= \sum_{p,l}\hat{U}({}_p\hat{t}(\hat{y}),\hat{y})h_l({}_p\hat{t})(\hat{y})\boxed{h}^{-1}_{lq}(\hat{y})\\ &= \sum_p \hat{U}({}_p\hat{t}(\hat{y}),\hat{y})\delta_{pq}\\ &= \hat{U}({}_q\hat{t}(\hat{y}),\hat{y}).\end{aligned} \tag{4.7}$$

In this way we can recover $\hat{U}$ on V from $\mathbf{R}_V U$.

We denote by $\mathbf{R}_{Vl}$ the "Radon transform"

$$\mathbf{R}_{Vl}\hat{U}(\hat{y}) = \mathbf{R}_V U(\hat{y},l).$$

The companion to (4.6) is

$$\mathbf{R}^*_{Vl}\mathbf{R}_{Vl}U(\hat{t},\hat{y}) = h_l(\hat{t})\sum_p U({}_p\hat{t}(\hat{y}),\hat{y})h_l({}_p\hat{t}(\hat{y})). \tag{4.6*}$$

As in the case of the ordinary Radon transform we want to recover $\hat{U}$ on V from $\{\mathbf{R}^*_{Vl}\mathbf{R}_{Vl}\hat{U}\}$. We regard $\mathbf{R}^*_{Vl}\mathbf{R}_{Vl}\hat{U}$ as a matrix whose ql component is

$$h_l({}_q\hat{t}(\hat{y}))\sum_p \hat{U}({}_p\hat{t}(\hat{y}),\hat{y})h_l({}_p\hat{t}(\hat{y})).$$

We apply the inverse matrix $\boxed{h}^{-1}_{mq}$ on the left where we sum over q, which results in the matrix

$$\sum_p \hat{U}({}_p\hat{t}(\hat{y}),\hat{y})h_l({}_p\hat{t}(\hat{y}))\delta_{ml}.$$

This is a diagonal matrix; we consider the diagonal values as the components of a row vector with index m. Now we apply $\boxed{h}^{-1}_{mu}$ on the right. We obtain the vector

$$\sum_p \hat{U}({}_p\hat{t}(\hat{y}),\hat{y})\delta_{pu}$$

which can be interpreted as $\hat{U}({}_u\hat{t}(\hat{y}),\hat{y})$. Let us denote the operator $\boxed{h}^{-1}$ when operating on a diagonal matrix in the manner indicated by $\widetilde{\boxed{h}}^{-1}$. Then we can express the passage from $\mathbf{R}^*_V\mathbf{R}_V\hat{U}$ to $\hat{U}$ schematically as

$$\hat{U}({}_u\hat{t}(\hat{y}),\hat{y}) = \sum_m \boxed{h}^{-1}_{mu}\sum_q \boxed{h}^{-1}_{mq}\mathbf{R}^*_{Vl}\mathbf{R}_{Vl}\hat{U}({}_q\hat{t}(\hat{y}),\hat{y}) \tag{4.8}$$

or

$$\hat{U} = \widetilde{\boxed{h}}^{-1} \boxed{h}^{-1} \mathbf{R}_V^* \mathbf{R}_V \hat{U} \qquad \text{on } V.$$

We conclude:

The operator $\widetilde{\boxed{h}}^{-1} \boxed{h}^{-1}$ *is the analog for* $\mathbf{R}_V$ *of the powers of the Laplacian for* $\mathbf{R}$ *(see Section 2.1).*

Problem 4.1 *What is the conceptual relation between* $\widetilde{\boxed{h}}^{-1} \boxed{h}^{-1}$ *and the power of the Laplacian?*

We can use the same method to compute the coefficients $\hat{S}_m(\vec{j}(\hat{y}), \hat{y}) = \hat{S}_m(\hat{y})$. By (4.3)

$$\begin{aligned} \hat{S}_l(\hat{y}) &= \sum \boxed{h}_{lp}^{-1} \hat{U}({}_p\hat{t}(\hat{y}), \hat{y}) \\ &= \boxed{h}^{-1}(\hat{y}) \hat{U}(\vec{\hat{t}}(\hat{y}), \hat{y}). \end{aligned} \tag{4.9}$$

One can think of (4.9) in slightly different terms. Since (4.3) holds on V it holds on $\mathcal{O}_{\hat{y}}$ for each $\hat{y}$. For generic $\hat{y}$ there are N values of ${}_p\hat{t} \in \mathcal{O}_{\hat{y}}$. This gives N equations for the N values $\hat{S}_l(\hat{y})$. (4.9) is the solution of this system of equations.

It is important to study the Fourier transform of $\mathbf{R}_V$. Suppose $u \in \mathcal{E}'(t, y)$. The difference of the two sides of (4.3) belongs to the ideal generated by the P_j so the Fourier transform of (4.3) is

$$U(t, y) = \sum S_l(y) h_l \left(i \frac{\partial}{\partial t} \right) \delta_{t=0} + \vec{P} \left(i \frac{\partial}{\partial t}, i \frac{\partial}{\partial y} \right) v(t, y) \tag{4.10}$$

where under suitable hyperbolicity conditions (to be discussed below) $S_l \in \mathcal{E}'(y)$ and $v \in \mathcal{E}'(t, y)$.

Equation (4.10) is dual to the Cauchy map (Section 1.4)

$$\begin{aligned} &\mathcal{C}' : f \to \left[\vec{h} \left(i \frac{\partial}{\partial t} \right) f|_{t=0}, \vec{P} \left(i \frac{\partial}{\partial t}, i \frac{\partial}{\partial y} \right) f \right] \\ &\mathcal{C}' : \mathcal{E}(t, y) \to [\mathcal{E}(y)]^N \oplus \mathcal{E}^{\vec{P}} \end{aligned}$$

where $\mathcal{E}^{\vec{P}}$ is the subspace of $\mathcal{E}^r$ defined by the compatibility of the P_j.

The components $\hat{S}_l$ of $\hat{\mathcal{C}}\hat{U}$ are given by (4.9).

Remark. *An analogous procedure can be used in some cases when the orbits* $\mathcal{O}_{\hat{y}}$ *are infinite. In the infinite case our results are restricted to* $\mathcal{O}_{\hat{y}}$ *which are orbits of suitable Lie groups (see Sections 4.2 and 4.3).*

The opposite of pulling apart is *pushing together*. Suppose that the orbits in the $\hat{t}$ plane defined by $\vec{\imath}(\hat{t}) = \vec{c}$ pull apart to orbits $\mathcal{O}_{\hat{y}}$ in V over $\hat{y}$ and also orbits in the $\hat{y}$ plane defined by $\vec{j}(\hat{y}) = \vec{c_1}$ pull apart to orbits $\widetilde{\mathcal{O}}_{\hat{t}}$ in V. Then we can introduce a *correspondence* between $\hat{t}$ and $\hat{y}$ orbits as follows. Start with an orbit $\mathcal{O}_{\hat{y}} \subset V$. The point $\hat{y}$ lies in an orbit $\widetilde{\mathcal{O}}(\hat{y})$ in $\hat{Y}$ which we say *corresponds to* $\mathcal{O}_{\hat{y}}$ *under* V. We call $\widetilde{\mathcal{O}}(\hat{y})$ the *push down* of $\mathcal{O}_{\hat{y}}$.

This concept is useful when, as in (4.1),

$$\mathcal{O}_{\hat{y}^1} = \mathcal{O}_{\hat{y}}$$

for any $\hat{y}^1 \in \widetilde{\mathcal{O}}(\hat{y})$. In this case we say V has a *double orbit structure*.

Let us give some illustrations of the above ideas.

Let $r = 1$, $\vec{P} = P = \hat{y}^2 - \hat{t}_1^2 - \hat{t}_2^2$. (We have adjusted the notation to conform with later developments.) As an example of $\vec{\imath}$ we take

$$i = \hat{t}_1^2 + \hat{t}_2^2.$$

(This example violates the condition that for each $\hat{y}$ the number of $\hat{t}(\hat{y})$ is finite.)

A harmonic function for i is an ordinary harmonic function in $\hat{t}$. A basis for harmonics is $\{\hat{t}_1 \pm i\hat{t}_2\}^l$. Any entire function $F(\hat{t}, \hat{y})$ on V is uniquely expressible in the form

$$F(\hat{t}, \hat{y}) = \sum F_l^{\pm}(\hat{y})(\hat{t}_1 \pm i\hat{t}_2)^l. \tag{4.11}$$

This decomposition will be discussed in detail later and depends on the harmonic expansion discussed in Section 3.3. It corresponds to Fourier series on the orbits.

The $\hat{t}$ orbits are the complex spheres $\hat{t}_1^2 + \hat{t}_2^2 = c^2$. For $\hat{y} = \pm c$ these spheres pull apart to $\mathcal{O}_{\hat{y}}$. The orbit of $\hat{y}$ using $\hat{j}(\hat{y}) = \hat{y}^2$ is $\mathcal{O}(\hat{y}) = \pm\hat{y}$ so we have a double orbit structure.

For another example let us look at the same variety from a different point of view.

We now write

$$P = (\hat{t}^2 - \hat{y}_1^2 - \hat{y}_2^2)$$
$$i = \hat{t}^2.$$

This is also a double orbit structure. The harmonics are now linear functions of $\hat{t}$ which are independent of $\hat{y}$. Harmonic extension takes the form

$$F(\hat{t}, \hat{y}) = F_0(\hat{y}) + F_1(\hat{y})\hat{t} \qquad \text{on } V.$$

Let us pass to a precise theory. In conformity with (4.3) we search for a tensor product decomposition of "any" function $\hat{U}(\hat{t}, \hat{y})$ in the form

$$\hat{U}(\hat{t}, \hat{y}) = \sum \hat{S}_j(\hat{y})h_j(\hat{t}) \quad \text{on } V. \tag{4.12}$$

We consider varieties V which are of somewhat more general form than (4.1). We require

Assumption. *V is defined by equations*

$$i_m(\hat{t}) = \sum a_{mq}(\hat{y}) v_{mq}(\hat{t}) \quad m = 1, 2, \ldots, k \tag{4.13}$$

where $\vec{\imath}$ is strongly free and

$$\begin{aligned} deg\ v_{mq} &< deg\ i_m \\ deg\ a_{mq} + deg\ v_{mq} &\leq deg\ i_m. \end{aligned} \tag{4.14}$$

The assumption allows us to apply the theory of Chapter 3 to the present situation. For, when we fix $\hat{y} = \hat{y}_0$, the intersection $\mathcal{O}_{\hat{y}_0}$ of V with $\{\hat{y} = \hat{y}_0\}$ is defined by equations

$$i_m(\hat{t}) = \sum a_{mq}(\hat{y}_0) v_{mq}(\hat{t}) \tag{4.13*}$$

where $a_{mq}(\hat{y}_0)$ are constants and deg $v_{mq} <$ deg i_m. This means that $O_{\hat{y}_0}$ is of the form X_Y as in (3.14). Hence, we can think of $O_{\hat{y}_0}$ as "orbits" for $\vec{\imath}$ on $\hat{T}$. (These are not necessarily orbits of the form $\vec{\imath} = \vec{c}$ but rather, as (4.13) and (4.14) show, they are of the form $\vec{\imath}(\hat{t}) = \vec{\gamma}(\hat{t}, \hat{y})$ where $\vec{\gamma}$ is a polynomial in $\hat{t}$ of degree < degree $\vec{\imath}$ whose coefficients are polynomials in $\hat{y}$.) In particular, since $\vec{\imath}$ is strongly free, the DP for $\partial(\vec{\imath})$ is well defined on $O_{\hat{y}}$.

Equation (4.12) is central to our work. We want to show that it is a consequence of (4.13) and (4.14).

We demonstrate (4.12) for $\hat{U}$ a polynomial. More general functions are considered below.

Since $\vec{\imath}$ is strongly free, we can use the tensor product for $\vec{\imath}$ in $(\hat{t}, \hat{y})$ space (Proposition 3.5) to decompose $\hat{U}$ (note that we are now using $(\hat{t}, \hat{y})$ harmonic functions; they are $\vec{\imath}$ harmonic in $\hat{t}$ and arbitrary in $\hat{y}$) as

$$\hat{U}(\hat{t}, \hat{y}) = \sum \hat{A}_j(\hat{y}) B_j(\vec{\imath}) h_j(\hat{t}). \tag{4.15}$$

The degree of the right side in $\hat{t}, \hat{y}$ is the same as the degree of $\hat{U}$. (We assume for the present that $\hat{U}$ is homogeneous.)

Suppose there is a nontrivial term $B_{j_0}(\vec{\imath})$. Then we can replace $\vec{\imath}$ according to (4.13). This replaces $B_{j_0}(\vec{\imath}) h_{j_0}(\vec{t})$ by

$$B_{j_0}(\vec{\imath}) h_{j_0}(\hat{t}) = \sum \alpha^k_{j_0}(\hat{t}) \beta^k_{j_0}(\hat{y}) \quad \text{on } V. \tag{4.16}$$

By (4.14) the degree of $\alpha^k_{j_0}(\hat{t}) \beta^k_{j_0}(\hat{y})$ in $\hat{t}$ is less than that of $B_{j_0} h_{j_0}$ in $\hat{t}$. We can apply Proposition 3.5 to decompose, in the ring of polynomials in $\hat{t}$, the highest homogeneous part of $\alpha^k_{j_0}(\hat{t})$ using $\vec{\imath}$ harmonics. The degree in $\hat{t}$ of any $B'_j(\vec{\imath}) h_j(\hat{t})$ that occurs is less than the degree of $B_{j_0} h_{j_0}$ in $\hat{t}$. We can continue

this process as long as B'_j is not a constant. We then deal in a similar manner with the other homogeneous parts of $\alpha^k_{j_0}$.

Finally, if $\hat{U}$ is not homogeneous we treat each homogeneous part separately.

We conclude with an expression of the form (4.15) in which B_j are constants. We have verified (4.12).

Remark. *This is an instance in which the tensor product works even though the degree of the decomposition, i.e. the right side of (4.13), may be equal to the degree of i_m.*

At this point it is essential to distinguish three cases:

(1) *The number of points in $\mathcal{O}_{\hat{y}}$ is finite.*
(2) *Each $\mathcal{O}_{\hat{y}}$ is an irreducible algebraic variety whose real points form an orbit of a compact Lie group W.*
(3) *Each $\mathcal{O}_{\hat{y}}$ is an irreducible algebraic variety whose real points form an orbit of a noncompact Lie group.*

Of course, the cases are listed in ascending order of difficulty. We shall meet examples of all three in the sequel. We present here a sketch of our techniques.

In case (1) we can express the coefficients $\hat{S}_j(\hat{y})$ of $\hat{U}$ in terms of algebraic expressions in the values of $\hat{U}$ on $\mathcal{O}_{\hat{y}}$ (Watergate method). What makes (1) essentially simpler than (2) and (3) is that it is not difficult to study directly the analytic nature of the $\hat{S}_j$ in terms of these algebraic expressions.

In cases (2) and (3) we have expressions for $\hat{S}_j(\hat{y})$ only for $\hat{y}$ real; they involve integrals of $\hat{U}$ over $\mathcal{O}_{\hat{y}} \cap \{\text{real}\}$. One could attempt to extend such formulas into the complex plane but such an extension seems rather hard to handle. Instead we use the group structure and Fourier analysis on the group.

All this refers to strongly free $\vec{\imath}$ and nondegenerate orbits. Presumably there is an analog of the degenerate orbit ideas discussed at the end of Section 3.1, but we do not know how to proceed except in some examples which appear below.

Lemma 4.1 *Suppose that the harmonics form a finite dimensional space H. Then, on V, i.e. modulo the ideal of V, each $\hat{t}_l$ is integral over that ring of polynomials in $\hat{y}$.*

Proof Equation (4.12) shows that for any l, N we can write

$$\hat{t}_l^N = \sum \hat{S}_j^{N,l}(\hat{y}) h_j(\hat{t}) \quad \text{on } V. \tag{4.17}$$

Suppose

$$N > (\max \deg\, h_j)$$

Then the degree of any of the terms on the right side of (4.16) in $\hat{t}$ is $< N$. This means that all $\hat{t}_l$ are integral.

Actually we have the stronger

Theorem 4.2 *Suppose that the harmonics form a finite dimensional space. Then $t = 0$ is noncharacteristic for the system of differential equations defined by V.*

In the language of FA this means that

$$|\hat{t}| \leq c(1 + |\hat{y}|) \quad \text{for } (\hat{t}, \hat{y}) \in V. \tag{4.18}$$

Proof We may assume $\dim Y > 0$ for otherwise the result is trivial.

The proof would be simple if we could show directly from (4.13) and (4.14) that the degree of the right side of (4.16) is $\leq N$. We do not know how to accomplish this so we use a somewhat different method.

Equations (4.13) hold on V. By raising to suitable powers l_m we can assume that all $k_m = i_m^{l_m}$ have the same degree M and that equations like (4.13) with the degree conditions (4.14) hold for them.

Since $\dim H < \infty$ the origin is the only common zero of the k_m. Since the k_m are homogeneous of degree M it follows that on every real complex ray $\rho = \{r\alpha\}_{r\in\mathbb{R}}$ where α is a point in the complex unit sphere there is an m such that

$$|k_m(r\alpha)| \geq cr^M \tag{4.19}$$

for some $c > 0$ which is independent of α.

The noncharacteristic condition can be reformulated as saying that the points at infinity on $\hat{Y}$ do not belong to V. For, if $\hat{Y}$ is characteristic then there exist arbitrarily large $(\hat{t}, \hat{y})$ for which

$$|\hat{y}| \leq \varepsilon|\hat{t}|.$$

By a "compactness argument" this means that there exists a ray ρ in $\hat{T}$ above which there exist $\hat{y}$, i.e. $(\hat{t}, \hat{y}) \in V$, satisfying

$$|\hat{y}| \leq c|\hat{t}^{\gamma}|$$

for some $\gamma < 1$.

We examine (4.13), or rather its analog for k_m, for $\hat{t} \in \rho$ and these $\hat{y}$. By (4.19) there is an m for which

$$k_m(\hat{t}) \geq c|\hat{t}|^M.$$

On the other hand the right side of (4.13) is of degree $\leq M-1$ in $\hat{t}$ and of degree $\leq M$ in all variables so it is bounded by

$$c|\hat{t}|^{M-1+\gamma}.$$

Since $\gamma < 1$ this is a contradiction.

Theorem 4.2 is proven.

As in FA we denote by $\mathcal{H}$ the space of entire functions with the topology of uniform convergence on compact sets. $\mathcal{H}'$ is the dual of $\mathcal{H}$ and $\hat{\mathcal{H}}'$ is the Fourier

transform of $\mathcal{H}'$. Thus $\hat{\mathcal{H}}'$ consists of all entire functions of exponential type. The topology of $\hat{\mathcal{H}}'$ is defined by means of the semi-norms

$$\|\hat{S}\|_k = \max \frac{|\hat{S}(\hat{x})|}{k(\hat{x})}$$

for any continuous positive function $k(\hat{x})$ which is larger at infinity than any $\exp(c|\hat{x}|)$.

Theorem 4.3 *Assume that V is defined by equations of the form (4.13) with the degree conditions (4.14). Assume also that $\vec{\imath}$ is strongly free and the space H of harmonics is finite dimensional. Then $\hat{\mathcal{H}}'(V)$ is the tensor product $H \otimes \hat{\mathcal{H}}'(Y)$. Thus the CP for the system of differential equations corresponding to V with data on Y and "normal derivatives" $h_j(\partial/\partial t)$ is well posed for $\mathcal{H}$.*

Proof We need to derive bounds in (4.12) and (4.15).

Suppose that there is a generic point $\hat{y}^0$ for the fibers $O_{\hat{y}^0}$, meaning that $O_{\hat{y}^0}$ is a variety and not a multiplicity variety; the general case involves only a slight technical modification. Since $\vec{\imath}$ is strongly free the $h_j(\hat{t})$ form a basis for all functions on $O_{\hat{y}^0} = \{{}_p\hat{t}(\hat{y}^0)\}$. Denote by $\boxed{h}(\hat{y}^0)$ the matrix whose pj entry is $\{h_j({}_p\hat{t}(\hat{y}^0))\}$. By our construction the determinant $\det \boxed{h}(\hat{y})$ is an algebraic function which does not vanish identically.

Using the values ${}_p\hat{t}(\hat{y}^0)$ for $\hat{t}$ in (4.17) gives a (generically) determined system of linear equations for the unknowns $S_j^{N,l}(\hat{y})$. We can write for generic $\hat{y}$

$$\hat{S}_j^{N,l}(\hat{y}) = \sum_p \boxed{h}^{-1}_{jp}(\hat{y}) {}_p\hat{t}_l^N(\hat{y}). \tag{4.20}$$

Now, the noncharacteristic nature of $t = 0$ is the inequality (4.18). By the ideas of division as promulgated in FA, the matrix $\boxed{h}^{-1}$ of algebraic functions does not affect growth conditions by more than a factor $C(1 + |\hat{y}|)^\alpha$. Thus we can conclude that

$$|\hat{S}_j^{N,l}(\hat{y})| \leq c^N (1 + |\hat{y}|)^{N+\alpha}. \tag{4.21}$$

We start with a function $F \in \hat{\mathcal{H}}'(V)$ which, by the fundamental principle of FA, has an extension $\widetilde{F} \in \mathcal{H}(\hat{t}, \hat{y})$ and the growth conditions of $\widetilde{F}$ are not significantly more than those of F.

We write $\widetilde{F}$ as a power series

$$\widetilde{F}(\hat{t}, \hat{y}) = \widetilde{F}_m(\hat{y})\hat{t}^m \tag{4.22}$$

where m is a multi-index. Standard (and readily verified) estimates are

$$|\widetilde{F}_m(\hat{y})| \leq \frac{c\epsilon^{|m|}}{m!} e^{c|\hat{y}|}. \tag{4.23}$$

We write (4.17) in the form

$$\hat{t}^m = \sum \hat{S}_j^m(\hat{y}) h_j(\hat{t}) \quad \text{on } V.$$

In this notation (4.21) becomes

$$\left|\hat{S}_j^m(\hat{y})\right| \leq c^{|m|}(1+|\hat{y}|)^{|m|+\alpha}. \tag{4.24}$$

Inequality (4.24) is crucial for our purposes. Together with (4.23) it implies immediately that for each j the series

$$\mathcal{F}_j = \sum \widetilde{F}_m(\hat{y}) \hat{S}_j^m(\hat{y}) h_j(\hat{t}) \tag{4.25}$$

converges to an entire function in $\hat{y}$ of exponential type. It is clear that the function $\mathcal{F}(\hat{t},\hat{y}) = \sum \mathcal{F}_j(\hat{y}) h_j(\hat{t})$ is equal to F on V and $\mathcal{F} \in H \otimes \mathcal{H}(\hat{y})$.

The only point left to prove is the uniqueness on V of expressions of the form $\sum G_j(\hat{y}) h_j(\hat{t})$. Our derivation of (4.20) from (4.17) shows, in fact, that if $\sum G_j(\hat{y}) h_j(\hat{y}) = F(\hat{t},\hat{y})$ on V then G_j are given (generically) by

$$G_j(y) = \sum_p \boxed{h}^{-1}_{jp}(\hat{y}) F({}_p\hat{t},\hat{y}).$$

Hence they are uniquely determined by F.

Theorem 4.3 is thereby proven.

Example. Let $i_1 = \hat{t}_1, i_2 = \hat{t}_1\hat{t}_2$. Then the ring I generated by i_1 and i_2 is clearly free but not strongly free because $i_1 = i_2 = 0$ has dimension 1. The variety

$$i_1 = \hat{y}_1$$
$$i_2 = \hat{y}_2^2$$

contains the points

$$\hat{y}_1 = \hat{t}_1 = 1, \quad \hat{t}_2 = \hat{y}_2^2$$

which clearly violates (4.18).

This example shows why we need the degree conditions (4.14).

All the above refers to the space $\mathcal{H}$ of entire functions. For $\mathcal{H}'$ or, more generally, for AU spaces $\mathcal{W}$ for which the growth conditions in $\hat{\mathcal{W}}$ depend only on $|\hat{x}|$, the above power series agreements work nicely. But for spaces like the space $\mathcal{E}$ of C^∞ functions the growth conditions on $\hat{\mathcal{E}}'$ are different in real and imaginary directions. Power series methods alone do not suffice.

In Chapter VIII of FA we introduced the general concept of *hyperbolicity*. A system of partial differential equations is hyperbolic in t if solutions defined for all y and for $|t| < a$ (for any $a > 0$) extend uniquely to solutions for all t, y.

The criterion for hyperbolicity is that on the corresponding variety V we have the inequality

$$|\Im\ \hat{t}| \leq c(1 + |\Im\ \hat{y}|). \tag{4.26}$$

Theorem 4.4 *Suppose the hypotheses of Theorem 4.3 hold. Suppose moreover that $\Im\ \hat{t}$ is bounded on V when $\hat{y}$ is real. Then the system of differential equations corresponding to V is hyperbolic.*

Proof We show that (4.26) for $\hat{y}$ real and (4.18) imply (4.26) for all of V. Standard techniques show how to reduce this implication to the case when V is a curve (see e.g. Chapter IX of FA). We write the Puisseaux expansion at infinity for $\hat{t} = \hat{t}(\hat{y})$ in the form

$$_p\hat{t}_l = \sum_{m=-\infty}^{N} b_{plm}\hat{y}^{\alpha m}. \tag{4.27}$$

Here α is the reciprocal of a positive integer. From (4.18) it is clear that $\alpha N \leq 1$. Since (4.26) holds for real $\hat{y}$, the coefficient of $\hat{y}$ in (4.27) must be real since if the imaginary part of the coefficient of $\hat{y}$ is not 0 the term involving $\hat{y}$ (if it appears) dominates at infinity so (4.26) would be violated. Similarly, we cannot have a nontrivial term of the form $b_{plm}\hat{y}^{\alpha m}$ with $\alpha m > 0$ unless $\alpha m = 1$ because it is easily seen that there would then be real values of $\hat{y}$ which give values of $_p\hat{t}_l$ with unbounded imaginary part. Thus (4.21) takes the form

$$_p\hat{t}_l = b_{pl}\hat{y} + \sum_{m=-\infty}^{0} b_{plm}\hat{y}^{\alpha m} \tag{4.28}$$

with b_{pl} real; (4.26) is an immediate consequence of this.

This completes the proof of Theorem 4.4.

We have shown that V corresponds to a hyperbolic system. Our next step is to show that the system is Cauchy hyperbolic. Let $\vec{P}(\partial)$ be the Fourier transform of the generators of the ideal of V. Recall the definition in Section 1.4. The CP for $\vec{P}(\partial)$ is well posed in the C^∞ sense if the Cauchy map

$$\mathcal{C}'(f) = \left(\vec{h}(-i\partial/\partial t)f|_{t=0}; \vec{P}(\partial)f\right) \tag{4.29}$$

is a topological isomorphism of $\mathcal{E}$ onto $[\mathcal{E}(Y)]^N \oplus \mathcal{E}^r_{\vec{P}}$ where $\mathcal{E}^r_{\vec{P}}$ is the space of $\vec{g} \in \mathcal{E}^r$ which satisfy $\vec{Q}(\partial) \cdot \vec{g} = 0$ for all vectors $\vec{Q}$ of polynomials in the module of relations of $\vec{P}$, meaning $\vec{Q} \cdot \vec{P} \equiv 0$.

Theorem 4.5 *Let the assumptions be as in Theorem 4.4. Then the CP is well posed in the C^∞ sense.*

Proof The well posedness of the CP is equivalent to being able to solve

$$\hat{U}(\hat{t},\hat{y}) = \sum \hat{S}_j^{\hat{U}}(\hat{y}) h_j(\hat{t}) \quad \text{on } V \tag{4.30}$$

for $\hat{S}_j^U \in \hat{\mathcal{E}}'(\hat{Y})$ depending uniquely and continuously on $\hat{U} \in \hat{\mathcal{E}}'(V)$. For then the difference of the terms in (4.30) vanishes on V so is of the form $\vec{P}\cdot\vec{\hat{B}}$ where, by the results of FA, $\vec{\hat{B}} \in (\hat{\mathcal{E}}')^r$ can be chosen to depend continuously on $\hat{U}$. $\vec{\hat{B}}$ is clearly unique modulo $\vec{\hat{B}}_0$ where $\vec{P}\cdot\vec{\hat{B}}_0 = 0$. Using the results of FA again, $\vec{\hat{B}}_0$ is of the form $\hat{B}_1\vec{Q}$ where $\vec{Q}$ belongs to the polynomial module of relations of $\vec{P}$ and $\hat{B}_1 \in \hat{\mathcal{E}}$.

All this means that $\hat{\mathcal{C}}$ is a topological isomorphism which is equivalent to Theorem 4.5.

It remains to verify (4.30). We have already shown that $\hat{S}_j^{\hat{U}} \in \hat{\mathcal{H}}'(\hat{Y})$ and that $\hat{S}_j^{\hat{U}}$ are uniquely determined by $\hat{U}$. By the method developed in the proof of Theorem 4.3 (see (4.20)) we can write

$$\hat{S}_j^{\hat{U}}(\hat{y}) = \left\{ \boxed{h}^{-1}(\hat{y})\hat{U}(\vec{\hat{t}}(\hat{y}),\hat{y}) \right\}_j. \tag{4.31}$$

Now, $\hat{U} \in \hat{\mathcal{E}}^{\mathcal{C}}$ satisfies the inequality

$$\left|\hat{U}(\vec{\hat{t}}(\hat{y}),\hat{y})\right| \le c(1+|\hat{t}(\hat{y})|+|\hat{y}|)^c \exp\left\{c[|\Im\vec{\hat{t}}(\hat{y})|+|\Im\ \hat{y}|]\right\}. \tag{4.32}$$

By Theorem 4.4 we can use (4.26) to remove the term $\Im\vec{\hat{t}}(\hat{y})$ from the right side of (4.32). Standard division theory applied to the matrix $\boxed{h}^{-1}$ in (4.31) which is generically nonsingular shows that $\boxed{h}^{-1}$ can only affect such inequalities by a factor $c(1+|\hat{y}|)^c$. Thus

$$|\hat{S}_j^{\hat{U}}(\hat{y})| \le c(1+|\hat{y}|)^c \exp(c|\Im\ \hat{y}|). \tag{4.33}$$

This shows that $\hat{S}_j^{\hat{U}} \in \hat{\mathcal{E}}^{\mathcal{C}}(\hat{Y})$. By a simple modification of the above we see that $\hat{S}_j^{\hat{U}}$ depends continuously on $\hat{U}$.

This completes the proof of Theorem 4.5.

Remark. *The same method of proof, in a somewhat simpler form, shows that when Y is noncharacteristic the CP is well posed for $\mathcal{H}$.*

Example. Think of $\hat{Y} = R^m$ as the space of symmetric k by k matrices (so $m = k(k+1)/2$). We define $\vec{ch}(\hat{y})$ to be the coefficients of the characteristic polynomial of $\hat{y} \in R^m$ and $\vec{\imath}$ to be the elementary symmetric functions in $\hat{T} = R^k$.

Clearly $\vec{\imath}(\hat{t})$ and $\vec{ch}(\hat{y})$ have the same degrees, namely $1, 2, \ldots, k$. We define the variety V by

$$\vec{\imath}(\hat{t}) = \vec{ch}(\hat{y}). \tag{4.34}$$

The assumptions of (4.13) and (4.14) are clearly satisfied. For generic $\hat{y}$ the points in $\mathcal{O}_{\hat{y}}$ are the characteristic roots of $\hat{y}$ in some order. The number of points in $\mathcal{O}_{\hat{y}}$ is $k!$. The additional hypothesis of Theorem 4.5, namely the validity of (4.26) for real $\hat{y}$ is the assertion that the characteristic roots of a real symmetric matrix are real. Thus Theorem 4.5 applies and the system is Cauchy hyperbolic.

The theorems presented thus far depend on the fact that $\vec{\imath}$ is strongly free. What happens when this is not the case?

We shall continue with our assumption, i.e. (4.13) and (4.14). We also assume there are "enough" i_j to guarantee the finiteness of the number of $\hat{t}$ on $0_{\hat{y}}$.

As the situation is rather complicated, let us begin with an example which is important for the nonlinear Fourier transform (see Chapter 5) and has significant consequences in automorphic function theory (see [62]).

Let $\hat{T} = R^k$ and $\hat{Y} = R^m$ with $m = k(k+1)/2$. The variety V is defined by

$$\hat{t}_p \hat{t}_q = \hat{y}_{pq} \tag{4.35}$$

for $(\hat{y}_{pq})$ a symmetric matrix; $\vec{\imath} = \{\hat{t}_p \hat{t}_q\}$.

The quadratic monomials $\hat{t}_p \hat{t}_q$ are not independent. They satisfy the Plücker relations (see (1.88))

$$(\hat{t}_{p_1}\hat{t}_{q_1})(\hat{t}_{p_2}\hat{t}_{q_2}) = (\hat{t}_{p_1}\hat{t}_{p_2})(\hat{t}_{q_1}\hat{t}_{q_2}) \tag{4.36}$$

for all pairs $(p_1, q_1), (p_2, q_2)$. Equation (4.36) has as consequence that $O_{\hat{y}}$ is empty unless $\hat{y}$ satisfies

$$\hat{y}_{p_1 q_1}\hat{y}_{p_2 q_2} = \hat{y}_{p_1 p_2}\hat{y}_{q_1 q_2} \tag{4.37}$$

for all such pairs. The variety defined by (4.37) in $\hat{Y}$ is denoted by V_2.

Proposition 4.6 *Equation (4.37) is necessary and sufficient for $O_{\hat{y}}$ to be nonempty. If (4.37) holds then $O_{\hat{y}}$ consists of two points which are an orbit of reflection in the origin unless $\hat{y} = 0$, in which case $\underline{O}_0$ is the variety*

$$\underline{O}_0 = (0, \textit{identity}; 0, \partial/\partial\hat{t}_1; \ldots; 0, \partial/\partial\hat{t}_k). \tag{4.38}$$

Remark. *Proposition 4.6 means that V_2 is the projection of V on $\hat{Y}$; that is, the ideal of V_2 is the intersection of the ideal of V with the ring of polynomials in $\hat{y}$. For, the ideal of V_2 is contained in the ideal of V since both sides of (4.37) are congruent to $\hat{t}_{p_1}\hat{t}_{p_2}\hat{t}_{q_1}\hat{t}_{q_2}$ modulo this ideal. Conversely if $P(\hat{y})$ vanishes on V then by Proposition 4.6 it vanishes on V_2.*

Proof We have already explained the necessity of (4.37). As for its sufficiency, suppose first that $\hat{y} \neq 0$. Define $\hat{t}_j = \sqrt{\hat{y}_{jj}}$. To determine the square root, we choose the sign in $\hat{t}_1$ arbitrarily. (Some $\hat{y}_{jj}$ is not 0, otherwise, by (4.37), $(\hat{y}_{ij})^2 = \hat{y}_{ii}\hat{y}_{jj}$ would vanish for all i, j. We can suppose that $\hat{y}_{11} \neq 0$.) Then we choose the square root $\hat{t}_j$ of $\hat{y}_{jj}$ for $j > 1$ to conform to $\hat{y}_{1j} = \hat{t}_1\hat{t}_j$. This is possible since $(\hat{y}_{1j})^2 = \hat{y}_{11}\hat{y}_{jj} = \hat{t}_1^2\hat{t}_j^2$. We claim that (4.35) is satisfied for all i, j. For, if $i \neq 1, j \neq 1$,

$$\hat{y}_{ij}\hat{y}_{1j} = \hat{y}_{1i}\hat{y}_{jj} = \hat{t}_1\hat{t}_i\hat{t}_j^2.$$

It follows that if $\hat{y}_{ij}$ vanishes then $\hat{t}_i = 0$ or $\hat{t}_j = 0$ since $\hat{t}_1 \neq 0$ so (4.35) holds.

On the other hand, suppose $\hat{y}_{ij} \neq 0$. It cannot be the case that either $\hat{y}_{1i}$ or $\hat{y}_{1j}$ vanish since $\hat{y}_{ij}\hat{y}_{11} = \hat{y}_{1i}\hat{y}_{1j}$ and $\hat{y}_{11} \neq 0$. Thus $\hat{t}_i\hat{t}_j \neq 0$ and, dividing by $\hat{y}_{1j} = \hat{t}_1\hat{t}_j$, we arrive at (4.35).

The only ambiguity in the construction is the choice of sign of $\hat{t}_1$. Thus there are two points in $O_{\hat{y}}$ which are reflections in the origin.

The orbit $\underline{O}_0$ is defined by the ideal generated by $\{\hat{t}_i\hat{t}_j\}$. For a polynomial $R(\hat{t})$ the conditions $R(0) = \partial R(0)/\partial\hat{t}_j = 0$ for all j are exactly the conditions that R belong to this ideal. These are exactly the conditions that R vanishes on the multiplicity variety (4.38) which verifies that $\underline{O}_0$ has the form (4.38). Proposition 4.6 is thereby verified.

It is clear from our last remark that the harmonics are spanned by $1, \hat{t}_1, \ldots, \hat{t}_k$. (We shall sometimes write $\hat{t}_0 \equiv 1$.) Thus there are two differences between the present situation and the one corresponding to $\vec{\imath}$ strongly free:

(1) $O_{\hat{y}}$ is empty for some $\hat{y}$.
(2) When $O_{\hat{y}}$ is nonempty and $\hat{y} \neq 0$ then $O_{\hat{y}}$ has two points whereas H has dimension $k + 1 > 2$ (unless $k = 1$, in which case $\vec{\imath}$ is strongly free).

A consequence of (1) is that the $\hat{S}_j(\hat{y}) = \hat{S}_j^{\hat{U}}(\hat{y})$ are only determined modulo $I(V_2)$ which is the ideal generated by the relation (4.37). (2) leads to further nonuniqueness of the $\hat{S}_j$. To describe the extra lack of uniqueness we have

Theorem 4.7 *Any $\hat{U} \in \hat{\mathcal{E}}'$ can be written in the form*

$$\hat{U}(\hat{t}, \hat{y}) = \sum \hat{S}_l(\hat{y})\hat{t}_l + \sum \hat{B}_{rs}(\hat{t}, \hat{y})[\hat{t}_r\hat{t}_s - \hat{y}_{rs}]. \tag{4.39}$$

The functions $\hat{S}_l(\hat{y})$ and $\hat{B}_{rs}(\hat{t}, \hat{y})$ are entire functions of exponential type which belong to the Fourier transform of the dual of the Gevrey 2 space (see Chapter 5 below or Chapter 5 of FA) in $\hat{y}$ while $\hat{B}_{rs}(\hat{t}, \hat{y})$ belongs to $\hat{\mathcal{E}}'$ in $\hat{t}$.

The coefficients $\hat{S}_l, \hat{B}_{rs}$ are unique modulo the following relations:

(a) *We can add to $\hat{S}_{l_0}$ a term $\hat{S}^0_{l_0}$ which vanishes on V_2 and for which $S^0_{l_0} \in$ dual Gevrey provided that we compensate by adding a corresponding term to the*

second sum in (4.39). (By (4.35) the ideal of V_2 is contained in the ideal of V.)

(b) *If $\hat{A}(\hat{y})$ is any function in the Fourier transform of the dual of Gevrey 2 we can add $\hat{y}_{il}\hat{A}(\hat{y})$ to $\hat{S}_j(\hat{y})$ and $-\hat{y}_{jl}\hat{A}(\hat{y})$ to $\hat{S}_i$ provided that we add*

$$\hat{t}_j[\hat{t}_i\hat{t}_l - \hat{y}_{il}]\hat{A}(\hat{y}) - \hat{t}_i[\hat{t}_l\hat{t}_j - \hat{y}_{lj}]\hat{A}(\hat{y})$$

to the second sum in (4.39).

(c) *We can add to $\hat{B}_{rs}$ an element of the module of relations of $\{\hat{t}_r\hat{t}_s - \hat{y}_{rs}\}_{r\leq s}$ with coefficients in the dual of Gevrey 2.*

Proof We expand $\hat{U}$ in a power series in $\hat{t}$:

$$\hat{U}(\hat{t},\hat{y}) = \sum \hat{U}^1_{ml}(\hat{y})\hat{t}^m\hat{t}_l. \tag{4.40}$$

We have written monomials in the nonunique form $\hat{t}^m\hat{t}_l$ where $|m| = \sum m_j$ is even. (Recall that $\hat{t}_0 \equiv 1$.) We can clearly write $\hat{t}^m$ as a monomial in the quadratic powers $\hat{t}_i\hat{t}_j$. On the variety V, meaning modulo the ideal generated by $\{\hat{t}_i\hat{t}_j - \hat{y}_{ij}\}$, we can replace $\hat{t}_i\hat{t}_j$ by $\hat{y}_{ij}$ which is entered in $\hat{U}^1_{ml}$; we compensate by terms in the second sum of (4.39). We continue this process until $m = 0$. We arrive at an expression of the form (4.39).

There are two aspects of nonuniqueness in this construction.

(1) The choice of l when $|m|$ is odd > 1.
(2) The way of writing $\hat{t}^m$ as a monomial in quadratics when $|m|$ is even.

The nonuniqueness in (1) comes from writing

$$\hat{t}^m = \hat{t}_i\hat{t}^{m'} = \hat{t}_j\hat{t}^{m''}$$

where $|m'| = |m''|$ is even.

We write

$$\hat{t}^m = \hat{t}_i\hat{t}_j\hat{t}^{m'''}.$$

Since $|m'|$ and $|m''|$ are even and ≥ 2 we can pair $\hat{t}_j$ in $\hat{t}^{m'}$ and $\hat{t}_i$ in $\hat{t}^{m''}$ with the same $\hat{t}_l$ in $\hat{t}^{m'''}$ when decomposing $\hat{t}^{m'}$ and $\hat{t}^{m''}$ into monomials in quadrics. Thus

$$\begin{aligned}\hat{t}^m &= (\hat{t}_j\hat{t}_l)\hat{t}^{m'''}\hat{t}_i = \hat{y}_{jl}\hat{t}^{m''''}\hat{t}_i \quad \text{on } V\\ &= (\hat{t}_i\hat{t}_l)\hat{t}^{m''''}\hat{t}_j = \hat{y}_{il}\hat{t}^{m''''}\hat{t}_j \quad \text{on } V.\end{aligned}$$

Since $|m''''|$ is even the nonuniqueness in (1) can be accounted for by the nonuniqueness (b) of the statement of Theorem 4.7.

Let us study the nonuniqueness (2).

To understand how things work, we examine the monomial $t_1^{10}t_2^8$:

$$\begin{aligned}\hat{t}_1^{\,10}\hat{t}_2^{\,8} &= \hat{t}_1^{\,2}(\hat{t}_1\hat{t}_2)^8\\ &= (\hat{t}_1^{\,2})^2(\hat{t}_1\hat{t}_2)^6\hat{t}_2^{\,2}\\ &= (\hat{t}_1^{\,2})^3(\hat{t}_1\hat{t}_2)^4(\hat{t}_2^{\,2})^2\\ &= (\hat{t}_1^{\,2})^4(\hat{t}_1\hat{t}_2)^2(\hat{t}_2^{\,2})^3\\ &= (\hat{t}_1^{\,2})^5(\hat{t}_2^{\,2})^4.\end{aligned}$$

We replace $\hat{t}_i\hat{t}_j$ by $\hat{y}_{ij}$ to obtain, on V,

$$\begin{aligned}\hat{t}_1^{\,10}\hat{t}_2^{\,8} &= \hat{y}_{11}\hat{y}_{12}^8\\ &= \hat{y}_{11}^2\hat{y}_{12}^6\hat{y}_{22}\\ &= \hat{y}_{11}^3\hat{y}_{12}^4\hat{y}_{22}^2\\ &= \hat{y}_{11}^4\hat{y}_{12}^2\hat{y}_{22}^3\\ &= \hat{y}_{11}^5\hat{y}_{22}^4.\end{aligned}$$

We can interpret going from each line to the next as an application of the identity

$$y_{12}^2 = y_{11}y_{22}$$

which holds on V so these changes can be absorbed in the nonuniqueness part (a) of the statement of Theorem 4.7.

In general we can see that the expressions we obtain from (4.40) by replacing $\hat{t}_i\hat{t}_j$ by $\hat{y}_{ij}$ in various ways all amount to iterated application of the identities (4.37) (which hold on V). This proves our assertion about the lack of uniqueness in (4.39).

Conversely these relations represent the total lack of uniqueness of $\{\hat{S}_l\}$. For the ideal of V is generated by polynomials which are quadratic in $\hat{t}$ so for $\sum \hat{S}_j(\hat{y})\hat{t}_j$ to vanish on V it must be the case that each $\hat{S}_j(\hat{y})$ vanishes on V, i.e. $\hat{S}_j(\hat{y})$ vanishes on V_2 by the remark following the statement of Proposition 4.6.

The only possibility for further nonuniqueness is the modification of $\hat{B}_{rs}$ by an element of the module of relations of $\{\hat{t}_r\hat{t}_s - \hat{y}_{rs}\}_{r\le s}$.

Finally there remains the question of bounds on the terms in (4.39). Since $\hat{U}$ is an entire function of exponential type belonging to $\hat{\mathcal{E}}'(\hat{t},\hat{y})$ the standard application of Cauchy's integral formula to (4.40) shows that

$$|U^1_{ml}(\hat{y})| \le \frac{c^{|m|+1}}{|m|!}(1+|\hat{y}|)^c e^{c|\Im\hat{y}|}.$$

When we replace $\hat{t}^m$ by a monomial in $\hat{y}_{ij}$ the degree μ of this monomial satisfies $\mu = |m|/2$. (This assumes m is even; the odd case requires only trivial modifications.) Hence

$$\left|\sum U^1_{ml}(\hat{y})\hat{t}^m(\hat{y})\right| \le c(1+|\hat{y}|)^c e^{c|\Im\hat{y}|}\sum \frac{c^{|m|}}{m!}|\hat{y}|^{|m|/2}$$
$$= c(1+|\hat{y}|)^c e^{c(|\Im\hat{y}|+|\hat{y}|^{1/2})}.$$

On the real $\hat{y}$ space this grows like $\exp(c|\hat{y}|^{1/2})$ which is the growth associated to Gevrey 2.

Once having established that the S_l belong to the dual of Gevrey 2 the structure of $\hat{B}_{rs}$ is established via the theory of FA.

We have verified the statement about bounds in Theorem 4.7 so the proof is complete.

Note that if $\hat{U}(\hat{t},\hat{y})$ is in the Fourier transform of the dual of Gevrey 2 the above estimates show that the same theorem is valid except that $\hat{B}_{rs}$ is in the Fourier transform of dual Gevrey 2 in all variables. It can be shown, by making the above argument more precise, that the decomposition (4.39) is a topological isomorphism modulo the compatibility conditions.

Using the treatment of the Cauchy map that we have used on several occasions we deduce

Theorem 4.8 *The CP for the variety V of (4.35) defined by the map*

$$\mathcal{C}' : f \to \left\{ f(0,y); \frac{\partial f}{\partial t_j}(0,y); \left(\frac{\partial^2}{\partial t_r \partial t_s} - \frac{\partial}{\partial y_{rs}}\right) f(t,y) \right\} \tag{4.41}$$

is a topological isomorphism for $f \in$ Gevrey 2 onto the space of functions in Gevrey 2 which are defined on the Fourier transform of the space of $\{\hat{S}_j\}$, $\{\hat{B}_{rs}\}$ of (4.39) with the compatibility conditions (a), (b), (c) *as described in Theorem 4.7. In particular if $\partial(V)f = 0$ then the "Cauchy data" $\{(\partial f/\partial t_j)\}$ satisfies*

$$\left(\frac{\partial^2}{\partial y_{p_1q_1}\partial y_{p_2q_2}} - \frac{\partial^2}{\partial y_{p_1p_2}\partial y_{q_1q_2}}\right)\frac{\partial f}{\partial t_j}(0,y) = 0 \tag{4.42}$$

for all j. (Recall $\partial/\partial t_0 =$ identity.)

All this is proven except the last statement. $\partial(V)f = 0$ means

$$\frac{\partial f}{\partial y_{rs}} = \frac{\partial^2 f}{\partial t_r \partial t_s}$$

for all r, s so

$$\frac{\partial^2 f(t,y)}{\partial y_{p_1q_1}\partial y_{p_2q_2}} = \frac{\partial^2 f(t,y)}{\partial y_{p_1p_2}\partial y_{q_1q_2}}$$

for all t, y in particular for $t = 0$. (Similarly for $j \neq 0$.)

This completes the proof of Theorem 4.8.

The decomposition (4.39) is an analog of (4.3). One might wonder if it is possible to obtain more precise results using our "Lagrange interpolation" methods, meaning (4.9), in place of power series. These methods solve for $\hat{S}_j(y)$ in terms of values of $\hat{U}$ on V. They represent a change from the harmonic base to the δ functions base on the orbits $\mathcal{O}_{\hat{y}}$. We wish to point out the difficulties involved in carrying out such a program.

Let us examine the possibility of solving (4.12) for $\hat{S}_j = \hat{S}_j^U$ in terms of $\hat{U} \in \hat{\mathcal{E}}'$. We begin with the region $Y_1 \subset V_2$ where $\hat{y}_{11} \neq 0$. On an orbit $O_{\hat{y}}$ above $\hat{y} \in Y_1$ we know that $\hat{t}_1 \neq 0$. This means that we can solve

$$\hat{U}(\hat{t}, \hat{y}) = \hat{S}_{01}(\hat{y}) + \hat{S}_{11}(\hat{y})\hat{t}_1 \quad \text{on } O_{\hat{y}} \tag{4.43}$$

for $\hat{S}_{01}, \hat{S}_{11}$ since $\hat{t}_1$ takes two values $\hat{t}_1(\hat{y}) = \pm\sqrt{\hat{y}_{11}}$ on $O_{\hat{y}}$. (The second subscript 1 of $\hat{S}$ refers to the region Y_1.) In the proof of Proposition 4.6 we showed that $t_j = \sqrt{y_{jj}}$ for all j. Thus

$$\begin{aligned}
\hat{S}_{01}(\hat{y}) &= \tfrac{1}{2}\left[\hat{U}(\hat{y}_{11}^{1/2}, \hat{y}_{22}^{1/2}, \ldots, \hat{y}_{kk}^{1/2}, \hat{y}) + \hat{U}(-\hat{y}_{11}^{1/2}, -\hat{y}_{22}^{1/2}, \ldots, -\hat{y}_{kk}^{1/2}, \hat{y})\right] \\
\hat{S}_{11}(\hat{y}) &= \frac{1}{2\hat{y}_{11}^{1/2}}\left[\hat{U}(\hat{y}_{11}^{1/2}, \hat{y}_{22}^{1/2}, \ldots, \hat{y}_{kk}^{1/2}, \hat{y}) - \hat{U}(-\hat{y}_{11}^{1/2}, -\hat{y}_{22}^{1/2}, \ldots, -\hat{y}_{kk}^{1/2}, \hat{y})\right].
\end{aligned} \tag{4.44}$$

The choice of $\hat{y}_{11}^{1/2}$ is arbitrary. As in the proof of Proposition 4.6 the choices of square roots of $\hat{y}_{jj}^{1/2}$ are then determined. Our notation is such that when we pick one square root which we call $+y_{11}^{1/2}$ then the corresponding square roots of y_{jj} are denoted by $+y_{jj}^{1/2}$.

As for growth conditions, $\hat{t}_j(\hat{y}) = \sqrt{\hat{y}_{jj}}$ so that

$$|\hat{t}(\hat{y})| \leq c(1 + |\hat{y}|^{1/2}). \tag{4.45}$$

Inequality (4.45) combined with (4.44) mean that for the range of $\hat{y}$ considered

$$|\hat{S}_{j1}(\hat{y})| \leq c(1 + |\hat{y}|)^c \exp[c(|\Im \hat{y}^{1/2}| + |\Im\ \hat{y}|)]. \tag{4.46}$$

Inequality (4.46) represents an improvement on Theorem 4.7 because the $\hat{S}_{j1}(\hat{y})$ are bounded when $\hat{y}$ is real and near the positive orthant. The growth $\exp(c|\hat{y}|^{1/2})$ occurs in the rest of the real space. In the language of Chapter 5: the positive orthant does not belong to the Gevrey wave front set of $\hat{S}_{j1}$. However, $\hat{S}_{j1}$ are defined only when $\hat{y}_{11} \neq 0$.

If we pass from the region where $\hat{t}_1 \neq 0$ to a region $Y_l \subset V_2$ where $\hat{y}_{ll} \neq 0$ then we can solve (4.12) and obtain functions $\hat{S}_{0l}, \hat{S}_{1l}$ which satisfy (4.46).

We have thus obtained a "piecewise" solution of (4.12) satisfying (4.46). The functions $\hat{S}_{0l}$ clearly agree on the overlap. But the functions $\hat{S}_{1l}$ satisfy relations of the form

$$\hat{t}_m \hat{S}_{1m}(\hat{y}) - \hat{t}_l \hat{S}_{1l}(\hat{y}) = 0 \quad \text{on} \quad Y_l \cap Y_m. \tag{4.47}$$

Near the origin in V_0 we can use power series to obtain a local solution $\tilde{\hat{S}}_{kl}$ of (4.12).

The solutions of (4.12) are not unique because of relations described in Theorem 4.7.

Problem 4.2 *Is it possible to piece these solutions together?*

The analogous problem for monomials of degree $m > 2$ can be treated in exactly the same manner. $\hat{Y}$ is a space of dimension $\binom{m+k-1}{m}$. The coordinates are $\hat{y}_{j_1 \cdots j_m}$ which are completely symmetric in their indices. $V = V^m$ is defined by the equations

$$\hat{y}_{j_1 \ldots j_m} = \hat{t}_{j_1} \ldots \hat{t}_{j_m}. \tag{4.48}$$

The Plücker equations become

$$\hat{y}_{j_1 \ldots j_m} \hat{y}_{k_1 \ldots k_m} = \hat{y}_{k_1 j_2 \ldots j_m} \hat{y}_{j_1 k_2 \ldots k_m}. \tag{4.49}$$

The Plücker variety (4.47) is denoted by V_m. The analog of Proposition 4.6 goes as follows. The Plücker conditions are necessary and sufficient for $O_{\hat{y}}$ to be nonempty. When $\hat{y} \neq 0$ and $O_{\hat{y}}$ is nonempty then it contains exactly m points which form an orbit of the group of scalar multiplication by the m-th roots of unity. The harmonics consist of all polynomials of degree $< m$ and the multiplicity of $\underline{O}_0$ is dim H. In fact

$$\mathcal{O}_0 = \{0; h_j(\partial/\partial \hat{t})\}.$$

For the proof: when $\hat{y} \neq 0$ we claim first that some $y_{jj\ldots j} \neq 0$. For by iteration of (4.47) and symmetry we can write

$$(\hat{y}_{j_1,\ldots,j_m})^m = \hat{y}_{j_1,\ldots,j_1} \hat{y}_{j_2,\ldots,j_2} \hat{y}_{j_m,\ldots,j_m}.$$

If all $\hat{y}_{jj\ldots j}$ vanish then so does $\hat{y}_{j_1,\ldots,j_m}$ contrary to our assumption that $\hat{y} \neq 0$. We can assume that $\hat{y}_{11\ldots1} \neq 0$. Define $\hat{t}_j = (\hat{y}_{jj\ldots j})^{1/m}$. Once the m-th root of unity for $\hat{t}_1$ is chosen, all $\hat{t}_j$ are determined by

$$\hat{y}_{11\ldots1j} = \hat{t}_1^{m-1} \hat{t}_j.$$

Again it is readily verified that all the Plücker conditions are satisfied. The only ambiguity is the root of unity in the choice of $\hat{t}_1$.

The description of the harmonics and the description of $\underline{O}_0$ are clear.

We can formulate the extension of Theorem 4.7 to arbitrary m. Relation (4.39) becomes

$$\hat{U}(\hat{t},\hat{y}) = \sum \hat{S}_j(\hat{y})\hat{t}^j + \sum \hat{B}_r(\hat{t},\hat{y})(\hat{t}^r - \hat{y}_r). \tag{4.50}$$

j and r are multi-indices with $|j| < m$ and $|r| = m$. For $\hat{U} \in \hat{\mathcal{E}}'(\hat{t},\hat{y})$ the functions $\hat{S}_j$ and $\hat{B}_r$ are in the tensor product of $\hat{\mathcal{E}}'(\hat{t})$ and the Fourier transform of the dual of Gevrey m. If $\hat{U} \in$ Fourier transform dual Gevrey m in $\hat{t},\hat{y}$ then so are $\hat{S}_j$ and $\hat{B}_r$ and (4.50) is a topological isomorphism modulo the following relations:

(a_m) We can add to $\hat{S}_{j_0}$ a term $\hat{S}^m_{j_0}$ which vanishes on V_m provided that we add a compensating term to the second sum in (4.50).

(b_m) If $\hat{A}(\hat{y}) \in$ Fourier transform of dual Gevrey m then we can add $\hat{y}_r\hat{A}(\hat{y})$ to $\hat{S}_j(\hat{y})$ provided that we add $-\hat{y}_{r'}\hat{A}(\hat{y})$ to $\hat{S}_{j'}(\hat{y})$ where $(r''j')$ is a permutation of (rj) and compensate as in the case $m = 2$.

(c_m) We can add an element of the module of relations of $\{\hat{t}^r - \hat{y}_r\}$ to $\{\hat{B}_r\}$.

This assertion is proven in the same manner as Theorem 4.7.

We deduce

Theorem 4.9 *For any m there is an exotic CP (the CD is given on the linear space Y) for the variety $V = V^m$ defined by (4.48). This CP is well posed for data of class Gevrey m. The data is given on Y and the "normal derivatives" correspond to the monomials of degree $< m$. The compatibility conditions on the CD are defined by (a_m), (b_m), (c_m) by duality. In particular if $\partial(V)f = 0$ then the CD is annihilated by $\partial(V_m)$.*

For general varieties V of the form (4.13) with the degree conditions (4.14) there is probably a set of subvarieties $\{V_\alpha\}$ of V such that for $\hat{y} \in$ projection on $\hat{Y}$ of V_α the harmonics on $O_{\hat{y}}$ satisfy relations with coefficients in $\hat{Y}$; the type of relations depends on α. However, unless the V_α are nested, meaning $V_1 \supset V_2 \supset \ldots$, we do not know how to translate such conditions into conditions on the CD. Moreover we do not know how to prove such a result.

Problem 4.3 *Carry out this program.*

In case $k = 1, m = 2$, we have given the explicit solution to the CP in FA.

Problem 4.4 *Find an explicit solution to the CP for general k, m.*

Remark. *In Section 5.1 we shall make a detailed study of the CP for (4.49). In the simplest nontrivial case $m = k = 2$, $\partial(V_m)$ is the wave operator in $k = 3$ dimensions.*

4.2 The compact Watergate problem

At the end of the previous section we described one type of exotic CP—the CD is given on a space Y whose dimension is larger than the dimension of V.

In the present section we discuss the opposite case—the CD given on a lower dimensional variety. As explained in Section 1.4, we call this the Watergate problem (WP).

In the present section we treat the situation in which the real points on $O_{\hat{y}^0}$ for $\hat{y}^0$ real are the orbits of a group W which acts on the variety V. Unlike the considerations of Section 4.1 the variety $V_{\hat{y}^0}$ has dimension > 0.

We start with the wave equation in $(2+1)$ variables with data on the time axis. We shall show how to reduce this problem to Theorem 3.14 which deals with C^k in place of V. We shall then give a more general approach to such problems.

In our notation V is defined by

$$\hat{y}^2 = \hat{t}_1^2 + \hat{t}_2^2. \tag{4.51}$$

We use $\vec{\imath} = i = \hat{t}_1^2 + \hat{t}_2^2$. The harmonics are just the ordinary harmonics in the $\hat{t}$ plane thought of as being constant in $\hat{y}$. A basis is $\{(\hat{t}_1 \pm i\hat{t}_2)^p\}$.

In accordance with the ideas of the previous section we want to express any $F \in \hat{\mathcal{E}}^{\mathcal{C}}(V)$ as

$$F(\hat{t}, \hat{y}) = \sum (\hat{t}_1 \pm i\hat{t}_2)^p \hat{S}_{\pm}^p(\hat{y}) \quad \text{on } V. \tag{4.52}$$

We pick some extension of $\widetilde{F}$ to a function in $\hat{\mathcal{E}}^{\mathcal{C}}$; (4.52) is derived (formally), as usual, by expanding $\widetilde{F}$ in a power series in $\hat{t}$ with coefficients which are functions of $\hat{y}$, then applying Fischer's idea (Theorem 2.2) to write the harmonic expansion of the monomials $\hat{t}^j$ using the "invariant" $i = \hat{t}_1^2 + \hat{t}_2^2$, and then replacing the invariants by functions of $\hat{y}$ using (4.51).

Of course this procedure is only formal; we need inequalities on the Fischer expansion to establish convergence. It is easier to apply a somewhat different approach.

(4.52) is one expression for F on V. A second expression when $\hat{t}$ is real and $\hat{y} > 0$ is

$$\begin{aligned} F(\hat{t}, \hat{y}) &= \widetilde{F}(\hat{y}\cos\theta, \hat{y}\sin\theta, \hat{y}) \\ &= \sum \hat{T}_{\pm}^p(\hat{y}) e^{\pm ip\theta}. \end{aligned} \tag{4.52*}$$

It is the Fourier series expansion on the circle $(\hat{t}, \hat{y}) = (\hat{y}\cos\theta, \hat{y}\sin\theta, \hat{y})$.

We can identify the two expansions. This means that $\hat{T}_{\pm}^p(\hat{y}) = \hat{y}^p \hat{S}_{\pm}^p(\hat{y})$ for $\hat{y} > 0$. The Fourier inversion formula then implies

$$\hat{S}_{\pm}^p(\hat{y}) = \hat{y}^{-p} \int F(\hat{y}\cos\theta, \hat{y}\sin\theta, \hat{y}) e^{\mp ip\theta}\, d\theta. \tag{4.53}$$

(We have chosen a normalization in which the usual factor 2π disappears.) We could use (4.53) to give another definition of $\hat{S}_{\pm}^p(\hat{y})$. The right side of (4.53)

clearly extends to a function which is holomorphic in $\hat{y} \neq 0$. We shall soon see that actually the $\hat{S}^p_{\pm}$ are entire.

The evaluation of $\hat{S}^p_{\pm}$ in terms of the values of F on V is termed the *Watergate method*.

However, one must take care: *ab initio* the expressions (4.52*) and (4.53) for F on V are defined only when $\hat{y} > 0$. When $\hat{y} < 0$ we can write

$$(\hat{y}\cos\theta, \hat{y}\sin\theta, \hat{y}) = ((-\hat{y})\cos\theta', (-\hat{y})\sin\theta', \hat{y})$$

where $\theta' = \theta + \pi$ is the antipode of θ. The expression on the right is still meaningful in the sense of polar coordinates because we only need the coefficient of $\cos\theta', \sin\theta'$ to be nonnegative since this coefficient represents the distance coordinate in the real $\hat{t}$ plane.

In this way we can give meaning to (4.53) when $\hat{y} < 0$.

It is important to observe that the expression given by the right side of (4.52) generally does not belong to $\hat{\mathcal{E}}^{\mathcal{C}}$. For, if $\hat{y}$ is fixed and real then the series is harmonic in the usual sense in the real $\hat{t}$ plane. Hence, by Liouville's theorem it is of polynomial growth only when it is a polynomial. Thus the right side of (4.52) does not represent an extension of F from V to all of C^n which is in $\hat{\mathcal{E}}^{\mathcal{C}}$. (The fundamental principle tells us that such an extension exists.)

According to the previous section, (4.52) is related to a parametrization problem for the wave equation

$$\Box f = \left(\frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial t_1^2} - \frac{\partial^2}{\partial t_2^2}\right) = 0$$

with data given on the Y axis (the time axis) and with "normal derivatives" given by Fourier transforms of $(\hat{t}_1 \pm i\hat{t}_2)^p$, namely $(\partial/\partial t_1 \pm i\partial/\partial t_2)^p$. On the other hand, the expansion (4.53) looks very much like the expansion of functions in $\hat{\mathcal{E}}^{\mathcal{C}}(\hat{t}_1, \hat{t}_2)$ in Fourier series as given in Section 3.2. We want to use Theorem 3.14 to characterize the $\hat{S}^p_{\pm}(\hat{y})$ corresponding to $F \in \hat{\mathcal{E}}^{\mathcal{C}}(V)$. In this way we can determine the exact nature of this parameterization problem.

In order to coordinate (4.52) and (4.52*) with the ideas of Theorem 3.14, we want to identify (4.52) with the Fourier series expansion of a function belonging to $\hat{\mathcal{E}}(\hat{t})$. This involves the pushing together process introduced in the previous section. Now, the usual CP for the wave operator corresponding to (4.51) has data on the t plane. It corresponds to the harmonic expansion using $i' = \hat{y}^2$, namely

$$F(\hat{t}, \hat{y}) = \hat{U}^0(\hat{t}) + \hat{y}\hat{U}^1(\hat{t}) \quad \text{on } V. \tag{4.54}$$

We claim that $\hat{U}^0$ and $\hat{U}^1 \in \hat{\mathcal{E}}^{\mathcal{C}}$. In fact (we write $\hat{t}^2 = \hat{t}_1^2 + \hat{t}_2^2$)

$$\hat{U}^0(\hat{t}) = \tfrac{1}{2}[F(\hat{t}, (\hat{t}^2)^{1/2}) + F(\hat{t}, -(\hat{t}^2)^{1/2})]$$

$$\hat{U}^1(\hat{t}) = \frac{(\hat{t}^2)^{-1/2}}{2}[F(\hat{t}, (\hat{t}^2)^{1/2}) - F(\hat{t}, -(\hat{t}^2)^{1/2})].$$

Since $F \in \hat{\mathcal{E}}'$

$$|F(\hat{t}, \hat{y})| \leq c(1 + |\hat{t}| + |\hat{y}|)^c e^{c(|\Im \hat{t}| + |\Im \hat{y}|)}.$$

We have

$$|\Im(\hat{t}^2)^{1/2}| \leq c(1 + |\Im \hat{t}|).$$

This can be easily seen directly or it follows from Theorem 4.4 since clearly $\Im(\hat{t}^2)^{1/2} = 0$ when $\hat{t}$ is real.

Combining this with the inequality for F

$$\left|\hat{U}^{(j)}(t)\right| \leq c(1 + |\hat{t}|)e^{c|\Im \hat{t}|}.$$

Since (4.54) is related to the usual CP it is easily seen that the power series approach as described following (4.52) can be used to show that $\hat{U}^{(j)}$ are entire functions of exponential type. (For a detailed treatment see Chapter IX of FA.)

Since $\hat{U}^{(j)}$ are entire functions which satisfy the inequalities relevant to $\hat{\mathcal{E}}'$ they are in $\hat{\mathcal{E}}'$. Thus the right side of (4.54) extends F to $\widetilde{F} \in \hat{\mathcal{E}}^{\mathcal{C}}$. (We can use this for an explicit form of $\widetilde{F}$.) The fact that $\hat{U}^j \in \hat{\mathcal{E}}^{\mathcal{C}}$ is the hyperbolicity of the wave equation with space-like surface $\{\hat{y} = 0\}$.

Remark. *The hyperbolicity of the wave operator is crucial to our construction. To understand this point, suppose that we wanted to study the $\hat{S}_{\pm}^p$ directly from (4.52) and (4.53). We should like to characterize the set of $\hat{S}_{\pm}^p(\hat{y})$ which correspond to an $F \in \hat{\mathcal{E}}'(V)$. To follow the same lines as the proof of Theorem 3.15 would require an analog for V of the Phragmén–Lindelöf theorem. This would involve great complications. (The theory of extensions of Phragmén–Lindelöf theorems to varieties is taken up in detail in [120], [121], [122].)*

We now apply the expansion of Theorem 3.14 to $\hat{U}^j$ (we replace r by $\hat{y}$). As we have seen, the harmonics are spanned by $\{(\hat{t}_1 \pm i\hat{t}_2)^p\}$. The invariants are generated by $\hat{t}^2$ which we can replace by $\hat{y}^2$. By (4.54) these expansions of $\hat{U}^{(j)}$ yield an expansion for F on V of the form (4.52). The $\hat{S}_{\pm}^p$ are entire; from (4.53) and the fact that $F = \widetilde{F} \in \hat{\mathcal{E}}'$ we deduce immediately that the $\hat{S}_{\pm}^p$ satisfy the bounds that guarantee that $\hat{S}_{\pm}^p \in \hat{\mathcal{E}}'$.

We can quickly extend these ideas to the wave equation in $k+1$ variables where now W is the orthogonal group $SO(k)$. The harmonics are the homogeneous harmonic polynomials $h_l^p(\hat{t})$ whose restrictions to the sphere are the usual spherical harmonics $h_l^p(\theta)$. Here p is the degree and l parametrizes a suitably normalized basis for spherical harmonics of degree p. The analog of (4.52) is

$$F(\hat{t},\hat{y}) = \sum h_l^p(\hat{t})\hat{S}_l^p(\hat{y}) \quad \text{on } V \tag{4.55}$$

which is derived, formally, as before.

Again for $\hat{y} > 0$ we can find a second expression for $F(\hat{t},\hat{y})$ on the real sphere $\mathcal{O}_{\hat{y}}^R = \{(\hat{y}h_1^1(\theta),\dots,\hat{y}h_k^1(\theta),\hat{y}\}$:

$$F[\hat{y}h_1^1(\theta),\dots,\hat{y}h_k^1(\theta),\hat{y}] = \sum \hat{T}_l^p(\hat{y})h_l^p(\theta). \tag{4.56}$$

We are once more in the favorable situation in which we can identify the two expansions. More precisely, $h_l^p(\hat{t}) = \hat{y}^p h_l^p(\theta)$ because $V_{\hat{y}}$ has "radius" $\hat{y}$. We can thus compute $\hat{S}_l^p(\hat{y}) = \hat{y}^{-p}\hat{T}_l^p(\hat{y})$ using the integral inversion formula to obtain

$$\hat{S}_l^p(\hat{y}) = d_p\hat{y}^{-p}\int F[\hat{y}h_1^1(\theta),\dots,\hat{y}h_k^1(\theta),\hat{y}]\overline{h}_l^p(\theta)\,d\theta. \tag{4.57}$$

d_p is the dimension of the space of spherical harmonics of degree p.

Formula (4.57) is very much in the spirit of (3.120) except that F now depends explicitly on $\hat{y}$. As in case $k=2$ we apply Theorem 3.14 to the terms $\hat{U}^j$ of (4.54) to obtain

Theorem 4.10 *The $\{\hat{S}_l^p\}$ are characterized by being entire functions in $\hat{\mathcal{E}}'$ for which*

$$\{\hat{y}^p\hat{S}_l^p(\hat{y})\} \text{ is bounded in } \hat{\mathcal{E}}'. \tag{4.58}$$

Note that the vanishing conditions of Theorem 3.14 show that the $\hat{S}_l^p$ are entire. Although $\hat{y}^{-p}A_l^p(\hat{y})$ of Theorem 3.14 are even functions, the $\hat{S}_l^p$ are *not* even functions because of the factor $\hat{y}$ of U^1 in (4.54).

Theorem 3.18 gives us the topology of the Fourier series coefficients of functions in $\hat{\mathcal{E}}'$. It is readily translated, using (4.54), into

Theorem 4.11 *The topology of $\{\hat{S}_l^p\}$ for $F \in \hat{\mathcal{E}}'(V)$ is given by the semi-norms*

$$\max \frac{|\hat{y}^p|\,|\hat{S}_l^p(\hat{y})|}{k_2(\hat{y})} \tag{4.59}$$

where $\{k_2\}$ forms an AU structure for $\mathcal{E}(Y)$.

Theorem 3.19 states that $\mathbb{C}^{\mathbb{R}}$ is sufficient for the space $\mathcal{E}$. Again, using (4.54), we derive

Theorem 4.12 *The union of the complexifications of the real generators of the light cone is sufficient for* $\hat{\mathcal{E}}'(V)$.

A generator of V is a line of the form $(\hat{y}h_1^1(\theta), \ldots, \hat{y}h_k^1(\theta), \hat{y})$ for θ fixed. The generator is real if $\{h_j^1(\theta)\}$ are real.

Let us observe that the series (4.52) and (4.55) converge on compact sets in $\mathbb{C}^n$. For (3.123) implies that

$$|\hat{S}_l^p(\hat{y})| \leq \frac{c^{p+1}}{p!} e^{c_1|\hat{y}|}.$$

Since

$$|h_l^p(\hat{t})| \leq c_2^{p+1}|\hat{t}|^p$$

the series (4.52) and (4.55) define entire functions of exponential type. As mentioned above, these functions cannot belong to $\hat{\mathcal{E}}'$ unless the series are finite.

We have studied the general CP in terms of the Cauchy map

$$\mathcal{C}' : f \to (\{h_j(D)f|_Y\}, \vec{P}(D)f)$$

where $h_j(D)$ are the "normal derivatives." The Fourier transform $\hat{C}$ of the adjoint of $\mathcal{C}'$ is given by

$$\hat{C} : [\{\hat{S}_j(\hat{y})\}, \vec{G}(\hat{t}, \hat{y})] \to \sum h_j(\hat{t})\hat{S}_j(\hat{y}) + \vec{P} \cdot \vec{G}.$$

Remark. *According to (4.53) the* $\hat{S}_j$ *(or* $\hat{S}_\pm^p$*) (except for the normalizing factor* $\hat{y}^p$*) are integrals of* Fh_j *over the leaf whose spread parameter is* $\hat{y}$*. Thus* $\{\hat{S}_j\}$*, which is the Fourier transform of the dual of the CD, is the multiplicity Radon transform of* F *on the spread whose leaves are* $\{\hat{y} = const.\}$ *and whose multiplicities are the harmonics* $\{h_j\}$.

In our case $\vec{P} = P = \hat{t}_1^2 + \hat{t}_2^2 - \hat{y}^2$ and the harmonics are denoted by h_l^p. The well-posedness of the CP means that $\mathcal{C}'$ and $\hat{\mathcal{C}}$ are topological isomorphisms. In setting

$$F(\hat{t}, \hat{y}) = \sum h_l^p(\hat{t})\hat{S}_l^p(\hat{y}) + PG$$

for F in some space $\hat{\mathcal{W}}'$ we are led to

$$F(t, \hat{y}) = \sum h_l^p(\hat{t})\hat{S}_l^p(\hat{y}) \quad \text{on } V.$$

Our above arguments show, for $\mathcal{W} = \mathcal{E}$, that $\{\hat{S}_l^p\}$ with topology (4.59) depend continuously on $F \in \hat{\mathcal{E}}'$. Moreover the map $F \to \{\hat{S}_l^p\}$ is a topological isomorphism on $\hat{\mathcal{E}}'(V)$.

On the other hand the map $\hat{\mathcal{C}}$ is *not* a topological isomorphism on $\hat{\mathcal{E}}'$ because $\sum h_l^p(\hat{t})\hat{S}_l^p(\hat{y})$ is not, in general, in $\hat{\mathcal{E}}'$. Our above arguments show that, for

$\mathcal{W} = \mathcal{H}$, $\hat{\mathcal{C}}$ is a topological isomorphism on $\hat{\mathcal{H}}'$ if $\{\hat{S}_l^p\}$ is given the topology defined by the semi-norms

$$\max \frac{|\hat{y}|^p \, |\hat{S}_l^p(\hat{y})|}{k_1(\hat{y})} \tag{4.60}$$

where $\{k_1\}$ is an AU structure for $\mathcal{H}$.

The WP is thus well-posed in the space $\mathcal{H}$. But in order to interpret the WP directly in the space $\mathcal{H}$ we have to translate the norms (4.60) from $\hat{\mathcal{H}}'$ to $\mathcal{H}$. Since $\{\hat{y}^p \hat{S}_l^p(\hat{y})\}$ is bounded in the space $\hat{\mathcal{H}}'(\hat{Y})$ it is reasonable that if $\{f_l^p\}$ is the Cauchy (Watergate) data of a solution of $Pf = 0$ then $\{\hat{y}^{-p} \hat{f}_l^p(\hat{y})\}$ should be bounded in the topology of $\hat{\mathcal{H}}(\hat{Y})$. (There is no problem at $\hat{y} = 0$ since the complement of $|\hat{y}| \leq 1$ is sufficient for $\mathcal{H}(Y)$.) This is also consonant with the fact that (formally) if we write f in accordance with the fundamental principle

$$f(t, y) = \int e^{iy\hat{y}} \, d\hat{y} \int_{(\hat{t}^2)^{1/2} = \hat{y}} \hat{f}(\hat{t}, \hat{y}) e^{it\cdot\hat{t}} \, d\hat{t}$$

then $\mathrm{CD}(f)$ is given by

$$\begin{aligned} f_l^p(y) &= h_l^p(-i\partial/\partial t) f(0, y) \\ &= \int e^{iy\hat{y}} \, d\hat{y} \int_{(\hat{t}^2)^{1/2} = \hat{y}} h_l^p(\hat{t}) \hat{f}(\hat{t}, \hat{y}) \, d\hat{t}. \end{aligned}$$

This leads to the interpretation

$$\hat{f}_l^p(\hat{y}^0) = \int_{V \cap \{\hat{y} = \hat{y}^0\}} h_l^p(\hat{t}) \hat{f}(\hat{t}, \hat{y}^0) \, d\hat{t}.$$

The function $\hat{f}(\hat{t}, \hat{y}^0)$ (more precisely $\hat{f}(\hat{t}, \hat{y}^0) d\hat{t}$ should be thought of as a measure on V) decreases at infinity more rapidly than $c \exp[-N(|\hat{t}| + |\hat{y}|)]$ for any N (see Section 1.4). Moreover, as we have explained in Chapter 3, $|h_l^p| \leq 1$ on the unit sphere so

$$|h_l^p(\hat{t})| \leq |\hat{t}|^p.$$

Since $|\hat{t}| = |\hat{y}|$ on V we find

$$\begin{aligned} |\hat{f}_l^p(\hat{y}^0)| &\leq c \int_{V \cap \{\hat{y} = \hat{y}^0\}} |\hat{t}|^p e^{-N(|\hat{t}| + |\hat{y}|)} \, d\hat{t} \\ &\leq c'(1 + |\hat{y}^0|)^p e^{-N|\hat{y}^0|}. \end{aligned}$$

It follows that

$$\{|\hat{f}_l^p(\hat{y})(1 + |\hat{y}|)^{-p}\}$$

is bounded in $\hat{\mathcal{H}}'(\hat{y})$, meaning that these measures are bounded by $c_N \exp(-N|\hat{y}|)$ where c_N is independent of p, l.

Since the complement of the unit ball is sufficient (see Section 1.4) for $\hat{\mathcal{H}}'$ we can assume support $\hat{f}_l^p$ omits $|\hat{y}| \leq 1$. Our condition can be rephrased as

$$\{\hat{f}_l^p(\hat{y})\hat{y}^{-p}\} \text{ is bounded in } \hat{\mathcal{H}}'(\hat{y}).$$

Conversely this boundedness condition implies via (4.60) or, rather, its analog for $\hat{\mathcal{H}}'$ which is somewhat easier to establish, that $\{\hat{f}_l^p(\hat{y})\}$ defines a continuous linear function on $\{\hat{S}_l^p(\hat{y})\}$. There is a slight problem because, although the $\hat{f}_l^p(\hat{y}) \cdot \hat{S}_l^p(\hat{y}) = \hat{y}^p \hat{S}_l^p(\hat{y}) \cdot \hat{y}^{-p} \hat{f}_l^p(\hat{y})$ are uniformly bounded in p, l, and $\{\hat{S}_l^p\}$ in a neighborhood of 0 by our inequalities, we need an additional decrease in p, l to guarantee that the series

$$\sum_{p,l} \hat{f}_l^p(\hat{y}) \cdot \hat{S}_l^p(\hat{y})$$

is uniformly bounded for $\{\hat{S}_l^p\}$ in a neighborhood of 0. This entails an additional polynomial decrease in p which can be established by the argument of (3.129)ff.

The Fourier transform of division by $\hat{y}$ is, up to a factor i which is unimportant, the integral starting from 0. Thus the formal condition that we obtain is that the p-fold integrals

$$\int_0^y dy_p \int_0^{y_p} dy_{p-1} \cdots \int_0^{y_2} f_l^p(y_1)\, dy_1 = I_p f_l^p(y) \tag{4.61}$$

form a bounded set in $\mathcal{H}(Y)$. A standard argument can be used to rigorize this formalism.

We have thus proven

Theorem 4.13 *The WP for $\Box$ with data $\{h_l^p(D)f\}$ on the Y axis is well-posed in the space $\mathcal{H}$ for data $\{f_l^p\}$ on the Y axis satisfying*

$$\{I_p f_l^p\} \text{ is bounded in } \mathcal{H}(Y).$$

The topology of $\{f_p^l\}$ is the natural topology for such bounded sequences.

For the space $\mathcal{E}$ a weaker result holds because the series $\sum h_l^p(\hat{t})\hat{S}_l^p(\hat{y})$ converges only in $\hat{\mathcal{E}}'(V)$ but, as we mentioned above, it does not generally converge in $\hat{\mathcal{E}}'(\mathbb{C}^n)$. As a consequence we can only study the *semi-Cauchy map*

$$\mathcal{C}_s : f \to \{h_l^p(D)f|_Y\} \tag{4.62}$$

for $f \in \text{kernel } P(D)$. Thus we deal only with solutions of the equation $P(D)f = 0$.

The Fourier transform of the dual of this kernel is $\hat{\mathcal{E}}'(V)$ by the main results (fundamental principle) of FA. Thus $\mathcal{C}_s$ is a topological isomorphism on the dual of the space of $\{\hat{S}_l^p\}$ as given by Theorem 4.11. The same arguments as those used in the proof of Theorem 4.13 yield

Theorem 4.14 *The WP for* $\Box$ *with data* $\{h_l^p(D)\}$ *on the* Y *axis is semi-well posed in the space* $\mathcal{E}$ *for data* $\{f_l^p\}$ *on the* Y *axis satisfying*

$$\{I_p f_l^p\} \text{ is bounded in } \mathcal{E}(Y).$$

Semi-well-posed means $\mathcal{C}_s$ is a topological isomorphism.

Problem 4.5 *What is the image* $\mathcal{CE}$*?*

The ideas of Section 1.4 together with Theorem 4.12 show that any solution $f \in \mathcal{E}$ of $\Box f = 0$ admits a Fourier representation

$$f(x) = \int e^{ix\cdot\hat{x}} \frac{d\mu(\hat{x})}{k(\hat{x})} \tag{4.63}$$

where μ is a measure on $\mathbb{C}^{\mathbb{R}}(V) = \mathbb{C}^{\mathbb{R}} \cap V$ and k is in an AU structure for $\mathcal{E}$.

The origin of Theorem 4.12 is Theorem 3.19 which is the analog for euclidean space in place of V. Theorem 3.21 sharpens the Fourier representation of functions in $\mathcal{E}$ coming from Theorem 3.19 by allowing the representing measures to be smooth. Following the same line of reasoning we deduce (we write $x = (t, y)$)

Theorem 4.15 *Any* $f \in \mathcal{E}$ *satisfying* $\Box f = 0$ *has a Fourier representation*

$$f(x) = \int e^{ix\cdot\hat{x}} \hat{f}(\hat{x})\, d\hat{t}. \tag{4.64}$$

For any N, $\hat{f}$ *can be chosen to be a* $\mathbb{C}^N$ *function on* $\mathbb{C}^{\mathbb{R}}(V)$ *which vanishes in the neighborhood of the origin. Moreover there is a* k *in an AU structure for* $\mathcal{E}$ *such that* kf *is bounded on* $\mathbb{C}^{\mathbb{R}}(V)$.

Since $\hat{f}$ is smooth we can decompose the integral in (4.64) in polar coordinates $\hat{r}, \hat{\theta}$. This means we integrate first over the lines $\theta = \hat{\theta}$ fixed and then over $\hat{\theta}$ as in the proof of Theorem 3.22. We write $\hat{x} = \hat{r}\hat{x}_{\hat{\theta}}$ where $\hat{x}_{\hat{\theta}}$ is a unit vector on the line through $\hat{x}$; $\exp(ix \cdot \hat{x})$ is a spread function for the spread (which we call $\hat{\theta}$) of hyperplanes whose Minkowski normal is $\hat{x}(\hat{\theta})$. Thus the integral over the complex line determined by $\hat{\theta}$ is a spread function for $\hat{\theta}$.

We have proven

Theorem 4.16 *Any* $f \in \mathcal{E}$ *satisfying* $\Box f = 0$ *has a Radon representation*

$$f(x) = \int f_{\hat{\theta}}(x)\, d\hat{\theta} \tag{4.65}$$

where $f_{\hat{\theta}}$ *is a spread function for* $\hat{\theta}$*. The* $f_{\hat{\theta}}$ *depend smoothly on* $\hat{\theta}$*. The integral is over the unit sphere in* $\{\hat{t}\}$.

Theorem 4.16 is the general analog for solutions of the wave equation in the space $\mathcal{E}$ of the Whittaker representation of harmonic functions in $\mathbb{C}^3$ (see Section 1.5). We now deal with this topic in more detail.

We want to formulate the WP for the wave operator in terms of orbits of groups in the sense of our discussion in Chapter 3.

The relevant group is $W = SO(n) \times SO(2)$ which acts in the usual way on $(\hat{t}, \hat{y})$. A cross-section for the real orbits is the product $\{\hat{t}_1 \geq 0\} \times \{\hat{y} \geq 0\}$. According to the ideas of Chapter 3 it is better to use the euclidean space $\{\hat{t}_1\} \times \{\hat{y}\} = \mathbb{R}^2$ which is a four sheeted covering of the orbits.

Remark. *If $f(x_1, x_2, \ldots, x_n)$ is harmonic then it is real analytic so we can form $f(ix_1, x_2, \ldots, x_n)$ which is a solution of the wave equation. Thus Theorem 4.16 contains the Whittaker representation of harmonic functions.*

We now impose the algebraic equation $\hat{t}^2 = \hat{y}^2$; that is, we restrict ourselves to this algebraic variety. Since $\hat{t}^2 - \hat{y}^2$ is W invariant this defines an equation on the orbits of W. This means that if $(\hat{t}^0)^2 = (\hat{y}^0)^2$ then all points $\hat{t}, \hat{y}$ on the W orbit of $\hat{t}^0, \hat{y}^0$ satisfy $\hat{t}^2 = \hat{y}^2$. In particular we can regard $\hat{t}^2 - \hat{y}^2 = 0$ as the equation

$$\hat{t}_1^2 - \hat{y}^2 = 0 \tag{4.66}$$

on $\mathbb{R}^2$. The fundamental domain for W acting on $V = \{\hat{t}^2 - \hat{y}^2 = 0\}$ is now "cut down" to the subset of $\{\hat{t}_1\} \times \{\hat{y}\}$ where $\hat{t}_1 = \hat{y}$.

The orbit of $\hat{t}_1 = \hat{y}$ can be described by

$$w_\theta^{\pm}(\hat{y}, 0, \ldots, 0; \hat{y}) = (\hat{y}h_1^1(\theta), \ldots, \hat{y}h_r^1(\theta); \pm\hat{y}). \tag{4.67}$$

Here $w_\theta^{\pm}$ represents a rotation by $\theta \in SO(k)/SO(k-1) = S^k = k$ sphere applied to $(\hat{y}, 0, \ldots, 0)$ in $\hat{t}$ space and $\hat{y} \to \pm\hat{y}$ on the $\hat{y}$ axis. Note that there are no ambiguities in the parametrization unlike the analogous situation dealt with in Chapter 3.

These ideas can be applied to a study of the ultrahyperbolic variety

$$V_{rs} = \{\hat{t}_1^2 + \cdots + \hat{t}_r^2 - \hat{y}_1^2 - \cdots - yh_r^2 = 0\}. \tag{4.68}$$

The relevant group is

$$W_{rs} = SO(r) \times SO(s).$$

A euclidean space covering of a cross-section for real orbits of W_{rs} is $\{\hat{t}_1\} \times \{\hat{y}_1\}$. On V_{rs} the analogous covering of a cross-section is

$$\{(\underbrace{\hat{y}_1, 0, \ldots, 0}_{r}; \underbrace{\hat{y}_1, 0, \ldots, 0}_{s})\}. \tag{4.69}$$

This suggests that an analog of Theorem 3.15 would involve using $\hat{y}$ in place of r and expansions on the orbits of $\hat{y}$ using representation functions for W. However, there is a serious difficulty in this method. The proof of Theorem 3.15 (and the corresponding topological result) depend on the Phragmén–Lindelöf theorem for $\mathbb{R}^n$.

Problem 4.6 *Is the space* $\mathbb{C}^{\mathbb{R}}(V_{rs})$ *of real complex lines in* V_{rs} *sufficient for* $\hat{\mathcal{E}}'(V_{rs})$*?*

Up to now we have dealt with C^∞ solutions. It is for such solutions that we have difficulties in extending Theorem 4.12 to ultrahyperbolic operators. If we replace the space $\mathcal{E}$ by $\mathcal{H}$ then the $\pm$ signs are insignificant for most purposes since we can change the $\pm\hat{y}_j^2$ to $\mp\hat{y}_j^2$ by replacing $\hat{y}_j$ by $i\hat{y}_j$. However, the meaning of "real complex line" depends on the real structure.

Theorem 4.17 *The set of real complex lines* $\mathbb{C}^{\mathbb{R}}(V_{rs})$ *is sufficient for* $\hat{\mathcal{H}}'(V_{rs})$.

Since the kernel of $\Box_{rs}$ on $\mathcal{H}$ depends only on $r+s$ this means that $\mathbb{C}^{\mathbb{R}}(V_{rs})$ is sufficient for $\hat{\mathcal{H}}'(V_{r's'})$ for any (r,s), (r',s') with $r+s=r'+s'$ and $rs\neq 0$.

Proof of Theorem 4.17 For $s=1$ the operator $\Box_{r1}$ is hyperbolic so Theorem 4.12 certainly applies to give the sufficiency of $\mathbb{C}^{\mathbb{R}}(V_{r1})$.

Suppose $s>1$. The direction cosines $(a_1,\dots,a_r;b_1,\dots,b_s)$ of a real line in V_{rs} satisfy

$$\begin{aligned}\sum a_j^2-\sum b_j^2&=0\\ \sum a_j^2+\sum b_j^2&=1.\end{aligned}\tag{4.70}$$

Since we are dealing with solutions of $\Box_{rs}f=0$, $f\in\mathcal{H}$, the CP is well posed on any noncharacteristic hyperplane, e.g. $\hat{y}_s=0$. According to (4.54)ff. $\mathcal{H}(V_{rs})$ can be identified with $\mathcal{H}^2(\{\hat{y}_s=0\})$ which is the space of pairs of functions in $\{\hat{y}_s=0\}$.

By (4.70) the real lines in $\hat{y}_s=0$ which correspond to real lines in V_{rs} have direction cosines $a_1,\dots,a_r$; $b_1,\dots,b_{s-1}$ satisfying

$$\sum_1^r a_j^2-\sum_1^{s-1}b_j^2>0.\tag{4.71}$$

Inequality (4.71) represents an open set of real lines in the hyperplane $\{\hat{y}_s=0\}$. By Theorem 3.23 this set of real lines is sufficient for $\mathcal{H}(\{\hat{y}_s=0\})$.

This completes the proof of Theorem 4.17.

As in the derivation of Theorems 4.15 and 4.16 from Theorem 4.12 we deduce the general Whittaker representation.

Theorem 4.18 *Let* r,s *be given nonnegative integers. Any solution* $f\in\mathcal{H}^{r+s}$ *of* $\Box_{rs}f=0$ *has a Radon representation as an integral over the unit sphere in* $\mathbb{R}^{r+s}$

$$f(x)=\int f_{\hat{\theta}}(x)\,d\hat{\theta}.\tag{4.72}$$

$f_{\hat{\theta}}$ *is an entire spread function for the spread of hyperplanes orthogonal to* $\hat{\theta}$. $\{\hat{\theta}\}$ *corresponds to (4.70) with* $b_s=0$.

We have discussed the line structure on the varieties V_{rs} (cones) in detail; this is relevant for the Whittaker representation of harmonic functions as discussed in Section 1.4. For the Bateman representation we need to analyze 2 planes in such cones.

To understand how things work let us begin with 2 planes in a cone in $\mathbb{R}^4$. The simplest equation which is suitable for real analysis is defined by the algebraic variety in $\mathbb{C}^4$

$$V = \left\{ \begin{pmatrix} \hat{u} & \hat{v} \\ \hat{s} & \hat{w} \end{pmatrix} \middle| \det \begin{pmatrix} \hat{u} & \hat{v} \\ \hat{s} & \hat{w} \end{pmatrix} = 0 \right\}. \tag{4.73}$$

For any $\hat{u}_0, \hat{s}_0$ the variety V contains the 2 plane

$$P_2(\hat{u}, \hat{s}) = \left\{ \begin{pmatrix} r_1\hat{u}_0 & r_2\hat{u}_0 \\ r_1\hat{s}_0 & r_2\hat{s}_0 \end{pmatrix} \right\} \tag{4.74}$$

where r_1 and r_2 are arbitrary complex numbers. Thought of as a subset of $\mathbb{C}^4$ this plane is spanned by $(\hat{u}, \hat{s}, 0, 0)$ and $(0, 0, \hat{u}, \hat{s})$ which are orthogonal for the quadratic form defined by det. In particular, real 2 planes can be parametrized (essentially) by $\{(\hat{u}_0, \hat{s}_0)\}$ where

$$(\hat{u}_0, \hat{s}_0) = (\cos\varphi, \sin\varphi) \quad \text{or} \quad (\hat{u}_0, \hat{s}_0) = (\tau, 1). \tag{4.75}$$

Note that any real point in V is contained in a 2 plane $P_2(\hat{u}, \hat{s})$. It follows that every real line is contained in a real 2 plane. Thus, by Theorem 4.17 the union of the complexifications of real 2 planes is sufficient for $\hat{\mathcal{H}}'(V)$.

The parametrization $\{(\hat{u}_0, \hat{s}_0)\} = \{(\cos\varphi, \sin\varphi)\}$ captures all real points but the parametrization $\{(\hat{u}_0, \hat{s}_0)\} = \{(\tau, 1)\}$ misses the points $\left(\begin{smallmatrix} \hat{u} & \hat{v} \\ 0 & 0 \end{smallmatrix}\right)$. We should really think of this parametrization in terms of the projective line rather than the affine line so $\left\{\left(\begin{smallmatrix} \hat{u}_0 & \hat{v}_0 \\ 0 & 0 \end{smallmatrix}\right)\right\}$ corresponds to $\tau = \infty$. We can, in fact, think of the parametrization $\{(\cos\varphi, \sin\varphi)\}$ as the projectivization of $\{(\tau, 1)\}$ with $\varphi = 0$ corresponding to $\tau = \infty$.

Before going to higher dimensional planes, let us explain how Bateman's representation of harmonic functions fits into our ideas.

We start with the second parametrization in (4.75) of real 2 planes. We make the complex change of variables

$$\begin{aligned} 2\hat{u} &= \hat{x}_1 + i\hat{x}_2 \\ 2\hat{w} &= \hat{x}_1 - i\hat{x}_2 \\ 2\hat{s} &= -\hat{x}_3 - i\hat{x}_4 \\ 2\hat{v} &= \hat{x}_3 - i\hat{x}_4. \end{aligned} \tag{4.76}$$

Then the plane $\{(\hat{u}, \hat{s}, \hat{v}, \hat{w})\} = \{(r_1\tau, r_1, r_2\tau, r_2)\}$ becomes

$$\hat{x} = (r_1\tau + r_2, -r_1 i\tau + r_2 i, -r_1 + r_2\tau, ir_1 + ir_2\tau) \tag{4.77}$$

and det becomes $\frac{1}{4}(\hat{x}_1^2 + \hat{x}_2^2 + \hat{x}_3^2 + \hat{x}_4^2)$.

We write a solution f of the ultrahyperbolic equation as a Fourier transform of $\hat{f}$ over the union of the complex 2 planes parametrized by real τ. Thus, formally,

$$\begin{aligned} f(x) &= \int d\tau \int \hat{f}(r_1, r_2; \tau) e^{ix\cdot\hat{x}}\, dr_1\, dr_2 \\ &= \int d\tau \int \hat{f}(r_1, r_2; \tau) e^{ir_1 x\cdot(\tau, -i\tau, -1, i) + ir_2 x\cdot(1, i, \tau, i\tau)}\, dr_1\, dr_2 \\ &= \int h[\tau x_1 - i\tau x_2 - x_3 + ix_4, x_1 + ix_2 + \tau x_3 + i\tau x_4; \tau)\, d\tau \end{aligned} \tag{4.78}$$

where h is the Fourier transform of $\hat{f}$ in the variables r_1, r_2 for each fixed τ.

This agrees exactly with Bateman's representation (1.82).

In order for this computation to be meaningful we have to know that $\hat{f}$ is exponentially decreasing where the exponential is complex. From (4.74) it follows that $\hat{x}_1$ and $\hat{x}_3$ are real while $\hat{x}_2$ and $\hat{x}_4$ are pure imaginary on real $(\hat{u}, \hat{v}, \hat{s}, \hat{w})$. Since $\hat{f}(\hat{u}, \hat{v}, \hat{s}, \hat{w})$ decreases exponentially in $\Im(\hat{u}, \hat{v}, \hat{s}, \hat{w})$ the convergence of the integrals in (4.74) with respect to r_1, r_2 is established.

There are two difficulties with Bateman's representation:

(1) It involves integration over the noncompact τ. It seems difficult to show directly that the τ integral in (4.75) converges.
(2) The point $\tau = \infty$ is missing. It corresponds to $\hat{s} = \hat{w} = 0$.

We can remove the difficulty (1) by changing variables from τ to ϕ where $\tau = \cot\phi$.

Difficulty (2) is not serious because, as the proof of Theorem 4.17 shows, we need only an open set of real complex lines to obtain a sufficient set. As for (2), the point $\hat{s} = \hat{w} = 0$ corresponds to $\hat{x}_1 - i\hat{x}_2 = \hat{x}_3 + i\hat{x}_4 = 0$. Thus

$$e^{i(x_1\hat{x}_1 + x_2\hat{x}_2 + x_3\hat{x}_3 + x_4\hat{x}_4)} = e^{i\bar{z}_1\hat{x}_1 + iz_2\hat{x}_3}$$

where $\bar{z}_1 = x_1 - ix_2$ and $z_2 = x_3 + ix_4$ so f is a holomorphic function of $\bar{z}_1$ and z_2.

Remark. *The Radon and Bateman representations of solutions are not unique. The deviation from uniqueness is discussed in Chapter 9.*

Let us return to the WP.

It is possible to give an explicit solution to the WP for the wave operator $\Box$. By this we mean that we can compute the value $f(t, y)$ of a solution of $\Box f = 0$ in terms of the WD on the y axis.

For simplicity we take $k = 2$; the case of $k > 2$ presents no essential extra difficulty. Let us compute

$$f(1, 0, 0) = \delta_{1,0,0} \cdot f. \tag{4.79}$$

Thus we want to write

$$\delta_{1,0,0} = \sum \left(\frac{\partial}{\partial t_1} \pm i \frac{\partial}{\partial t_2} \right)^p S_\pm^p + \Box T$$

where $S_\pm^p$ are distributions of y; this means

$$e^{i\hat{t}_1} = \sum (\hat{t}_1 \pm i\hat{t}_2)^p \hat{S}_\pm^p(\hat{y}) \quad \text{on } V. \tag{4.80}$$

According to the Watergate method we set $\hat{t}_1 = \hat{y}\cos\theta$ and $\hat{t}_2 = \hat{y}\sin\theta$ so (4.80) becomes

$$e^{i\hat{y}\cos\theta} = \hat{y}^p \sum e^{\pm ip\theta} \hat{S}_\pm^p(\hat{y}) \tag{4.81}$$

or

$$\begin{aligned} \hat{S}_\pm^p(\hat{y}) &= \hat{y}^{-p} \int e^{\mp ip\theta + i\hat{y}\cos\theta}\, d\theta \\ &= \hat{y}^{-p} J_p(\hat{y}) \end{aligned} \tag{4.82}$$

in the standard notation (except for an unimportant constant) for Bessel functions (see e.g. [11, vol. II p. 81 (2)]).

Note that J_p has a zero at the origin of order p so $\hat{S}_\pm^p$ is an entire function. Since $|\cos\theta| \le 1$, $J_p(\hat{y})$ is an entire function of exponential type 1. This means that $S_\pm^p$ has its support in the unit interval. Moreover it is clear from (4.82) that $\hat{S}_\pm^p(\hat{y}) = \mathcal{O}(|\hat{y}|^{-p})$ as $\hat{y} \to \infty$ so that the derivatives of S^p up to order $p-2$ are continuous.

If we replace the point $(1,0,0)$ in (4.80) by an arbitrary real point (t_1^0, t_2^0, y^0) with $t^0 \neq 0$ then (4.82) becomes

$$\begin{aligned} \hat{S}^p_{\pm(t_1^0,t_2^0,y^0)}(\hat{y}) &= \hat{y}^{-p} \int e^{\mp ip\theta + i\hat{y}|t^0|\cos(\theta-\theta^0) + i\hat{y}y^0}\, d\theta \\ &= \hat{y}^{-p} e^{i\hat{y}y^0} J_p(|t^0|\hat{y}) e^{\mp ip\theta^0}. \end{aligned} \tag{4.83}$$

We have written $t_1^0 = |t^0|\cos\theta^0, t_2^0 = |t^0|\sin\theta^0$.

$J_p(|t^0|\hat{y})$ is an entire function of exponential type $|t^0|$ so the support of $S_\pm^p(t_1^0, t_2^0, y^0)$ is an interval of length $|t^0|$ centered at y^0. Since $f(t_1^0, t_2^0, y^0) = \sum S_\pm^p(t_1^0, t_2^0, y^0) \cdot f_\pm^p(y)$ the value of f at (t_1^0, t_2^0, y^0) depends on the value of its WD on

$$\{|y - y^0| \le |t^0|\}.$$

Proposition 4.19 *The wave equation has speed 1 with respect to the WD, which we have shown.*

This means that the support of a solution of the wave equation expands in t at a speed at most 1 from the support of its WD, as we have seen.

It follows from (4.83) that $S^p_\pm(t^0, y^0)$ is $p-2$ times differentiable in y for $t^0 \neq 0$.

We can use (4.83) to compute some interesting solutions of the wave equation. Let us start with the WD $= (0, \dots 0, \delta_0, 0, \dots)$, meaning that all data except that corresponding to $(\partial/\partial t_1 + i\partial/\partial t_2)^p$ for some fixed large p are zero and that data is δ_0. By the regularity of $S^p_\pm$ described above, this data corresponds to a sufficiently differentiable solution f^p off the Y axis if p is large. Our calculation gives

$$\begin{aligned} f^p(t_1, t_2, y) &= \delta_{t_1,t_2,y} \cdot \mathrm{WD}(f^p) \\ &= S^p_{+(t_1,t_2,y)} \cdot \hat{\delta}_{000} \\ &= e^{-ip\theta} \int \hat{y}^{-p} e^{i\hat{y}y} J_p(|t|\hat{y})\, d\hat{y} \end{aligned} \tag{4.84}$$

by (4.83) since $\hat{\delta}_{000}(\hat{y}) = 1$. ($\delta_{000}$ is the WD of f^p which is the δ function at 0 on the y axis.) Thus $f^p(t, y)$ is $\exp(-ip\theta)$ times the Fourier transform in $\hat{y}$ of $\hat{y}^{-p} J_p(|t|\hat{y})$.

Our first observation is that

$$f^p(t, y) = 0 \quad \text{if } |y| > |t| \tag{4.85}$$

which follows (Paley–Wiener) from the fact that $\hat{y}^{-p} J_p(|t|\hat{y})$ is an entire function of exponential type $|t|$ (or from Proposition 4.19). This means that

$$f^p \text{ vanishes inside the light cone.} \tag{4.86}$$

To go further, we make the change of variables $|t|\hat{y} \to \hat{y}$ in the integral (4.84). We find

$$f^p(t, y) = e^{-ip\theta} |t|^{p-1} \int \hat{y}^{-p} e^{i\hat{y}y/|t|} J_p(\hat{y})\, d\hat{y} = |t|^{p-1} f^p(1, y/|t|) \tag{4.87}$$

so that

$$f^p(\alpha t, \alpha y) = \alpha^{p-1} |t|^{p-1} f^p(1, y/|t|) = \alpha^{p-1} f^p(t, y); \tag{4.88}$$

that is, f^p is homogeneous of degree $p-1$. Moreover, for y fixed, or even for y in a compact set excluding the origin, as $|t|/y \to 1$ the above integral $\to 0$ and similarly for its first $p-2$ derivatives since, as we have already noted, the integral represents a function of class C^{p-2} which vanishes when $|y| > |t|$ (see Figure 4.1).

The function f^p has certain interesting properties.

By (4.88) f^p is a homogeneous solution of the wave equation. Hence f^p can be thought of as a function on the hyperboloid $B = \{t_1^2 + t_2^2 - y^2 = 1\}$. Now, $\square$

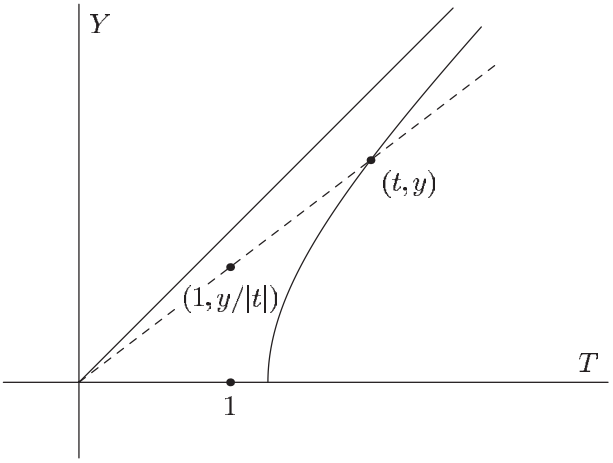

Figure 4.1

separates in hyperbolic spherical coordinates

$$\begin{aligned} t_1 &= r\cosh\zeta\cos\varphi \\ t_2 &= r\cosh\zeta\sin\varphi \\ y &= r\sinh\zeta \\ \Box &= \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} - \frac{1}{r^2}\Delta \end{aligned} \tag{4.89}$$

where Δ is the Laplacian on B; that is, the separated part of $\Box$ depending on B. (See [58] and also Chapter 7 for more details.) r is the Minkowski distance of (t_1, t_2, y) to the origin.

Thus the space of homogeneous solutions of the wave equation of fixed homogeneity degree $p-1$ can be identified with the space of functions u on B satisfying

$$[\Delta - p(p-1)]u = 0. \tag{4.90}$$

B corresponds to $r = 1$ in (4.89). Going to infinity on B means $\zeta \to \infty$ so that

$$\frac{|t|}{|y|} = |\coth\zeta| \to 1. \tag{4.91}$$

We have already remarked following (4.88) that (4.91) implies $f^p(t/|y|, 1)$ and its first $p-2$ derivatives in $t, y \to 0$. In (4.88) let $y = \alpha^{-1} = |\sinh\zeta|$.

Then (see Figure 4.1) $|t|/\sinh\zeta - 1 \to 0$ so that

$$\begin{aligned} f^p(t,y) &= |\sinh\zeta|^{p-1} f^p\left[\frac{t}{|\sinh\zeta|}, 1\right] \\ &= \mathcal{O}\left[|\sinh\zeta|^{p-1}\left(\frac{|t|}{|\sinh\zeta|} - 1\right)^{p-2}\right] \\ &= \mathcal{O}[e^{(p-1)\zeta}(|\coth\zeta| - 1)^{p-2}] \\ &= \mathcal{O}(e^{(p-1)\zeta - 2(p-2)\zeta}) \\ &= \mathcal{O}(e^{(-p+3)\zeta}) \end{aligned} \tag{4.92}$$

as $\zeta \to +\infty$. In particular, f^p is square integrable on B for $p > 5$ since the measure on B is $\sinh 2\zeta \; d\zeta \; d\varphi$. Since the space of solutions of (4.90) is a representation space for G (see Chapter 7 for details) we have shown

Theorem 4.20 *f^p belongs to the discrete series of representation of G for $p > 4$.*

An irreducible representation space for G is called a *discrete series representation* if some representation function is in L_2.

In a forthcoming work on group representations we shall show that this method constructs all discrete series for G. When properly modified it can be used to construct discrete series for general semisimple Lie groups.

Let us give some insight into why f^p belongs to the discrete series. The method we present is the crucial point in our general construction of discrete series.

If we combine (4.83) with (4.84) we see that f^p is represented as the Fourier transform of the measure

$$\hat{f}^p = e^{-ip\theta}\hat{y}^{-p}\, d\hat{y}\, d\theta$$

on the light cone. For, a point on the light cone has coordinates $(\hat{y}\cos\theta, \hat{y}\sin\theta, \hat{y})$ and the inner product of this point with (t_1^0, t_2^0, y^0) is

$$-|t^0|\hat{y}\cos(\theta - \theta^0) + \hat{y}y^0.$$

To make sense of the integral defining f^p we use a contour indented around 0 either positively or negatively (see Figure 4.2). For definiteness we choose the + contour.

We want to know that f^p is smooth and small at infinity on B. Smoothness of f^p depends on smallness of $\hat{f}^p$; this is governed by the factor $\hat{y}^{-p}$. The smallness of f^p depends on the smoothness of $\hat{f}^p$; the factor $\hat{y}^{-p}$ seems to preclude that.

We are saved by Schur orthogonality. In actuality the term $e^{-ip\theta}\hat{y}^{-p}$ has a type of smoothness at the origin. For, we are integrating it against $\exp[-i|t^0|\hat{y}\cos(\theta - \theta^0) + iy^0\hat{y}]$. If we expand $\exp[-i|t^0|\hat{y}\cos(\theta - \theta^0)]$ in a power series then the integrals with respect to θ of all powers $< p$ vanish because they are orthogonal to $\exp(-ip\theta)$. Thus for purposes of Fourier transformation we

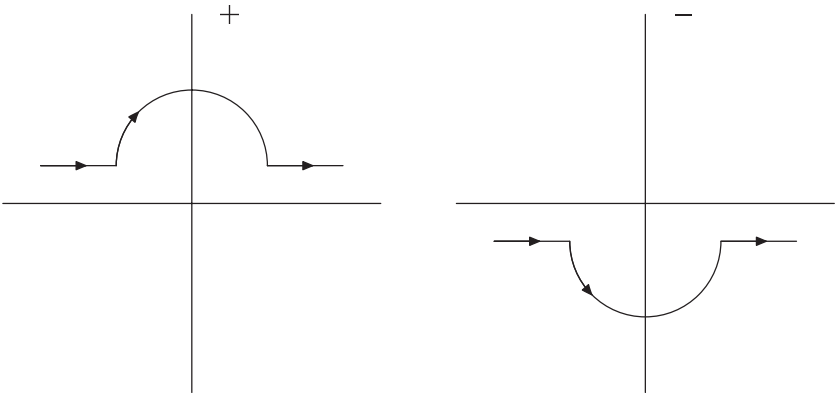

Figure 4.2

can consider the integrand to be smooth so the Fourier transform in $\hat{y}$ is small at infinity.

This is the conceptual reason for the existence of discrete series.

Let us now pass to another example, which is related to the quadratic map discussed at the end of Section 4.1. (We could also study the higher degree maps from a similar point of view.) In view of Theorem 4.9 it is important to study the Plücker variety V_2 defined by (4.37).

For $k = 2$ the Plücker variety is just the light cone in three dimensions. Our above study suggests that there might be a nice WP for V_k when $k > 2$. This means that we seek a linear space $\hat{Y}$ for which the orbits $\mathcal{O}_{\hat{y}}$ (or at least a "real part" of $\mathcal{O}_{\hat{y}}$) are orbits of some compact group W.

Let $G = SL(k, R)$ be the group of real $k \times k$ matrices of determinant 1. G acts on the space $\hat{S}$ of real symmetric matrices by

$$\hat{s} \to g\hat{s}g^t = g(\hat{s}). \tag{4.93}$$

(Note the change in notation: in Section 4.1 symmetric matrices were denoted by $\hat{y}$.) Let W be the orthogonal subgroup of G. Then W preserves the identity matrix $\hat{s} = I$ and, in fact, it is easily seen that the set of all $\hat{s}$ preserved by W consists of the scalar multiples $\{\hat{y}I\} = \hat{Y}$.

We want to find a WP with data on Y = the Fourier transform of $\hat{Y}$. In order to do this concretely we first need a nondegenerate quadratic form to define the Fourier transform on $\hat{S}$. In case $k = 2$ the determinant of $\hat{s}$ is a G invariant nondegenerate quadratic polynomial. But for $k > 2$ there are no invariant quadratic polynomials. To see this, observe that $\hat{s}$ can be diagonalized by a $w \in W$, meaning

$$w\hat{s}w^t = w\hat{s}w^{-1} = \hat{d} \tag{4.94}$$

where $\hat{d}$ is diagonal. If $\hat{d}_0 = |\det \hat{d}|^{1/k} \neq 0$ then $\hat{d}$ is of the form

$$\hat{d} = a(\hat{d}_0 I^0) = a\hat{d}_0 I^0 a \tag{4.95}$$

where a is a diagonal matrix in G and I^0 is diagonal with entries ± 1. Hence

$$\hat{s} = w^{-1} a \hat{d}_0 I^0 a (w^{-1})^t.$$

The orbits of G on the set of fixed determinant $\neq 0$ are determined by the finite set $\{I^0\}$, so there are open G orbits in the algebraic sets $\{\det \hat{s} = c\}$. Any G invariant polynomial must therefore be constant on the sets where det=c and hence is a polynomial in det. Unfortunately, such polynomials (unless they are trivial) are of degree $\geq k$.

The best we can do, therefore, is to use the nondegenerate W invariant quadratic polynomial

$$\|\hat{s}\|^2 = \text{tr } \hat{s}^2. \tag{4.96}$$

In case $k = 2$ we have a better choice than this, namely $\text{tr}(J\hat{s}J^{-1}\hat{s})$ where J is the Weyl element $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ (see Section 7.2) since

$$\text{tr}(J\hat{s}J^{-1}\hat{s}) = 2 \det \hat{s}$$

which is G invariant. But no such formalism exists for G when $k > 2$ so we content ourselves with (4.96). (There are other interesting groups for which invariant quadratic polynomials exist.)

$\hat{Y}$ is the linear space of scalar multiples of the identity. Denote by $\hat{T}$ the orthogonal space to $\hat{Y}$. By (4.96) $\hat{T}$ is the $k(k+1)/2 - 1$ dimensional plane of traceless matrices.

We have shown in (4.94) and (4.95) that the "generic" orbits of G, i.e. those consisting of points with nonvanishing determinant, are the orbits of $\hat{d}_0 I^0$ where $\hat{d}_0$ is a positive scalar and I^0 is a diagonal matrix with entries ± 1. The Weyl group $W_0 \subset W$ is the finite subgroup of permutation matrices; it acts as permutations on the diagonal matrices in $\hat{S}$. Thus the nondegenerate orbits of G are determined by $\hat{d}_0$ and the number of positive signs in I^0 or, what is the same thing, by the index. In particular, there are $k+1$ generic orbits for each $\hat{d}_0$.

The same argument can now be applied to $\hat{S}_0 = \{$symmetric matrices of determinant $0\}$. Using (4.94) we see that every orbit contains a diagonal element $\hat{d}$. The rank of $\hat{s}$ which is the number of nonzero entries of $\hat{d}$ is constant on the G orbit of $\hat{s}$. For $g \in G$ cannot increase the rank and we can go back from $g(\hat{s})$ to $\hat{s}$ by using g^{-1}. Thus each orbit is determined by the rank $r(\hat{s})$, and by the index of the nondegenerate minor of rank r. We do not need scalar multiplication by $\hat{d}_0$ as before because when $r < k$ the $k - r$ entries in a corresponding to the zeros in $\hat{d}$ are arbitrary; the condition $\det a = 1$ is irrelevant.

We now come to the G invariant set that is of interest to us. This is V_2 which consists of symmetric matrices of the form $\{\hat{b}_i\hat{b}_j\}$, i.e. symmetric matrices of rank 1. For $k = 2$, V_2 is the light cone; it is studied in Section 7.2. (V_2 is actually two orbits.) In general it contains the quadratic transform of $\mathbb{R}^k$ as an open set. We want to parametrize $\hat{\mathcal{E}}^{\mathcal{C}}(V_2)$ by functions on $\hat{Y}$ times polynomials in the orthogonal direction.

Since $\hat{Y}$ is W fixed and since W is norm preserving, W acts on $\hat{T}$ (which is obvious anyway). In order to define a WP we must first study the harmonic function theory on $\hat{T}$ using the invariants of W.

Let us examine the orthogonal projection of V_2 on $\hat{T}$. The points q in V_2 can be diagonalized by W. Since the rank of $q \neq 0 \in V_2$ is 1 and since $W_0 \subset W$, q lies in the W orbit of

$$\begin{pmatrix} b & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & 0 & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} = \begin{pmatrix} b/k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & b/k \end{pmatrix} + \begin{pmatrix} ((k-1)/k) \times b & 0 & 0 & 0 \\ 0 & -b/k & 0 & 0 \\ 0 & 0 & -b/k & 0 \\ 0 & 0 & 0 & -b/k \end{pmatrix}. \tag{4.97}$$

In general we can write the orthogonal decomposition of an element q of V_2 as

$$\begin{pmatrix} b_1^2 & b_1b_2 & \dots b_1b_k \\ b_2b_1 & b_2^2 & \dots \\ \vdots & \vdots & \\ b_kb_1 & . & \dots \end{pmatrix} = \begin{pmatrix} b/k & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & b/k \end{pmatrix} + \begin{pmatrix} b_1^2 - b/k & b_1b_2 & \dots b_1b_k \\ b_2b_1 & b_2^2 - b/k & \dots \\ \vdots & \vdots & \\ b_kb_1 & \dots & \end{pmatrix} \tag{4.98}$$

where $b = \sum b_j^2 = \operatorname{Tr} q$.

Since W preserves the trace, (4.98) shows that the projection on $\hat{Y}$ is independent of the point on the orbit and thus characterizes its orbit.

The invariants of W acting on traceless symmetric matrices $\hat{t}$ are

$$\vec{\imath} = (\operatorname{tr} \hat{t}^2, \operatorname{tr} \hat{t}^3, \dots, \operatorname{tr} \hat{t}^k) \tag{4.99}$$

since the invariants of W acting on all symmetric matrices M are spanned by the characteristic roots of M or, equivalently, by the traces of the powers of M.

The isotropy group in W of a generic point in $\hat{T}$ is finite. For, such a point can be diagonalized with distinct diagonal entries. The only matrices that commute with a diagonal matrix with distinct eigenvalues are diagonal. The only diagonal matrices in W have entries ± 1. Thus the isotropy group of a generic point is

finite. Hence if $\mathcal{O}$ is a generic orbit

$$\begin{aligned}\dim \mathcal{O} &= \dim W \\ &= \frac{k(k-1)}{2} \\ &= \frac{k(k+1)}{2} - 1 - (k-1) \\ &= \dim \hat{T} - (k-1). \end{aligned} \tag{4.100}$$

This, or other considerations, shows that $\vec{\imath}$ is strongly free.

Although the generic orbits can be described by $\vec{\imath} = \vec{c}$ with no multiplicity, this is not the case for the projection on $\hat{T}$ of the orbits in V_2. We see from (4.97) that the isotropy group of a point $\hat{t} \in \text{proj}_{\hat{T}} V_2$, $\text{Tr}\, \hat{t} \neq 0$, is $W \cap [O(1) \times O(k-1)]$ which has dimension $(k-1)(k-2)/2$. Thus the dimension of the W orbit of $\hat{t}$ is

$$\frac{k(k-1)}{2} - \frac{(k-1)(k-2)}{2} = k - 1 = \dim V_2 - 1. \tag{4.101}$$

These orbits are degenerate and must be treated by the methods of **A** ((3.52)ff).

Problem 4.7 *Develop the theory of harmonicity for degenerate orbits. Use it to formulate a WP for V_2 with data on the Y axis.*

In Chapter 5 we shall discuss a different type of parametrization problem for this system of equations.

4.3 The noncompact Watergate problem

We return to the wave equation, but we now change our viewpoint in that we want to give Watergate data (WD) on a space-like line. Thus we write the wave equation in the form

$$\Box \equiv \frac{\partial^2}{\partial t_1^2} - \frac{\partial^2}{\partial t_2^2} - \frac{\partial^2}{\partial y^2}.$$

The algebraic variety V associated to $\Box$ is defined by

$$\hat{t}_1^2 - \hat{t}_2^2 = \hat{y}^2. \tag{4.102}$$

For $\hat{y}$ fixed the orbit $\mathcal{O}_{\hat{y}}$ on V is the set of $\hat{t}$ satisfying (4.102).

We are in the situation studied in Section 3.4. We want to use the expansion theory of Section 3.4 to construct a WP with data on the Y axis. For the analogous problem in the case when the $\mathcal{O}_{\tilde{y}}$ are compact (Section 4.2) we consider the expansion basis $\{\exp(\pm ip\hat{\theta})\}$ to be restrictions of $\{(\hat{t}_1 \pm i\hat{t}_2)^p\}$ to circles. The extensions $(\hat{t}_1 \pm i\hat{t}_2)^p$ are "nice" because their Fourier transforms have support at the origin; this has the consequence that the compact WP can be formulated with data WD(f) given by restrictions to the Y axis of the local operators $(\partial/\partial t_1 \pm i\partial/\partial t_2)^p$ applied to f.

In the present section we seek "nice" functions to extend $\exp(ix\xi)$ from hyperbolas in $\hat{t}$ space to all $\hat{t}$. There are no polynomials which give the extension so we cannot define a WP in this way with local data on a line. Rather we shall use another idea that appears in Section 3.4 which indicates that we should multiply $\exp(is\xi)$ by a Bessel function $K_s(\hat{r})$ to obtain an extension whose Fourier transform has support on a hyperbola. (Actually we shall use a variation of this idea.)

We should deal with functions which belong to the space $\mathcal{S}$ if we hope to obtain a reasonable expansion theory. We begin by presenting some properties of the space $\hat{\mathcal{S}}(V)$. We define $\hat{\mathcal{S}}(V)$ as the space of functions on $\Re V$ which extend to functions in $\hat{\mathcal{S}}(\mathbb{R}^3)$. We can describe such functions $\hat{f}$ intrinsically by: $\hat{f}$ is C^∞ on V and its derivatives decrease at infinity faster than $|x|^{-N}$ for any N.

For this to have meaning we have to "normalize" the derivatives so as to give a precise meaning to "growth of derivatives at infinity" and also describe what happens at the origin. The normalization can be done in two ways:

(1) Since $\Re(V)$ is a circle times a line we can use $\partial^N/\partial\hat{\theta}^{N_1}\partial\hat{r}^{N_2}$.
(2) We can use the usual euclidean metric in $\mathbb{R}^3$ to project $\partial/\partial\hat{t}_j$ and $\partial/\partial\hat{y}$ on the tangent space to V (except at the origin).

$\partial/\partial\hat{r}$ and its powers are defined at the origin (not $\partial/\partial\hat{\theta}$). ($\hat{r}$ is regarded as a parameter on a whole—not half—line.) We have arrived at a position which is in the same spirit as that of Proposition 2.1. We have functions $\hat{f}(\hat{r},\hat{\theta})$ on each real line in V which depend in a C^∞ manner on $\hat{\theta}$ and satisfy COMP ORIGIN, meaning that there is a formal power series $\hat{f}^0(x)$ at the origin whose restriction to each line $\hat{\theta}^0$ in V agrees with the formal power series of $\hat{f}(\hat{r},\hat{\theta}^0)$ at the origin.

We claim that, conversely, if $\hat{f}$ is a function on V which satisfies these conditions, then $\hat{f}(\hat{t},\hat{\theta})$ is C^∞ near the origin, meaning that there is a C^∞ function $\widetilde{f}(x)$ which agrees with $\hat{f}$ near the origin.

To prove this we use the mapping (4.54) which identifies functions on V with pairs of functions $\hat{f}_0, \hat{f}_1$ on the $\hat{t}_2, \hat{y}$ plane. It is clear from the explicit formulas that $\hat{f}_0$ and $\hat{f}_1$ are C^∞ on every line $\hat{\phi}^0$ in this plane. Moreover the restriction of the formal power series $\hat{f}^0(x)$ to V corresponds to two formal power series $\hat{f}_0^0, \hat{f}_1^0$ which agree with the formal power series of $\hat{f}_0, \hat{f}_1$ respectively at the origin on each line $\hat{\phi}^0$. Thus COMP ORIGIN holds for $\hat{f}_0, \hat{f}_1$ so they agree locally with C^∞ functions $\hat{g}(\hat{t}_2,\hat{y}), \hat{g}_1(\hat{t}_2,\hat{y})$. Relation (4.54) shows that $\hat{g}_0 + \hat{t}_1\hat{g}_1$ is a C^∞ function which equals $\hat{f}$ on V near the origin.

We conclude:

Theorem 4.21 *The fundamental principle is valid for the space $\hat{\mathcal{S}}$ and the variety V. This means that the restriction map $\hat{f} \to \hat{f}|_V$ is a topological isomorphism from $\hat{\mathcal{S}}/(\hat{t}_1^2 - \hat{t}_2^2 - \hat{y}^2)\hat{\mathcal{S}}$ onto $\hat{\mathcal{S}}(V)$. $\hat{\mathcal{S}}(V)$ can be described intrinsically as the space of functions $\hat{f}(\hat{r},\hat{\theta})$ which are C^∞ and for all $N = N_1 + N_2$*

(a) $\left|\dfrac{\partial^{N_1+N_2}\hat{f}}{\partial\hat{r}^{N_1}\partial\hat{\theta}^{N_2}}(\hat{r},\hat{\theta})\right| \le C_N(1+\hat{r})^{-N}$.

(b) $\hat{f}$ *satisfies COMP ORIGIN.*

The topology of $\hat{\mathcal{S}}(V)$ *is defined by the norms associated to* (a).

By duality we have (compare Section 1.4)

Theorem 4.22 *Solutions* $U \in \mathcal{S}'$ *of the wave equation are represented by Fourier integrals of distributions* $\hat{U} \in \hat{\mathcal{S}}'(V)$.

What is the intrinsic description of the space $\mathcal{S}(V)$? If $f \in \mathcal{S}(V)$ then $f \in C^\infty$ because $\hat{f}$ is small at ∞; in fact, the (t_1, t_2, y) derivatives of any order are bounded. But f is not L_2 because $f \in L_2$ and $\hat{\Box}\hat{f} = 0$ implies $\hat{f} \equiv 0$ since the support of the L_2 function $\hat{f}$ cannot lie in V which has measure 0.

Note that

$$\begin{aligned} \widehat{\hat{r}\frac{\partial}{\partial\hat{r}}} &= -r\frac{\partial}{\partial r} - 3 \\ \widehat{\frac{\partial}{\partial\hat{\theta}}} &= \frac{\partial}{\partial\theta} \end{aligned} \tag{4.103}$$

It follows from Theorem 4.21 that

$\mathcal{S}(V)$ *is the space of solutions of* $\Box$ *for which all* (t_1, t_2, y) *derivatives of*

$$\left(r\frac{\partial}{\partial_r}\right)^{N_1}\left(\frac{\partial}{\partial\theta}\right)^{N_2} f$$

are bounded for any N_1, N_2.

This is stronger than the boundedness of all (t_1, t_2, y) derivations of f because the coefficients of $r\partial/\partial r$ and $\partial/\partial\theta$ thought of as operators in (t_1, t_2, y) are linear functions (are unbounded).

Let us examine the "orbits"

$$\mathcal{O}_{\hat{y}} = \{\hat{t}_1^2 - \hat{t}_2^2 = \hat{y}^2\}. \tag{4.104}$$

V is a two sheeted cover of $\hat{T}$. This means that we use the Lagrange interpolation method as in (4.54)ff. to identify functions $\hat{f} \in \hat{\mathcal{S}}(V)$

$$\hat{f} = \hat{u}_1(\hat{t}) + \hat{y}\hat{u}_2(\hat{t}) \qquad \text{on } V \tag{4.54*}$$

with pairs of functions

$$(\hat{u}_0, \hat{u}_1) \in \hat{\mathcal{S}}(|\hat{t}_1| \ge |\hat{t}_2|). \tag{4.105}$$

$\hat{u}_0, \hat{u}_1$ are the push down of $\hat{f}$ from V to $\hat{T}$. Here $\hat{\mathcal{S}}(|\hat{t}_1| \ge |\hat{t}_2|)$ is the space of functions which are C^∞ on $\{|\hat{t}_1| \ge |\hat{t}_2|\}$ whose derivatives of any order decrease

at infinity faster than $|\hat{t}|^{-N}$ for all N. (Alternatively $\hat{\mathcal{S}}(|\hat{t}_1| \geq |\hat{t}_2|)$ is the space of restrictions of functions of $\hat{\mathcal{S}}(\hat{T})$ to $|\hat{t}_1| \geq |\hat{t}_2|$.)

There is a technical point to show that $\hat{u}_j \in \hat{\mathcal{S}}(|\hat{t}_1| \geq |\hat{t}_2|)$. It is clear from the Lagrange interpolation formulas that $\hat{u}_j \in C^\infty$ on $\{|\hat{t}_1| > |\hat{t}_2|\}$ and $\hat{u}_j$ satisfies the desired inequalities there. The only difficulty occurs at the boundary.

In its simplest form the problem is that of showing that for a C^∞ function w on the unit circle in the x_1, x_2 plane the functions $w_j(x_1)$ defined by

$$\begin{aligned} w_0(x_1) &= \frac{1}{2}\left[w\left(x_1, \sqrt{1-x_1^2}\right) + w\left(x_1, -\sqrt{1-x_1^2}\right)\right] \\ w_1(x_1) &= \frac{1}{2\sqrt{1-x_1^2}}\left[w\left(x_1, \sqrt{1-x_1^2}\right) + w\left(x_1, -\sqrt{1-x_1^2}\right)\right] \end{aligned} \tag{4.106}$$

are C^∞. Note that $w(x_1, x_2) + w(x_1, -x_2)$ is an even function of x_2 and hence is a C^∞ function of x_2^2. Thus $w \in C^\infty$ except possibly at $x_1 = 0$. Similarly $w(x_1, x_2) - w(x_1, -x_2)$ is an odd function of x_2 so that $[w(x_1, x_2) - w(x_1, x_2)]/x_2$ is also a C^∞ function of x_2. This shows that $w_j \in C^\infty$ except possibly at the origin. We can use Taylor's formula with remainder as in the proof of Proposition 2.1 to establish C^∞ at the origin.

A second method (which avoids the problem at the origin) is to write w in a Fourier series

$$w(\theta) = \sum w_n^1 \cos n\theta + w_n^2 \sin n\theta. \tag{4.107}$$

We can use the idea of Tchebyshev to write

$$\begin{aligned} \cos n\theta &= p_n(\cos\theta) = p_n(x_1) \\ \sin n\theta &= \sin\theta q_n(\cos\theta) = x_2 q_n(x_1) \end{aligned}$$

where p_n, q_n are polynomials [11, vol. III].

Since the Fourier series of w converges uniformly on the circle as do the Fourier series of its derivatives of any order it follows that

$$w(\theta) = w_1(x_1) + x_2 w_2(x_1) \quad \text{on } |x| = 1 \tag{4.108}$$

where $w_j \in C^\infty(|x_1| \leq 1)$.

We can now use either of these ideas to show that $\hat{u}_j \in \hat{\mathcal{S}}(|\hat{t}_1| \geq |\hat{t}_2|)$. (We shall illustrate the use of the Fourier series method.) We have already shown, in the argument leading to Theorem 4.21, how to identify functions $\hat{f}$ in $\hat{\mathcal{S}}(V)$ with pairs of functions $\hat{f}_0, \hat{f}_1$ in $\hat{\mathcal{S}}(\hat{t}_2, \hat{y})$ (push down). We expand $\hat{f}_j$ in Fourier series in this plane and then reverse the push down by writing *à la* (4.108)

$$\hat{f} = \hat{f}_0(\hat{t}_2, \hat{y}) + \hat{t}_1 \hat{f}_1(\hat{t}_2, \hat{y}) \quad \text{on } V$$

to obtain a Fourier series expansion on V.

For any fixed $\hat{t}_1$ the orbit $\mathcal{O}_{\hat{t}_1}$ is the real circle in the $(\hat{t}_2, \hat{y})$ plane of radius $|\hat{t}_1|$, so we can use the above-described Fourier series method (4.108) (with $x \leftrightarrow (\hat{t}_2, \hat{y})$) to write

$$\hat{f} = \hat{u}_0(\hat{t}_2) + \hat{y}\hat{u}_1(\hat{t}_2) \quad \text{on } \mathcal{O}_{\hat{t}_1}$$

where $\hat{u}_0, \hat{u}_1 \in C^\infty(|\hat{t}_2| \leq |\hat{t}_1|)$.

This expansion can also be derived by the decomposition (4.54*) of $\hat{f}|_{O_{\hat{t}_1}}$ into functions which are odd or even in $\hat{y}$ as above.

We now let $\hat{t}_1$ vary. This results in

$$\hat{f} = \hat{u}_0(\hat{t}_1, \hat{t}_2) + \hat{y}\hat{u}_1(\hat{t}_1, \hat{t}_2) \quad \text{on } |\hat{t}_1| \geq |\hat{t}_2|. \tag{4.109}$$

$\hat{u}_j \in C^\infty(|\hat{t}_1| \geq |\hat{t}_2|)$ except possibly at $\hat{t}_1 = 0$. At $\hat{t}_1$ we can use the Fourier series and COMP ORIGIN as we have used above to show that $\hat{u}_j \in C^\infty$ there.

Finally we can show that $\hat{u}_j \in \hat{\mathcal{S}}(|\hat{t}_1| \geq |\hat{t}_2|)$ because we can use monomials in the operators $\partial/\partial\hat{\theta}, \partial/\partial\hat{t}_1$ to define the space $\hat{\mathcal{S}}(V)$ and these "transform nicely" in our construction.

By (4.109) and duality we have

Proposition 4.23 *There is a semi-well-posed CP for $\Box$ in the space $\mathcal{S}'$ if the CD lies in $\mathcal{S}'(T)$ and the support of the Fourier transform of the CD lies in $|\hat{t}_1| \geq |\hat{t}_2|$.*

5

THE NONLINEAR RADON AND FOURIER TRANSFORMS

The linear Fourier transform is founded on the function $\exp(i\hat{x} \cdot x)$ which is the exponential of the generic linear function in x. In nonlinear Fourier analysis the linear function is replaced by general polynomials of various kinds. In this work the polynomials are of the form $\hat{x} \cdot x + \hat{p} \cdot p$ where p is an enumeration of the monomials of fixed degree m and $\hat{p}$ is a vector of constants.

In Section 5.1 we show that the relation of the nonlinear Fourier transform to certain nonlinear Radon transforms parallels the relation of the ordinary Fourier transform to the usual (linear) Radon transform. An inversion formula for such Radon transforms is developed.

The remainder of Chapter 5 is, for the most part, concerned with applications of the nonlinear Fourier transform. The characterization of local regularity and local vanishing of functions by means of the nonlinear Fourier transform is given in Section 5.2. As a consequence we obtain new insights into edge-of-the-wedge theorems and lacunas. (More profound forms of the edge-of-the-wedge theorem appear in Sections 9.3 and 9.4.)

Various types of wave front sets are studied in Section 5.3. Some applications to partial differential equations are given. Usual wave front sets are related to local questions—microlocal analysis. Our methods allow applications to global problems—microglobal analysis, and hence, for example, to the characteristic Cauchy problem. This is introduced in Section 5.4.

5.1 Nonlinear Radon transform

Up to now our studies of the Radon transform have been confined to integration over linear varieties. We wish to extend our ideas so as to apply to nonlinear varieties.

Our study of the linear Radon transform was based on three components:

(1) *Spreads.* These allow for the introduction of analysis since the geometric spreads are the level sets of functions.
(2) *Projection slice mechanism.* The integrals over the leaves of a spread (projection) of a function f are transformed into the restrictions (slice) of some transform $\tilde{f}$ of f to a subset of an associated space.
(3) *Inversion.* There is an inverse transform leading from $\tilde{f}$ to f.

We shall present some examples in which the geometric spreads are the level sets of nonlinear functions and the transform from f to $\tilde{f}$ is a variation of the usual Fourier transform.

Suppose we have geometric spreads parametrized by a Grassmann parameter $\mathbf{g}$ whose leaves are parametrized by s. Suppose that the spreads are defined by functions $\alpha(\mathbf{g})(x)$ with the leaves being the sets $\alpha(\mathbf{g}) = s$. We could introduce a Fourier transform

$$\begin{aligned} F(\mathbf{g}, \hat{s}) &= \int e^{i\alpha(\mathbf{g})(x)\cdot\hat{s}} f(x)\, dx \\ &= \int e^{is\cdot\hat{s}}\, ds \int_{\alpha(\mathbf{g})=s} f(x)\, d\mu(x) \\ &= \int e^{is\cdot\hat{s}} \mathbf{R}_\alpha f(\mathbf{g}, s)\, ds \end{aligned}$$

provided that $d\mu(x)$ is such that dx is a Fubini measure for this spread, meaning that it decomposes as a product of ds and $d\mu(x)$ on the sets $\alpha_i(\mathbf{g}) = s$.

The one-dimensional Fourier inversion formula shows that $F(\mathbf{g}, \hat{s})$ encodes exactly the information that is contained in the set of integrals of f over the sets $\alpha(\mathbf{g}) = s$.

We certainly cannot expect to determine f from $F(\mathbf{g}, \hat{s})$ for a single $\mathbf{g}$. Somehow we need to have the functions $F(\mathbf{g}, \hat{s})$ to be defined on all the subsets (slices) $\{(\mathbf{g}, \hat{s})\}$ of a fixed space $\tilde{R} = U\{(\mathbf{g}, \hat{s})\}$. Moreover, on $\tilde{R}$ we have an inversion formula allowing us to reconstruct f from the set of $F(\mathbf{g}, \hat{s})$.

In what follows we shall present a comprehensive theory in case the functions $\alpha(\mathbf{g})$ are homogeneous polynomials of a fixed degree m. It should be possible to extend our theory to inhomogeneous polynomials; we have not attempted to do so because the algebra seems rather formidable.

There is one aspect of spreads which is puzzling and leads to interesting problems. Suppose we have a spread defined by a polynomial α of degree m which is not necessarily homogeneous. The leaves are defined by $\alpha(x) - s = 0$. The polynomials $\alpha(x) - s$ have all coefficients fixed except the constant coefficient which varies with s. What happens if we change things and vary other coefficients?

An example is

$$\begin{aligned} \alpha(s; x) &= x_1^2 + \cdots + x_n^2 - 2s_1x_1 - \cdots - 2s_nx_n \\ &= \sum (x_j - s_j)^2 - \sum s_j^2. \end{aligned}$$

The sets $\alpha(s; x) = 0$ are the spheres passing through the origin. They can be decomposed into spreads (coarse grains) on $\mathbb{R}^n \setminus \{0\}$. Namely, for each ray $\mathbf{g}$ we define the spread consisting of all spheres centered on $\mathbf{g}$ passing through the origin. We denote the corresponding Radon transform by $\mathbf{R}_\alpha f(\mathbf{g}, s)$. Although they are geometric spreads on $\mathbb{R}^n \setminus \{0\}$ they cannot be the level sets of functions $\beta_\mathbf{g}$ since for every c the set $\beta_\mathbf{g} - c = 0$ would contain the origin, so that in some limit sense $\beta_\mathbf{g}$ would take all values at $\{0\}$.

Note that for any ray $\mathbf{g}$ the spread defined by $\mathbf{g}$ consists of all multiples $rT_\mathbf{g}$ where $T_\mathbf{g}$ is the sphere of radius 1 centered on $\mathbf{g}$ and passing through the origin.

One might have some success dealing with the Radon transform $\mathbf{R}_\alpha$ by way of the Mellin transform in scalar multiplication.

Problem 5.1 *Is a function f of compact support on $\mathbb{R}^n \setminus \{0\}$ determined by $\mathbf{R}_\alpha f$? If so, find an inversion formula.*

In a somewhat different vein we can examine spheres of varying center and fixed radius (say 1). Since the set of such spheres is obtained by translation from the unit sphere S we can write the corresponding Radon transform as

$$\mathbf{R}_S f(x) = (f * \delta_S)(x).$$

The translates of S do not decompose into spreads (globally), but they do decompose locally. For if $\mathbf{g}$ represents a line through the origin then it is geometrically evident that the small translates of δ_S in the direction of $\mathbf{g}$ form a geometric spread near the points on $\mathbf{g}$ of distance ± 1 to the origin. In fact suitable translates define a spread locally near any point. Given any point x and any spread $\tilde{\mathbf{g}}$ of affine hyperplanes there are translates of S which approximate $\tilde{\mathbf{g}}$ near x. This leads us to expect that there should be local injectivity and local invertibility near any point, i.e. for functions of small support.

Indeed this is the case; in fact if f is of compact support then $f * \delta_S$ cannot vanish identically. For the Fourier transform of $f * \delta_S$ is $\hat{f}\hat{\delta_S}$; both $\hat{f}, \hat{\delta_S}$ being the Fourier transforms of functions of compact support are entire functions so their product cannot vanish identically.

However, there are functions f which are not of compact support for which $f * \delta_S \equiv 0$. For example, if $f(x) = \exp(ix\hat{x})$ where $\hat{\delta_S}(-\hat{x}) = 0$ (since $\hat{\delta}_S$ is not an exponential it has zeros) then

$$\begin{aligned}(f * \delta_S)(y) &= \int_S e^{i(y-x)\cdot\hat{x}}\, dx \\ &= e^{iy\cdot\hat{x}}\hat{\delta}_S(-\hat{x}) \\ &= 0.\end{aligned}$$

For $n = 2$ it is proven by Delsarte [30] (see also Weil [157]) that if r_1, r_2 are two radii whose ratio is not the zero of a suitable Bessel function then the map

$$f \to [f * \delta_{|x|=r_1}, f * \delta_{|x|=r_2}]$$

is injective.

We can put $\mathbf{R}_S$ in a form which is consenent with the other Radon transforms we have studied. The sphere of radius ρ centered at s is defined by

$$x \cdot x - 2s \cdot x + s \cdot s - \rho^2 = 0.$$

This equation differs from the equation for $\mathbf{R}_\alpha$ because the constant term for $\mathbf{R}_\alpha$ is 0 while for $\mathbf{R}_S$ it is $s \cdot s - 1$. In general we can imagine a situation in which we have spreads defined by polynomial equations $\mathbf{P}(x, \beta) = 0$ with coefficients β in some parameter space defined by equations $Q_j(\beta) = 0$.

It would be of interest to find some general theory of such Radon transform.

Let us turn to the type of Radon transform which we understand; this was introduced in Section 1.3.

Let $p = (p_1, \ldots, p_N)$ be an enumeration of the monomials of degree m on $\mathbf{R}^n$, so $N = \binom{m+n-1}{m}$. We define the Radon transform of degree m by

$$\mathbf{R}_m f(\hat{p}, s) = \int_{\hat{p} \cdot p = s} f(x) dx. \tag{5.1}$$

Here $\hat{p} = (\hat{p}_1, \ldots, \hat{p}_N)$ is a vector with constant components so $\hat{p} \cdot p$ represents the most general polynomial which is homogeneous of degree m. The measure dx is not the euclidean measure but rather the euclidean measure divided by $|\nabla(\hat{p} \cdot p)|$.

In keeping with the ideas of Chapter 1, each $\hat{p}$ defines a spread and s is the spread parameter. Actually the whole line through the origin containing $\hat{p}$ defines a single spread since $t\hat{p} \cdot p = s$ is the same as $\hat{p} \cdot p = \frac{s}{t}$. Thus the spreads form projective space $\hat{P}$ of dimension $N - 1$.

The spread corresponding to $\hat{p}$ can be defined by differential equations in much the same way as in the linear theory. We have

$$d\,(\hat{p} \cdot p) = \hat{p} \cdot d\,p$$

where $dp = (dp_1, \ldots, dp_N)$. The tangent space to the hypersurface $\hat{p} \cdot p = c$ at x is the orthogonal complement of $\hat{p} \cdot dp(x)$.

We want to invert $\mathbf{R}_m$. In accordance with the case $m = 1$ we should expect to write any f as some operator applied to an integral over $\{\hat{p}\}$ of $\mathbf{R}_m^* \mathbf{R}_m f(\hat{p}, x)$ where for each fixed $\hat{p}$

$$\mathbf{R}_m^* \mathbf{R}_m f(\hat{p}, x) = \int (\mathbf{R}_m f(\hat{p}, s)) \delta_{\hat{p} \cdot p = s}\, ds.$$

The measure on $\hat{p} \cdot p = s$ is, as in (5.1), the euclidean measure divided by $|\nabla(\hat{p} \cdot p)|$.

In analogy with the linear case ($m = 1$) we deal with $\mathbf{R}_m$ by means of the *nonlinear Fourier transform*

$$F(\hat{p}, \hat{x}) = \int e^{i\hat{p} \cdot p + i\hat{x} \cdot x} f(x) dx. \tag{5.2}$$

The relation to $\mathbf{R}_m$ occurs when $\hat{x} = 0$. For then we can write

$$\begin{aligned} F(t\hat{p}^0, 0) &= \int e^{it\hat{p}^0 \cdot p} f(x) dx \\ &= \int e^{its} ds \int_{\hat{p}^0 \cdot p = s} f(x) dx \\ &= \int e^{its} (\mathbf{R}_m f)(\hat{p}^0, s) ds. \end{aligned} \tag{5.3}$$

The measure on $\hat{p} \cdot p = s$ was chosen as the Fubini measure so as to make the integral over all x into an iterated integral in accordance with our general spread philosophy (see Section 1.1).

Equation (5.3) shows that the nonlinear Fourier transform on $\hat{x} = 0$ bears the same relation to $\mathbf{R}_m$ as the ordinary Fourier transform does to $\mathbf{R}_1$, namely that the linear Fourier transform of the nonlinear Radon transform on a nonlinear spread is the restriction of the total nonlinear Fourier transform to a line in $\hat{p}$ space. Thus (5.3) can be regarded as the projection–slice theorem for the nonlinear Radon transform.

One might wonder at the usage of both $\hat{p} \cdot p$ and $\hat{x} \cdot x$ which are homogeneous of degrees m and 1 respectively. What happens to the intermediate degrees? The answer is that they would occur in any complete theory but the algebra involved in dealing with them is very complicated so we shall restrict our attention to problems that can be handled exclusively using degrees m and 1, except for some remarks at the end of Section 5.1.

The relation between $\hat{p} \cdot p$ and $\hat{x} \cdot x$ is an example of *hierarchy*. In the hierarchy method we introduce functions of several variables. These functions satisfy differential equations or more general functional equations; an important aspect of these equations is that there is a Cauchy surface S on which we can interpret the significant properties of our functions in a useful way. In the present case $S = \{\hat{p} \text{ axis}\}$ since $\hat{p} \cdot p$ is naturally associated to the integral geometry as it defines the spreads.

We also need a second Cauchy surface U which in the present case is the $\hat{x}$ axis, from which we can return to f. The intertwining between S and U allows us to use the given information about f as translated to S to obtain more information on f.

In the next sections we shall deal with other properties of f which are translated into properties of F on some surface S lying between the $\hat{p}$ and $\hat{x}$ axes. Again $U = \{\hat{x} \text{ axis}\}$. These properties involve regularity; they are translated into growth conditions on S. Apparently growth conditions are more easily handled in certain ways than regularity. In addition to usual function theory the hierarchy is useful in other areas. In [62] it is used to give a new interpretation of the Weil representation of $SL(2, \mathbb{R})$ and its finite field analogs.

Naturally there are many variations of the hierarchy method. Instead of intertwining a whole space of functions we may intertwine a single function (or dual element). Such is the idea that underlies many identities such as the Rogers–Ramanujan identities (see [57]). It is the basis for one of the proofs of the index theorem (see [78]).

Equation (5.3) gives

Proposition 5.1 *For any $\hat{p}_0$ the information conveyed by the Radon transform $\mathbf{R}_m f(\hat{p}_0, s)$ for all s is equivalent to the knowledge of $F(a\hat{p}_0, 0)$ for all scalars a; that is, to the knowledge of F on the line through $\hat{p}_0$ in the $\hat{p}$ space.*

This proposition shows the relation between the natural geometric grouping $\{\hat{p}_0 \cdot p = s\}$ for varying s and the corresponding analytic grouping defined by lines in $\hat{p}$ space (i.e. $\hat{x} = 0$).

Another approach to the nonlinear Fourier transform occurs in relation to the Albanese map $A : \mathbf{R}^n \to \mathbf{R}^N$ given by

$$A : x \to p(x) = (p_1(x), \ldots, p_N(x)). \tag{5.4}$$

We form the algebraic variety

$$V_A = \{(A(x), x)\}. \tag{5.5}$$

Each function $f(x)$ is "lifted" to the measure $f(x)dx$ on V_A. We then take the ordinary Fourier transform in (p, x) space of this lift and obtain the nonlinear Fourier transform of f. The Albanese map adds an additional hierarchy to our structure since now U is embedded in $\mathbb{R}^{N+n}$. This second hierarchy plays an essential role in Section 10.3 but in the present chapter we shall deal with the nonlinear Fourier transform per se.

It is essential for us that the nonlinear Fourier transform $F(\hat{p}, \hat{x})$ of a nice function f satisfies the system of partial differential equations

$$\left[\frac{\partial}{\partial \hat{p}_j} - ip_j\left(-i\frac{\partial}{\partial \hat{x}_j}\right)\right] F = 0. \tag{5.6}$$

We refer to (5.6) as the "heat equation" although "Schroedinger equation" would probably be a better appellation. From the heat equations, or directly, one deduces equations in $\hat{p}$ only; we call these Plücker relations:

$$\left[\frac{\partial^{\ell_1+\cdots+\ell_N}}{\partial \hat{p}_1^{\ell_1} \ldots \partial \hat{p}_N^{\ell_N}} - \frac{\partial^{k_1+\cdots+k_N}}{\partial \hat{p}_1^{k_1} \ldots \partial \hat{p}_N^{k_N}}\right] F = 0 \tag{5.7}$$

whenever

$$p_1^{\ell_1} \ldots p_N^{\ell_N} = p_1^{k_1} \ldots p_N^{k_N} \tag{5.8}$$

as polynomials in x. (Actually we could reduce all relations to quadratic relations, i.e. $\sum \ell_j = \sum k_j = 2$.)

Moreover F is the solution of (5.6) with the initial data

$$F(0, \hat{x}) = \hat{f}(\hat{x}) \tag{5.9}$$

so F determines $\hat{f}$ and hence f.

It seems that the first usage of the nonlinear Fourier transform as a sharpening of the linear Fourier transform occurs in FA in Chapter V and some later places. The nonlinear Fourier transform is used to help characterize the Fourier transforms of some function spaces and also to prove uniqueness theorems.

Considerations somewhat in the same spirit as those of this chapter appear for the first time in the papers [66] and [59] of the author and P. Malliavin. In that work we were concerned with determining the actual support of a function f in terms of its (linear) Fourier transform $\hat{f}$. (By the Paley–Wiener theory, growth conditions determine the convex hull of the support.) The simplest example of that idea takes the following form: suppose $n = 1$ and support $f \subset [-1, 1]$. What conditions on $\hat{f}$ imply that $f = 0$ on $[-a, a]$?

To answer this question we consider the lift Af of $f(x)\,dx$ on the variety V_A for $m = 2$, i.e. the curve $p = x^2$. The image of the set $[-1, -a] \cup [a, 1]$ in V_A is contained in the *convex* set in R^2 bounded by $p = x^2$ and the lines $p = 1$ and $p = a^2$. Thus we can determine if support $f \subset [-1, -a] \cup [a, 1]$ using the Paley–Wiener theorem for the two-dimensional Fourier transform $\hat{Af}$ of Af; that is, the nonlinear Fourier transform of f.

To express this condition in terms of $\hat{f}$ directly we note that $\widehat{Af}$ satisfies the Schroedinger equation

$$\left[\frac{\partial}{\partial \hat{p}} + i\frac{\partial^2}{\partial \hat{x}^2}\right] \widehat{Af}(\hat{p}, \hat{x}) = 0 \tag{5.10}$$

with initial condition

$$\widehat{Af}(0, \hat{x}) = \hat{f}(\hat{x}). \tag{5.11}$$

The fundamental solution to the Schroedinger equation leads to a condition which involves $\hat{f}$ directly. (See Section 5.2 for details.)

Remark. *As we have noted the Paley–Wiener theory depends on convexity. The modification presented here depends on mappings. More precisely, we start with a domain $\Omega \subset R^n$. Suppose that there is a polynomial map (preferably by homogeneous polynomials)*

$$\alpha : \mathbb{R}^n \longrightarrow \mathbb{R}^n \times \mathbb{R}^N$$
$$\alpha(x) = (x, A(x))$$

so that $\alpha(\Omega)$ is convex, meaning that it is the intersection of a convex set $\tilde{\Omega}$ in $\mathbb{R}^n \times R^N$ with $\alpha(R^n)$. Then we can deal with functions supported on Ω exactly as we did in the above simple case.

Instead of sets Ω we could apply the same idea to functions. Thus we start with a function φ on R^n and we want $\alpha(\varphi)$ to be convex, meaning the restriction to $\alpha(\mathbb{R}^n)$ of a convex function on $\mathbb{R}^n \times R^N$. (Dealing with functions is more general than dealing with sets Ω as the latter can be subsumed by a modified characteristic function

$$\tilde{\chi}(x) = \begin{cases} 1 & x \in \Omega \\ \infty & x \notin \Omega \end{cases}$$

whose convexity is equivalent to that of Ω.) In this way we can define a new type of Young conjugate of nonconvex functions φ and hence we can characterize, by growth conditions and differential inequalities, the space of functions which are $\mathcal{O}(\exp(-\varphi))$. We can regard this as the Paley–Wiener theorem for the growth $\exp(-\varphi)$.

Problem 5.2 *Characterize those functions φ for which there is an α such that $\alpha(\varphi)$ is convex.*

Much more can be done with the nonlinear Fourier transform from the point of view of classical Fourier analysis. Thus, for example, a slight modification of the above can be applied to detect whether f is analytic near zero or is C^∞ near zero. The case of analyticity was also developed (independently of the work of Malliavin and the author) by Bros and Iagolnitzer [25] (and is usually called the FBI transform). Roughly speaking, the FBI transform is the nonlinear Fourier transform for $m = 2$. The ideas were applied to partial differential equations by many authors.

These aspects of the nonlinear Fourier transform are discussed in Section 5.2.

Let us also point out that if $f = \sum \delta_j$ is the sum of the δ functions at the lattice points then, for $m = 2$, $F(\hat{p}, \hat{x})$ is the theta function. This theory is developed in [62] and in detail in Section 10.3.

In the present section we study the relation of nonlinear Fourier analysis to the Radon transform.

For the study of the injectivity of the nonlinear Radon transform on sets of spreads of the form $\{\hat{p}^0 \cdot p = s\}$ we are left with the task of finding sets in the $\hat{p}$ spaces which are sets of determination for solutions of the system of partial differential equations (5.6) or (5.7). For (5.9) shows that F determines $\hat{f}$ and hence determines f.

Let us begin with the simplest example: $m = 2, n = 1$. Note that $\mathbf{R}_2(f)$ does not determine f because if f is odd then $\mathbf{R}_2(f) \equiv 0$. (This is true for any $\mathbf{R}_m$ for m even and any n.)

$F(\hat{p}, \hat{x})$ is a function of the single variables $\hat{p}$ and $\hat{x}$ which satisfies the heat equation (5.10) (for F in place of $\widehat{Af}$). The heat equation is second order in $\hat{x}$, which means that the data $F(\hat{p}, 0)$ which, by Proposition 5.1, represents $\mathbf{R}_2(f)$, is not sufficient to determine F, in conformity with our above remark that $\mathbf{R}_2(f)$ does not determine f.

To determine F we must include the normal derivative $\partial F(\hat{p}, 0)/\partial \hat{x}$. How is this to be understood in terms of the Radon transform?

Using the expression (5.2) for F we have

$$\begin{aligned} \frac{\partial F}{\partial \hat{x}}(\hat{p}, 0) &= i \int x f(x) e^{i\hat{p}x^2}\, dx \\ &= i \int_0^\infty e^{i\hat{p}c} (\delta_{x^2=c} \cdot xf) \frac{dc}{2\sqrt{|c|}}. \end{aligned} \tag{5.12}$$

Thus we need to evaluate the integrals of f and xf on $x^2 = c$. (Of course, in the present case the set $x^2 = c$ consists of two points so the "integral over $x^2 = c$" means the sum of values of the respective functions f or xf on $x^2 = c$.) The measure $dc/2\sqrt{c}$ is the Fubini measure for this spread.

Let us pass to the next case: $m = n = 2$. Now $N = 3$ and $p = (x_1^2, x_1x_2, x_2^2)$. The heat equations become

$$\begin{aligned} i\frac{\partial F}{\partial \hat{p}_1} &= \frac{\partial^2 F}{\partial \hat{x}_1^2} \\ i\frac{\partial F}{\partial \hat{p}_2} &= \frac{\partial^2 F}{\partial \hat{x}_1 \partial \hat{x}_2} \\ i\frac{\partial F}{\partial \hat{p}_3} &= \frac{\partial^2 F}{\partial \hat{x}_2^2}. \end{aligned} \tag{5.13}$$

The Plücker equations are the compatibility conditions in $\hat{p}$ for this system. In this case the Plücker equations reduce to the wave equation

$$\frac{\partial^2 F}{\partial \hat{p}_1 \partial \hat{p}_3} = \frac{\partial^2 F}{\partial \hat{p}_2^2}. \tag{5.14}$$

Note again that any odd function $f(x) = -f(-x)$ has vanishing $\mathbf{R}_2(f)$. This is reflected in the fact that solutions of the system (5.13) are not uniquely determined by their values on the $\hat{p}$ axis.

To understand this more clearly we must use the ideas of Chapter IX of FA and Chapter 4 above. A good Cauchy problem for (5.14) can be posed on a space-like linear space for the wave equation. For example, call $\hat{q}_1 = (\hat{p}_1 + \hat{p}_3)$

and $\hat{q}_3 = (\hat{p}_1 - \hat{p}_3)$. Then (5.14) becomes

$$\frac{\partial^2 F}{\partial \hat{q}_1^2} = \frac{\partial^2 F}{\partial \hat{q}_3^2} + \frac{\partial^2 F}{\partial \hat{p}_2^2}. \tag{5.14*}$$

Thus for the heat equations (5.13) we can give the data

$$\begin{aligned}
&F|_{\hat{q}_1=\hat{x}_1=\hat{x}_2=0}\\
&\frac{\partial F}{\partial \hat{q}_1}\Big|_{\hat{q}_1=\hat{x}_1=\hat{x}_2=0}\\
&\frac{\partial F}{\partial \hat{x}_1}\Big|_{\hat{q}_1=\hat{x}_1=\hat{x}_2=0}\\
&\frac{\partial^2 F}{\partial \hat{q}_1 \partial \hat{x}_1}\Big|_{\hat{q}_1=\hat{x}_1=\hat{x}_2=0}\\
&\frac{\partial F}{\partial \hat{x}_2}\Big|_{\hat{q}_1=\hat{x}_1=\hat{x}_2=0}\\
&\frac{\partial^2 F}{\partial \hat{q}_1 \partial \hat{x}_2}\Big|_{\hat{q}_1=\hat{x}_1=\hat{x}_2=0}.
\end{aligned} \tag{5.15}$$

The first two Cauchy data in (5.15) determine $F(\hat{p}, 0)$ via the wave equation (5.14*). In a similar manner the third and fourth CD determine $\partial F(\hat{p}, 0)/\partial \hat{x}_1$ and the last two CD determine $\partial F(\hat{p}, 0)/\partial \hat{x}_2$ ($\partial F(\hat{p}, 0)/\partial \hat{x}_1$ and $\partial F(\hat{p}, 0)/\partial \hat{x}_2$ satisfy the wave equation). Once we know $F(\hat{p}, 0), \partial F(\hat{p}, 0)/\partial \hat{x}_1, \partial F(\hat{p}, 0)/\partial \hat{x}_2$ then we can use the CP for (5.13) with CD on the $\hat{p}$ axis to solve for $F(\hat{p}, \hat{x})$. This CP for (5.13) has been discussed at the end of Section 4.1 (see Theorem 4.8). We recall that the CD are not independent because the corresponding harmonic expansion is related to the nonreflection group $x \to -x$.

We can pass from the CD in (5.15) to the Radon transform $\mathbf{R}_2$ as in (5.12)f. From (5.2) we see that the nonlinear Fourier transform of $(x_1^2 + x_2^2)f$ is

$$\begin{aligned}
\int e^{i\hat{p}\cdot p + i\hat{x}\cdot x}(x_1^2 + x_2^2) f(x)\, dx &= -\left(\frac{\partial^2}{\partial \hat{x}_1^2} + \frac{\partial^2}{\partial \hat{x}_2^2}\right) F(\hat{p}, \hat{x})\\
&= -2i \frac{\partial F(\hat{p}, \hat{x})}{\partial \hat{q}_1}.
\end{aligned} \tag{5.16}$$

Moreover the nonlinear Fourier transform of $x_j f$ can be represented as $-i\partial F/\partial \hat{x}_j$.

Taking into account the fact that (5.15) are CD for solutions of the heat equations we have

Proposition 5.2 *A smooth function f of compact support in the plane is uniquely determined by the Radon transforms*

$$\begin{gathered}\mathbf{R}_2 f|_{\hat{q}_1=0}\\ \mathbf{R}_2[(x_1^2+x_2^2)f]|_{\hat{q}_1=0}\\ \mathbf{R}_2(x_1 f)|_{\hat{q}_1=0}\\ \mathbf{R}_2[x_1(x_1^2+x_2^2)f]|_{\hat{q}_1=0}\\ \mathbf{R}_2(x_2 f)|_{\hat{q}_1=0}\\ \mathbf{R}_2[x_2(x_1^2+x_2^2)f]|_{\hat{q}_1=0}.\end{gathered} \tag{5.17}$$

(5.16) and (5.17) follow the general ideas of Chapter 4. We use the $\hat{p}$ plane as an exotic Cauchy surface. The exotic CD itself can be determined by the nonlinear Radon transform.

Let us pass to a direct method of finding a CP for the heat equations which is, in fact, more efficient. According to Chapter IX of FA or Chapter 4 above we can give a CP for (5.13) by prescribing a linear space $\hat{Y}$ in the (linear) Fourier transform variables $\hat{\hat{p}}_j, \hat{\hat{x}}_j$ dual to $\hat{p}_j, \hat{x}_j$ in the usual inner product, and a finite set of polynomials $(s_1, \dots, s_k)$ in $(\hat{\hat{p}}, \hat{\hat{x}})$ not depending on the variables of $\hat{\hat{Y}}$, so that the variety $\hat{\hat{V}}$ associated to the system of equations (5.13) is parametrized by the values of the $\hat{\hat{Y}}$ coordinates and corresponding values of the polynomials s_j on the points $(\hat{\hat{p}}, \hat{\hat{x}}) \in \hat{\hat{V}}$ above each $\hat{\hat{t}} \in \hat{\hat{Y}}$.

In our case the variety $\hat{\hat{V}}$ is defined by $\hat{\hat{p}}_1 = \hat{\hat{x}}_1^2, \hat{\hat{p}}_2 = \hat{\hat{x}}_1\hat{\hat{x}}_2, \hat{\hat{p}}_3 = \hat{\hat{x}}_2^2$. $\hat{\hat{Y}}$ is defined by $\hat{\hat{q}}_1 = \hat{\hat{x}}_1 = \hat{\hat{x}}_2 = 0$. We can use coordinates $\hat{\hat{p}}_2$ and $\hat{\hat{q}}_3$ on $\hat{\hat{Y}}$ and $\hat{\hat{q}}_1, \hat{\hat{x}}_1$, and $\hat{\hat{x}}_2$ on the orthogonal space $\hat{\hat{T}}$. Thus, again, our Cauchy surface $\hat{Y}$ is the subset of the $\hat{p}$ axis (i.e. $\hat{x} = 0$) defined by $\hat{q}_1 = 0$. We now use our ideas on harmonicity to give a CP for this Cauchy surface.

When we fix $\hat{\hat{p}}_2$ and $\hat{\hat{q}}_3$ then the $\hat{\hat{T}}$ coordinates of the points in $\hat{\hat{V}}$ are defined by

$$\begin{aligned}\hat{\hat{p}}_2 &= \hat{\hat{x}}_1\hat{\hat{x}}_2\\ \hat{\hat{q}}_3 &= \hat{\hat{x}}_1^2 - \hat{\hat{x}}_2^2\\ 0 &= \hat{\hat{x}}_1^2 + \hat{\hat{x}}_2^2 - \hat{\hat{q}}_1.\end{aligned} \tag{5.18}$$

We are now in the situation described at the beginning of Chapter 4. The notation $\hat{y}, \hat{t}$ there is now: $\hat{y}$ corresponds to $(\hat{\hat{p}}_2, \hat{\hat{q}}_3)$ and $\hat{t}$ to $(\hat{\hat{q}}_1, \hat{\hat{x}}_1, \hat{\hat{x}}_2)$. Equations (5.18) are the analog of (4.13) with

$$\vec{\imath} = (\hat{\hat{x}}_1\hat{\hat{x}}_2, \hat{\hat{x}}_1^2 - \hat{\hat{x}}_2^2, \hat{\hat{x}}_1^2 + \hat{\hat{x}}_2^2 - \hat{\hat{q}}_1).$$

There is a slight difficulty in applying the theory of Chapter 3 because i_3 is not homogeneous. But it is homogeneous of degree 2 in $\hat{x}$ and of degree 1 in q and that is sufficient to apply the theory of harmonicity of Chapter 3. A simple computation shows that the harmonics are spanned by

$$1, \hat{\hat{x}}_1, \hat{\hat{x}}_2, \hat{\hat{x}}_1^2 + \hat{\hat{x}}_2^2 + 4\hat{\hat{q}}_1. \tag{5.19}$$

The fact that there are four linearly independent harmonics corresponds to the fact that when $\hat{p}_2, \hat{q}_3$ are fixed then there are, generically, four solutions of (5.18). In our discussion following (5.16) we showed that the data (5.17) corresponds to the operator's identity, $\partial/\partial\hat{x}_1$, $\partial/\partial\hat{x}_2$, $\partial/\partial\hat{q}_1$, $\partial^2/\partial\hat{x}_1\partial\hat{q}_2$, $\partial^2/\partial\hat{x}_2\partial\hat{q}_1$, applied to F and evaluated on $\hat{T}$. It is a simple matter to express, on $\hat{\hat{V}}$, the harmonic decomposition of the corresponding polynomials $\hat{\hat{q}}_1, \hat{\hat{x}}_1\hat{\hat{q}}_1$, and $\hat{\hat{x}}_2\hat{\hat{q}}_1$ in terms of the harmonics (5.19). We write, on $\hat{\hat{V}}$,

$$\begin{aligned}\hat{\hat{x}}_1\hat{\hat{q}}_1 &= \hat{\hat{x}}_1(\hat{\hat{x}}_1^2 + \hat{\hat{x}}_2^2)\\ &= \hat{\hat{x}}_1(\hat{\hat{q}}_3 + 2\hat{\hat{x}}_2^2)\\ &= \hat{\hat{q}}_3\hat{\hat{x}}_1 + 2\hat{\hat{p}}_2\hat{\hat{x}}_2\end{aligned}$$

which is the harmonic decomposition of $\hat{\hat{x}}_1\hat{\hat{q}}_1$. The expression for $\hat{\hat{x}}_2\hat{\hat{q}}_1$ is similar. Finally

$$5\hat{\hat{q}}_1 = (\hat{\hat{x}}_1^2 + \hat{\hat{x}}_2^2 + 4\hat{\hat{q}}_1) - (\hat{\hat{x}}_1^2 + \hat{\hat{x}}_2^2 - \hat{\hat{q}}_1).$$

The harmonically defined CP with data on $\hat{Y}$ can be translated, via Proposition 5.1, into the sharpened form of Proposition 5.2:

Proposition 5.2* *A smooth function f of compact support in the plane is uniquely determined by the Radon transform*

$$\begin{aligned}&\mathbf{R}_2 f|_{\hat{q}_1=0}\\ &\mathbf{R}_2(x_1 f)|_{\hat{q}_1=0}\\ &\mathbf{R}_2(x_2 f)|_{\hat{q}_1=0}\\ &\mathbf{R}_2[(x_1^2 + x_2^2)\, f]|_{\hat{q}_1=0}.\end{aligned} \tag{5.17*}$$

Expressions (5.17) and (5.17*) are capable of an interesting geometric interpretation. The hyperplane $\hat{q}_1 = 0$ is the $\hat{p}_2, \hat{q}_3$ plane. A point in this plane corresponds (since $\hat{q}_1 = \hat{p}_1 + \hat{p}_3, \hat{q}_3 = \hat{p}_1 - \hat{p}_3$) to the conic

$$\hat{p}_2 x_1 x_2 + \hat{p}_1(x_1^2 - x_2^2) = c.$$

This is a two-parameter family of hyperbolas (since we can normalize $c = 1$ or $c = 0$). Proposition 5.2* can be translated to

Proposition 5.2** *$f \in \mathcal{D}(\mathbb{R}^2)$ is determined by the integrals of $f, x_1 f, x_2 f, (x_1^2 + x_2^2) f$ over the hyperbolas*

$$\hat{p}_2 x_1 x_2 + \hat{p}_1 (x_1^2 - x_2^2) = 1, 0$$

for varying $\hat{p}_2, \hat{q}_3$.

Let us now pass to the case of general m, n. We can proceed in two ways as in the case $m = n = 2$. We can first find a good CP for the Plücker equations and then use the exotic CP of Chapter 4 to pass from the $\hat{p}$ space to the whole $\hat{p}, \hat{x}$ space. Finding a "good" CP for the Plücker equations is complicated. It has been carried out in some cases by J. Wang in his Temple University Ph.D. thesis. We shall, therefore, deal directly with a CP for the heat equations with data on a subspace of $\hat{p}$ space. Let $\hat{\hat{V}}$ be the algebraic variety associated to the system of heat equations, so $\hat{\hat{V}}$ is defined by $\hat{\hat{p}}_j = p_j(\hat{\hat{x}})$. We claim that a Cauchy–Kowalewski (CK) problem (see Chapter IX of FA) for $\hat{\hat{V}}$ can be defined by the linear space $\hat{\hat{Y}}$:

$$\hat{\hat{Y}} = \text{ subspace of } \hat{\hat{p}} \text{ axis spanned by } \hat{\hat{p}}_1, \ldots, \hat{\hat{p}}_n \tag{5.20}$$

where our notation is such that $p_1 = x_1^m, \ldots, p_n = x_n^m$. (It should be noted that this $\hat{\hat{Y}}$ differs from the Cauchy surface used to illustrate $m = 2 = n$; of course there are many possible choices for Cauchy surfaces.)

For $j = 1, \ldots, n$ we have $\hat{\hat{p}}_j = \hat{\hat{x}}_j^m$ on $\hat{\hat{V}}$ from which it follows that

$$|(\hat{\hat{p}}, \hat{\hat{x}})| \leq c(1 + |\hat{\hat{p}}_1| + \cdots + |\hat{\hat{p}}_n|) \quad \text{on } \hat{\hat{V}}. \tag{5.21}$$

In the language of FA this means that $\hat{\hat{Y}}$ is a principal noncharacteristic for the heat equation, so we can set up a CP with data on $\hat{\hat{Y}}$.

The same $\hat{\hat{Y}}$ can also be used as a Cauchy surface for the Plücker equations. As mentioned above this CP together with the exotic CP discussed in Chapter 4 leads to an analog of Proposition 5.2.

For the present we want to set up a CP directly for the heat equation (without using the Plücker equations) with data on $\hat{\hat{Y}}$. Denote by $\hat{\hat{T}}$ the orthogonal to $\hat{\hat{Y}}$ in the usual inner product. Thus a point $\hat{\hat{t}} \in \hat{\hat{T}}$ has coordinates $\hat{\hat{p}}_\ell, \hat{\hat{x}}_k$ where p_ℓ is not of the form x_j^m and $k = 1, \ldots, n$.

Call $\hat{\hat{y}}_j = \hat{\hat{p}}_j$ for $j = 1, 2, \ldots, n$. For each $\hat{\hat{y}} \in \hat{\hat{Y}}$ the set $\hat{\hat{T}}(\hat{\hat{y}})$ of $\hat{\hat{t}}$ coordinates of points $(\hat{\hat{y}}, \hat{\hat{t}}) \in \hat{\hat{V}}$ is defined by the equations

$$\begin{aligned} \hat{\hat{y}}_j &= \hat{\hat{x}}_j^m \\ 0 &= \hat{\hat{p}}_l - p_l(\hat{\hat{x}}) \quad \text{for all } l > n. \end{aligned} \tag{5.22}$$

(This is the generalization of (5.18).)

Note that for fixed generic $\hat{y}$ the set $\hat{\hat{T}}(\hat{y})$ consists of m^n points. Thus the polynomials on the right side of (5.22) form a strongly free system. (As in (5.18) there is different homogeneity in $\hat{x}$ and in $\hat{p}$.) A computation which we leave to the reader can be used to show that the harmonics are spanned by

$$h_k(\hat{\hat{x}}) = \hat{\hat{x}}_1^{a_1} \ldots \hat{\hat{x}}_n^{a_n} + \chi_k(\hat{\hat{x}}, \hat{\hat{p}}) \tag{5.23}$$

where each $a_j < m$. Here χ_k is a polynomial in $\hat{\hat{x}}$ and $\hat{\hat{p}}$ depending on $a_1, \ldots, a_n$ satisfying

$$(\text{degree } \chi_k \text{ in } \hat{\hat{x}}) + m(\text{degree } \chi_k \text{ in } \hat{\hat{p}}) = \sum a_j.$$

Also, clearly, $\chi_k = 0$ when $\sum a_j < m$.

We illustrate the case $m = 4, n = 2$. The 16 harmonics are

$$\begin{gathered} 1, \hat{\hat{x}}_1, \hat{\hat{x}}_2, \hat{\hat{x}}_1^2, \hat{\hat{x}}_1\hat{\hat{x}}_2, \hat{\hat{x}}_2^2, \hat{\hat{x}}_1^3, \hat{\hat{x}}_1^2\hat{\hat{x}}_2, \hat{\hat{x}}_1\hat{\hat{x}}_2^2, \hat{\hat{x}}_2^3, \\ \hat{\hat{x}}_1^3\hat{\hat{x}}_2 + 6\hat{\hat{p}}_{31}, \hat{\hat{x}}_1^2\hat{\hat{x}}_2^2 + 4\hat{\hat{p}}_{22}, \hat{\hat{x}}_1\hat{\hat{x}}_2^3 + 6\hat{\hat{p}}_{13}, \\ \hat{\hat{x}}_1^3\hat{\hat{x}}_2^2 + 12\hat{\hat{p}}_{31}\hat{\hat{x}}_2 + 12\hat{\hat{p}}_{22}\hat{\hat{x}}_1, \hat{\hat{x}}_1^2\hat{\hat{x}}_2^3 + 12\hat{\hat{p}}_{22}\hat{\hat{x}}_2 + 12\hat{\hat{p}}_{13}\hat{\hat{x}}_1, \\ \hat{\hat{x}}_1^3\hat{\hat{x}}_2^3 + 16\hat{\hat{p}}_{31}\hat{\hat{x}}_2^2 + 36\hat{\hat{p}}_{22}\hat{\hat{x}}_1\hat{\hat{x}}_2 + 16\hat{\hat{p}}_{13}\hat{\hat{x}}_1^2. \end{gathered}$$

Our notation is

$$p_{b_1 b_2 \ldots b_n} = x_1^{b_1} x_2^{b_2} \ldots x_n^{b_n}.$$

Since, by (5.21), $\hat{\hat{Y}}$ is a principal noncharacteristic for the heat equation, we have

Theorem 5.3 *The CP for the heat equation with data*

$$\{\partial(h_k)F|_{\hat{\hat{Y}}}\} \tag{5.24}$$

is well posed in the CK sense.

In particular, Theorem 5.3 implies the uniqueness of the CP. Using the ideas of the proof of Proposition 5.2* allows us to derive from Theorem 5.3

Theorem 5.4 *A smooth function f of compact support is uniquely determined by*

$$\{\mathbf{R}_m(h_k f)|_{\hat{\hat{Y}}}\}. \tag{5.25}$$

We must clarify the meaning of $h_k f$ since h_k depends on $\hat{\hat{x}}$ and $\hat{\hat{p}}$. But, since F satisfies the heat equations we can replace all $\hat{p}$ derivatives in (5.24) by $\hat{x}$ derivatives. This is the meaning of $h_k f$ in (5.25).

We can translate Theorem 5.4 geometrically into

Theorem 5.4* *A function $f \in \mathcal{D}$ is determined by the integrals of $\{h_j f\}$ over the hypersurfaces*

$$\sum \hat{p}_j \hat{x}_j^m = 1, 0.$$

In particular, for $m = 2$, f is determined by the integrals of $\{h_j f\}$ over hyperboloids and ellipsoids whose principal axes are the coordinate axes.

Let us go further and compute the explicit inversion formula for $\mathbf{R}_m$. We may express the value of the restriction $F(\hat{y}, 0)$ of F to the $\hat{Y}$ axis as

$$\begin{aligned} F(\hat{y}, 0) &= \int e^{i\hat{y}\cdot y(x)} f(x)\, dx \\ &= \int e^{i\hat{y}\cdot y} \sum f(x_j(y)) D(x_j(y))\, dy. \end{aligned} \tag{5.26}$$

$x_j(y)$ are the (generically 2^n for even n and 1 for odd n) points x corresponding to a fixed y under the Albanese map and D is the Jacobian $\partial(x)/\partial(y)$. (It can be shown that D is locally integrable.)

The second equality in (5.26) expresses $F(\hat{y}, 0)$ as a linear Fourier transform; the first equality via (5.3) expresses $F(\hat{y}, 0)$ as a linear Fourier transform of $\mathbf{R}_m f$. Setting $\hat{y} = t\hat{y}^0$ for $|\hat{y}^0| = 1$ leads to

$$\begin{aligned} \int e^{its} (\mathbf{R}_m f)(\hat{y}^0, s)\, ds &= F(t\hat{y}^0, 0) \\ &= \int e^{its} \mathbf{R}_1 \left[\sum f(x_j(y)) D(x_j(y)) \right] (\hat{y}^{0\perp}, s)\, ds. \end{aligned} \tag{5.27}$$

In the second equation we consider the sum as a function of y; $\mathbf{R}_1$ is the linear hyperplane Radon transform in y.

We can apply the same formula to $h_k f$ in place of f.

Now, by the formulas for the inversion of the Radon transform discussed in Section 2.1 we can express $\sum f(x_j(y)) D(x_j(y))$ in terms of its $\mathbf{R}_1$, hence by (5.27) in terms of $\mathbf{R}_m f$. To obtain the individual summands $f(x_j(y))$ we use the same formula for $\sum h_k(x_j(y)) f(x_j(y)) D(x_j(y))$; this allows us to express these in terms of $\mathbf{R}_m h_k f$. If we think of $f(x_j(y)) D(x_j(y))$ as the components of a vector $(f\vec{\partial})(y)$ then we can write the sum $\sum h_k(x_j(y)) f(x_j(y)) D(x_j(y))$ as $\boxed{h}(f\vec{\partial})(y)$ where the matrix $\boxed{h} = \{h_k(x_j(y))\}$ is a general Vandermonde matrix as in Chapter 4.

We now apply (2.6*) to obtain the inversion formula for $\mathbf{R}_m$:

Theorem 5.5 *The inversion of* $\mathbf{R}_m$ *is given by*

$$(f\vec{\partial})(y) = c_0 \boxed{h}^{-1} \Delta_y^{\frac{n-1}{2}} \int_{|\hat{y}|=1} \mathbf{R}_m(h_k f)(\hat{y}, \hat{y}\cdot y)\, d\hat{y}. \tag{5.28}$$

Remark 1 *The integrand in (5.28) for* $y = y(x)$ *can be interpreted as* $\mathbf{R}_m^* \mathbf{R}_m(h_k f)$. *For when* $\hat{y}$ *is fixed the set of* x *with* $\hat{y}\cdot y(x) = s$ *(fixed) is exactly the leaf for the spread defined by* $\hat{y}$ *whose spread parameter value is* s. *Any function* $u(\hat{y}\cdot y) = u(\hat{y}\cdot y(x))$ *can be interpreted as a distribution*

$$u(\hat{y}\cdot y) = \int u(s)\delta_{\hat{y}\cdot y = s}\, ds$$

so long as the measures on the sets $\{\hat{y}\cdot y\}$ *and the measure* ds *are Fubini for* dx.

We have denoted the distribution

$$\int \mathbf{R}_m(h_k f)(\hat{y}, s)\delta_{\hat{y}\cdot y = s}\, ds = \mathbf{R}_m^* \mathbf{R}_m(h_k f)(\hat{y}).$$

Thus (5.28) is an integral over $\hat{y}$ of $\mathbf{R}_m^* \mathbf{R}_m(h_k f)(\hat{y})$.

Remark 2 *There are natural moment conditions related to* $\mathbf{R}_m$. *The simplest comes from the fact that*

$$\int \mathbf{R}_m f(\hat{p}^0, s)\, ds = \int f(x)\, dx$$

is independent of $\hat{p}^0$ *for* $|\hat{p}^0| = 1$. *Other moment conditions are related, as usual, to the smoothness of* $F(\hat{p}, 0)$ *at the origin. (We have written* $\hat{p} = (\hat{p}^0, s)$ *with* $|p^0| = 1$. *Of course* $\mathbf{R}_m f(p^0, s) = \mathbf{R}_m f(-p^0, -s)$.)

Suppose we are given a function $H(\hat{p}, 0)$ which satisfies the moment conditions; we want to know if H is of the form $\mathbf{R}_m f(\hat{p}^0, s)$.

As in Chapter 2, the moment conditions imply that the (linear) Fourier transform $\hat{H}(\hat{p}^0, \hat{s})$ in s defines a smooth function of $\hat{p}$ which we denote by $\hat{H}(\hat{p})$.

If we assume that H also satisfies the compatibility conditions which correspond to the Plücker conditions on $\hat{H}$ then we should expect H to be of the form $\mathbf{R}_m f$. But H alone is not enough to determine f; we need other functions corresponding to $\mathbf{R}_m(h_k f)$. Moreover we need some regularity to be able to solve the heat equations with data on the $\hat{p}$ axis.

Problem 5.3 *Characterize* $\mathbf{R}_m(\mathcal{S})$.

Remark 3 *The use of* Δ *and the inversion of* $\boxed{h}$ *could be avoided if we knew how to solve the CP for the heat equation. For that would lead to a direct determination of* $F(0, \hat{x}) = \hat{f}(\hat{x})$ *and hence of* f. *For* $m = 2, n = 1$ *such a solution is*

given in FA, p. 168:

$$F(\hat{p},\hat{x}) = \sum \frac{\partial^j}{\partial \hat{p}^j} F(\hat{p},0)(-i)^j \frac{\hat{x}^{2j}}{(2j)!} + \sum \frac{\partial^{j+1}}{\partial \hat{p}^j \partial \hat{x}} F(\hat{p},0)(-i)^j \frac{\hat{x}^{2j+1}}{(2j+1)!}.$$

This example shows that we need regularity on the $\hat{p}$ axis to solve the CP.

Remark 4 *It seems that our success in this Radon transform is due to the fact that we can form suitable spreads such as $\{\hat{p}\cdot p = s\}$ which fill R^n without overlap. Such is not the case if we use, for example, sets like $\{(\hat{p}\cdot p)(x-a) = s_0\}$ for varying a. The simplest form of this is the set of spheres with varying center and fixed radius. This problem was studied for the first time by Delsarte [30]. He proved that one needs two different radii s_0 whose ratio is not a zero of a suitable Bessel function for injectivity of the Radon transform. Subsequently much work has been done on this and related questions (see e.g. [12], [13]).*

Remark 5 *By using nonlinear varieties of dimension $n-\ell$ in $\hat{p}$ space as in the case $m=1$, we could deal with $\mathbf{R}_m^\ell$ which is the Radon transform on intersections of ℓ homogeneous varieties $\hat{p}^1\cdot p = s_1, \ldots, \hat{p}^\ell \cdot p = s_\ell$.*

5.2 Nonconvex support and regularity

If f is a function on $\mathbb{R}^n$ then classical Fourier analysis cannot be applied to study local regularity of f unless f is small at infinity and the regularity is uniform over all of $\mathbb{R}^n$. For functions which are large at infinity or only locally defined there are two procedures that can be applied:

(1) *AU spaces and duality* (see Section 1.4). Regularity can often be described by duality. For example, if f is C^∞ on $(-1,1)$ then $f \in \mathcal{W}'$ where $\mathcal{W}$ is the space of distributions of compact support on $(-1,1)$. Since $\mathcal{W}$ is AU we can obtain a Fourier representation for f in spite of the fact that f may be defined only on $(-1,1)$.

A similar procedure can be applied if f is holomorphic, e.g. in $|\Im x| < \epsilon$ near the origin.

(2) *Cut-offs and nonlinear Fourier transform.* If f is defined on $(-1,1)$ and we want to study its local regularity then we can "cut off" f by multiplying by a smooth function φ of compact support. If "regularity" means real analytic then this requires certain delicate constructions which can be accomplished in the framework of nonlinear Fourier analysis.

In certain situations (2) has an advantage over (1) because the "Fourier transform" can be written directly in terms of f.

Remark. *The nonlinear Fourier transform depends on the cut-off function ϕ. In general we shall be interested in local properties of f so we require that $\phi \equiv 1$ on some neighborhood of 0 (or the point being studied). For most of our work the explicit ϕ is of no importance and will be ignored. But for some problems the*

cut-offs depend on various spaces so we shall be more precise in our usage of the nonlinear Fourier transform.

The strongest regularity is vanishing. We mentioned in the above section how to use the nonlinear Fourier transform F of f to determine whether f vanishes near the origin. One application of this method is to the problem of lacunas, which usually means: when does there exist a hole in the support of the fundamental solution of a partial differential equation?

Let us illustrate how things work for the classical wave equation (note change in notation)

$$\Box f \equiv \frac{\partial^2 f}{\partial t^2} - \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j^2} = 0.$$

According to the ideas of Chapter 4 above or of Chapter IX of FA (Watergate method), we can evaluate $f(1;0,\dots,0)$ in terms of $f(0;x_1,\dots,x_n)$ and $f_t(0;x_1,\dots,x_n)$ by means of the equation

$$\delta_{1;0,\dots,0} = S_0(x) + \frac{\partial}{\partial t} S_1(x) + \Box T. \tag{5.29}$$

Here S_0 and S_1 are distributions on X space, i.e. $t = 0$, and T is a distribution in (t,x). Precisely

$$f(1;0,\dots,0) = S_0 \cdot f(0;x_1,\dots,x_n) + S_1 \cdot f_t(0;x_1,\dots,x_n) + T \cdot \Box f.$$

The Fourier transform of (5.29) is

$$e^{i\hat{t}} = \hat{S}_0(\hat{x}) + i\hat{t}\hat{S}_1(\hat{x}) \qquad \text{on } V \tag{5.30}$$

where V is the algebraic variety $\hat{t}^2 = \hat{x}_1^2 + \cdots + \hat{x}_n^2$. We can solve (5.30) explicitly for $\hat{S}_0$ and $\hat{S}_1$. For any $\hat{x}$ there are two values (counting multiplicity) of $\hat{t}$ for which $(\hat{t},\hat{x}) \in V$. This leads to the solution

$$\begin{aligned} \hat{S}_0(\hat{x}) &= \cos\sqrt{\hat{x}^2} \\ \hat{S}_1(\hat{x}) &= \frac{\sin\sqrt{\hat{x}^2}}{\sqrt{\hat{x}^2}}. \end{aligned} \tag{5.31}$$

We have written $\hat{x}^2 = \sum \hat{x}_j^2$.

The problem of lacunas is the determination as to whether or not $S_j(x)$ vanishes near the origin. (The problem is usually stated in terms of the support of the fundamental solution; the two approaches are easily seen to be equivalent.) We apply the method sketched at the beginning of Section 5.1. This means that we have to solve the Schroedinger equation (5.10) (with $\frac{\partial^2}{\partial x^2}$ interpreted as $\sum \frac{\partial^2}{\partial \hat{x}_j^2}$) with initial value $\hat{S}_j(\hat{x})$. The solution is, up to a constant,

$$U_j(\hat{p},\hat{x}) = \hat{p}^{-\frac{n}{2}} \int e^{\frac{i(\hat{x}-\hat{y})^2}{\hat{p}}} \hat{S}_j(\hat{y}) d\hat{y}.$$

We have to study $U_j(\hat{p}, \hat{x})$ as $\Im\, \hat{p} \to +\infty$. To simplify the notation, we restrict our considerations to $\hat{x} = 0$ and to $\hat{p}$ pure imaginary; the case of general $\hat{x}, \hat{p}$ presents no essential difficulty. Since we are interested in exponential growth, polynomials in $\hat{p}$ are unimportant so we shall make our computation ignoring powers of $\hat{p}$.

We start with $\hat{S}_0$. Using polar coordinates in $\hat{y}$

$$U_0(\hat{p}, 0) \cong \int_0^\infty e^{-\frac{ir^2}{\hat{p}}} (\cos r) r^{n-1}\, dr.$$

Since $\hat{p} \to +i\infty$ the last integral is, after a simple change of variables, like the Fourier transform of a polynomial times $\exp(-r^2)$ except that we integrate from 0 to ∞ instead of from $-\infty$ to $+\infty$. Thus (writing $\hat{p} = is$)

$$\begin{aligned} U_0(\hat{p}, 0) &\cong \int_0^\infty e^{-r^2} (e^{ir\sqrt{s}} + e^{-ir\sqrt{s}}) r^{n-1}\, dr \\ &\cong \int_{-\infty}^\infty e^{-r^2} |r|^{n-1} e^{ir\sqrt{s}}\, dr. \end{aligned} \tag{5.32}$$

For this integral to be exponentially decreasing for $s \to +\infty$ (which is the condition for lacunas as explained in Section 5.1) we claim that it must be the case that $|r|^{n-1}$ is a polynomial in r, i.e. *n is odd*. For in this case we obtain the actual Fourier transform of a polynomial times $\exp(-r^2)$. On the other hand, if $U_0(is, 0)$ is exponentially decreasing for $s \to +\infty$ then, by taking the inverse Fourier transform, we find that $\exp(-r^2)|r|^{n-1}$ has an analytic extension to the upper half-plane (in r) which means that n is odd.

The same argument works for U_1 except that $\hat{S}_1$ produces a denominator of r as is seen from (5.31). Instead of (5.32) we find

$$U_1(\hat{p}, 0) \cong \int_0^\infty e^{-r^2} (e^{ir\sqrt{s}} - e^{-ir\sqrt{s}}) r^{n-2}\, dr. \tag{5.33}$$

We make this an integral from $-\infty$ to ∞; this replaces r^{n-2} by $(\operatorname{sgn} r)\, |r|^{n-2}$. The exponential decrease in s requires $n - 2$ to be odd and ≥ 0.

We have arrived at the classical criterion: *lacunas for the wave equation exist only in an odd number (≥ 3) of space dimensions.*

This argument is due to Ehrenpreis and Malliavin [66] (see also [63]). Some profound applications of the ideas were developed by Shapiro [140].

So much for actual support. What about some finer structure of functions such as regularity?

The classical (linear) Fourier transform is usually an excellent tool for determining such structure for functions supported on *convex sets* $\Omega \subset \mathbb{C}^n$ in terms of growth conditions. The reason for the restriction to convex sets comes from the fact that for fixed $\hat{x}$ the inequality $|\exp(ix \cdot \hat{x})| \leq 1$ (for x complex) holds in a half-plane.

Of course, the half-plane condition comes from the linear nature of $x \cdot \hat{x}$. If we use the general nonlinear Fourier transform then "half-plane" should be generalized to "Runge domain" (or "polynomial convex" domain), meaning a domain defined by a limit of polynomial inequalities, i.e. $|P_\alpha(x)| \leq 1$ where P_α are polynomials.

It should be pointed out that polynomial convexity of a domain is a serious condition only for complex domains; any reasonable real domain is polynomially convex.

The fundamental question in this direction is

Problem 5.4 *Let Ω be a connected bounded Runge domain in the complexification of $\mathbf{R}^n$ such that $\Omega \cap \mathbf{R}^n$ is open in $\mathbf{R}^n$. Characterize by means of the nonlinear Fourier transform (hopefully by growth conditions) those functions on $\mathbf{R}^n$ which have holomorphic extensions from $\Omega \cap \mathbf{R}^n$ to Ω.*

Remark. *Problem 5.4 propounds the question as to what extent we can use the nonlinear Fourier transform to intertwine regularity, meaning holomorphicity on Ω, with suitable growth conditions in $(\hat{p}, \hat{x})$, in accordance with the hierarchy method.*

We begin our "assault" on this problem with the simplest case of $n = 1$ and Ω a small "half"-ball around the origin, meaning

$$\Omega = \{|x| < \tfrac{1}{2} + \varepsilon, \Im x > 0\}.$$

We consider functions f whose support is in $[-1, 1]$ and which have holomorphic extensions from $(-\frac{1}{2}, \frac{1}{2})$ to Ω.

We search for polynomial $P_0(x)$ (not homogeneous) and contours γ lying in $[-1, 1] \cup \Omega$ (see Figure 5.1) such that

$$\Im P_0(x) > 0 \quad \text{on } \gamma. \tag{5.34}$$

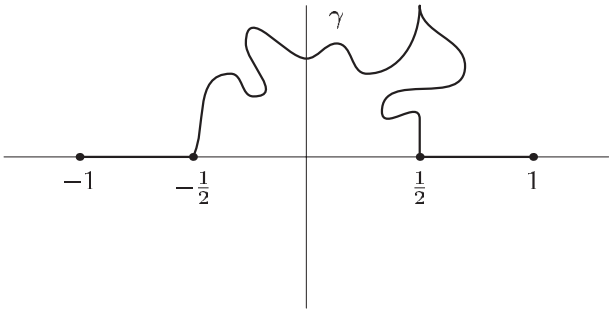

Figure 5.1

If this is the case then

$$F(tP_0) = \int e^{itP_0(x)} f(x)dx \tag{5.35}$$

is exponentially decreasing in t.

For $n > 1$ and for general Ω the idea is similar. We want to deform the contour from $\mathbf{R}^n$ to an n chain in $\mathbb{R}^n \cup \Omega$ on which a polynomial $P_0(x)$ has positive imaginary part.

Conjecture *The exponential inequalities thus obtained are equivalent to f having a holomorphic extension to Ω.*

If the conjecture is true then, to a certain extent, the relation of nonlinear Fourier analysis to arbitrary polynomially convex Ω would parallel the relation of linear Fourier analysis to those Ω which are tubes with convex bases.

My student T. Banh [8] has solved this conjecture for the case $n = 1$ for general classes of domains Ω (see Theorems 2.7 and 3.2 of [8]). For $n > 1$ Banh (unpublished) has verified the conjecture for Ω of the form $U + i\Gamma$ where U is an open set in $\mathbb{R}^n$ and Γ is open and star shaped with respect to the origin. The proof is based on Theorem 5.1 of [8] and Theorem 5.8 below.

How can we construct suitable P_0? Let us return to the case $n = 1$. The simplest P_0 is of the form

$$P_0(x) = \hat{x}x + \hat{p}x^2. \tag{5.36}$$

We choose γ as the semicircle $|x| = \frac{1}{2}, \Im x \geq 0$ (Figure 5.1). For $\Im P_0$ to be positive on γ we require first of all that $\Im \hat{p} > 0$. For simplicity let $\hat{p}$ be (positive) pure imaginary, say $-i\hat{p} = \beta > 0$. This makes the term $\hat{p}x^2$ have positive imaginary part near $\gamma \cap R$ and on the part of R with $|x| \geq \frac{1}{2}$.

But as x goes around the circular part of γ, especially near the imaginary axis, the imaginary part of $\hat{p}x^2$ becomes negative. Thus we need to adjust $\hat{x}$ properly. If $\hat{x}$ is real and positive, or close to real and positive, $\Im \hat{x}x > 0$ in this part of γ. We must, however, insure that $\Im \hat{x}x$ is large enough to cancel the negative imaginary part of $\hat{p}x^2$. This can be accomplished by making $\hat{x}$ large enough.

Precisely

$$|\Im(\hat{p}x^2)| \leq |\beta x^2| = \tfrac{1}{4}\beta.$$

On the other hand, for $\frac{\pi}{4} \leq \arg x \leq \frac{3\pi}{4}$, which is the region where $\Re x^2 < 0$,

$$\Im \hat{x}x \geq \tfrac{\sqrt{2}}{4}\hat{x}.$$

This means we are all right so long as

$$\hat{x} > \tfrac{\sqrt{2}}{2}\beta.$$

Actually this can be seen best (with the slightly better inequality $\hat{x} > \beta/2$) from

$$\begin{aligned}\Im(\hat{x}x + \hat{p}x^2) &= \hat{x}\Im x + \beta[(\Re x)^2 - (\Im x)^2] \\ &\geq (\hat{x} - \beta/2)\Im x + \beta(\Re x)^2 \end{aligned} \tag{5.37}$$

since

$$\begin{aligned} 0 \leq &\Im x \leq \tfrac{1}{2} \\ &\Im x \geq 2(\Im x)^2. \end{aligned}$$

This means that for $\hat{x} > \beta/2$ the imaginary part of $P_0(x)$ is > 0 on the whole of γ. But more is true, namely, when we are in the region $\hat{x} \geq \beta/\lambda$ with $\lambda < 2$:

$$\Im P_0(x) > c\beta \tag{5.38}$$

on all of γ as is clear from (5.37). Inequality (5.38) for the nonlinear Fourier transform is the analog of the exponential decrease that is characteristic of the linear Fourier transform of functions which are holomorphic in strips around the real axis and small at infinity in the strip.

Theorem 5.6 *Let f be a sufficiently differentiable function with*

$$\textit{support } f \subset [-1, 1].$$

f extends to a function which is holomorphic in $\Im x > 0$ near 0 if and only if the nonlinear Fourier transform $F(\hat{p}, \hat{x})$ and enough derivatives are $O(\exp(-c\beta))$ in the region of the positive quadrant in $\beta, \hat{x}$ bounded by the positive $\hat{x}$ axis and another ray ρ. Moreover F and enough derivatives are polynomially decreasing in $\hat{x}$ in the region bounded by $\beta = 0$ and ρ.

Remark 1 *The condition support $f \subset [-1, 1]$ is basically irrelevant since we can apply our condition to $f\chi_{[-1,1]}$ where $\chi_{[-1,1]}$ is a smoothing of the characteristic function of $[-1, 1]$.*

Remark 2 *The argument preceding the statement of Theorem 5.6 shows that the nonlinear terms in the nonlinear Fourier transform serve the function of a cut-off of the part of f lying outside $[-\frac{1}{2}, \frac{1}{2}]$. This is the part of f that accounts for our inability to apply linear Fourier analysis to this problem.*

Remark 3 *The region $\hat{x} \geq \beta/\lambda$ with $\lambda < 2$ that we met above came from the fact that γ was chosen as a semicircle of radius $1/2$. If we choose a smaller radius then, as is readily seen from a modification of (5.37), the region of exponential decrease is larger! However, this is not a contradiction because the amount of exponential decrease, i.e. the constant c in $\exp(-c\beta)$, is smaller for the smaller circle.*

Proof of Theorem 5.6 We have already proven the direct part of Theorem 5.6 except for the statements about growth and regularity in the angular region. There is no difficulty in differentiating the integral as many times as one wants with respect to $\hat{x}$ and $\hat{p}$ and thus establish the regularity of $F(\hat{p}, \hat{x})$. However, the usual way of showing that F is small, namely by applying integration by parts to derivatives of f, does not work because integration by parts of f' introduces the factor $i\hat{x} + 2ix\hat{p}$.

To rectify this we integrate by parts[1]

$$\int e^{i\hat{x}x}[e^{i\hat{p}x^2} f(x)]' \, dx = -i\hat{x}F(\hat{p}, \hat{x}). \tag{5.39}$$

The derivative $[\exp(i\hat{p}x^2)f(x)]'$ is bounded by $c|\hat{p}|$. This bound is not good enough for us. However, if we shift the contour to γ as before, the exponential factor becomes exponentially decreasing in β in the region of interest, that is, between $\beta = 0, \hat{x} \geq 0$, and ρ. Thus the factor $2i\hat{p}x$ which arises from differentiation can be absorbed in the exponential so the above integral is bounded. This shows that $F = \mathcal{O}(|\hat{x}|^{-1})$. By integration by parts of higher order derivatives we derive the rapid decrease. This completes the proof of the direct part of Theorem 5.6.

Let us pass to the converse. We know that $\hat{f} = F(0, \hat{x})$ so we can express f in terms of $F(\hat{x}, 0)$.

In the case of the linear Fourier transform we use the analyticity of $\hat{f}$ to shift the contour to a region where we have some special information about $\hat{f}$. This shifting of contours is possible by Cauchy's theorem.

Thus we want to shift the contour to a region in $(\hat{x}, \hat{p})$ where we have exponential decrease. Shifting contours for holomorphic functions is based on the Cauchy–Riemann equations; we now replace the Cauchy–Riemann equation by the heat (Schroedinger) equation satisfied by F.

To shift the contour we use integration by parts. Note that $\exp(-i\hat{x}x - i\hat{p}x^2)$ is the solution of the backwards (adjoint) of the Schroedinger equation (5.13) with initial data $\exp(-i\hat{x}x)$ on the $\hat{x}$ axis. Writing Green's formula in the region between the positive real $\hat{x}$ axis and the ray ρ allows us to shift the contour from the positive $\hat{x}$ axis to this ray.

Actually we use the heat equation rather than the Schroedinger equation because we use Green's formula in the plane $\hat{x}$ real, $\hat{p} = i\beta$. In this plane $F(i\beta, \hat{x})$ satisfies $\partial F = 0$ where ∂ is the heat operator $\frac{\partial}{\partial \beta} - \frac{\partial^2}{\partial \hat{x}^2}$. Also

$$\partial' e^{-i\hat{x}x - i\hat{p}x^2} = \partial' e^{-i\hat{x}x + \beta x^2} = 0$$

where $\partial' = -\frac{\partial}{\partial \beta} - \frac{\partial^2}{\partial \hat{x}^2}$ is the adjoint of ∂.

[1]This method of integration by parts also appears in the work of T. Banh.

We write f as an inverse Fourier transform

$$\begin{aligned} f(x) &= \int_{-\infty}^{0} \hat{f}(\hat{x}) e^{-i\hat{x}x} d\hat{x} + \int_{0}^{\infty} \hat{f}(\hat{x}) e^{-i\hat{x}x} d\hat{x} \\ &= f^+(x) + f^-(x) \end{aligned} \tag{5.40}$$

where $f^{\pm}$ is holomorphic in the upper (lower) half-plane. Thus we want to show that f^- is holomorphic in $\Im x > 0$ near $x = 0$.

Now we use Green's formula (integration by parts) in the real $\beta, \hat{x}$ plane in the region Δ bounded by the positive $\hat{x}$ axis and the real ray ρ: $\beta = \epsilon\hat{x}, \hat{x} \geq 0$. That gives,

$$\begin{aligned} 0 &= \iint \left[\partial' e^{-i\hat{x}x+\beta x^2}\right] \chi_\Delta F(\beta, \hat{x})\, d\hat{x}\, d\beta \\ &= \int_0^\infty e^{-i\hat{x}x} F(0, \hat{x})\, d\hat{x} \\ &\quad + \int_{\beta=\varepsilon\hat{x}} e^{-i\hat{x}x+\beta x^2} Q(\hat{x}, \beta, x) \tilde{\partial} F(\beta, \hat{x})\, d\hat{x}. \end{aligned} \tag{5.41}$$

Here χ_Δ is the characteristic function of Δ, Q is a simple polynomial, and $\tilde{\partial}$ is a first-order differential operator. The integrals in (5.41) make sense when x is small and $\Im x < 0$, since for such x the exponential factor is bounded and, in fact, is exponentially decreasing for $x\hat{x} \neq 0$. The function $F(\beta, \hat{x})$ is polynomially decreasing. Thus the integrals exist and there is no contribution from infinity, i.e. the integral over large arcs joining $\beta = 0$ and $\beta = \varepsilon\hat{x}$ tends to zero.

The last equality in (5.41) means that the inverse Fourier transform of $F(\hat{x}, 0)$ on $[0, \infty)$, i.e. f^-, can be expressed in terms of an integral along $\beta = \varepsilon\hat{x}$. But $F(\beta, \hat{x})$ and also $\tilde{\partial} F(\beta, \hat{x})$ are exponentially decreasing there. Thus the integral converges for small $|x|$. This shows that f^- has an analytic continuation to $|x| < \epsilon, \Im x > 0$. Since f^+ is also holomorphic there, so is $f = f^+ + f^-$.

This completes the proof of Theorem 5.6.

Remark. *The exponential decrease $\mathcal{O}(\exp(-c\beta))$ in Theorem 5.6 is proven by shifting contours in x using the Cauchy–Riemann equation. In Chapter 9 we develop a different procedure based on the fundamental principle (see Section 1.4). It seems that the fundamental principle works better for general systems of linear constant coefficient equations.*

The crucial point we just observed is that if f is holomorphic in $\Im x > 0$ (locally) then the same is true of $f^{\pm}$ where $f^{\pm}$ represents the (inverse) Fourier transform of the restriction of $\hat{f}$ to $\hat{x} < 0$ (resp. $\hat{x} > 0$). (In any case it is clear that f^+ is holomorphic in $\Im x > 0$.)

Putting things somewhat differently, f^- is holomorphic in a whole neighborhood of zero.

Let us pass to $n > 1$. For simplicity we shall deal only with the quadratic nonlinear Fourier transform. As before we write $\beta = -i\hat{p}$.

If $\vec{e} = (e_j)$ is a vector in $\hat{x}$ space then we define $\boxed{e^2}$ as the matrix $(e_i e_j)$ which we identify with a point β. In this identification

$$\begin{aligned}\boxed{e^2} \cdot (x_i x_j) &= \sum e_i x_i e_j x_j \\ &= (\vec{e} \cdot x)^2.\end{aligned}$$

There are several versions of *contracted nonlinear Fourier transform* that are significant:

(1) *One-dimensional contraction.* Let L be a line in $\hat{x}$ space and let $\vec{e}$ be a unit vector on L. For each $\hat{x}^0 \in L^\perp$ there is a natural plane $P_{\vec{e}}(\hat{x}^0)$ in $(\beta, \hat{x})$ "lying over" $L + \hat{x}^0$, namely

$$P_{\vec{e}}(x^0) = \{t\boxed{e^2}, \hat{x}^0 + u\vec{e}\}_{t,u}.$$

We write

$$F(t\boxed{e^2}, \hat{x}^0 + u\vec{e}) = \int e^{-t\boxed{e^2} \cdot x^2 + iu\vec{e}\cdot x + i\hat{x}^0 \cdot x} f(x)\, dx$$

so the heat equation

$$\left(\frac{\partial}{\partial t} - \frac{\partial^2}{\partial u^2}\right) F(t\boxed{e^2}, \hat{x}^0 + u\vec{e}) = 0 \tag{5.42}$$

holds on $P_{\vec{e}}(\hat{x}^0)$. Moreover $P_{\vec{e}}(\hat{x}^0)$ is the only 2 plane in $(\beta, \hat{x})$ containing $L + \hat{x}^0$ on which F satisfies a differential equation.[2]

(2) *Circular contraction.* We set

$$F_c(\beta, \hat{x}) = F(\beta, \ldots, \beta, 0, \ldots; \hat{x}) = \int e^{ix\cdot\hat{x} - \beta \sum x_j^2} f(x)\, dx.$$

F_c satisfies the usual heat equation

$$\left(\frac{\partial}{\partial \beta} - \Delta_{\hat{x}}\right) F_c = 0. \tag{5.43}$$

Instead of the quadratic form $\sum \hat{x}_j^2$ we could have defined F_c using any positive definite quadratic form.

(3) *Rectangular contraction*

$$F_r(\beta_1, \ldots, \beta_n, \hat{x}) = \int e^{ix\cdot\hat{x} - \sum \beta_j x_j^2} f(x)\, dx.$$

[2]This contraction was used in Section 5 of [8].

F_r satisfies the equations

$$\left(\frac{\partial}{\partial\beta_j} - \frac{\partial^2}{\partial x_j^2}\right) F_r = 0 \quad j = 1, \ldots, n. \tag{5.44}$$

Remark. *In what follows we shall write F for F_r unless specifically stated otherwise.*

We want to formulate an analog of Theorem 5.6 in several complex variables. We can generalize "holomorphic in $\Im x > 0$" to "holomorphic in a tube over a proper convex cone $\Gamma \subset$ imaginary $z = x + iy$ space." (All our considerations are local.) This tube is $\{x + iy\}_{y\in\Gamma}$.

When Γ is open, or the closure of an open cone, the meaning of "holomorphic" is clear. But we shall meet cones which are (closures of) open cones lying in linear subspaces L of $\{y\}$. In this case "holomorphic in the tube over Γ" means holomorphic in the tube over Γ in the complexification $L^{\mathbb{C}}$ of L and in the real translates of $L^{\mathbb{C}}$. (This is the same as lying in the kernel of the induced Cauchy–Riemann equations on $L^{\mathbb{C}}$ and on its real translates, which is a concept that has meaning for more general systems of partial differential equations.)

If f^0 is a function on $\mathbb{R}^n$ we say that f^0 extends to a holomorphic function in the tube over Γ ($\Gamma \subset L$) if there is a function f defined locally on $L^\perp + L^{\mathbb{C}}$ which satisfies the induced Cauchy–Riemann equations on $u + L^{\mathbb{C}}$ for each $u \in L^\perp$. The size of the domain of holomorphicity is uniform over $u \in L^\perp$.

Our methods lend themselves more conveniently to Γ which are convex polygonal cones. Since any convex Γ can be approximated by convex polygonal cones our results can easily be extended to general Γ. We shall not deal with the necessary "epsilonics" except when they are significant, leaving them to the reader.

As mentioned in Section 1.1 there are two ways of describing a set. In the case of a convex polygon A they become:

(1) (Parametric). Express A in terms of its extreme points.
(2) (Equations). Express A in terms of its supporting planes.

In Section 2.3, which involved linear Fourier transform, we used (2). Nonlinear Fourier transform seem more geared to (1).

If we are given a set σ of points $\{y^j\}$ we can construct the convex hull $H(\sigma)$ by first connecting the points $\{y^j\}$ by linear segments, then connecting the points on the resulting line segments, etc. $H(\sigma)$ results after $\leq n$ steps. (In this context a vector is considered as two points, namely the origin and its endpoint.)

For convex polygonal cones $\sigma = \{\gamma_j\}$ where γ_j can be chosen to be points on the unit sphere or lying in an affine hyperplane not passing through the origin. Γ is obtained from $H(\sigma)$ by homothety, i.e.

$$\Gamma = \{rH(\sigma)\}_{r\geq 0}.$$

We begin our treatment of general Γ by the simplest case in which $n = 2$ and Γ is the positive quadrant, so we can use $\gamma_j = \vec{e}_j$ which are the unit vectors along the positive axes. Note that the hypothesis of the Hartogs–Bernstein theorem (Section 1.5) is that we are given a function f^0 on $\mathbb{R}^2$ which extends locally to the tubes over $H(\vec{e}_1)$ and $H(\vec{e}_2)$.

Theorem 5.7 *Let f^0 be defined (locally) on $\mathbb{R}^2$. The following properties are equivalent:*

(a) *f^0 extends to holomorphic functions in the tubes over $H(\vec{e}_1)$ and $H(\vec{e}_2)$.*
(b) *$F(\beta_1, 0; \hat{x}_1, \hat{x}_2)$ is exponentially decreasing in the region in the positive $\{(\beta_1, \hat{x}_2)\}$ quadrant bounded by $\{\beta_1 = 0\}$ and by another ray ρ_1. The measure of exponential decrease and ρ_1 are independent of $\hat{x}_2$. Similarly when the roles of $\beta_1, \hat{x}_1$ and $\beta_2, \hat{x}_2$ are reversed.*
(c) *$F(\beta_1, \beta_2; \hat{x}_1, \hat{x}_2)$ is exponentially decreasing in β_1 in the same region as in (b); the exponential decrease and ρ_1 are independent of $\beta_2, \hat{x}_2$. Similarly when the roles of $\beta_1, \hat{x}_1$ and $\beta_2, \hat{x}_2$ are reversed.*
(d) *Let $\vec{e}$ be any unit vector lying in the first quadrant of the (y_1, y_2) plane. Then $F(u\boxed{e^2}, v\vec{e})$ is exponentially decreasing in u in the part of the positive quadrant of the (u, v) plane bounded by $u = 0$ and $u = \alpha v$, $\alpha > 0$. The amount of exponential decrease and α can be chosen to be independent of $\vec{e}$.*
(e) *The circular contraction $F_c(\beta; \hat{x})$ is exponentially decreasing in β as long as $\hat{x}$ is in the complement of a conical neighborhood of the negative quadrant and $0 \leq \beta \leq c|\hat{x}|$.*
(f) *f^0 extends to a function which is holomorphic in the tube over the positive quadrant.*

Remark. *The standard regularity and polynomial decrease should be added to the hypotheses; they will be used implicitly in the proof.*

Proof The proof of Theorem 5.6 certainly shows the required decrease in the $\beta_1, \hat{x}_1$ plane independently of $\beta_2, \hat{x}_2$ if f is holomorphic in the tube over $H(\vec{e}_1)$. In this way (a) implies (c). Conversely (b) which is weaker than (c) implies (a) by the method of the proof of Theorem 5.6. Thus (a), (b), (c) are equivalent.

The same method shows that (f) implies (d). For, f being holomorphic in a tube over the positive quadrant implies that it is holomorphic in the tube over any ray in the quadrant. The uniformity is readily established. Since (a) is a special case of (d), (f) implies (a), (b), (c), (d). To complete the proof of Theorem 5.7 we show:

(c) implies (e).
(a) implies (f).
(e) implies (f).

Proof that (c) implies (e) Suppose we are in the complement of a conical neighborhood of the negative quadrant. Then (at least) one of $\hat{x}_1, \hat{x}_2$, say $\hat{x}_1$, is positive and $|\hat{x}| \leq c''\hat{x}_1$. Let $\beta_1^0 = \beta_2^0 = \beta^0$ and suppose $0 \leq \beta^0 \leq c'|\hat{x}|$ so

$0 \le \beta^0 \le c\hat{x}_1$. Hence by (c) we have

$$|F_c(\beta^0; \hat{x})| \le c^0 e^{-c^0 \beta^0}. \tag{5.45}$$

Since $0 \le \beta^0 \le c'|x|$ the same is true for $\alpha\beta^0$ for $0 \le \alpha \le 1$ so the same exponential decrease is valid for all $r\beta^0, r \ge 0$. This verifies (e).

Proof that (a) implies (f) As in the case $n = 1$ we start with the inverse Fourier transform

$$f(x_1, x_2) = \int \hat{f}(\hat{x}_1, \hat{x}_2) e^{-ix\cdot\hat{x}}\, d\hat{x}. \tag{5.46}$$

The part of the integral coming from the negative quadrant is holomorphic in the tube over the positive quadrant since $y \cdot \hat{x} < 0$ when $\hat{x} \in$ negative quadrant and $y \in$ (slightly diminished) positive quadrant. (As mentioned we leave the "epsilonics" of such diminuations to the reader.)

For the rest of the integral we shift the contour using a refinement of the idea of proof of Theorem 5.6. To accomplish this we modify (5.46) to

$$f(x_1, x_2) = \int_{\beta=0} F(\beta_1, \beta_2; \hat{x}_1, \hat{x}_2) e^{-ix\cdot\hat{x} + \beta_1 x_1^2 + \beta_2 x_2^2}\, d\hat{x}. \tag{5.47}$$

We use the fact that the exponential satisfies the adjoint heat equations in $(\beta_1, \hat{x}_1)$ and in $(\beta_2, \hat{x}_2)$. (By "shifting contours" we mean using Green's formula as in (5.41).)

We first shift the $\hat{x}_1$ contour from the positive $\hat{x}_1$ axis to $\beta_1 = \gamma\hat{x}_1$. As in (5.41) this gives us exponential decrease in $\hat{x}_1$. Unfortunately we need exponential decrease in both $\hat{x}_1$ and $\hat{x}_2$. Exponential decrease in $\hat{x}_1$ is fine so long as $|\hat{x}_2| \le \varepsilon|\hat{x}_1|$. But when $\hat{x}_2$ is large compared to $\hat{x}_1$ it is of no value.

We are saved by the fact that the shifting in $\beta_1, \hat{x}_1$ is independent of $\beta_2, \hat{x}_2$. Thus in the first $\hat{x}$ quadrant with $\beta_1, \hat{x}_1$ on $\beta_1 = \gamma\hat{x}_1$ we shift the contour in $\hat{x}_2$ to $\beta_2 = \gamma\hat{x}_2$. We now have exponential decrease in $|\hat{x}_1| + |\hat{x}_2|$ so, as in the case of $n = 1$, this integral is holomorphic in x in the tube over the first quadrant.

In the second quadrant we shift only in $\hat{x}_2$ while in the fourth quadrant we shift in $\hat{x}_1$. The exponential factor in the integrand for the fourth quadrant for complex $x_1 + iy_1$, $x_2 + iy_2$ is

$$\exp(-ix_1\hat{x}_1 + y_1\hat{x}_1 + \beta_1 x_1^2 - ix_2\hat{x}_2 + y_2\hat{x}_2). \tag{5.48}$$

On the contour the function F is exponentially decreasing in β_1 which is of the order of $\hat{x}_1$. Thus for y_1, x_1 small this decrease overcomes the exponential growth of $\exp y_1\hat{x}_1$ and $\exp \beta_1 x_1^2$. Since $\hat{x}_2 < 0$ the terms involving $\hat{x}_2$ are bounded for $y_2 \ge 0$.

We conclude that the contribution from the fourth quadrant (and similarly from the second quadrant) is holomorphic in $y_1 > 0, y_2 > 0$.

This completes the proof that (a) implies (f).

Proof that (e) implies (f) We want to use the heat equation in the region between the complement of a neighborhood N of the negative $\hat{x}$ quadrant and $0 \le \beta \le c|\hat{x}|$. There is a difficulty because the "top" of this region, meaning $\{\beta = c|\hat{x}|\}$, is not connected to the "bottom" which is $\{\beta = 0\}$. This is not a serious problem because we start with a slightly larger neighborhood N_2 of the negative quadrant and then connect the top over N_2 to the bottom as the boundary of N_1 (see Figure 5.2).

For $z = x + iy$ we write the "circular" analog of (5.47) as

$$f(z_1, z_2) = \int_{\beta=0} F_c(\beta, \hat{x}) e^{-ix\cdot\hat{x} + y\cdot\hat{x} + \beta\hat{x}\cdot\hat{x}}\, d\hat{x}. \tag{5.47c}$$

We can shift the contour and use the exponential decay of F_c as long as $y \cdot \hat{x} \le 0$, i.e. $y \in -N_2'$ (meaning the dual of N_2). N_2 differs epsilonically from the negative quadrant so we obtain holomorphicity as long as y is in a cone which is epsilonically smaller than the positive quadrant.

As we mentioned, we shall ignore these epsilonics so Theorem 5.7 is proven.

Let us pass to general n and general polyhedral cones. For $n = 2$ the only convex cones are angular regions, or rays. We have treated the cone which is the first quadrant. Suppose we had a different nondegenerate proper convex cone which in parametric (birth) form is $H(\vec{e}_1^*, \vec{e}_2^*)$ where $\vec{e}_j^*$ are linearly independent unit vectors.

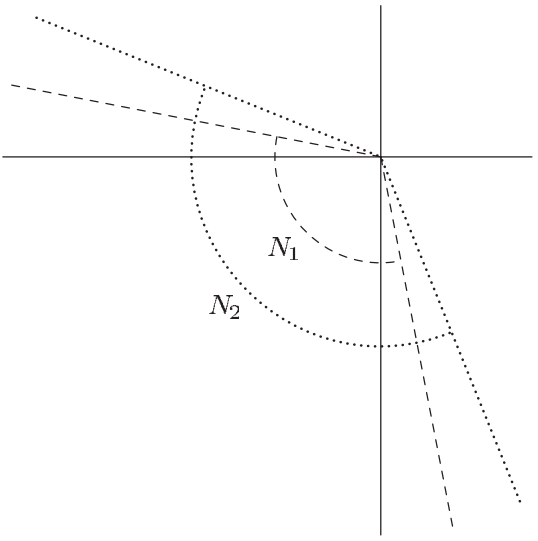

Figure 5.2

We write the cartesian coordinates x of a point in the form $x = \Lambda(x^*)$ where x^* represents the coordinates in the $\vec{e}^*$ basis and Λ is a linear transformation. Then the inner product becomes

$$x \cdot \hat{x} = \Lambda x^* \cdot \hat{x} = x^* \cdot \Lambda' \hat{x}$$

where Λ' is the adjoint of Λ. We regard $\hat{x}^* = \Lambda' \hat{x}$ as the coordinates of $\hat{x}$ in the dual basis to $\{\vec{e}_1^*, \vec{e}_2^*\}$. If we define the coordinates of β in terms of this basis then everything goes as before.

For $n > 2$ and Γ the positive orthant we can argue exactly as in the case $n = 2$. Using the linear transformation as above we see that the same is true if Γ is defined by n linearly independent vectors. However, when Γ is a general polygonal convex cone we need a new twist.

Γ is the set of positive scalar multiples of a "base" which is a convex polygon. We can choose the base as the convex hull $H(\{\gamma_j\})$ of a set of points $\{\gamma_j\}$ lying in some affine hyperplanes not passing through the origin. We may assume that each γ_j is an extreme point for the base.

The general form of Theorem 5.7 is

Theorem 5.8 *Let f^0 be defined (locally) in $\mathbb{R}^n$. The following properties are equivalent:*

(A) *f^0 extends to a holomorphic function in the tube over the rays through each γ_j.*

(B) *For each j, $F(u\boxed{\omega_j^2}, v\vec{\omega}_j)$ is exponentially decreasing in the part of the positive quadrant of the (u, v) plane bounded by the positive v axis and $u = \alpha v$, $\alpha > 0$. The exponential decrease and a can be chosen independent of $\hat{x} \in \vec{\omega}_1^{\perp}$. Here $\vec{\omega}_j$ is the unit vector pointing to γ_j.*

(C) *$F(\beta, \hat{x})$ is exponentially decreasing when $\beta = u\boxed{\omega_j^2} + \beta^1, \hat{x} = v\vec{\omega}_j + \hat{x}^{\perp}$ in the same u, v region as in* (B). *$\hat{x}^{\perp}$ is an arbitrary point in $\omega^{\perp}$ and β^1 defines a positive quadratic form of $x^{\perp}$.*

(D) *Let $\vec{e}$ be the unit vector on the ray joining the origin with some point in $H(\{\gamma_j\})$. $F(u\boxed{e^2}, v\vec{e})$ is exponentially decreasing in the same u, v region as in* (B).

(E) *The circular contraction $F_c(\beta, \hat{x})$ is exponentially decreasing in β as long as $\hat{x}$ is in the complement of a neighborhood of $-\Gamma'$ and $0 \leq \beta \leq c|\hat{x}|$.*

(F) *f^0 extends to a function which is holomorphic in the tube over Γ.*

Proof All implications amongst (A), (B), (C), (D), (E), (F) are proven exactly is the same way as their counterparts in Theorem 5.7 except for (A) implies (F). The difficulty in verifying this implication comes from the fact that Γ may not be the linear transform of an orthant.

Given one γ_k we can find, perhaps in several ways, n linearly independent vectors amongst the vectors joining the origin to γ_k and to the $\tilde{\gamma}_{jk}$ which are those γ_l such that the line joining γ_k to $\tilde{\gamma}_{jk}$ is an edge of $H(\{\gamma_l\})$ (see Figure 5.3).

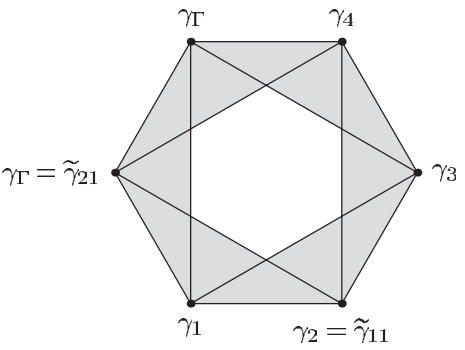

Figure 5.3

For fixed k, $\vec{\omega}_k$ and the vectors $\tilde{\omega}_{jk}$ joining the origin to $\tilde{\gamma}_{jk}$ are linearly independent and form a basis for $\mathbb{R}^n$. By Theorem 5.7 and the remarks following the proof of Theorem 5.7 we know that f^0 extends to a function which is holomorphic over the tube spanned by $\vec{\omega}_k$ and the $\tilde{\omega}_{jk}$. Applying this idea to each γ_k leads to an extension of f^0 to a function holomorphic in the tube over the union of rays through an inside neighborhood of the boundary of $H(\{\gamma_j\})$. This is the shaded area in Figure 5.3 so we refer to it as *shaded* $H(\{\gamma_j\})$.

It is clear that for any point p in $H(\{\gamma_j\})\backslash$ *shaded* $H(\{\gamma_j\})$ we can find n points $\{\lambda_l\}$ in *shaded* $H(\{\gamma_j\})$ such that p lies in the interior of the convex hull of $\{\lambda_l\}$. Since f^0 extends to a holomorphic function in the tube over each λ_l it follows that f^0 extends to a holomorphic function in the tube over the convex cone $\{rH(\{\lambda_l\})\}_{r>0}$ which clearly contains a conical neighborhood of p. Thus f^0 extends to a function which is holomorphic in a tube over Γ.

This completes the proof of (A) implies (F) and hence of Theorem 5.8.

Note that in the hypothesis (A) of Theorem 5.8 f is defined on a set of real dimension $n+1$ whereas in the conclusion it is holomorphic on a set of real dimension $2n$.

We can rephrase the implication (A) implies (F) (which is the main result) in Theorem 5.8 in the following terms which have significance for more general overdetermined systems of partial differential equations. We start with a function f^0 on $\mathbb{R}^n$, which is a Cauchy surface for $\bar{\partial}$. We want to know if f^0 is the CD of a solution of $\bar{\partial}$. Theorem 5.8 tells us that if f^0 extends to a solution (globally or locally) on $\cup(\mathbb{R}^n \oplus I_j)$ where I_j is the imaginary x_j axis then f^0 extends to, i.e. is the CD of, a solution (locally) on $\mathbb{C}^n = \mathbb{R}^{2n}$. (By a solution on $\cup(\mathbb{R}^n \oplus I_j)$ we mean a function which satisfies the induced Cauchy–Riemann equations on this set.)

The Korevaar–Wiegerinck theorem which we have mentioned in various places is of a similar nature. This theorem involves the exotic CP for $\bar{\partial}$ with data f at the origin. Thus f^0 is a formal power series and f^0 extends to a solution of $\bar{\partial}$ if f^0 extends on every complex line through the origin.

In Chapter 9 we discuss various types of extension properties. In particular we present an extension of Theorem 5.8 to more general systems of linear partial differential equations with constant coefficients.

Remark. *In the next section we shall explain how to formulate Theorem 5.8 in terms of wave front sets. It is the wave front formalism which provides the above-mentioned extension of Theorem 5.8 to partial differential equations.*

A second consequence is the famous edge-of-the-wedge theorem (see [105]) which we state in local form (there is no difficulty in extending our argument to the global theorem):

Theorem 5.9 *Let f be defined on R^n near 0. Suppose f has (locally) holomorphic extensions f_1, f_2 to the tubes over Γ_1, Γ_2. Then f has a holomorphic extension to the tube over the convex hull of $\Gamma_1 \cup \Gamma_2$.*

Remark. *For the original edge-of-the-wedge theorem Γ_1 = positive orthant and Γ_2 = negative orthant. The (global) conclusion is that f is entire.*

Theorem 5.9 follows easily from Theorem 5.8.

So much for the holomorphic theory. How about C^∞? We shall assume that $n = 1$ as the technique of passing to $n > 1$ has already been clarified in our treatment of the holomorphic case. Thus, let f be a distribution with support in $[-1, 1]$ as before and let f be C^∞ on $[-\frac{1}{2}, \frac{1}{2}]$. We write $f = f_1 + f_2$ where $f_1 \in \mathbb{C}^\infty$ with support $f_1 \subset [-\frac{1}{2}, \frac{1}{2}]$ and support f_2 lies outside $[-\frac{1}{2} + \epsilon, \frac{1}{2} - \epsilon]$.

Theorem 5.10 *A necessary and sufficient condition that f be C^∞ on $(-\frac{1}{2}, \frac{1}{2})$ is: there is an a such that for all ϵ, M*

$$|F(\beta, \hat{x})| \leq c_{M,\epsilon}(1 + |\hat{x}|)^a \exp(-(\tfrac{1}{4} - \epsilon)\beta) \tag{5.49}$$

for $0 \leq \beta \leq M \log(1 + |\hat{x}|)$, with similar inequalities for the derivatives of F. ($c_{M,\varepsilon}$ and a depend on the order of the distribution.)

Proof Call F_1, F_2 the respective quadratic nonlinear Fourier transforms of f_1, f_2. Since f_2 is a distribution of compact support whose support is contained in $|x| \geq \frac{1}{2} - \varepsilon$ we can write $f_2 = \sum d^j \mu_j / dx^j$ for some measures μ_j of compact support whose supports are contained in $|x| \geq \frac{1}{2} - \varepsilon'$ for $\varepsilon' > \varepsilon$ and arbitrarily close to ε. (We shall not distinguish between ε and ε'.)

Since

$$F_2(\beta, \hat{x}) = \sum \int_{|x| \geq \frac{1}{2} - \varepsilon} \frac{d^j}{dx^j} [e^{ix\hat{x} - \beta x^2}] \, d\mu_j(x)$$

it is clear that F_2 satisfies

$$|F_2(\beta, \hat{x})| \leq c(1 + |\hat{x}| + |\beta|)^a e^{-(\frac{1}{4} - \varepsilon)\beta}$$

for all $\beta \geq 0$ and all $\hat{x}$. Since we are interested in a region where $\beta \leq |\hat{x}|$, the term $(1 + |\hat{x}| + \beta)$ can be replaced by $(1 + |\hat{x}|)$.

Since $f_1 \in C^\infty$ its linear Fourier transform

$$\hat{f}_1(\hat{x}) = F_1(0, \hat{x}) = \mathcal{O}(1 + |\hat{x}|)^{-M} \quad \text{all } M.$$

We now use the fact that $F_1(\beta, \hat{x})$ satisfies the heat equation. This means that (in our normalization)

$$F_1(\beta, \hat{x}) = \beta^{-1/2} \int e^{-(\hat{x}-\hat{y})^2/\beta} F_1(0, \hat{y})\, d\hat{y}. \tag{5.50}$$

There is no difficulty in establishing (5.49) when $\beta \leq 1$ so we assume $\beta > 1$. The main contribution to the integral comes from values of $\hat{y}$ near $\hat{x}$. For example, when $|\hat{x} - \hat{y}| \leq |\hat{x}|/2$ the integrand is bounded for large $\hat{x}$ by

$$c_N |\hat{x}|^{-N}$$

for all N since $|\hat{y}| \geq |\hat{x}|/2$. We want this to be $\mathcal{O}(\exp(-(\frac{1}{4} - \varepsilon)\beta))$. This happens when

$$\begin{aligned} |\hat{x}|^{-N} &= \mathcal{O}(e^{-(\frac{1}{4}-\varepsilon)\beta)}) \\ -N \log |\hat{x}| &\leq c - (\tfrac{1}{4} - \varepsilon)\beta, \end{aligned} \tag{5.51}$$

i.e.

$$\beta \leq c + N \log |\hat{x}|.$$

This shows the inequality for the part of the integral with $|\hat{x} - \hat{y}| \leq |\hat{x}|/2$. As $\hat{y}$ moves away from $\hat{x}$ the function $F_1(0, \hat{y})$ can grow if $|\hat{y}|$ diminishes. But this is more than compensated by the decrease in $\exp(-(\hat{x} - \hat{y})^2/\beta) \leq \exp(-\hat{x}^2/4\beta)$. It is easy to carry out the details and verify the necessity of (5.49).

For the converse we want to write the inverse Fourier transform and shift the contour as we did in the case of holomorphicity. One problem with this is that F may not be integrable on $\beta = 0$ if the a in (5.49) is nonnegative. We can overcome this difficulty by convolving f with a function g of support $\subset (-\frac{\epsilon}{2}, \frac{\epsilon}{2})$ which is of class $C^{|a|+3}$. It is clear that if $g * f$ is C^∞ on $(-\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon)$ for all such g then f is C^∞ on $(-\frac{1}{2}, \frac{1}{2})$.

All this means that we can assume that $a = -3$ in (5.49). We now apply (5.41) to the region Ω' bounded by $\beta = 0$ and $\beta = N' \log(1 + |\hat{x}|)$. It is clear from (5.49) that all integrals in (5.41) converge absolutely and that the integration by parts is valid. We have thus an expression for the inverse Fourier transform of $F(\hat{x}, 0)$ in terms of an integral on $\beta = N \log(1 + |\hat{x}|)$. So long as $x^2 < \frac{1}{4} - 2\epsilon$, the exponential factor in (5.41) satisfies, on this contour,

$$e^{\beta x^2} \leq (1 + |\hat{x}|)^{(\frac{1}{4} - 2\epsilon)N}.$$

Thus the product of this with the right side of (5.51) is bounded by

$$c(1 + |\hat{x}|)^c c(1 + |\hat{x}|)^{-\epsilon N}.$$

This is enough to absorb differentiation of the last term of (5.41) to any order since N is arbitrary large. (Of course c depends on N.)

This completes the proof of Theorem 5.10.

Remark 1 *This formulation of Theorem 5.10 is due to T. Banh. The author's original formulation was less elegant.*

Remark 2 *The main difference between the present situation and the one encountered in support (i.e. does f vanish on $[-\frac{1}{2}, \frac{1}{2}]$?) is that in the latter case the "lift" of f to $y = x^2$ vanishes in the half-plane $y \leq \frac{1}{4}$. This condition can be studied by Paley–Wiener (convex) techniques. But if f is regular in $[-\frac{1}{2}, \frac{1}{2}]$ then the lifted measure is regular only on the curve $y = x^2$. It is* not *regular in the variables x, y.*

There are several ways to get around this problem. For one thing we could replace the single curve $y = x^2$ by a family of curves.

Instead of using a family of curves we can use the translates of f. This is the standard form of the FBI transform [25]

$$G(z, \hat{x}) = \int e^{-i\hat{x}x - |\hat{x}|(z-x)^2} f(x) \cdot dx. \tag{5.52}$$

The FBI transform lives on $\beta = |\hat{x}|$ so that no use is made of heat equations which play a central role in our work.

(5.52) can be also be considered from the viewpoint of a family of curves. To this end we think of $\exp(i\hat{x}z)$ *as a "normalizing" factor. Then the integrals are the Fourier transforms evaluated at $(-\hat{x}, i|\hat{x}|)$ of the lifts of $f(x)dx$ to $y = (z - x)^2$ for varying z.*

Remark 3 *We discussed local vanishing of functions at the beginning of this section, then local analyticity (Theorem 5.6), then local C^∞ (Theorem 5.10). In the next section we shall discuss functions which are locally between C^∞ and analytic (e.g. Gevrey functions). We could also characterize functions which extend from a small neighborhood of 0 to "very regular" functions, e.g. entire functions of order ρ.*

We have seen that C^∞ corresponds to exponential decrease in logarithmic regions, analyticity to angular regions, and vanishing to "vertical regions," i.e. $\Im \hat{p} \to \infty$. Thus it seems reasonable that being an entire function of order ρ should correspond to the quadratic nonlinear Fourier transform being exponentially decreasing in regions of the form (roughly)

$$\beta \sim c\hat{x}^\sigma.$$

Indeed, this is the case with

$$\frac{1}{\rho} + \frac{1}{\sigma} = 1.$$

This checks with our previous cases if we interpret $\rho = \infty$ as analytic, and if we interpret $\rho = 1, \sigma = \infty$ in a slightly different fashion. We leave the details to the reader.

A more ambitious program is to extend theorems of linear Fourier analysis, e.g. gap and density theorems, to nonlinear Fourier analysis. We leave this to the reader.

Remark 4 *One of the original impetuses for introducing nonlinear Fourier analysis was the hope that it would lead to a method of solving partial differential equations with polynomial coefficients much as linear Fourier analysis leads to an algebraic theory for constant coefficient operators. Of course, one could not hope for a reduction to algebraic equations, but at least to first-order equations, which could then be solved using ordinary differential equations.*

Unfortunately, up to now we have success only with equations each of whose terms is of rank ≤ 2. The *rank* of $p\partial(q)$ where p and q are homogeneous polynomials is defined as degree $p +$ degree q.

5.3 Wave front set

If f is small at infinity then $\mathbf{R}f$ (hyperplane Radon transform) describes, via the Fourier transform, the restrictions of $\hat{f}$ to lines through the origin. Closely allied to $\mathbf{R}f$ is the concept of *wave front set* which describes a certain structure on the restrictions of $\hat{f}$ to rays through the origin. The wave front set is (roughly) the set of rays where $\hat{f}$ is "large" at infinity, the meaning of "large" depending on the theory in question. (The passage between rays and lines does not involve any significant difficulties.) Of course if f is large at infinity or only locally defined then we have to use cut-off functions. The use of cut-off functions represents the cornerstone of our theory.

The mathematical definition of wave fronts arose out of the study of hypersurfaces H along which f can have discontinuities. If H is linear (affine), this is related by the projection–slice theorem to the fact that $\hat{f}$ can be large in the orthogonal to $\hat{H}$. For $\hat{f}|_{\hat{H}^\perp}$ is the Fourier transform of $\mathbf{R}f$ on the spread $\mathbf{g}$ of hyperplanes parallel to H. The fact that $\hat{f}$ is large at infinity corresponds to the nonsmoothness of $\mathbf{R}f(\mathbf{g})$ in the spread parameter. If f describes water waves then there is an orientation on $\hat{H}^\perp$; this accounts for the change from lines to rays.

We shall give brief mention to wave fronts corresponding to nonlinear H.

Let us put these ideas in an abstract setting. We start with two spaces $\mathcal{W}_1 \subset \mathcal{W}$ of functions or distributions which are small at infinity. Assume that the Fourier transforms of functions or distributions in $\mathcal{W}$ and $\mathcal{W}_1$ have restrictions to certain subsets Γ of $\mathbb{R}^n$. We say that Γ is *not in the* $(\mathcal{W}_1, \mathcal{W})$ *wave front set* of $f \in \mathcal{W}$ if

$$\hat{f}|_\Gamma \in \hat{\mathcal{W}}_1(\Gamma).$$

The meaning of $\hat{\mathcal{W}}_1(\Gamma)$ needs clarification. In the cases of interest $\hat{\mathcal{W}}_1$ is a space of functions on $\mathbb{R}^n$ which is defined by growth conditions and regularity. $\hat{\mathcal{W}}_1(\Gamma)$ means the space of functions on Γ satisfying the same conditions there.

In the standard cases $\mathcal{W}$ is a space of distributions and $\mathcal{W}_1$ is either a space of C^∞ functions or a space of holomorphic functions. In the former case we search for rays ρ in which $\hat{f}$ decreases faster than $|\hat{x}|^{-N}$ for all N and in the latter case for the directions in which $\hat{f}$ is exponentially decreasing in a suitable sense which is clarified in Theorem 5.15. Γ is a cone containing ρ in its interior.

Another important case is for $\mathcal{W}_1$ a Gevrey space, in which case the relevant decrease is $\exp(-\varepsilon|\hat{x}|^\alpha)$ for a fixed $\alpha < 1$ determined by the $\mathcal{W}_1$. (We shall be more precise below.)

One can go even further in the direction of abstraction. The Fourier and Radon transforms represent averages of a function f. The restriction of $\hat{f}$ to a line is the Fourier transform of the set of averages of f over a spread (projection–slice theorem). Spreads represent coarse grains in $\mathcal{W}'$. In general we are given a decomposition of $\mathcal{W}'$ into subspaces $\mathcal{W}'(S_{\mathbf{g}})$ (coarse grains) which may have large overlaps. We say that $\mathcal{W}'(S_{\mathbf{g}_0})$ *is not in the* $(\mathcal{W}_1, \mathcal{W}; \{S_{\mathbf{g}}\})$ *wave front set of* $f \in \mathcal{W}$ if the behavior of f on $\mathcal{W}'(S_{\mathbf{g}_0})$ is comparable to the behavior of functions in $\mathcal{W}_1$ on $\mathcal{W}'(S_{\mathbf{g}_0})$.

The concept "comparable" needs clarification. Choose a fixed coarse grain $\mathcal{W}'(S_{\mathbf{g}_0})$. We assume that it is equipped with a basis $\{w'(s,\alpha)\}$. (In applications $w'(s,\alpha)$ might be the δ function of a geometric object where s represents the usual Radon type of parameter, e.g. belonging to a line or $n-l$ dimensional plane, and α represents the base parameter of a cone or similar structure. Another possibility is that $\{w'(s,\alpha)\}$ is the linear or nonlinear Fourier basis.) Thus for $f \in \mathcal{W}$ we can examine

$$F(s,\alpha) = f \cdot w'(s,\alpha).$$

We want $F(s,\alpha)$ to satisfy the same regularity and growth as that of functions $f \in \mathcal{W}_1$.

Let us contrast the usual analytic wave front set with the notion of holomorphic function. If $h(x)$ is holomorphic in $\Im x_1 > 0$ for each real $x_2, \ldots, x_n$ and is suitably small at infinity, then $\hat{h}(\hat{x})$ is exponentially decreasing in $\hat{x}_1 \to \infty$ for all real $\hat{x}_2, \ldots, \hat{x}_n$. Thus the sets $\{\tilde{\Gamma}\}$ used in the definition of "wave front set for actual analyticity" are half-spaces. Since the Γ used in the definition of analytic wave front set (for analyticity of averages) are arbitrarily thin cones, we can think of Γ as being like a ray in $\{\hat{x}\}$. Such rays represent codirections in x space; this is in conformity with the fact that the wave front set is related to averages of f over hyperplanes. For this reason wave front sets are considered to be in the cotangent bundle (when the definition is extended to manifolds).

On the other hand, $\Im x_1 > 0$ is a direction, so when we say f is holomorphic in $\Im x_1 > 0$ then we are in the tangent bundle (complexified). This is consistent with the fact that the condition on $\hat{f}$ is exponential decrease in a half-space, i.e. a codirection condition in the dual space.

Put in other terms, the wave front condition for the codirection $\hat{x}_1 > 0$ refers to the *averages* of f on the planes $x_1 = \text{const}$. The direction condition $\Im x_1 > 0$ refers to the *values* at the points $(x_2, \dots, x_n) = \text{const}$. as x_1 moves.

The definition of (usual) wave front set to which we have alluded involves averages over hyperplanes, while actual analyticity involves the unaveraged function. In keeping with the spirit of the Radon transform we can define "l plane wave front set" in terms of averages of f over spreads of l planes, meaning $\mathbf{R}^l f$. However, as pointed out in Section 2.3, the analog of the passage from "line" to "ray" for $\mathbf{R}^l$ can only be made if we assume some spectral condition. The spectral condition tells us that *support* $\hat{f}$ meets "a large number" of $n - l$ planes $\mathbf{g}$ through the origin in the union of two cones $\Gamma^{\pm}(\mathbf{g})$ with disjoint interiors.

We shall not deal with l plane wave front sets in this work.

Earlier in this chapter we discussed the Radon transform on nonlinear varieties. This leads to the concept of *nonlinear wave front set*. Using our previous notation, we say (preliminary definition) that the m codirection $\hat{p}^0$ does not belong to the C^∞ nonlinear wave front set of order m if

$$\int e^{it\hat{p}\cdot p} f(x)\, dx = \mathcal{O}(t^{-N}) \quad \text{for all } N \text{ as } t \to \infty \tag{5.53}$$

for $\hat{p}$ in a suitable neighborhood of $\hat{p}^0$.

In order for (5.53) to be meaningful f must be small. Actually we want the wave front set to be locally defined so (5.53) should hold not for f itself but for χf where χ is a *cut-off function*, meaning $\chi \in C^\infty$ has compact support and is $\equiv 1$ on some neighborhood of 0.

We want to think of a wave front set as a regularity condition. As such it makes sense that if $\hat{p}^0$ is not in the C^∞ wave front set of f then neither is $\hat{p}^0$ in the wave front set of φf for any C^∞ function φ. This leads to the (second) definition

$\hat{p}^0$ is not in the C^∞ wave front set of f if and only if (5.53) holds for φf (in place of f) for all $\varphi \in C^\infty$ with support in some (small) compact set.

The reason for requiring (5.53) in a suitable neighborhood of $\hat{p}^0$ rather than $\hat{p}^0$ itself will be clarified in the linear case. In what follows we shall limit ourselves to linear wave front sets.

We have presented two essential aspects of the possible definition of "wave front set," one based on varying p^0 (more precisely, tp^0) in a suitable neighborhood, and another based on multiplying f by cut-offs. We want to coordinate the definitions. To accomplish this we must first understand: what is a suitable neighborhood? Classically one uses conical neighborhoods. In fact, we can sometimes use regions which are smaller than conical. We shall study regions of

the form

$$\hat{x}_1 \geq 0$$
$$|\hat{x}_2| + \cdots + |\hat{x}_n| \leq c\eta(\hat{x}_1) \tag{5.54}$$

where η is a smooth monotonically increasing function $\to \infty$ with bounded derivative.

Let us begin with C^∞ wave front sets. Assume that f is a distribution of compact support and the positive $\hat{x}_1$ axis does not belong to the η C^∞ wave front set, meaning that $\hat{f}(\hat{x}) = O(|\hat{x}|^{-N})$ for all N in the region (5.54). Under what conditions on η does this imply that the same is true for ϕf whenever $\phi \in \mathcal{D}$? When this property is true we say that the η C^∞ *wave front set is locally defined.*

Theorem 5.11 *Suppose η satisfies*

(a) $\hat{x}_1^\alpha \leq c\eta(\hat{x}_1) \leq c'\hat{x}_1^{\alpha'}$ *for* $\hat{x}_1 \to +\infty$

(b) $\eta(c''\hat{x}_1) \leq c'''\eta(\hat{x}_1)$ (5.55)

for some constants $\alpha, \alpha', c, c', c'', c'''$.

Then the $\eta\, C^\infty$ wave front set is locally defined.

Proof We may clearly assume that $\alpha < 1$. We shall use the notation c for a constant which may vary in different inequalities, providing that no confusion is possible. Similarly ε is a small constant which may change according to the context.

Let us examine

$$\widehat{\phi f} = \hat{\phi} * \hat{f}$$

in the region (5.54). By definition

$$(\hat{\phi} * \hat{f})(\hat{x}) = \int \hat{\phi}(\hat{x} - \hat{y})\hat{f}(\hat{y})\, d\hat{y}. \tag{5.56}$$

Since $\phi \in \mathcal{D}$ we certainly have

$$\hat{\phi}(\hat{x} - \hat{y}) \leq c|\hat{x} - \hat{y}|^{-N} \tag{5.57}$$

for all N. Fix $\hat{x}$ in the region (5.54). The right side of (5.57) is $O(|\hat{x}|^{-M}\, |\hat{y}|^{-n-2})$ for all M as long as

$$|\hat{x} - \hat{y}| \geq |\hat{x}|^\beta |\hat{y}|^\beta \tag{5.58}$$

for some β. Thus this part of the integral (5.56) satisfies the desired regularity $= \mathcal{O}(|\hat{x}|^{-M})$ for all M.

Suppose that $\hat{x}$ is in the region (5.54). If (5.58) is violated then, in fact, for suitably small ε

$$\begin{aligned} |\hat{x} - \hat{y}| &\leq |\hat{x}|^\varepsilon\, |\hat{y}|^\varepsilon \\ &\leq c|\hat{x}_1|^{\varepsilon'}\, |\hat{y}|^\varepsilon \end{aligned} \tag{5.59}$$

if ε' is suitably chosen, because of (5.54) and (5.55).

We claim that (5.59) implies

$$\begin{aligned} |\hat{y}_1| &\geq c\hat{x}_1 \\ |\hat{y}_j| &\leq c\eta(y_1) \quad \text{for } j > 1. \end{aligned} \tag{5.60}$$

To verify the first inequality of (5.60), we apply the triangle inequality to (5.59) and use the fact that $|\hat{x}_1 - \hat{y}_1| \leq |\hat{x} - \hat{y}|$. We find

$$|\hat{y}_1| \geq \hat{x}_1 - \hat{x}_1^{\varepsilon}\,|\hat{y}|^{\varepsilon}.$$

(We have replaced ε' by a new ε which is the larger of $\varepsilon, \varepsilon'$ of (5.59).) We must show first that for $\hat{y}$ large

$$c|\hat{x}| \leq |\hat{y}| \leq c'|\hat{x}|. \tag{5.61}$$

The first inequality in (5.59) shows that

$$|\hat{y}| \leq |\hat{x}| + |\hat{x}|^{\varepsilon}\,|\hat{y}|^{\varepsilon}.$$

Suppose $|\hat{y}| \geq |\hat{x}|$. Then

$$|\hat{y}| \leq |\hat{x}| + |\hat{y}|^{2\varepsilon}.$$

But $|\hat{y}|^{2\varepsilon}$ is insignificant compared to $|\hat{y}|$ so

$$|\hat{y}| \leq 2|\hat{x}|.$$

If $|\hat{x}| \geq |\hat{y}|$ we can reverse the roles of $\hat{x}$ and $\hat{y}$; thus (5.61) is verified. In particular by (5.59) and (5.61)

$$\begin{aligned} |\hat{y}_j| &\leq |\hat{x}_j| + |\hat{x}|^{\varepsilon}\,|\hat{y}|^{\varepsilon} \\ &\leq |\hat{x}_j| + c\hat{x}_1^{2\varepsilon} \quad \text{by (5.54) and (5.55a).} \end{aligned}$$

For $j > 1$ this gives

$$\begin{aligned} |\hat{y}_j| &\leq c\eta(\hat{x}_1) + c\hat{x}_1^{2\varepsilon} \\ &\leq c'\eta(\hat{x}_1) \end{aligned} \tag{5.62}$$

by (5.54) as long as $2\varepsilon < \alpha$. On the other hand

$$\begin{aligned} |\hat{y}_1| &\geq \hat{x}_1 - c\hat{x}_1^{\varepsilon}|y|^{\varepsilon} \\ &\geq \hat{x}_1 - c\hat{x}_1^{2\varepsilon} \\ &\geq c\hat{x}_1. \end{aligned} \tag{5.63}$$

Since η is monotonically increasing we have thus verified (5.60).

Our hypothesis is that in the region (5.60) the function $\hat{f}$ is rapidly decreasing. Thus for $\hat{x}$ in (5.54) and $\hat{y}$ in the part of (5.60) where (5.59) holds

$$\begin{aligned}|\phi(\hat{x}-\hat{y})\hat{f}(\hat{y})| &\le c|\hat{f}(\hat{y})| \\ &\le c|\hat{y}|^{-N-2-n} \\ &\le c|\hat{x}|^{-N}\,|\hat{y}|^{-2-n} \quad \text{by (5.61).}\end{aligned}$$

(We are using N to denote a large constant (which may differ in the various inequalities).) Hence the integral (5.56) over this set of $\hat{y}$ is $\mathcal{O}(|\hat{x}|^{-N})$ in the region (5.54).

This completes our proof of Theorem 5.11.

The converse is:

Theorem 5.12 *Let $\eta(\hat{y}_1)$ be a smooth function with bounded derivative defined for $\hat{y}_1 > 0$. Suppose that*

$$\eta(\hat{y}_1) = \mathcal{O}(\hat{y}_1^{\alpha})$$

for all $\alpha > 0$. There exists an $\hat{f} \in \mathcal{S}'$ which is rapidly decreasing in the region (5.54) and a $\hat{\varphi} \in \hat{\mathcal{D}}$ such that $\hat{\varphi} * \hat{f}$ is not rapidly decreasing on the positive $\hat{y}_1$ axis. The support of φ can be chosen arbitrarily small.

Proof We can clearly assume that $\eta \ge c > 0$. We choose $\hat{f}$ as the δ function of the surface

$$\hat{y}_1 > 0, \quad |\hat{y}_2| + \cdots + |\hat{y}_n| = \eta(\hat{y}_1). \tag{5.64}$$

Thus $\hat{f}$ vanishes on (5.54) for $c < 1$ and *a fortiori* is rapidly decreasing there.

We choose for $\hat{\varphi}$ a positive function of $\hat{r} = |\hat{x}|$. Then the value $(\hat{\varphi} * \hat{f})(\hat{x}_1, 0, \ldots, 0)$ is positive and is an integral of $\hat{\varphi}(\hat{x}_1 - \hat{y}_1, -\hat{y}_2, \ldots, -\hat{y}_n)$ over the surface (5.64). To make this integral large we have to use small values of $|(\hat{x}_1 - \hat{y}_1, -\hat{y}_2, \ldots, -\hat{y}_n)|$ because $\hat{\varphi}$ falls off rapidly at infinity. On (5.64), for fixed $\hat{x}_1 > 0$,

$$\min |(\hat{x}_1 - \hat{y}_1, -\hat{y}_2, \ldots, -\hat{y}_n)| \le \eta(\hat{x}_1) \tag{5.65}$$

since we can take $\hat{y}_1 = \hat{x}_1$ and $(\hat{y}_2, \ldots, \hat{y}_n)$ satisfying (5.64). This is the "smallness" that we need.

We now show that if $\eta(\hat{x}_1) = \mathcal{O}(\hat{x}_1^{\alpha})$ for all $\alpha > 0$ then we can find a positive radial function $\hat{\varphi} \in \hat{\mathcal{D}}$ and a sequence of points $\hat{x}_{1m} \to \infty$ such that

$$|\hat{\varphi}(\hat{x}_{1m} - \hat{y}_1, -\hat{y}_2, \ldots, -\hat{y}_n)| \ge |\hat{x}_{1m}|^{-p} \tag{5.66}$$

for some fixed p for all $\hat{y}$ satisfying (5.64) and

$$|\hat{x}_{1m} - \hat{y}_1| \le 1. \tag{5.67}$$

We construct $\hat{\varphi}$ as follows. We start with a lacunary sequence $\{\hat{x}_{1m}\}$ with the property that

$$\eta(\hat{x}_1) \leq \hat{x}_1^{1/m} \quad \text{for } \hat{x}_1 \geq \hat{x}_{1m}. \tag{5.68}$$

We can choose the sequence $\{\hat{x}_{1m}\}$ lacunary enough so that there is a positive function $\hat{\psi} \in \hat{\mathcal{D}}(\mathbb{R}^1)$ satisfying

$$\hat{\psi}(\eta(\hat{x}_{1m})) = [\eta(\hat{x}_{1m})]^{-\sqrt{m}}.$$

We may assume that $\hat{\psi}$ is an even function so it defines a radial function $\hat{\varphi} \in \hat{\mathcal{D}}$.

By construction, when $\hat{y}_1 = \hat{x}_{1m}$ and $(\hat{y}_2, \ldots, \hat{y}_n)$ satisfy (5.64)

$$\begin{aligned} \hat{\varphi}(\hat{x}_{1m} - \hat{y}_1, -\hat{y}_2, \ldots, -\hat{y}_n) &= \hat{\psi}(\|(\hat{y}_2, \ldots, \hat{y}_n)\|) \\ &= [\eta(\hat{x}_{1m})]^{-\sqrt{m}} \\ &\geq \hat{x}_{1m}^{-\sqrt{m}/m}. \end{aligned} \tag{5.69}$$

By a slight modification we can demonstrate a similar inequality for all $\hat{y}_1$ lying within a distance ≤ 1 from $\hat{x}_{1m}$. Thus (5.66) and (5.67) are verified with $p = 1$.

Since the derivative of η is bounded the fact that $\hat{\varphi} * \hat{f}$ is not $\mathcal{O}(\hat{x}_1^{-2})$ on the positive $\hat{x}_1$ axis is readily established.

This completes the proof of Theorem 5.12.

Remark. *Theorem 5.12 does not precisely contradict the definition of locality, for that requires f to be of compact support.*

Problem 5.5 *Can one choose f in Theorem 5.12 to be of compact support?*

There is a classical theory of spaces lying between C^∞ and real analytic. The theory was developed to a large extent by Denjoy and by Carleman (see [126]). It is presented in detail in Chapter XIII of FA. We assume for the present that $n = 1$ and we use y as a variable in $\mathbb{R}^1$.

Let $\{m_j\} = M$ be a sequence of positive numbers. (Sometimes some regularity is imposed on M.) We define the space $\mathcal{E}_M$ to consist of all C^∞ functions f on $\mathbb{R}$ such that f and all its derivatives ∂f satisfy

$$|(\partial f)^{(j)}(y)| \leq c\varepsilon^j m_j \tag{5.70}$$

for some $\varepsilon > 0$ on every interval. ε and c may depend on ∂ and the interval. This definition is somewhat different from the one used in FA. (Formula (5.108) of FA requires that (5.70) must hold for all ε rather than for some ε which may depend on the interval. The difference between the two types of conditions is minor.)

Let

$$\lambda_M(y) = \sum \frac{|y|^j}{m_j} \tag{5.71}$$

where $\lambda_M(y) = \infty$ if the series diverges.

The main result of Denjoy and Carleman is that the space $\mathcal{E}_M$ is quasi-analytic (QA), meaning it contains no functions of compact support if and only if

$$\int_{-\infty}^{\infty} \frac{\log \lambda_M(y)}{1+y^2}\, dy = \infty. \tag{5.72}$$

When $\mathcal{E}_M$ is nonquasianalytic we denote by $\mathcal{D}_M$ the space of $f \in \mathcal{E}_M$ of compact support. $\mathcal{D}_M$ is given the topology of $\mathcal{LF}$ space (inductive limit of the subspaces defined by support $f \subset K$ for a fixed cube K).

We have defined the spaces $\mathcal{E}_M$ for $n = 1$. For $n > 1$ we shall only consider sequences $\{m_j\}$ $(j = j_1, \ldots, j_n)$ for which

$$m_j = m^0_{j_1} m^0_{j_2} \ldots m^0_{j_n} \tag{5.73}$$

for some sequence $\{m^0_l\}$ depending on a single index. The QA properties of $\{m_j\}$ are the same as those of $\{m^0_l\}$.

An important class of nonquasianalytic spaces are the *Gevrey spaces* $\mathbf{G}_\alpha$ defined by

$$\mathbf{G}_\alpha : \quad m_j = (\alpha j)! \tag{5.74}$$

for $\alpha \geq 1$. It is readily verified from (5.72) that $\mathbf{G}_\alpha$ is nonquasianalytic unless $\alpha = 1$. In fact the function $\lambda = \lambda_\alpha \sim \exp |y|^{1/\alpha}$ where $\sim$ means = modulo "unimportant factors."

When $\mathcal{E}_M$ is nonquasianalytic it contains the dense subspace $\mathcal{D}_M$ of functions of compact support. We use the notation $\mathbf{G}^0_\alpha$ for the space of functions of compact support in $\mathbf{G}_\alpha$.

If $f \in \mathcal{D}_M$ then the Fourier transform $\hat{f}$ satisfies an inequality of the form

$$|\hat{f}(\hat{y})| \leq c \left[\lambda\left(\frac{\hat{y}}{\varepsilon}\right)\right]^{-1} e^{b|\Im \hat{y}|}(1+|\hat{y}|)^{-N} \tag{5.75}$$

for some $b, c > 0$ and some ε, and all $N > 0$. The constant c can depend on N.

The simplest heuristic way to understand (5.75) is that (5.70) means that the series

$$\sum m_j^{-1} \varepsilon^{-j} \left(\frac{d}{dy}\right)^j f \tag{5.76}$$

converges in the topology of the space $\mathcal{D}$ for any $f \in \mathcal{D}_M$. Taking the Fourier transform leads to (5.75).

We define M *wave front sets* formally as before. Thus for the Gevrey spaces $\hat{x}_1 > 0$ is not in the (formal) $\mathbf{G}_\alpha$ wave front set at $x = 0$ if there exists a

neighborhood of the $\hat{x}_1$ axis on which

$$\hat{f}(\hat{y}) = \mathcal{O}(\exp(-b|\hat{y}|^{1/\alpha})) \tag{5.77}$$

for some $b > 0$ (This is meaningful when $\alpha > 1$.)

How large should this neighborhood be in order for the wave front set to have a local definition? If we examine the proofs of Theorems 5.11 and 5.12 we deduce

Theorem 5.13 *Let $\eta(\hat{y}_1)$ be a smooth concave function with bounded derivative which is monotonically increasing to ∞. A necessary and sufficient condition that if (5.77) holds in the region (5.54) then it also holds for $\hat{\varphi} * \hat{f}$ for any $\varphi \in \mathbf{G}_\alpha^0$ is*

$$\eta(\hat{y}_1) \geq \gamma \hat{y}_1 \tag{5.78}$$

for some $\gamma > 0$.

Thus cones are the right sets for Gevrey classes. Similarly cones work for all $\mathcal{E}_M$ as long as M has some regularity. For this reason we define M wave front sets using cones.

The analog of Theorem 5.10 for Gevrey spaces, i.e. is the nonlinear Fourier transform characterization of functions belonging to $\mathbf{G}_\alpha$ on $|y| \leq 1/2$, can be proven along the same lines as the proof of Theorem 5.10.

We find that the local characterization of $\mathbf{G}_\alpha$, in analogy to (5.51), is

$$|F(\beta, \hat{y})| \leq ce^{-c'\beta} \tag{5.79}$$

for some c, c' in the region

$$0 \leq \beta \leq c''|\hat{y}|^{1/\alpha}. \tag{5.80}$$

Moreover F grows at most polynomially in this region.

A similar result holds for all nonquasianalytic spaces $\mathcal{E}_M$; the term $|\hat{y}|^{1/\alpha}$ in (5.80) is replaced by $\log \lambda_M(\hat{y})$. For our methods to be applicable we need to assume that $\log \lambda_M(\hat{y})$ is a concave function which is monotonically increasing to ∞. (This is a mild restriction on M.)

We now give a more precise definition of M (or $\mathcal{E}_M$) wave front sets:

Definition *The ray $\hat{\rho} \subset \mathbb{R}^n$ is not in the M wave front set of f at $x = 0$ if there is a $b > 0$ so that*

$$\widehat{\varphi f}(\hat{x}) = \mathcal{O}(e^{-\lambda_M(b\hat{x})}) \tag{5.77M}$$

in a conical neighborhood of $\hat{\rho}$ for any $\varphi \in \mathcal{D}_M$ of support in a sufficiently small neighborhood at the origin.

Thus the wave front set consists of pairs $x, \hat{x}$ where x is a base point and $\hat{x}$ is a codirection; that is, it is a subset of the cotangent bundle.

It is important to study analytic (and QA) wave front sets. Unfortunately there are no functions which are both real analytic and of compact support so another approach is needed.

It is proven in [37] that

$$\mathbf{A} = \bigcap_{\text{nonquasianalytic}} \mathcal{E}_M \tag{5.81}$$

where $\mathbf{A}$ is the space of real analytic functions. It therefore seems reasonable to say that the ray $\hat{\rho}$ is not in the analytic wave front set of h if it is not in the M wave front set for all M which are nonquasianalytic. We shall make this definition a little more precise below.

Actually we shall present four approaches to the concept of analytic wave front set:

(1) Relation (5.81) and nonquasianalytic wave front sets.
(2) Nonlinear Fourier transform.
(3) Sato's definition (see [103]).
(4) AU spaces.

The definition of M wave front sets was given in terms of the Fourier transform. We shall translate this condition into x space.

Let us determine, for $n = 1$, the condition that the positive $\hat{x}$ axis does not belong to the M wave front set of h. If M is nonquasianalytic then this means that $(\widehat{\chi h})(\hat{x}) = \mathcal{O}[\lambda_M(\hat{x})]^{-1}$ as $\hat{x} \to +\infty$ for suitable cut-off functions $\chi \in \mathcal{D}_M$. (We have ignored polynomial factors and we have normalized $\varepsilon = 1$.)

We write

$$\begin{aligned} \chi h(x) &= \int_{-\infty}^{\infty} \widehat{\chi h}(\hat{x}) e^{-ix\hat{x}}\, d\hat{x} \\ &= \int_{-\infty}^{0} + \int_{0}^{\infty} \\ &= h^+(x) + h^-(x). \end{aligned} \tag{5.82}$$

$h^{\pm}$ is clearly holomorphic in the upper (lower) half-plane. They have boundary values in a sense which depends on M. (See Chapter 9 for a detailed investigation of the meaning of such boundary values.) It is clear that $h^- \in \mathcal{E}_M$.

Since we could have chosen $\chi \equiv 1$ near the origin it follows that, locally,

$$h(x) = h_1(x) + h_2(x) \tag{5.83}$$

where $h_1(x)$ is (locally) holomorphic in $\Im x > 0$ and $h_2 \in \mathcal{E}_M$.

Claim *The splitting (5.83) is equivalent to the positive $\hat{x}$ axis not being in the $\mathcal{E}_M$ wave front set of h.*

We can assume that $\mathcal{E}_M$ is closed under multiplication, since it is sufficient to use such M in (5.81). Thus $\chi h_2 \in \mathcal{D}_M$.

To verify our claim we introduce a construction which will be useful for other purposes. Suppose for simplicity that h_1 is holomorphic in

$$U = \{|\Re x| < 1,\ 0 < \Im x < 1\} \tag{5.84}$$

and is smooth on the closure of U. We want to apply the fundamental principle (see Section 1.4) to determine the growth properties of $\hat{h}_1$. More precisely we represent h_1 as a Fourier transform of a measure $\hat{h}_1$ in $\hat{\mathcal{E}}_{\bar\partial}(U)$ where $\mathcal{E}_{\bar\partial}(U)$ is the kernel of $\bar\partial$ on $\mathcal{E}(U)$. (U is now considered as a rectangle in real (x, y).)

By the fundamental principle we can describe $\hat{\mathcal{E}}'_{\bar\partial}(U)$ as the space of holomorphic functions on the variety $\hat{x} + i\hat{y} = 0$ with growth conditions induced from $\hat{\mathcal{E}}'(U)$. Since U is a product set the growth in $\hat{\mathcal{E}}'(U)$ is the product of growths in $\hat{x}$ and in $\hat{y}$ corresponding to the spaces $\mathcal{E}(\{|x| < 1\})$, $\mathcal{E}(\{0 < y < 1\})$.

The growth in $\hat{\mathcal{E}}'(\{|x| \le 1\})$ is (modulo polynomial and ε factors) of the form $\exp(|\Im\hat{x}|)$ while for $\hat{\mathcal{E}}'(\{0 < y < 1\})$ it is

$$\begin{cases} e^{\Im\hat{y}} & \Im\hat{y} \to -\infty \\ 1 & \Im\hat{y} \to \infty. \end{cases} \tag{5.85}$$

Hence on $\{\hat{x} + i\hat{y} = 0\}$ using the parameter $\hat{x}$ we find on the real and imaginary axes

$$\text{Real axis: } \begin{cases} 1 & \hat{x} \to \infty \\ e^{\hat{x}} & \hat{x} \to -\infty \end{cases} \tag{5.86}$$

$$\text{Imaginary axis: } e^{|\hat{x}|}. \tag{5.87}$$

It is not difficult to use standard techniques to show that the union of the real and imaginary axes is sufficient (Section 1.4). (As we shall see later this is not important.)

From our description of the space $\hat{\mathcal{E}}'_{\bar\partial}(U)$ we deduce, by duality, that we can find a representative measure $\hat{h}_1$ for the Fourier transform of $h_1(-x)$, where h_1 is considered as an element of the dual of $\mathcal{E}'(U)$, of the form

$$\hat{h}_1 = \hat{h}^+_{1\Re} + \hat{h}^-_{1\Re} + \hat{h}^+_{1\Im} + \hat{h}^-_{1\Im}. \tag{5.88}$$

The support of $\hat{h}^{\pm}_{1\Re}$ (resp. $\hat{h}^{\pm}_{1\Im}$) is on the real (resp. imaginary) positive (negative) axis with growth reciprocal to that of $\hat{\mathcal{E}}'_{\bar\partial}(U)$.

At this point we must clarify a certain (apparent) notational inconsistency. If f is a function then $\hat{f}$ is defined by

$$\hat{f}(\hat{x}) = \int f(x) e^{ix\cdot\hat{x}}\, dx. \tag{5.89}$$

If $S \in \mathcal{E}'$ we define

$$\hat{S}(\hat{x}) = S \cdot e^{ix\cdot\hat{x}}. \tag{5.90}$$

This is consistent with the usual embedding of functions into distributions.

The inverse Fourier transform of $\hat{f}$ is

$$f(x) = \int e^{-ix\cdot\hat{x}}\hat{f}(\hat{x})\,d\hat{x}. \tag{5.91}$$

In this formula we think of f and $\hat{f}$ as functions. This means that if we consider $\hat{f}$ as an element of $\hat{\mathcal{E}}$ (dual of $\hat{\mathcal{E}}'$) then we have to write

$$\hat{f}\cdot\hat{S} = \int \hat{f}(-\hat{x})\hat{S}(\hat{x})\,d\hat{x} \tag{5.92}$$

in order to make $\hat{f}\cdot\hat{S} = S\cdot f$ since this would be the correct formula if $S = \delta_{x^0}$. For then $\hat{S}(\hat{x}) = \exp(ix^0\cdot\hat{x})$ so

$$\delta_{x^0}\cdot f = f(x^0) = \int \hat{f}(-\hat{x})e^{ix^0\cdot\hat{x}}\,d\hat{x}. \tag{5.93}$$

Thus the function $\hat{f}$ is identified with the measure $\hat{f}(-\hat{x})\,d\hat{x}$ (supported on the real axis) as an element of $\hat{\mathcal{E}}$.

On the other hand $\hat{h}^{\pm}_{1\Re}, \hat{h}^{\pm}_{1\Im}$ are linear functions on $\hat{\mathcal{E}}'$ given by

$$\hat{h}^{\pm}_{1\Re}\cdot\hat{S} = \int \hat{h}^{\pm}_{1\Re}(\hat{x})\hat{S}(\hat{x})\,d\hat{x}. \tag{5.94}$$

In particular they define

$$h^{\pm}_{1\Re}(x) = \hat{h}^{\pm}_{1\Re}\cdot e^{ix\cdot\hat{x}} = \int \hat{h}^{\pm}_{1\Re}(\hat{x})e^{ix\cdot\hat{x}}\,d\hat{x} \tag{5.95}$$

(similarly for $\hat{h}^{\pm}_{1\Im}$). It is clear that $h^{\pm}_{1\Im}$ and $h^{-}_{1\Re}$ are holomorphic in full neighborhoods of 0. They certainly belong to $\mathcal{E}_M$ and so they contribute nothing to the M wave front set.

On the other hand $\hat{h}^{+}_{1\Re}$ vanishes on the negative $\hat{x}$ axis. Thus its convolution with χ satisfies

$$|(\hat{\chi} * \hat{h}^{+}_{1\Re})(\hat{x})| \le c_N[\lambda_M(\hat{x})]^{-1}(1+|\hat{x}|)^{-N} \quad \hat{x}\to-\infty. \tag{5.96}$$

We have accounted for the change from $\hat{x}$ to $-\hat{x}$ so the M wave front set of h_1 omits the positive axis. By (5.83) the same is true of h.

We have established our claim that (5.83) is equivalent to the property that the positive axis does not belong to the M wave front set of h.

All this works when M is nonquasianalytic. What happens when M is QA? Under suitable regularity conditions there is an analog of (5.81):

$$\mathcal{E}_M = \bigcap_{\substack{\mathcal{E}_{M'}\supset\mathcal{E}_M \\ M' \text{ nonquasianalytic}}} \mathcal{E}_{M'}. \tag{5.97}$$

We shall restrict most of our study to $\mathcal{E}_M = \mathbf{A} = \{\text{real analytic }\}$ but essentially everything we say regarding $\mathbf{A}$ applies to M which are QA and satisfy (5.97).

Definition 1 *($n = 1$) The positive $\hat{x}$ axis does not belong to the analytic wave front set of f if it does not belong to the $\mathcal{E}_M$ wave front set for any nonquasi-analytic M. The cut-off functions φ_M are $\equiv 1$ on some fixed neighborhood of the origin.*

It is reasonable to expect that the positive $\hat{x}$ axis not being in the $\mathbf{A}$ wave front set of h should be equivalent to h being holomorphic in $\Im x > 0$. The above argument shows that if h is holomorphic in $\Im x > 0$ then for any nonquasianalytic M, the M wave front set avoids $\{\hat{x} > 0\}$. Hence $\{\hat{x} > 0\}$ is not in the $\mathbf{A}$ wave front set of h (according to Definition 1).

Conversely suppose $\{\hat{x} > 0\}$ is not in the $\mathbf{A}$ wave front set of h. We should like to say that h is locally holomorphic in $\Im x > 0$. The above argument shows that we can write h in the form (5.83) or, more precisely,

$$h(x) = h_1^M(x) + h_2^M(x). \tag{5.98}$$

We should like to deduce (5.83) which amounts to "piecing" the h_j^M together to obtain an $h_2 \in \cap\mathcal{E}_M = \mathbf{A}$.

A word of caution: h_1^M and h_2^M depend on the cut-off so the piecing together is not obvious.

We do not know how to perform this piecing process directly. By translating to nonlinear Fourier transforms we shall be able to verify our contention.

Theorem 5.6 asserts that local analyticity can be expressed in terms of (5.79) and (5.80) with $\alpha = 1$. We can relate this to (5.81) by examining the proof of (5.81). It is seen that that result is equivalent to the following.

Let us restrict M by requiring that

$$\begin{aligned} m_{j+1} &\geq (j+1)m_j \\ m_1 &= 1. \end{aligned} \tag{5.99}$$

Relation (5.81) states that, with this restriction,

$$\max_{\text{nonquasianalytic}} \log \lambda_M(y) = |y|. \tag{5.100}$$

This is a "two-sided" inequality, meaning that it applies to $y > 0$ and $y < 0$. On the other hand, wave fronts are "one sided". Definition 1 represents a one-sided analog of (5.100). We shall examine its relation to other one-sided concepts of wave fronts.

We want to show that Definition 1 implies that the nonlinear Fourier transform $F(\beta, \hat{x})$ is exponentially decreasing in an angular region $0 \leq \beta \leq a\hat{x}$. The nonlinear Fourier transform has the advantage over the above definition of M wave front sets in that it does not require any "serious" regularity in the

cut-off function χ; the only requirement is that $\chi \equiv 1$ on some neighborhood of the origin.

The proof of (5.81), when formulated in the language of the nonlinear Fourier transform, asserts that if there is no a for which $|F(\beta, \hat{x})| \leq \exp(-a\hat{x})$ in $0 \leq \beta \leq a\hat{x}$, then there is an M for which for no a do we have

$$|F(\beta, \hat{x}| \leq e^{-a\beta} \tag{5.101}$$

for

$$0 \leq \beta \leq a \log \lambda_M(\hat{x}).$$

This is the one-sided analog of (5.81) alluded to above.

We can now apply Theorem 5.6 to conclude

For $n = 1$, if the positive $\hat{x}$ axis does not belong to the M wave front set of f for all M which are nonquasianalytic then f extends to an analytic function locally in $\Im x > 0$.

T. Banh has proposed a simplified proof that Definition 1 ($n = 1$) leads to the above analyticity of f, as follows.

We have shown that if the M wave front set of f omits the positive $\hat{x}$ axis then

$$(\varphi_M f)(x) = f_1^M(x) + f_2^M(x)$$

where f_1^M is holomorphic in some fixed neighborhood of the origin in $\Im x > 0$ independent of M and $f_2^M \in \mathcal{E}_M$. The functions φ_M are $\equiv 1$ on some fixed neighborhood of the origin, say $[-1, 1]$. Let $\varphi \in \mathcal{D}$ be a fixed function which is $\equiv 1$ on $[-1, 1]$. Thus $\varphi_M - \varphi \equiv 0$ on $[-1, 1]$ for any M. The nonlinear Fourier transforms, which we write more precisely as F_M, F, differ by the nonlinear Fourier transform of $(\varphi_M - \varphi)f$ which vanishes on $[-1, 1]$. As such, by the technique of Ehrenpreis and Malliavin discussed in Sections 5.1 and 5.2, this nonlinear Fourier transform is exponentially decreasing in β in any angular region in $\hat{x} \geq 0, \beta \geq 0$, the constants being independent of M.

We now use a variant of the proof of (5.85) in which we establish (5.83) or (5.98) using the nonlinear Fourier transform in place of the fundamental principle. This means we express $\varphi_M f$ in terms of its linear Fourier transform and then shift the contour for $\hat{x} \geq 0$ as in the proof of Theorem 5.6.

Now we write

$$F = F_M + (F - F_M)$$

where F, F_M are defined using the respective cut-offs φ, φ_M and shift the contour in the inverse linear Fourier transform as above. The part f_1 of F coming from the negative $\hat{x}_1$ axis is a fixed function which is holomorphic in a neighborhood of the origin in $\Im x > 0$. $f_1^M \in \mathcal{E}_M$ and the same is true of the part of $F - F_M$ which comes from the positive $\hat{x}$ axis by the result of Ehrenpreis and Malliavin described above. (In fact this part is analytic, as it must be for the result to be

true. One might think that we could shift the contour and make it vanish, but that is not possible because the Ehrenpreis–Malliavin shifting requires a shifting of the whole $\hat{x}$ axis, not only $\{\hat{x} \geq 0\}$.) Thus f_2 is also in $\mathcal{E}_M$.

We can now apply the two-sided theorem (5.81) to conclude that $f_2 \in \mathbf{A}$. Hence f is holomorphic in a neighborhood of the origin in $\Im x > 0$.

We now pass to $n > 1$. We have already explained how to define the $\mathcal{E}_M$ wave front set when M is nonquasianalytic. This leads to

Definition 1 *($n \geq 1$) The ray $\hat{\rho}^0$ does not belong to the analytic wave front set of f at $x = 0$ if there is a cone $\hat{\Gamma} \supset \hat{\rho}^0$ such that for every M which is nonquasianalytic there is a function $\varphi_M \in \mathcal{E}_M$ with $\varphi_M \equiv 1$ on some neighborhood of 0 which is independent of M, such that*

$$|\widehat{\varphi_M f}(\hat{x})| \leq c\lambda_M^{-1}(b|\hat{x}|)$$

on $\hat{\Gamma}$.

For the nonlinear Fourier transform definition of analytic wave front set we use a natural modification of Theorem 5.7.

Definition 2 *The ray $\hat{\rho}^0$ does not belong to the analytic wave front set of f if there is a cone $\hat{\Gamma}$ containing $\hat{\rho}^0$ such that the circular contraction $F_c(\beta, \hat{x})$ of the nonlinear Fourier transform is exponentially decreasing in β as long as $\hat{x} \in \hat{\Gamma}$ and $0 \leq \beta \leq c|\hat{x}|$.*

The proof that Definition 1 implies Definition 2 goes exactly as in the case $n = 1$. We now establish the converse by following along the lines of the proof of Theorem 5.7. Instead of using the full force of Definition 2 we show that if $F_c(\beta, \hat{x})$ is exponentially decreasing in β for $\hat{x} \in \hat{\Gamma}$ and $0 \leq \beta \leq \log \lambda_M(b|\hat{x}|)$ then $\hat{\rho}^0$ is not in the M wave front set of f.

To prove our contention we shift the contour defining the inverse Fourier transform of f (i.e. on $\beta = 0$) to

$$\beta_M = \beta_M(\hat{\Gamma}_\varepsilon) \cup B_M(\hat{\Gamma} - \hat{\Gamma}_\varepsilon) \tag{5.102}$$

where $\hat{\Gamma}_\varepsilon$ is a cone contained in $\hat{\Gamma}$ and close to $\hat{\Gamma}$. $\beta_M(\hat{\Gamma}_\varepsilon)$ is defined as the set of $(\beta, \hat{x})$ such that

$$\beta = \log \lambda_M(b|\hat{x}|), \quad \hat{x} \in \hat{\Gamma}_\varepsilon. \tag{5.103}$$

$B_M(\hat{\Gamma} - \hat{\Gamma}_\varepsilon)$ is defined as the set of $(\beta, \hat{x})$ such that $\hat{x} \in \hat{\Gamma} - \hat{\Gamma}_\varepsilon$ with β decreasing linearly from $\log \lambda_M(b|\hat{x}|)$ on the boundary of $\hat{\Gamma}_\varepsilon$ to zero on the boundary of $\hat{\Gamma}$.

Also $\beta = 0$ on $c(\hat{\Gamma})$. This leads to

$$\begin{aligned} f(x) &= \int e^{-ix\hat{x}} F(0,\hat{x})\, d\hat{x} \\ &= \int_{\beta_M(\hat{\Gamma}_\varepsilon)} e^{-ix\cdot\hat{x}+\beta x^2} QF(\beta,\hat{x})\, d\hat{x} + \int_{B_M(\hat{\Gamma}-\hat{\Gamma}_\varepsilon)} e^{-ix\cdot\hat{x}+\beta x^2} QF(\beta,\hat{x})\, d\hat{x} \\ &\quad + \int_{c(\hat{\Gamma})} e^{-ix\cdot\hat{x}} F(0,\hat{x})\, d\hat{x}. \end{aligned} \tag{5.104}$$

Q is a differential operator which comes from integration by parts.

As usual, the integral over $\hat{\Gamma}_\varepsilon$ lies in $\mathcal{E}_M$. The sum of the second and third integrals is a *Sato function for the codirection* $\hat{\rho}^0$. This is a function h of the form

$$h(x) = \sum h_j(x) \tag{5.105}$$

where h_j is holomorphic (locally) in a tube over a proper cone Γ_j such that

$$\hat{\rho}^0 \subset \text{ interior } \hat{\Gamma}_j'. \tag{5.106}$$

(Recall that $\hat{\Gamma}_j'$, the dual of Γ_j, is the set of $\hat{x}$ such that $\hat{x}\cdot x \geq 0$ for $x \in \Gamma_j$.)

It is easy to cut up $c(\hat{\Gamma}_\varepsilon)$ into such proper cones $-\hat{\Gamma}_j'$. If $\hat{h}_j$ has its support in $-\hat{\Gamma}_j'$ then the inverse Fourier transform h_j of $\hat{h}_j$ is holomorphic in the tube over Γ_j (up to an ε diminuation). (The additional quadratic exponential in $B_M(\hat{\Gamma}-\hat{\Gamma}_\varepsilon)$ only helps matters.) This verifies our claim that the sum of the second and third integrals in (5.104) is a Sato function.

We illustrate this cutting for $n = 2$ in Figure 5.4.

Since support $\hat{h}_j \subset -\hat{\Gamma}_j'$ it follows that h_j is holomorphic in a tube over Γ_j.

The method of proof of the claim following (5.83) can be used to show that, for each j,

$$(\hat{\varphi} * \hat{h}_j)(\hat{x}) = \mathcal{O}(\lambda_M^{-1}(c\hat{x})) \tag{5.107}$$

on a cone containing $\hat{\rho}^0$ in its interior.

We have verified the equivalence of Definitions 1 and 2.

Our methods can be used to show that Definition 2 of analytic wave front set implies

Definition 3 (Sato) *The function f in (5.103) is a Sato function for the codirection $\hat{\rho}^0$.*

The proof is identical with our proof that Definition 2 implies Definition 1 except that the term coming from the integral over $\hat{\Gamma}_\varepsilon$ is now holomorphic in a full neighborhood of the origin.

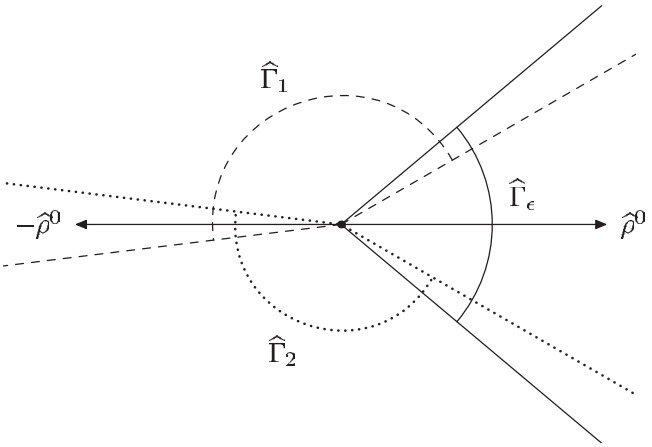

Figure 5.4

We want to prove that Sato's criterion is equivalent to Definitions 1 and 2. We shall show that Definition 3 implies

Definition 4 *(AU spaces). There exists a Fourier representation of f of the form*

$$f(x) = \int e^{ix\cdot\hat{x}}\, d\mu(\hat{x}) \tag{5.108}$$

where μ is a measure on $\mathbb{C}^n$ which satisfies the usual inequalities for $\hat{\mathcal{E}}^0$ (Fourier transform of C^∞ functions on some small neighborhood of 0)

$$d\mu(\hat{x}) = \mathcal{O}[(1+|\hat{x}|)^{-N} e^{-\alpha|\Im\hat{x}|}] \tag{5.109}$$

for $\hat{x} \in \mathbb{C}^n$, for all N for some fixed $\alpha > 0$, and also

$$d\mu(-\hat{x}) = \mathcal{O}(e^{-\alpha|\hat{x}|}) \tag{5.110}$$

for $\hat{x}$ in some conical neighborhood $\Gamma^{\mathbb{C}}$ of $\hat{\rho}^0$ in $\mathbb{C}^n$.

To show that Definition 3 implies Definition 4 it suffices to check the result for any function f_j which is holomorphic in the tube over Γ_j where $\hat{\rho}^0 \subset$ interior Γ_j'. This is an easy consequence of the fundamental principle applied to the kernel of the Cauchy–Riemann operator on C^∞ functions on this tube domain (see Chapter V of FA).

We show how Definition 4 implies Sato's criterion. For this purpose we write (5.108) as a sum of terms $f^1 + f^2$. f^1 is the integral over the set of $\hat{x}$ with $\Re\hat{x} \in (-\Gamma^{\mathbb{C}} \cap \mathbb{R}^n)_\varepsilon$ where the subscript ε means that we diminish this cone somewhat.

Note that when $\Re\hat{x} \in (-\Gamma^{\mathbb{C}} \cap \mathbb{R}^n)_\varepsilon$ but $\hat{x} \notin -\Gamma^{\mathbb{C}}$ then $|\Im\hat{x}| \geq \varepsilon'|\hat{x}|$. Using (5.110) we deduce that $d\mu$ is exponentially decreasing on this part of the tube over $(-\Gamma^{\mathbb{C}} \cap \mathbb{R}^n)_\varepsilon$ so its contribution to f^1 is holomorphic in a full neighborhood of the origin. On the other hand the contribution to f^1 of the integral (5.108) when $\hat{x} \in -\Gamma^{\mathbb{C}}$ is, by (5.110), holomorphic in a neighborhood of the origin.

It is easy to cut up the complement of the tube over $(-\Gamma^{\mathbb{C}} \cap \mathbb{R}^n)_\varepsilon$ as we did above to show that f^2 is a Sato function for the codirection $\hat{\rho}^0$. Thus f is a Sato function.

Although it is not necessary we have verified directly that Definition 4 implies Definition 3. We must still show how Definition 4 implies Definition 1.

We change our notation and write $\hat{z}$ in place of $\hat{x}$ and $\hat{z} = \hat{x} + i\hat{y}$ with real $\hat{x}, \hat{y}$. We want to show that $\widehat{\varphi f}(\hat{x}) = \mathcal{O}(\lambda_M^{-1}(a\hat{x}))$ near $\hat{\rho}^0$ for any $\varphi \in \mathcal{D}_M$ of sufficiently small support. Formally if f is the Fourier transform of the measure μ thought of as an element of $\hat{\mathcal{E}}_M$

$$\widehat{\varphi f} = \hat{\varphi} * \mu. \tag{5.111}$$

To make sense of the right side of (5.111) we have to regard $\hat{\varphi}, \widehat{\varphi f}$ as the measures $\hat{\varphi}(\hat{x})d\hat{x}, \widehat{\varphi f}d\hat{x}$ supported on $\mathbb{R}^n \subset \mathbb{C}^n$, so $\hat{\varphi} * \mu$ is the convolution of two measures on $\mathbb{C}^n$.

As we have mentioned before, there is some inconsistency in our notation since when we think of φ as a function we use $\exp(-ix \cdot \hat{x})$ as the kernel of the inverse Fourier transform whereas for $\mu \in \hat{\mathcal{E}}$ we use $\exp(+ix \cdot \hat{x})$. We regard $\hat{\varphi} * \mu$ as a function so for the inverse Fourier transform in (5.111) we use $\exp(-ix \cdot \hat{x})$. For this reason we shall see that the argument of $\hat{\varphi}$ is $(\hat{x} + \hat{z})$ rather than $\hat{x} - \hat{z}$.

We want to show that the consistent definition of the convolution (5.111) is

$$(\hat{\varphi} * \mu)(\hat{x}) = \int \hat{\varphi}(\hat{x} + \hat{z})\, d\mu(\hat{z}). \tag{5.112}$$

To verify that this definition is correct, we take the inverse Fourier transform of the right side of (5.112). There are no convergence problems so

$$\int e^{-ix\cdot\hat{x}}\, d\hat{x} \int \hat{\varphi}(\hat{x}+\hat{z})\, d\mu(\hat{z}) = \int e^{ix\cdot\hat{z}}\, d\mu(\hat{z}) \int e^{-ix\cdot(\hat{x}+\hat{z})}\hat{\varphi}(\hat{x}+\hat{z})\, d\hat{x}. \tag{5.113}$$

Since $\hat{\varphi}$ is entire and small as $\hat{x} \to \infty$, we can apply Cauchy's theorem to the last integral and shift the contour for each $\hat{z}$ to $\hat{x} + \hat{z}$ real. We obtain

$$\int e^{ix\cdot\hat{z}}\, d\mu(\hat{z}) \int e^{-ix\cdot\hat{x}}\hat{\varphi}(\hat{x})\, d\hat{x} = \varphi f \tag{5.114}$$

as required by (5.111).

At first sight it might seem strange that the right side of (5.111) should be a measure on $\mathbb{R}^n$ since we are translating $\hat{\varphi}$ by complex $\hat{z}$. It is Cauchy's theorem which enables us to put everything on the real plane.

We can put the convolution in a somewhat different framework. Regard the measure μ as an integral of δ functions. For any fixed $\hat{z}^0$

$$\widehat{\delta_{\hat{z}^0}}(x) = e^{ix\cdot\hat{z}^0}. \tag{5.115}$$

We claim that

$$\hat{\varphi} * \delta_{\hat{z}^0}(\hat{x}) = \hat{\varphi}(\hat{x} + \hat{z}^0) \tag{5.116}$$

so that convolution by $\delta_{\hat{z}^0}$ is translation by $-\hat{z}^0$.

To verify (5.116) we apply Cauchy's theorem to obtain

$$\int \hat{\varphi}(\hat{x} + \hat{z}^0)e^{-ix\cdot\hat{x}}\, d\hat{x} = e^{ix\cdot\hat{z}^0} \int \hat{\varphi}(\hat{x} + \hat{z}^0)e^{-ix\cdot(\hat{x}+\hat{z}^0)}\, d\hat{x} = e^{ix\cdot\hat{z}^0}\varphi(x) \tag{5.117}$$

which agrees with (5.115).

If we think of $\{\delta_{\hat{z}^0}\}$ as a basis for $\{\mu\}$ as in (1.17)ff. then (5.116) indicates

$$(\hat{\varphi} * \mu)(\hat{x}) = \int \hat{\varphi}(\hat{x} + \hat{z})\, d\mu(\hat{z}) \tag{5.118}$$

in agreement with (5.112).

If we think of $\hat{\varphi} * \mu$ as a measure on $\mathbb{R}^n$ then

$$(\hat{\varphi} * \mu) \cdot \hat{h} = \int \hat{\varphi}(\hat{x} + \hat{z})\hat{h}(\hat{x})\, d\hat{x}\, d\mu(\hat{z}). \tag{5.119}$$

We want to show that if $d\mu$ is exponentially decreasing in a proper cone $\Gamma^{\mathbb{C}}$ in $\mathbb{C}^n$ containing $\hat{\rho}^0$ in its interior and $\varphi \in \mathcal{D}_M$ then $\hat{\varphi} * d\mu = \mathcal{O}(\lambda_M^{-s}(c|\hat{x}|))$ in a proper cone in $\mathbb{R}^n$ containing $\hat{\rho}^0$ in its interior.

Suppose $\hat{x}$ lies in a small, proper real cone $\hat{\Gamma}_0$ containing $\hat{\rho}^0$ in its interior and properly contained in $\Gamma^{\mathbb{C}} \cap \mathbb{R}^n$. There are two possibilities:

(1) $\hat{x} + \hat{z}^0$ is small. Then $\hat{z}^0$ is close to $-\hat{x}$ so $d\mu(\hat{z}^0)$ is small by (5.110).
(2) $\hat{x} + \hat{z}^0$ is large. If $\Im(\hat{x} + \hat{z}^0)$ is large then $d\mu(\hat{z}^0)$ is small by (5.109), while if $\Re(\hat{x} + \hat{z}^0)$ is large and $\Im(\hat{x} + \hat{z}^0)$ is small then $\hat{\varphi}(\hat{x} + \hat{z}^0)$ is small since $\varphi \in \mathcal{D}_M$.

To make things precise suppose that $\hat{\Gamma}_0$ is chosen with the property that if $\hat{x} \in \hat{\Gamma}_0$ and $|\hat{x} + \hat{z}^0| \le \frac{1}{2}|\hat{x}|$ then $\hat{z}^0 \in -\Gamma^{\mathbb{C}}$. For example,

$$\hat{\Gamma}_0 = \left\{\hat{x} : \quad |\Im\hat{x}| < \tfrac{1}{4}|\hat{x}|\right\}.$$

Thus $d\mu(\hat{z}^0)$ is exponentially small by (5.110). The contribution of all such points to the measure (5.117) is easily seen to be $\mathcal{O}(\exp(-c|\hat{x}|))$.

In case (2) $|\hat{x} + \hat{z}^0| \ge \frac{1}{2}|\hat{x}|$ so

$$|\varphi(\hat{x} + \hat{z}^0)| \le c\lambda_M^{-s}(|\Re(\hat{x} + \hat{z}^0)|)e^{c|\Im\hat{z}^0|} \tag{5.120}$$

(times unimportant factors); $\exp(c|\Im\hat{z}^0|)$ is compensated by the exponential

decrease of μ. If

$$|\Im \hat{z}^0| \geq c|\hat{x}|$$

then the exponential decrease of μ gives an additional factor of $\exp(-c'|\hat{x}|)$. On the other hand if $|\Im \hat{z}^0| \leq \frac{1}{8}|\hat{x}|$ then $|\Re(\hat{x} + \hat{z}^0)| \geq \frac{1}{8}|\hat{x}|$ so

$$\lambda_M(|\Re(\hat{x} + \hat{z}^0)|) \geq \lambda_M\left(\tfrac{1}{8}|\hat{x}|\right).$$

In all cases the integrand in (5.118) decreases like $\lambda_M^{-1}(c|\hat{x}|)$ in a real cone containing $\hat{\rho}^0$ in its interior. It is easy to modify these inequalities to derive the same for the integral itself.

We have proven

Theorem 5.14 *Definitions 1, 2, 3, 4 of analytic wave front set are equivalent.*

Remark 1 *The nonlinear Fourier transform definition of wave fronts was based on the quadratic nonlinear Fourier transform. The work of Banh [8] shows how to use higher order polynomials to obtain precise domains of analyticity for $n = 1$. It is hoped that such methods can be used to derive more precise results on the holomorphicity properties of Sato functions.*

Remark 2 *We have defined the wave front set in a microlocal fashion, meaning that it is local in both x space and in codirections. We shall discuss globalizing x below. As for globalization in codirection space, a simple compactness argument using the ideas of the proof of Theorems 5.11 and 5.13 shows that if the M wave front set of f (at $x = 0$) omits all rays in a convex cone $\hat{\Gamma}$ then, for all $\varphi \in \mathcal{D}_M$ of sufficiently small support, (5.75) holds for φf in place of f when $\Im \hat{y}$ lies in a conical neighborhood of $\hat{\Gamma}$.*

Definition 2 has an obvious extension to global $\hat{\Gamma}$. Definitions 3 and 4 also have natural modifications which are readily established.

Theorem 5.15 *Let Γ be a proper cone in $\mathbb{R}^n$. Suppose the analytic wave front set of f omits a conical neighborhood of $c(-\Gamma')$. Then f is holomorphic (locally) in the tube over Γ.*

Proof We use Definition 2. We shift the contour exactly as in our above proof that Definition 2 implies Definition 1, except that we now shift in β up to $\beta = \varepsilon|\hat{x}|$. The analyticity is clear.

Remark. *The proof of Theorem 5.15 used Definition 2 in an essential manner. Definition 1 can be used to prove that a function whose wave front set is empty is entire. However, the stronger statement involving convex cones $\hat{\Gamma}$ seems to require some analog of the nonlinear Fourier transform. In fact we shall meet a situation below (Φ wave front sets) for which we have analogs of Definition 1 but no form of nonlinear Fourier transform and for which we are unable to prove the analog of Theorem 5.15.*

One of the main applications of wave front sets is to partial differential equations. Let $P(x, D)$ be a partial differential operator of order m with analytic

coefficients defined near $x = 0$. Call $P_m(x, D)$ the principal part of $P(x, D)$, meaning the sum of terms of order m. The characteristic set of P is defined by

$$\text{Char } P = \{(x, \xi) \,|\, P_m(x, \xi) = 0\}. \tag{5.121}$$

One of the main results of the theory (see [98 vol. I]) is

$$WF(h) \subset [\text{Char } P \cup WF(Ph)] \tag{5.122}$$

where WF means the analytic wave front set. We shall show how to prove (5.122) using the nonlinear Fourier transform.

To simplify our proof let us assume that P has polynomial coefficients. The passage to analytic coefficients can be made using the Fourier representation of the holomorphic functions; we refer the reader to [54] where this representation is used in a similar manner. We shall also restrict our considerations to the quadratic nonlinear Fourier transform.

If f is a smooth function of a single variable y then the nonlinear Fourier transform (FT) of f' is given by

$$\begin{aligned}\int f'(y) e^{iy\hat{y} - \beta y^2}\, dy &= \int (-i\hat{y} + 2\beta y) f(y) e^{iy\hat{y} - \beta y^2}\, dy \\ &= \left(-i\hat{y} - 2i\beta \frac{d}{d\hat{y}}\right) F(\beta, \hat{y}).\end{aligned} \tag{5.123}$$

By iteration we see that

$$\text{nonlinear FT } (f^{(m)})(\hat{y}, \beta) = \left\{(-i\hat{y})^m + R\left(\beta, \hat{y}, \frac{d}{d\hat{y}}\right)\right\} F(\beta, \hat{y}). \tag{5.124}$$

We write R as an operator in $d/d\hat{y}$ whose coefficients are polynomials in $\beta, \hat{y}$ written on the left. We have

$$\begin{aligned}\text{degree }_{\hat{y}} R &< m \\ \text{degree }_{\beta} R &= m \\ \text{degree }_{\beta, \hat{y}} R &= m.\end{aligned} \tag{5.125}$$

We can regard $(-i\hat{y})^m + R$ as the coefficient of $\exp(iy\hat{y} - \beta y^2)$ in $(d^m/dy^m) \exp(iy\hat{y} - \beta y^2)$; we call $[(-i\hat{y})^m + R]$ the *nonlinear Fourier transform of* d^m/dy^m.

Let us write P in the form

$$P(x, D) = P_m(0, D) + [P_m(x, D) - P_m(0, D)] + Q(x, D) \tag{5.126}$$

where Q is of order $< m$. $[P_m(x, D) - P_m(0, D)]$ is of order m (it may be $\equiv 0$) but the coefficients of all $\partial^m / \partial x_1^{m_1} \dots \partial x_n^{m_n}$ are polynomials with no constant terms. We now examine the nonlinear Fourier transform of $P(x, D)$.

Remark. *In this proof we use the circular contraction (5.43) of the nonlinear Fourier transform.*

As in the computation of (5.124) we replace multiplicative factors of x_j by $-id/d\hat{x}_j$. We find

$$\text{nonlinear FT } [P_m(0,D)](\beta,\hat{x}) = P_m(0,i\hat{x}) + R_0\left(\beta,\hat{x},\frac{d}{d\hat{x}}\right) \tag{5.127}$$

where R_0 is a polynomial satisfying (5.125) (with $\hat{y}$ replaced by $\hat{x}$). $R_1 =$ nonlinear FT(Q) satisfies (5.125) with the improvements that degree $_{\beta,\hat{x}}R_1 < m$. Finally $R_2 =$ nonlinear FT $[P_m(x,D) - P_m(0,D)]$ satisfies (5.125) with the additional proviso that any monomial in $\beta,\hat{x}$ of degree m has a (right) coefficient which is a polynomial in $\partial/\partial\hat{x}$ with no constant term.

Suppose that the $\hat{x}_1$ axis belongs to neither the characteristic set of P nor to the wave front set of Ph at $x = 0$. We call $f = \varphi h$ as usual. By Theorem 5.14 the nonlinear Fourier transform of Pf is exponentially decreasing in an angle Ω in the $\beta,\hat{x}$ space "lying over" a cone Γ containing the $\hat{x}_1$ axis.

Since the $\hat{x}_1$ axis is noncharacteristic for $P(x,D)$ we can assume, by making Γ smaller, if necessary, that

$$|P_m(0,i\hat{x})| \geq c|\hat{x}|^m \tag{5.128}$$

for $\hat{x} \in \Gamma$ since P_m is homogeneous. If the nonlinear Fourier transform of $P(x,D)$ were reduced to $P_m(0,i\hat{x})$ then clearly the exponential decrease of $G(\beta,\hat{x}) = P_m(0,i\hat{x})F(\beta,\hat{x})$ in Ω would imply the exponential decrease of F there. Our task is to show that $P_m(0,i\hat{x})$ dominates R_0, R_1, R_2 in a suitable sense so that the result persists for the actual nonlinear Fourier transform of Pf.

Why does $P_m(0,i\hat{x})$ dominate R_0, R_1, R_2? The R_j contain three operators:

(1) β
(2) $\hat{x}$
(3) $\frac{\partial}{\partial\hat{x}}$.

For angles Ω in which $\hat{x} \gg \beta$ inequality (5.128) shows that $P_m(0,i\hat{x})$ dominates all monomials in $\beta,\hat{x}$ which are of degree $\leq m$ except possibly monomials in $\hat{x}$ of degree m, which can occur only in R_2.

The crucial point is that the coefficients of the monomials in $\hat{x}$ of degree m involve the $\partial/\partial\hat{x}_j$ in a nontrivial (nonconstant) manner. The $\partial/\partial\hat{x}_j$ are "small" operators. This might seem puzzling at first because one usually thinks of differentiation as "large," especially when compared to multiplication. But f is a function with compact support so $\hat{f}$ and $\widehat{Pf}$ are entire functions of exponential type which can be made as small as we want by further diminishing the size of the support of f.

Call F, G the respective nonlinear Fourier transforms of f, Pf. We want to apply Bernstein-like inequalities to G in Ω. The classical Bernstein inequality

(see [18]) states that, if $\hat{u}(\hat{x})$ is an entire function of exponential type $\leq A$ in the single variable $\hat{x}$ bounded on the real axis, then

$$\sup |\hat{u}'(\hat{x})| \leq A \sup |\hat{u}(\hat{x})|. \tag{5.129}$$

The sup is taken over the real axis.

The *raison d'être* for inequalities of the form (5.129) is the fact that the Fourier transform of $d/d\hat{x}$ is multiplication by x which is bounded by A on support u. Thus we should expect similar inequalities for $\partial/\partial\hat{x}_j$ in Ω since x_j is small on the curves like γ which appear at the beginning of Section 5.2; it is from these γ that the structure of the nonlinear Fourier transform on Ω is derived.

Because $\partial/\partial\hat{x}_j$ are small operators the operator $P_m(0, i\hat{x})$ dominates R_2. Hence $P_m(0, i\hat{x})$ dominates $R = R_0 + R_1 + R_2$.

We are thus left to solve the equation

$$[P(0, i\hat{x}) + R(\beta, \hat{x}, \partial/\partial\hat{x})]F(\beta, \hat{x}) = G(\beta, \hat{x}). \tag{5.130}$$

The solution is given by

$$F(\beta, \hat{x}) = \{1 + [P(0, i\hat{x})]^{-1} R(\beta, \hat{x}, \partial/\partial\hat{x})\}^{-1} [P(0, i\hat{x})]^{-1} G(\beta, \hat{x}). \tag{5.131}$$

We can expand the term $\{\quad\}^{-1}$ in a geometric series (Neumann series). There remains the problem of estimating the resulting function and showing that it is still exponentially decreasing in Ω. For this we need the precise form of the Bernstein inequalities.

We shall not give the details as they are completely straightforward and because the result is wellknown.

Our proof has the advantage of being essentially conceptual.

We wish to explain how the Kashiwara–Kawai–Hörmander (KKH) theorem [103] fits into our theory. The KKH theorem asserts that if f vanishes (locally) for $x_1 \leq 0$ and the analytic wave front set of f omits either the positive or negative $\hat{x}_1$ axis then f actually vanishes for $x_1 \leq \varepsilon$ for some $\varepsilon > 0$.

In the simplest case of $n = 1$ this means that if $f(x) = 0$ for $x \leq 0$ and f is holomorphic in the upper or lower half-plane for $\Re x \leq \eta$ then f vanishes for $x \leq \eta$.

Let $n > 1$. To get an intuition for the result let us examine the global analog of the KKH theorem. Thus suppose f is globally defined and has a classical Fourier transform $\hat{f}$ which is exponentially decreasing in a cone containing the positive $\hat{x}_1$ axis (the negative axis is treated in a symmetric way) and f vanishes for $x_1 \leq 0$. Then for each fixed ${}_1\hat{x}^0 = (\hat{x}_2^0, \ldots, \hat{x}_n^0)$ the function $\hat{f}(\hat{x}_1, {}_1\hat{x}^0) = \mathcal{O}(\exp(-b\hat{x}_1))$ as $\hat{x}_1 \to +\infty$. This means that the one-dimensional Fourier transform $\tilde{f}(x_1, {}_1\hat{x}^0)$ of $\hat{f}(\hat{x}_1, {}_1\hat{x}^0)$ is holomorphic in $\Im x_1 \leq b'$. (b' is independent of ${}_1\hat{x}^0$.)

The projection–slice theorem asserts that

$$\tilde{f}(x_1, {}_1\hat{x}^0) = \mathbf{R}[f \exp(i\, {}_1x \cdot {}_1\hat{x}^0)](x_1). \tag{5.132}$$

The Radon transform is taken on the planes $x_1 = \text{const}$. Hence $\tilde{f}(x_1, {}_1\hat{x}^0;)$ vanishes for $x_1 \leq 0$ for any ${}_1\hat{x}^0$. The holomorphicity of $\tilde{f}(x_1, \hat{x}^0)$ implies the vanishing of $\tilde{f}(x_1, {}_1\hat{x}^0)$ for all x_1, ${}_1\hat{x}^0$. Thus, in fact, $f \equiv 0$.

This method cannot work for the local problem for $n > 1$ without serious modification. To understand this point, suppose support $f \subset \{|x_j| \leq 1 \text{ all } j\}$ and $f = 0$ for $x_1 \leq 0$. The Radon transform $\mathbf{R}f(x_1)$ depends on the cut-off p near the boundary of the unit cube. It is unlikely that we could make $\mathbf{R}f(x_1)$ holomorphic in $\Im x_1 > 0$.

In terms of Fourier transform, $\hat{f}$ (or more precisely $\hat{\varphi}_M * \hat{f}$, where $\hat{\varphi}_M \in \mathcal{D}_M$) is not exponentially decreasing as $\hat{x}_1 \to \infty$. Although for each M it decreases faster than $\lambda_M^{-1}(\hat{x}_1)$, the multiplier φ_M depends on M so it is not a single function which has this uniform decrease.

Instead of $\varphi_M f$ we could study $\hat{f}$ as given by AU space theory. By Definition 4 of analytic wave front set $\hat{f}$ can be chosen to be exponentially decreasing in a complex cone containing the positive $\hat{x}_1$ axis. But we cannot restrict $\hat{f}$ to a line so this method does not work without serious modification.

The problem in these approaches is that restriction of $\hat{f}$ to a line depends on cut-offs of f or on values of $\hat{f}$ where $\hat{f}$ does not exist. Thus we need a modification of $\mathbf{R}f$ to integrals of f over contours γ which do not meet the boundary of the unit cube. Instead of lines $x_1 = \text{const}$. we could use (e.g. $n = 2$) $\gamma = \gamma_\eta$ of the form illustrated in Figure 5.5.

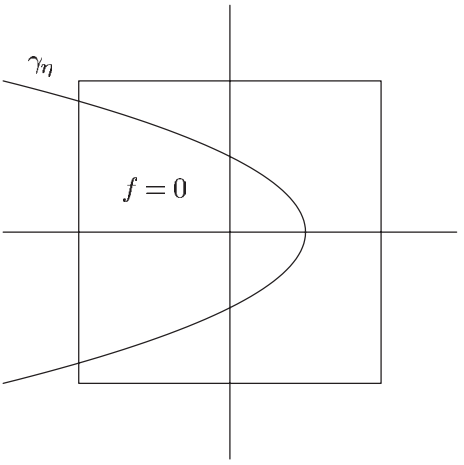

Figure 5.5

Such γ_η have been studied in connection with Holmgren's uniqueness theorem in Section IX.9 of FA. In particular $\hat{\delta}_\gamma$ is large only in the cone where $\hat{f}$ (Definition 4) is small. Thus $\hat{\delta}_{\gamma_\eta}$ depends analytically on η and vanishes with all its derivatives at $\eta = 0$.

Moreover linear combination of polynomials times δ_γ are dense. This gives a proof of KKH.

We shall not present any more details as the interested reader can complete this proof or refer to the proof in [103].

Up to now we have discussed linear wave fronts. We wish to extend these ideas to the nonlinear case.

We have discussed two tentative definitions for nonlinear wave front sets in (5.53)ff. As in the case of linear wave fronts we coordinate the definitions by requiring (5.53) for $\hat{p}$ in a suitable neighborhood of $\hat{p}^0$ and for f replaced by any multiple φf with $\varphi \in \mathcal{D}$ of small support.

It is possible to sharpen the definition by replacing the exponential $it\hat{p} \cdot p$ by $i\hat{q} \cdot p$ and requiring (5.53) (for φf) for $\hat{q}$ in a neighborhood of $\{t\hat{p}\}$. It is also possible to define nonlinear analytic wave front sets in analogy to the four definitions given following (5.81) and to show their equivalence. We shall not enter in these matters as the theory is still in its infancy.

We mention one possible use of nonlinear analytic wave front sets. The KKH theorem can be extended, via change of variables, to the case when the hypersurface $x_1 = 0$ is replaced by any hypersurface which is regular near the origin. But when the hypersurface is singular then the method breaks down.

Problem 5.6 *Extend the KKH theorem to hypersurfaces* $\{\hat{p} \cdot \hat{p}^0 = 0\}$ *which have singularities at the origin. Also give a nonlinear version of (5.122).*

One essential aspect of wave front sets is that it studies f in terms of "one-dimension constructs," i.e. the restriction of $\hat{f}$ to rays. As we have remarked this is essentially dual to the hyperplane Radon transform of f. Theorem 5.15 shows that suitable behavior on "enough" one-dimensional objects (rays) implies holomorphicity, which we regard as an n-dimensional property.

Another example of this phenomenon is the Korevaar–Wiegerinck theorem. We start with a formal power series f^0 whose restriction to (one-dimensional) lines through the origin has small coefficients (hence defines a holomorphic function, locally, on this line). Then f^0 extends to a holomorphic function f on a full neighborhood of 0.

We shall explain presently that f^0 (more precisely, the power series coefficients $\tilde{f}^0$ of f^0) can be regarded as the exotic CD of a solution of the Cauchy–Riemann equations with data at the origin. We can regard the smallness defined by holomorphicity of the restrictions of the power series coefficients of f^0 to a ray (in this case a line) as an analytic wave front condition on f^0. This type of analytic wave front sets uses the base of monomials rather than the Fourier basis which is standard for wave front sets. The Korevaar–Wiegerinck theorem states that if the monomial wave front set of $\tilde{f}^0$ is empty then the solution of

the exotic CP

$$\begin{aligned} \bar{\partial} f &= 0 \\ \mathrm{CD}(f) &= \tilde{f}^0 \end{aligned} \tag{5.133}$$

exists near the origin, i.e. $\tilde{f}^0$ is "admissible" CD for a solution of $\bar{\partial} f = 0$.

The Korevaar–Wiegerinck theorem can be regarded from a slightly different point of view. The antiholomorphic polynomials

$$\prod_{j=1}^{m} (\hat{x}_j - i\hat{x}_{m+j})^{\alpha_j} \quad (\alpha_j \geq 0) \tag{5.134}$$

for varying $\alpha = (\alpha_1, \ldots, \alpha_n)$ provide a basis for polynomials on the Cauchy–Riemann variety

$$V = \{\hat{x}_k + i\hat{x}_{m+k} = 0\} \tag{5.135}$$

in $\mathbb{C}^n$ $(n = 2m)$. According to the ideas promulgated in Chapters 3 and 4 this means that there is an exotic CP for $\bar{\partial}$ with data at the origin given by

$$\left\{ \prod_j \left(\frac{\partial}{\partial x_j} - i \frac{\partial}{\partial x_{j+m}} \right)^{\alpha_j} f_0 \right\} = \left\{ \prod_j \frac{\partial^{\alpha_j}}{\partial z^{\alpha_j}} f(0) \right\}. \tag{5.136}$$

The Korevaar–Wiegerinck theorem gives one-dimensional criteria for "potential CD" to be actual local CD.

For an exotic CP in which $Y = \mathrm{CS}$ has positive dimension $< m$ it seems that one would need a wave front set involving the Fourier transform in Y and monomials in $Y^{\perp}$ to determine which potential CD are actual CD.

Problem 5.7 *Find interesting exotic CPs in which the CS has positive dimension for which the admissibility of potential CD is governed by such Fourier–monomial wave fronts.*

Some related problems are studied in Section 9.3.

5.4 Microglobal analysis

All the wave front analysis that we have given up to now can be thought of as belonging to the realm of *microlocal* analysis. "Local" refers to local in x and "micro" to local in directions in $\hat{x}$. Microlocal analysis is an excellent tool in the study of problems for partial differential equations related to noncharacteristic hypersurfaces such as Holmgren's theorem as we have described above, and the noncharacteristic edge-of-the-wedge theorem discussed in Section 9.3.

However the characteristic CP does not lend itself to local analysis. For the heat equation

$$\frac{\partial f}{\partial t} - \frac{\partial^2 f}{\partial x^2} = 0 \tag{5.137}$$

there are solutions which vanish for $t \leq 0$. They can be eliminated by imposing a growth condition at ∞ in x. The most classical condition is

$$|f(t,x)| \leq ce^{cx^2}. \tag{5.138}$$

The precise result was proven by Täeklind [147]. Solutions satisfying

$$|f(t,x) \leq ce^{\phi(cx)}| \tag{5.139}$$

for suitable convex ϕ are uniquely determined by their CD if

$$\int \frac{\psi(\hat{x})}{1+|\hat{x}|^3}\, d\hat{x} = \infty. \tag{5.140}$$

ψ is the Legendre (Young) conjugate of ϕ

$$\psi(\hat{x}) = \max_x[(\hat{x} \cdot x - \phi(x)]. \tag{5.141}$$

This formulation is found in Section IX.9 of FA for general systems of linear partial differential equations with constant coefficients.

The main ingredient of the proof is the existence of averages of f on the planes $t =$ const. which depend quasianalytically on t. These averages can be thought of as attenuated Radon transforms on the spread $\{t =$ const.$\}$. Uniqueness is proven when the set of attenuations forms a spanning set for the space of CD.

It is shown [43] that the restriction on the x axis of any entire function f^0 is the CD for a solution f of the heat equation. f is certainly not unique unless we impose the growth conditions (5.139) and (5.140). However, if f is an entire function of such growth then there is an analog of the Cauchy–Kowalewski theorem as we have existence in addition to the uniqueness. This leads us to *microglobal analysis*.

The example of the heat equation as well as other examples in Chapter IX of FA indicate that the correct setting of the Cauchy–Kowalewski theorem is in spaces $\mathcal{H}(\phi)$ of entire functions g satisfying

$$g(x) = \mathcal{O}(e^{\phi(\alpha x)}) \tag{5.142}$$

for all $\alpha > 0$. ϕ is a smooth convex function which is larger than $|x|$ at infinity. Under mild assumptions on ϕ the space $\mathcal{H}(\phi)$ is closed under differentiation and multiplication by polynomials.

It is shown in Chapter V of FA that the space $\mathcal{H}'(\phi)$ is LAU (see also Section 1.5). The Fourier transform $\hat{\mathcal{H}}'(\phi)$ of $\mathcal{H}'(\phi)$ consists of entire functions $u(\hat{x})$ satisfying

$$u(\hat{x}) = \mathcal{O}(e^{\psi(c|\hat{x}|)}) \tag{5.143}$$

for some c. ψ is the conjugate of ϕ. By duality any $g \in \mathcal{H}(\phi)$ has a Fourier representation

$$g(x) = \int e^{ix\hat{x}} \frac{d\mu(\hat{x})}{k(\hat{x})} \tag{5.144}$$

where μ is a bounded measure on $\mathbb{C}^1$ and $k(\hat{x})$ dominates the right sides of (5.143) for all c. We shall often denote $d\mu/k$ by $\hat{g}$.

We denote by $\mathcal{E}(\phi)$ the space of C^∞ functions $h(x)$ satisfying

$$h^{(j)}(x) = \mathcal{O}(e^{\phi(\alpha x)}) \tag{5.145}$$

for all α, j. From the results of Chapter V of FA we can represent any $h \in \mathcal{E}(\phi)$ in the form

$$h(x) = \int e^{ix\hat{x}} \frac{d\mu(\hat{x})}{k_1(\hat{x})}. \tag{5.146}$$

Now $k_1(\hat{x})$ dominates all functions $\hat{v}$ in $\hat{\mathcal{E}}'(\phi)$. They satisfy

$$\hat{v}(\hat{x}) \leq c(1 + |\hat{x}|)^c e^{\psi(c\Im\hat{x})} \tag{5.147}$$

for some c.

The difference between k and k_1 occurs near the real axis. According to our ideas on wave front sets we should search for conditions on $h \in \mathcal{E}(\phi)$ so it should satisfy some of the requirements of $\mathcal{H}(\phi)$. As we have seen above, one way of accomplishing this is to replace h by multiples φh for a large enough set of φ so that we can somewhat recover h from $\{\varphi h\}$. We also want φ to be small at infinity so that we can form the Fourier transform $\widehat{\varphi h}$ in the usual way.

The natural φ to consider should satisfy

$$\varphi(x) = \mathcal{O}(e^{-\phi(c\Re x)+\phi(c\Im x)}).$$

For $\phi(x) = x^2$ such φ exist since

$$|e^{-x^2}| = e^{-(\Re x)^2+(\Im x)^2}.$$

However, they do not exist for essentially any other ϕ.

The problem we meet here is very similar to the problem we met in formulating a definition of (ordinary) analytic wave front sets: There are no analytic functions of compact support. We were able to overcome this difficulty by using nonquasianalytic classes and representing the space of analytic functions as the intersection of the spaces of nonquasianalytic functions.

In the present case we introduce the spaces $\mathcal{H}(\tilde{\phi}, \phi)$ where $\tilde{\phi}$ and ϕ are smooth convex functions. $\mathcal{H}(\tilde{\phi}, \phi)$ consists of all entire functions $\varphi(x)$ such that

$$\varphi(x) = \mathcal{O}(e^{-\phi(c\Re x)+\tilde{\phi}(c'\Im x)}). \tag{5.148}$$

In general $\tilde{\phi}$ is larger than ϕ. Under "reasonable" conditions $\mathcal{H}(\tilde{\phi}, \phi)$ is closed under differentiation and multiplication by polynomials. We define the topology in the usual manner. We call $\tilde{\phi}$ *admissible for* ϕ if $\mathcal{H}(\tilde{\phi}, \phi)$ is nontrivial. Since $\mathcal{H}(\tilde{\phi}, \phi)$ is invariant under complex translation and real scalar multiplication of x, once it is nontrivial the functions in it have no common zeros and it is generally dense in the space $\mathcal{H}$ of entire functions.

Problem 5.8 *Describe the set of* $\tilde{\phi}$ *which are admissible for* ϕ.

We denote by $\psi, \tilde{\psi}$ the respective conjugates of $\phi, \tilde{\phi}$. If $\varphi \in \mathcal{H}(\tilde{\phi}, \phi)$ then $\hat{\varphi}$ is an entire function satisfying

$$\hat{\varphi}(\hat{x}) = \mathcal{O}(e^{\psi(c_1 \Im \hat{x}) - \tilde{\psi}(c_2 \Re \hat{x})}). \tag{5.149}$$

Let $h \in \mathcal{E}(\phi)$. We say that the positive $\hat{x}_1$ axis does not belong to the ϕ wave front set of h if for every ϕ admissible $\tilde{\phi}$ and any $\varphi \in \mathcal{H}(\tilde{\phi}, \phi)$

$$\widehat{\varphi h}(\hat{x}) = \mathcal{O}(e^{-\tilde{\psi}(\gamma \hat{x})}) \quad \hat{x}_1 \to +\infty \tag{5.150}$$

for some γ.

Suppose $n = 1$ and that both the positive and negative $\hat{x}$ axis do not belong to the ϕ wave front set of h. Taking the inverse Fourier transform of (5.150) we find that φh is an entire function satisfying

$$(\varphi h)(x) = \mathcal{O}(e^{\tilde{\phi}(c_3 \Im x)}) \tag{5.151}$$

for all $\tilde{\phi}$ which are ϕ admissible. Since φ is entire and the set of φ has no common zero we conclude that h is entire.

For many interesting $\tilde{\phi}$ the minimum modulus results developed, for example, in [39] allow us to derive from (5.151)

$$h(x) = \mathcal{O}(e^{\tilde{\phi}(c_4 |x|)}). \tag{5.152}$$

If this is the case we say $\mathcal{H}(\tilde{\phi})$ is a *division space for* $\mathcal{H}(\tilde{\phi}, \phi)$. This inequality does not place $h \in \mathcal{H}(\tilde{\phi})$ because (5.142) is valid for some c_4. However, the c_4 in (5.152) can often be taken arbitrarily small if c in (5.148) is small so the c_2 in (5.149) is arbitrarily large. Hence we can often make c_4 in (5.152) arbitrarily small.

In any case the most we could hope for is

$$h \in \bigcap \mathcal{H}(\tilde{\phi}), \tag{5.153}$$

the intersection being taken over all ϕ admissible $\tilde{\phi}$ for which $\mathcal{H}(\tilde{\phi})$ is a division space for $\mathcal{H}(\tilde{\phi}, \phi)$. Let us denote the right side of (5.153) by $\mathcal{H}^*(\phi)$.

We have proven

Proposition 5.16 *Under the above assumptions, if the ϕ wave front set of $h \in \mathcal{E}(\phi)$ is empty then h is an entire function in $\mathcal{H}^*(\phi)$.*

Problem 5.9 *Calculate $\mathcal{H}^*(\phi)$. In particular when is $\mathcal{H}^*(\phi) = \mathcal{H}(\phi)$?*

Remark. *If we replace $\mathcal{E}(\phi)$ by $\mathcal{E}$ and $\mathcal{H}(\phi)$ by $\mathcal{H}$ we can think of ϕ as*

$$\phi(t) = \begin{cases} 0 & t \in [-1, 1] \\ \infty & |t| > 1. \end{cases} \tag{5.154}$$

This ϕ corresponds to $\psi(\hat{t}) = |\hat{t}|$. The only multipliers φ that we can take must be of compact support. This is in contradistinction to our previous condition which required that φ can be holomorphic. However, we can choose $\hat{\varphi}$ to satisfy (5.149) with $\tilde{\psi}(\hat{x}) = \log \lambda_M(\hat{x})$ for nonquasianalytic classes M. Our previous results show that the positive $\hat{x}$ axis does not belong to the ϕ wave front set of h if and only if it does not belong to the real analytic wave front set of h.

Thus the present process can be regarded as a direct analog of our previous process for ordinary analytic wave front sets.

Let us pass to $n > 1$. As usual we can restrict our considerations to $n = 2$ as all the essential distinctions between $n = 1$ and $n > 1$ are already present when $n = 2$.

We suppose that

$$\phi(x_1, x_2) = \phi_1(x_1) + \phi_1(x_2) \tag{5.155}$$

where ϕ_1 is the function of one variable which we have previously denoted by ϕ. Up to constants this is the same as

$$\phi(x_1, x_2) = \phi_1(|x|) \tag{5.156}$$

so the decomposition (5.155) is rotationally invariant.

We want to give meaning to the statement that the wave front set of $h \in \mathcal{E}(\phi)$ omits the positive $\hat{x}_1$ axis. We have noted the close relation between wave front sets and the attenuated Radon transform. Thus it makes sense to form

$$\mathbf{R}\varphi h(e_1, s) = \int h(s, x_2)\varphi(s, x_2)\, dx_2 \tag{5.157}$$

for each $\varphi \in \mathcal{H}(\tilde{\phi}, \phi)$ for all ϕ admissible $\tilde{\phi}$. (In this notation e_1 is the point in the Grassmannian representing the hyperplane $x_1 = 0$.)

We could now say that the positive $\hat{x}_1$ axis does not belong to the ϕ wave front set of h if the Fourier transform of $\mathbf{R}\varphi h(e_1, s)$ decreases on the positive $\hat{x}_1$ axis like $\exp(-\psi(c\hat{x}_1))$ for all such φ. We have seen in the case of the ordinary wave front set that this definition is not sufficient; we need to assume the same is true in some neighborhood of the positive $\hat{x}_1$ axis. How large should this neighborhood be?

We think of the wave front set's missing a region as a "regularity" property. Thus as in the case of ordinary wave front sets we should like to know that if the positive $\hat{x}_1$ axis does not belong to the ϕ wave front set of h and $u \in \mathcal{H}(\phi)$ then the same is true of the uh.

Theorem 5.17 *Suppose the ϕ wave front set of h omits an angle around the positive $\hat{x}_1$ axis. Then the same is true of uh for any $u \in \mathcal{H}(\phi)$.*

Proof We must examine the Fourier transform of $\mathbf{R}(\varphi uh)(e, s)$ for any $\varphi \in \mathcal{H}(\tilde{\phi}, \phi)$ and for e in a neighborhood of $(1, 0)$. The projection–slice theorem tells us that this Fourier transform is the restriction of $\widehat{\varphi uh}$ to the line through e.

Formally

$$\widehat{\varphi uh} = \hat{u} * \widehat{\varphi h}. \tag{5.158}$$

The meaning of this convolution is explained in the discussion preceding Theorem 5.14. There is no difficulty in extending that argument to prove Theorem 5.17 and also

Theorem 5.18 *Let $h \in \mathcal{E}(\phi)$. Suppose we can find a representative measure $\hat{h} \in \hat{\mathcal{E}}(\varphi)$ for the Fourier transform of h such that*

$$\hat{h}(\hat{x}) = \mathcal{O}(e^{-\psi(c\hat{x})}) \tag{5.159}$$

for $\hat{x}$ in a proper cone in $\mathbb{C}^n$ containing the positive $\hat{x}_1$ axis in its interior. Then the ϕ wave front set of h omits the positive $\hat{x}_1$ axis.

We are now in a position to formulate our main result on nonlinear wave front sets.

Theorem 5.19 *If $h \in \mathcal{E}(\phi)$ has an empty ϕ wave front set then $h \in \mathcal{H}^*(\phi)$.*

The proof of Theorem 5.19 goes along the same lines as the proof of Theorem 5.15.

6

THE PARAMETRIC RADON TRANSFORM

In this chapter we study various parametric Radon transforms of functions f on $\mathbb{R}^n$. They are defined as integrals of f over the images of a manifold Λ under a family $\mathbf{G}$ of maps $\Lambda \to \mathbb{R}^n$; the integrals are taken with respect to a fixed measure on Λ. The parametric Radon transform often leads formally to two types of differential equations, called the *John equations*. In Section 6.1 we construct these equations; we also give criteria for making these formalities rigorous.

We prove in various cases that the John equations characterize spaces of certain classical parametric Radon transforms. This is carried out by means of the Fourier transform in Section 6.2 and directly in Section 6.3. These classical Radon transforms are closely related to the general linear group. In Section 6.4 we examine analogous questions for some other classical Lie groups.

In Section 6.2 we give an example of the use of the Watergate method (Section 1.4 and Chapter 4) to give an explicit solution of the John equations in a case which is of interest to magnetic resonance imaging.

6.1 The John and invariance equations

Let us clarify the concept "parametric Radon transform." We begin with an abstract version and later we shall apply our ideas to the usual Radon transform and some ramifications.

We start with a manifold Λ with some fixed measure μ and a manifold $\mathbf{G}$ of mappings of Λ into $\mathbb{R}^n$ which depend continuously on $\mathbf{g} \in \mathbf{G}$. The *parametric Radon transform* of a function f on $\mathbb{R}^n$ is defined as

$$F(\mathbf{g}, b) = \int_\Lambda f(\mathbf{g}\lambda + b)\, d\mu(\lambda) \tag{6.1}$$

for $b \in \mathbb{R}^n$.

Remark. *For the present we regard this integral in a formal sense. (The meaning of the integral is discussed below.) It may not exist for some values of* $\mathbf{g}$ *even if* f *has compact support. In particular it may happen that* $\mathbf{g}\Lambda = \{0\}$.

We can think of $\mathbf{G}$ as a sort of Grassmannian and b as a "spread parameter," although when $\mathbf{g}\Lambda$ is linear the actual spread parameter depends only on the component of b in $(\mathbf{g}\Lambda)^\perp$.

Let f be a function on $\mathbb{R}^n$; we lift f to the function

$$F_\lambda(\mathbf{g}, b) = f(\mathbf{g}\lambda + b) \tag{6.2}$$

on $\mathbf{G} \times \Lambda \times \mathbb{R}^n$.

In this chapter we search for conditions on a function $F(\mathbf{g}, b)$ in order that it be of the form (6.1). We shall also study what happens when we vary μ. In particular we shall examine the possibility of determining both μ and f from F. We begin our analysis of (6.1) when μ is fixed.

There are three natural slice (coarse grain) decompositions of the set of functions F:

(1) Fix $\mathbf{g}$.
(2) Fix λ.
(3) Fix $\hat{b}$ = dual variable to b.

Decomposition (3) is meaningful in the context of a type of Fourier transform on $\{F\}$ which we shall introduce below.

Each of the decompositions leads to a way of determining some structure of the set of F.

(1) Fixing $\mathbf{g}$ leads naturally to forming the Fourier transform in b.
(2) Having taken the Fourier transform in b we can write (formally)

$$F_\lambda(\mathbf{g}, b) = \int \hat{f}(\hat{b}) e^{i(\mathbf{g}\lambda + b)\cdot \hat{b}}\, d\hat{b}.$$

Suppose that $\mathbf{G}$ is a linear group acting on the linear space Λ. Then we can write

$$\mathbf{g}\lambda \cdot \hat{b} = \mathbf{g} \cdot \hat{b}\lambda'.$$

(We think of $\mathbf{g}$ as a matrix and $\hat{b}, \lambda$ as column vectors; $\hat{b}\lambda'$ is the matrix $(\lambda_i \hat{b}_j)$.) This suggests that we introduce a "dual variable" $\hat{\mathbf{g}}$ so that the above integral becomes

$$\int \hat{f}(\hat{b}) \delta_{\hat{\mathbf{g}} = \hat{b}\lambda'} e^{i\mathbf{g}\cdot\hat{\mathbf{g}} + ib\cdot\hat{b}}\, d\hat{b}\, d\hat{\mathbf{g}}. \tag{6.3}$$

$\mathbf{G}$ is a matrix group but $\mathbf{G}$ may not be a linear space. If $\mathbf{G}$ is a linear space, e.g. $\mathbf{G}$ is the additive group of the Lie algebra of a linear group, then $\hat{\mathbf{g}}$ is the usual linear dual variable.

The expression (6.3) suggests that any F of the form (6.1) can be represented as the Fourier transform in the variables $\hat{\mathbf{g}}, \hat{b}$ of a measure (precisely $\hat{f}(\hat{b})\, d\hat{b}$) on the set

$$V^1 = \bigcup_{\lambda, \hat{b}} \{\hat{\mathbf{g}} = \hat{b}\lambda'\} = \bigcup_{\lambda} V_\lambda. \tag{6.4}$$

(3) It is natural to think of V^1 as a fiber space over the base Λ, the fiber over λ being V_λ. We reverse our position and consider V^1 as a fiber space over $\hat{B} = \{\hat{b}\}$. For fixed $\hat{b}$ the set of $\hat{\mathbf{g}} \in V^1$ is $\{\hat{b}\lambda'\}_{\lambda \in \Lambda}$. Since the functions we deal with are functions $f(b)$ we might expect that the space $\mathcal{F}^0$ of functions on V^1 which correspond to functions $f(b)$ satisfies conditions which guarantee that $\mathcal{F}^0$ restricted to the fiber over $\hat{b}$ is one dimensional. This could be accomplished

if, for example, the functions in $\mathcal{F}^0$ satisfy, for each $\hat{b}$, a system of differential equations in λ whose solution space is one dimensional.

We can understand this "oversize" nature of V^1 from another point of view. Note that the construction of V^1 is independent of the measure $d\lambda$; any measure on Λ would lead to the same V^1. Thus the set of Fourier transforms on V^1 should correspond to the tensor product of functions on b with measures on Λ. This is consistent with the fact that $V^1 = \{\hat{b}\lambda'\}_{\lambda,\hat{b}}$. Picking one-dimensional subsets of functions on $\{\hat{b}\lambda'\}_\lambda$ corresponds by "duality" to choosing a measure $d\mu(\lambda)$ in (6.1).

The simplest measure is $d\lambda$. It is characterized up to scalar multiplication by the Haar property. Using (6.1) this means

$$F(\mathbf{g}, b) = F(\mathbf{g}, b + \mathbf{g}\lambda^0)$$

for any λ^0. By (6.3) this implies

$$\begin{aligned} F_\lambda(\mathbf{g}, b) &= \int \hat{f}(\hat{b}) e^{i[\mathbf{g}(\lambda+\lambda^0)+b]\cdot\hat{b}}\, d\hat{b} \\ &= F_{\lambda+\lambda^0}(\mathbf{g}, b). \end{aligned} \tag{6.5}$$

Thus

$$F_\lambda \text{ is independent of } \lambda.$$

Putting things in other terms, the Fourier transform of F_λ in the variables $\mathbf{g}, b$ which by (6.3) can be thought of as $\hat{f}(\hat{b})\delta_{\hat{\mathbf{g}}=\hat{b}\lambda'}$ is independent of λ on the fiber $\{\hat{\mathbf{g}} = \hat{b}\lambda'\}_{\lambda'}$. This shows how the choice of measure μ cuts down $\mathcal{F}^0\big|_{\{\lambda'\hat{b}\}}$ to a one-dimensional space.

We shall discuss other choices of $d\mu$ below.

We can often make contact with V through differential equations when $\mathbf{G}$ is a connected linear Lie group; we denote the Lie algebra of $\mathbf{G}$ by $\mathfrak{g}$. By (6.2) F_λ is constant on the set $M_\lambda(c)$ of $\mathbf{g}, b$ defined by

$$M_\lambda(c) = \{\mathbf{g}\lambda + b = c\}.$$

This means that for any $\mathbf{g}, \mathbf{g}^1, b, \lambda$

$$F_\lambda(\mathbf{g}, b) = F_\lambda[\mathbf{g}^1, (\mathbf{g} - \mathbf{g}^1)\lambda + b]. \tag{6.6}$$

We want to replace (6.6) by a differential equation. To this end we must distinguish between an additive or multiplicative matrix group. If $\mathbf{G}$ is an additive group then $\mathfrak{g} = \mathbf{G}$ and we can set $\mathbf{g}^1 = \mathbf{g} + \epsilon E$ where $E \in \mathfrak{g}$. If $\mathbf{G}$ is a multiplicative group then we set $\mathbf{g}^1 = \mathbf{g}(1 + \epsilon E)$ so that

$$\mathbf{g}^1 - \mathbf{g} = \begin{cases} \epsilon E & \text{additive} \\ \epsilon \mathbf{g} E & \text{multiplicative.} \end{cases}$$

Let $\epsilon \to 0$. From (6.6) there results

$$\overrightarrow{\mathbf{J}}_\lambda F_\lambda \equiv \left[\partial_{\mathbf{g}}^E - \partial_b^E \cdot \lambda\right] F_\lambda = 0. \tag{6.7}$$

$\partial_{\mathbf{g}}^E$ represents differentiation in the $\mathbf{g}$ variables in the direction E and $\partial_b^E \cdot \lambda$ represents the same differentiation in b times appropriate λ variables. To clarify the meaning of $\partial_b^E \cdot \lambda$ we write E as a linear combination of matrix units E_{jk} which are defined by

$$(E_{jk})_{j'k'} = \begin{cases} 1 & \text{if } j = j', k = k' \\ 0 & \text{otherwise.} \end{cases}$$

We can write $\partial_b^E \cdot \lambda$ as the corresponding linear combination of the $\partial_b^{E_{jk}} \cdot \lambda$. In the additive case

$$\partial_b^{E_{jk}} \cdot \lambda = \lambda_k \frac{\partial}{\partial b_j}.$$

For multiplicative groups

$$\partial_b^{E_{jk}} \cdot \lambda = \lambda_k \sum_l g_{lj} \frac{\partial}{\partial b_l}.$$

If we use the Lie algebra coordinates in $\mathbf{G}$ near the identity then the equations (6.7) for the multiplicative group $\mathbf{G}$ at $\mathbf{g}$ = identity and for the additive group $\mathfrak{g}$ are the same.

The problem with (6.7) is that these equations depend on λ. We search for an *enveloping equation*, meaning a system of differential equations

$$\overrightarrow{\mathbf{J}} F = 0 \tag{6.8}$$

where $\overrightarrow{\mathbf{J}}$ is an operator in the variables $\mathbf{g}, b$ and the solutions of (6.8) essentially coincide with the linear combinations (integrals over λ) of solutions of the equations $\overrightarrow{\mathbf{J}}_\lambda F_\lambda = 0$. (We say "essentially" because there may be a "lower dimensional" set of solutions of (6.8) which are not so represented.) Thus the single enveloping equation (6.8) is essentially equivalent to the whole system (6.7).

In order to understand the concept of enveloping equation let us describe one situation in which we can explain how to construct it. Let $P_j(D)$ be a finite number of systems of partial differential operators with constant coefficients. The fundamental principle (Section 1.4) relates each P_j to an algebraic (multiplicity) variety V_j such that solutions of $P_j f = 0$ are represented as Fourier transforms of measures supported on V_j. (We ignore the multiplicities in this heuristic description.) Then Fourier integrals over the variety

$$V = \bigcup V_j$$

are sums of solutions of the individual $P_j f_j = 0$. Since V is an algebraic variety it defines a system of equations $\mathbf{P}(D)f = 0$. This is the enveloping equation of the $P_j f = 0$.

Sometimes we can find an enveloping equation for an infinite number of equations; (6.8), when it exists, is such an example.

When a system of equations has an enveloping equation we call the system *envelopable*.

When the $\{P_\alpha\}$ are infinite in number the usual structure of an envelopable system $\mathbf{P}$ of $\{P_\alpha\}$ is that the algebraic varieties V_α form a parametrization of a "large part" V^1 of the algebraic variety V which defines $\mathbf{P}$. In most of the examples we treat V^1 is actually a Zariski open subset of V, meaning $V = V^1 \cup V^0$ where V^0 is an algebraic subvariety of V of lower dimension.

This description of envelopable systems is meaningful within the context of AU spaces (Section 1.4). But for spaces like $\mathcal{S}$ we have to deal with the real parts of V and V_α. This generally causes difficulties that do not appear in the AU theory.

The simplest example is $V^{\mathbb{R}} = \mathbb{R}^n$ and $\{V_\alpha^{\mathbb{R}}\} = \{\text{line through origin}\}$. The equation $P_\alpha f_\alpha = 0$ related to V_α defines the hyperplane spread associated to $V_\alpha^\perp$; P_α is formed by a basis for the tangential derivatives to $V_\alpha^\perp$. The equation related to $V^{\mathbb{R}}$ is

$$0f = 0$$

(every f is a solution). We know (Section 2.1) that every function is a generalized integral (allowing derivatives) of the $\mathbf{R}_\mathbf{g}^* \mathbf{R}_\mathbf{g} f$ which are spread functions for $\mathbf{g}$.

The Whittaker–Bateman representations of harmonic functions (Section 1.5) are assertions that the Laplacian is the enveloping equation of the first-order equations $P_\alpha f_\alpha = 0$ defined by the lines or planes in the complex cone $\sum \hat{x}_j^2 = 0$.

We call (6.8) the (first) *John equation* as they generalize the equations introduced by John [99].

We have shown in equation (6.5) how to interpret the invariance of the Haar measure $d\lambda$. We write the infinitesimal form of (6.5) as

$$\overrightarrow{\mathbf{I}} F = 0. \tag{6.5*}$$

This is called the *second John equation* or the *invariance equation*.

Remark. *It seems that for measures μ other than $d\lambda$ there are equations $\overrightarrow{\mathbf{I}}^\mu$ which replace $\mathbf{I}$ and essentially characterize "attenuated" Radon transforms (6.1). We discuss some of these in (6.49)ff.*

The essential content of this chapter is the verification of some cases of

Ansatz. *Any suitable solution of the John and invariance equations is a parametric Radon transform (6.1) with $\mu = d\lambda$.*

We now assume that the John equation has constant coefficients and that V is the variety associated to it as in Section 1.4. By the fundamental principle any

solution F of the John equations can be expressed as the Fourier transform of a measure ξ on V. We want to write

$$\xi = \int d\xi_\lambda(\hat{b})\, d\nu(\lambda)$$

where support $\xi_\lambda \subset V_\lambda$. We shall see that this would verify the ansatz.

However, we are faced with two problems:

(1) ξ may not be smooth so the meaning of such a decomposition is vague.
(2) ξ may have mass on $V - V^1$. How do we get rid of such mass?

In order to overcome these difficulties we need to make the structures more precise.

We have to distinguish two theories which depend on the spaces of f:

(a) AU spaces.
(b) Spaces whose Fourier transform is defined on $\mathbb{R}^n$ rather than $\mathbb{C}^n$.

(a) We suppose that F lies in the dual $\mathcal{W}'$ of an AU space $\mathcal{W}$. The fundamental principle (Section 1.4) exhibits F as the Fourier transform of a measure ξ on V which is obtained from the Hahn–Banach theorem using the fact that the norms in $\hat{\mathcal{W}}$ are sup norms. Now, $\hat{\mathcal{W}}$ is a space of entire functions. Cauchy's formula for the derivative of a holomorphic function shows how to bound norms involving the sup of the function and of its first N derivatives by sup norms or even by distribution norms of any order (assuming some minor regularity of the AU structure). Since ξ is an element of the dual of $\hat{\mathcal{W}}$ the Hahn–Banach theorem produces a representation of ξ as an element of the dual of a distribution space; hence ξ can be chosen with arbitrary (finite) smoothness.

Of course this is meaningful only on the complement of the singular part of V. The singular part V^s of V is an algebraic subset. It is standard in AU theory that the complement of an algebraic subset is sufficient (see Section 1.4).

Thus ξ can be chosen to be a smooth function on $V - V^s$.

If $V - V^1$ is an algebraic subset of V then we can assume (again by sufficiency) that

$$\text{support}\, \xi \subset V^1 - V^s.$$

We have assumed that the mappings $\lambda \to \mathbf{g}\lambda + b$ depend continuously on $\mathbf{g} \in \mathbf{G}$. In the examples we study below the V_λ provide a smooth fibration of V^1. It is clear that we can decompose any smooth measure on $V^1 - V^s$ into an integral of measures supported by the fibers. Moreover any integral of the form

$$F_\lambda(\mathbf{g}, b) = \int_{V_\lambda} e^{i\mathbf{g}\cdot\lambda'\hat{b} + ib\cdot\hat{b}}\, d\xi_\lambda(\hat{b})$$

satisfies (6.6) and hence satisfies $\overrightarrow{\mathbf{J}}_\lambda F_\lambda = 0$. Also such an F_λ is of the form $f_\lambda(\mathbf{g}\lambda + b)$ for some function f_λ.

We have proven

Theorem 6.1 *Suppose F belongs to the dual of an AU space $\mathcal{W}$ and $V - V^1$ is a Zariski open subset of V. Suppose moreover that the V_λ form a smooth fibration of V^1. Then the ansatz is valid, meaning that any solution F of the John equation $\overrightarrow{\mathbf{J}}(D)F = 0$ is an integral of functions F_λ where $F_\lambda \in \mathcal{W}'$ and $\overrightarrow{\mathbf{J}}_\lambda(D)F_\lambda = 0$. Moreover*

$$F_\lambda(\mathbf{g}, b) = f_\lambda(\mathbf{g}\lambda + b)$$

where f_λ is a function on $\mathbb{R}^n$.

Remark. *The equations for F_λ are first order so, in conformity with our general ideas, all F_λ for fixed λ have the same level sets $M_\lambda(c)$ (which is clear from the representation of F_λ as $f_\lambda(\mathbf{g}\lambda + b)$) and so can be considered spread functions for the "geometric" spread M_λ whose leaves are $M_\lambda(c)$.*

(b) Suppose for simplicity that $f \in \mathcal{S}$. Unfortunately F cannot belong to $\mathcal{S}$ because F satisfies the constant coefficient system $\overrightarrow{\mathbf{J}}(D)F = 0$ so if $F \in \mathcal{S}$

$$\overrightarrow{\mathbf{J}}(i\hat{x})\hat{F}(\hat{x}) = 0.$$

This means that support $\hat{F} \subset V^{\mathbb{R}}$, which is impossible for a function ($\not\equiv 0$) which is smooth if V is a nontrivial algebraic variety (since V does not contain an open set).

Although F is not in $\mathcal{S}$, it is bounded. F is constant along the sets M_λ where $\mathbf{g}\lambda + b$ is constant. It turns out in the examples we study that F belongs to $\mathcal{S}$ in the directions "orthogonal" to $\{M_\lambda\}$. This allows us to understand the Fourier transform $\hat{F}$ of F in the variables of $\mathfrak{g} \times \{b\}$. The support of $\hat{F}$ lies in the real points $V^{\mathbb{R}}$ of V and $\{b\}$ is a cross-section of the sets $M_\lambda \cap (V^1)^{\mathbb{R}}$.

The main difficulty we shall face is the elimination of the points in $V^{\mathbb{R}} - (V^1)^{\mathbb{R}}$ from the support of $\hat{F}$. This is possible because F is small in the directions "orthogonal" to $\{M_\lambda\}$.

Finally we shall use the second John equation to determine the unique f whose parametric Radon transform is F, thereby inverting the parametric Radon transform.

Let us illustrate these ideas for the examples given in Chapter 1.

The usual parametric l plane Radon transform is defined by

$$J^l f(a^1, \ldots, a^l; b) = \int f(a^1\lambda_1 + \cdots + a^l\lambda_l + b)\, d\lambda \tag{6.9}$$

where $a^1, \ldots, a^l$ and b are arbitrary points in $\mathbb{R}^n$ thought of as column vectors and $(\lambda_1, \ldots, \lambda_l)$ is a variable in $\mathbb{R}^l$. (The reader should be careful not to confuse J^l, which is a map, with $\overrightarrow{\mathbf{J}}_\lambda$ or $\overrightarrow{\mathbf{J}}$, which are equations.) In particular

$$J^1 f(a, b) = \int f(a\lambda + b)\, d\lambda.$$

We denote the $n \times l$ matrix $(a^1, \dots, a^l)$ by $\boxed{a}$. Thus (6.10) can be written in the form

$$J^l f\left(\boxed{a}, b\right) = \int f\left(\boxed{a}\lambda + b\right)\, d\lambda. \tag{6.9*}$$

$\boxed{a}$ is a variable in the additive group of $n \times l$ matrices and λ and b are column matrices (vectors). Of course for things to be well defined we need rank $\boxed{a} = l$.

Note that when $a^1, \dots, a^l$ are orthonormal then $d\lambda = d\lambda_1 \dots d\lambda_l$ is the euclidean measure on the plane $\mathbf{g} = \{\boxed{a}\lambda\}$. (We sometimes denote this plane by $\boxed{a}$.) In the orthonormal case

$$J^l f(\boxed{a}, b) = \mathbf{R}^l f(\mathbf{g}, b). \tag{6.10}$$

This means that the parametric Radon transform contains all the information that is contained in the nonparametric Radon transform.

How is (6.10) to be modified when the a^j are not orthonormal? Suppose they are linearly independent. We change bases on $\boxed{a}$ from $\{a^j\}$ to an orthonormal basis $\{v^j\}$. We have to compute the Jacobian of the transformation.

In case $l = 1$ the Jacobian is clearly $|a|^{-1}$. In general the Jacobian is the same as the measure of the set of λ satisfying

$$|\boxed{a}\lambda| \leq 1$$

which can be written as

$$\begin{aligned} 1 &\geq |\boxed{a}\lambda|^2 \\ &= (\boxed{a}\lambda, \boxed{a}\lambda) \\ &= ((\boxed{a}^t\boxed{a})^{1/2}\lambda, (\boxed{a}^t\boxed{a})^{1/2}\lambda). \end{aligned}$$

Since $(\boxed{a}^t\boxed{a})^{1/2}$ is a square matrix the Jacobian is $\det(\boxed{a}^t\boxed{a})^{-1/2}$.[1]

This shows that $J^l f(\boxed{a}, b)$ is homogeneous in the sense

$$J^l f(\boxed{a}, b) = \det(\boxed{a}^t\boxed{a})^{-1/2}\mathbf{R}^l f(\mathbf{g}, b). \tag{6.10*}$$

Of course the Jacobian is infinite when the a^j are not linearly independent so the meaning of $J^l f$ and the equation it satisfies at such points is unclear. We shall show (Theorem 6.2 below) that $J^l f$ is locally integrable when $n > l$ and that the John equations are satisfied in the distribution sense when $n \geq l + 2$.

Remark. *In Chapter 2 we noted that the moment conditions played an essential role in the study of* $\mathbf{R}^{n-1}$. *Their counterpart for* J^{n-1} *is the John equations. We shall show below directly, i.e. without using* f, *that the John equations are equivalent to the moment conditions with the homogeneity (6.10*) adjoined.*

[1] This argument was pointed out to me by Sinai Robins.

We show how to apply our above ideas. The additive group $\boxed{A} = \{\boxed{a}\}$ is generated by the matrix units. Thus the equations $\overrightarrow{\mathbf{J}}_\lambda(D)F_\lambda = 0$ of (6.7) become, for $E = (E_{ij})$ a matrix unit,

$$\frac{\partial F_\lambda}{\partial a^i_j} = \lambda_i \frac{\partial F_\lambda}{\partial b_j}. \tag{6.11}$$

The compatibility condition in $\left(\boxed{\hat{a}}, \hat{b}\right)$ (which defines the enveloping equation) for the equations in λ

$$\hat{a}^i_j = \lambda_i \hat{b}_j \quad \text{for all j}$$

is clearly

$$\hat{a}^i_j \hat{b}_k = \hat{a}^i_k \hat{b}_j$$

which is the statement

$$\hat{a}^i \text{ is parallel to } \hat{b}.$$

The John equation is the Fourier transform of compatibility:

$$\frac{\partial^2 F}{\partial a^i_j \partial b_k} = \frac{\partial^2 F}{\partial a^i_k \partial b_j} \tag{6.12}$$

for all i, j, k.

For the invariance equation we can add $\epsilon\delta_i$ to λ in (6.9) where δ_i is the vector which is 1 in the i column and 0 otherwise. Invariance of $d\lambda$ under translation as in (6.5) leads to

$$a^i \cdot \nabla_b F = 0 \tag{6.13}$$

for all i.

Remark. *In the case $l = 1$ the additive Haar measure $d\lambda$ becomes the multiplicative Haar measure upon division by $|\lambda|$. Put in other terms, $d(\alpha\lambda) = |\alpha| d\lambda$ so that changing variables in (6.9) leads to the expression*

$$J^1(a, b) = \frac{1}{|a|} \int f\left(\frac{a}{|a|}\lambda + b\right) d\lambda.$$

Thus for $l = 1$ the homogeneity of J^1 of degree -1 in a is equivalent to (6.13).

We shall generally work with (6.13) as it has meaning for all l.

Next let us examine

$$J_1 f(x, b) = \int_K f(kx + b)\, dk.$$

$\Lambda = K$ is the orthogonal group in $\mathbb{R}^n$ and $\mathbf{G} = \mathbb{R}^n$ with variable x.

Since $\mathbf{G}$ is an additive group for which the matrix units (the vectors δ_i) belong to its Lie algebra $\mathfrak{g} = \mathbf{G}$, the equations (6.7) become

$$\frac{\partial F_k}{\partial x_i} - \sum k_{ij} \frac{\partial F_k}{\partial b_j} = 0.$$

(Our notation differs slightly from (6.7) because we write $K = \Lambda$ on the left.) The Fourier transform in x, b of these equations is

$$\left(\hat{x}_i - \sum k_{ij} \hat{b}_j\right) \hat{F}_k = 0.$$

Since k is an arbitrary orthogonal transformation this leads to the compatibility conclusion $\hat{x}^2 = \hat{b}^2$ which is the first John equation (Δ is the Laplacian)

$$\mathbf{J}_1 F \equiv (\Delta_x - \Delta_b) F(x, b) = 0. \tag{6.14}$$

For the invariance equation, either by applying the Lie algebra of K or else by direct inspection, we are led to

$$F(x, b) = J_1 f(x, b) \quad \text{depends only on } |x|, b.$$

For functions of x which depend only on $r = |x|$ we can replace Δ_x by its rotationally invariant part. Thus (6.14) becomes the Darboux equation

$$\left(\frac{\partial^2}{\partial r^2} + \frac{n-1}{r} \frac{\partial}{\partial r} - \Delta_b\right) F = 0. \tag{6.15}$$

Since polar coordinates are not defined at the origin we usually prefer to think of $F = F(x, b)$ as a function defined for all x, b which is rotationally invariant in x.

The group K is compact so there is no difficulty in defining $J_1 f$ for smooth functions f and verifying that they satisfy (6.14) and (6.15). But for J^l there are several difficulties. If we assume $f \in \mathcal{S}$ (the Schwartz space) then $J^l f(\boxed{a}, b)$ is defined and smooth in $\boxed{a}, b$ as long as rank $\boxed{a} = l$. But if rank $\boxed{a} = l' < l$ then there is a linear $l - l'$ plane of λ for which $\boxed{a}\lambda = 0$. Thus the integral (6.9) defining $J^l f$ cannot exist in the usual sense on the set of singular $\boxed{a}$.

Nevertheless $J^l f(\boxed{a}, b)$ does define a distribution which satisfies the John equation (6.12) for $n \geq l + 2$. Precisely

Theorem 6.2 *If $f \in \mathcal{S}$ then $J^l f(\boxed{a}, b)$ and its b derivatives of all orders are locally integrable in $\boxed{a}, b$ when $n > l$. For $n \geq l+2$ the b derivatives of all orders of $\partial F / \partial a_j^i$ are locally integrable; in this case the John equation (6.12) holds. If $n \leq l$ then there exist $f \in \mathcal{D}$ for which $J^l f$ are not locally integrable.*

Proof To understand how things work, consider first the case $l = 1$. The only singularity occurs at $a = 0$. Since $J^1 f(a, b)$ is homogeneous in a of degree -1, it is locally integrable when $n \geq 2$. Equations (6.12) are of order 1 in a. The

first derivatives of $J^1 f$ with respect to a cannot behave worse than $|a|^{-2}$ at the origin and hence are locally integrable when $n \geq 3$.

Although this argument is simple and treats the case $l = 1$ which we consider of central importance, it is difficult to generalize it to the case $l > 1$. We present a new approach.

Suppose (still for $l = 1$) that support $f \subset [-\alpha, \alpha]$. We can estimate

$$F(a, b) = \int f(a\lambda + b)\, d\lambda$$

by means of

$$\text{measure}\{\lambda \big| a\lambda + b \in [-\alpha, \alpha]\}.$$

In terms of local integrability in a at the origin b plays little role since we can replace α by $\alpha + |b|$, so we can assume $b = 0$. (We discuss the growth of F in b as $b \to \infty$ below.)

$a\lambda \in [-\alpha, \alpha]$ means

$$|\lambda| \leq \frac{\alpha}{|a|}.$$

The measure of this set is $2\alpha/|a|$. Thus

$$|F(a, b)| \leq \frac{c}{|a|}$$

uniformly for b in compact sets. In particular F is locally integrable if $n \geq 2$. The same argument shows that all b derivatives of F are locally integrable when $n \geq 2$.

Differentiation of F with respect to a_j multiplies the integrand by λ. Since $|\lambda| \leq \alpha/|a|$ it follows that

$$\left| \frac{\partial F}{\partial a}(a, b) \right| \leq \frac{c}{|a|^2}.$$

Thus F and its first derivatives in a and all their b derivatives are locally integrable if $n \geq 3$.

The same proof shows that $J^1 f$ satisfies the John equation when $n \geq 3$.

To treat the case of $f \in \mathcal{S}$ we break up the region of integration into sets

$$U_j = U_j(a, b) = \{\lambda \big| j \leq |a\lambda + b| \leq j + 1\}.$$

We call $U = U_0(a, 0)$ and $\mu(U_j)$ the measure of U_j. We set

$$m(U_j) = \max_{\lambda \in U_j} |\lambda|.$$

It is clear that

$$U_j \subset (j + 1 + |b|)U$$

so that

$$\begin{aligned}\mu(U_j) &\le (j+1+|b|)\mu(U)\\ &\le c(j+1+|b|)|a|^{-1}\end{aligned} \tag{6.16}$$

by our above calculation.

Since $f \in \mathcal{S}$, we have $|f| \le c_N(j+1)^{-N}$ for any N on U_j. It follows that $J^1 f$ is locally integrable if $n \ge 2$. Similarly b derivatives of all orders are locally integrable.

Note that $m(U) = |a|^{-1}$. When we differentiate J^1 with respect to a we multiply the integrand by λ. Since

$$\begin{aligned}|\lambda| &\le m(U_j)\\ &\le m[(j+1+|b|)U]\\ &= (j+1+|b|)m(U)\\ &\le c(j+1+|b|)|a|^{-1}\end{aligned}$$

it follows that $\partial F/\partial a_k$ (and similarly b derivatives of all orders) is locally integrable for $n \ge 3$.

The same argument shows that the John equation (6.12) is valid for $l = 1$, $n \ge 3$.

Let us pass to $l > 1$. As in the case $l = 1$ we define

$$U_j^{(l)} = U_j^{(l)}(\boxed{a}, b) = \{\lambda \big| j \le |\boxed{a}\lambda + b| \le j+1\}.$$

$\mu(U_j^{(l)})$ is the measure of $U_j^{(l)}$ and $U^{(l)} = U_0^{(l)}$. We define

$$m(U_j^{(l)}) = \max_{\substack{\lambda \in U_j^{(l)}\\ k}} |\lambda|_k.$$

For any scalar β it is clear that

(i) $\mu(\beta U^{(l)}) = |\beta|^l \mu(U^{(l)})$.
(ii) $m(\beta U^{(l)}) = |\beta| m(U^{(l)})$.

As in the case of $l = 1$ the problem of local integrability of $J^l f$ depends on the local integrability of $\mu(U^{(l)})$.

Our proof of Theorem 6.2 is by induction on l. We have already verified the sufficiency of the conditions $n \ge 2$ or $n \ge 3$ when $l = 1$. Since the induction is somewhat complicated let us examine the case $l = 2$.

We fix a vector α and an angle θ. We call $\Gamma(\alpha, \theta)$ the cone of vectors v making an angle θ with α. If $|\alpha \sin\theta| > 1$ then no vector $v \in \Gamma(\alpha, \theta)$ can be at a distance ≤ 1 from α. If $|\alpha \sin\theta| \le 1$ then the set of vectors v on any fixed line on $\Gamma(\alpha, \theta)$ which lie at a distance ≤ 1 from α has measure ≤ 2 (see Figure 6.1).

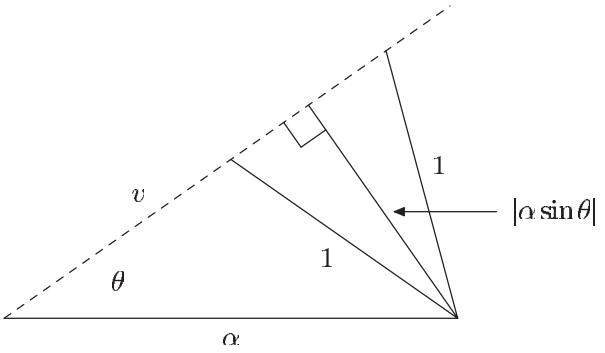

Figure 6.1

The total measure of all such v is ≤ 2 since the measure of $\{\theta\}$ is 1.

Now let α be chosen to be of the form $\lambda_1 a^1$. Our condition on α is

$$|\lambda_1 a^1| \leq \frac{1}{|\sin\theta|}$$

or

$$|\lambda_1| \leq \frac{1}{|a^1||\sin\theta|}.$$

Set $v = \lambda_2 a^2$ so θ is the angle between a^1 and a^2. For fixed α (i.e. λ_1) the measure of the set of λ_2 for which $v \in \Gamma(\alpha, \theta)$ and whose distance to $\alpha = \lambda_1 a^1$ is ≤ 1 is less than or equal to the measure of the set of λ_2 for which $\lambda_2 a^2$ lies in some interval (determined by $\lambda_1 a^1$) of length ≤ 2. This measure is bounded by the measure of the set of λ_2 satisfying

$$|\lambda_2 a^2| \leq 2$$

which is clearly $2/|a^2|$.

We have shown that the total measure of λ such that $|\lambda_1 a^1 - \lambda_2 a^2| \leq 1$ is bounded by

$$\frac{c}{|a^1||a^2||\sin\theta|}. \tag{6.16*}$$

(We have written $\lambda_1 a^1 - \lambda_2 a^2$ rather than $\lambda_1 a^1 + \lambda_2 a^2$ to make things more geometric.)

Our construction leads to the following coordinate structures on $\mathbb{R}^n \oplus \mathbb{R}^n = \{a^1, a^2\}$: We use the usual polar coordinates in a^1, that is, $|a^1|$ and ω where ω is a coordinate on the unit sphere. In a^2 we also use spherical coordinates but we specify a "base axis," namely $\{\lambda_1 a^1\}$ (which is the analog of the z axis, from which θ is measured). θ can be interpreted as the "latitude." A circle of latitude θ has measure $|\sin^{n-2}\theta|$ in θ and the measure in $|a^j|$ is $|a^j|^{n-1}\, d|a^j|$. ($|a^j|$ is the r coordinate.) We integrate in a^2 for fixed a^1 and then integrate in a^1. It is

clear from (6.16*) that F is locally integrable when f is of compact support and $n \geq 3$.

Similarly all b derivatives of F are locally integrable.

Differentiation of the integral for J^2 shows that to deal with derivatives in a we have to replace the measure of $\{\lambda\}$ by the weighted measures with weights λ_j. We have to estimate

$$m(U) = \max_{a,j}\{|\lambda_j| : \boxed{a}\lambda| \leq 1\}.$$

It is clear from Figure 6.1 that the max of such $|\lambda_j|$ is bounded by a constant multiple of

$$\max_{1,2}\left(\frac{1}{|a^1||\sin\theta|}, \frac{1}{|a^2||\sin\theta|}\right). \tag{6.17}$$

By (6.16*) we are left with the question of the local integrability of

$$\frac{1}{|a^1|^2|a^2|\sin^2\theta}, \frac{1}{|a^1||a^2|^2\sin^2\theta}$$

which requires $n \geq 4$.

If $f \in \mathcal{S}$ then we break up the λ region of integration into the sets $U_j^{(2)}(\boxed{a}, b)$. Note that

$$U_j^{(2)}(\boxed{a}, b) \subset (1 + j + |b|)U^{(2)}$$

so that

$$\mu[U_j^{(2)}(\boxed{a}, b)] \leq (1 + j + |b|)^2\mu(U^{(2)})$$
$$m[U_j^{(2)}(\boxed{a}, b)] \leq (1 + j + |b|)m(U^{(2)}).$$

We can proceed as in the case $l = 1$ to show that $J^2 f$ and all b derivatives are locally integrable for $n \geq 3$ and the same is true of $\partial J^2 f/\partial a_k$ for $n \geq 4$. The John equation (6.12) is valid when $n \geq 4$.

Our treatment of the case $l = 2$ suggests that to pass to $l > 2$ we should work with "fixed angles." We fix $\lambda_1, \ldots, \lambda_{l-1}$ and we denote by θ_l the angle between $\alpha = \lambda_1 a^1 + \cdots + \lambda_{l-1}a^{l-1}$ and a^l. As in case $l = 2$ if $|\alpha \sin\theta_l| > 1$ there is no point v on the cone $\Gamma(\alpha, \theta_l)$ at a distance ≤ 1 from α, and if $|\alpha \sin\theta_l| \leq 1$ the total measure of points v on $\Gamma(\alpha, \theta_l)$ which are at a distance ≤ 1 from α is ≤ 2. For v of the form $\lambda_l a^l$ the total λ_l measure is bounded by $2/|a^l|$.

The set of $\lambda_1, \ldots, \lambda_{l-1}$ for which $|\alpha| \leq 1/|\sin\theta_l|$ is contained in

$$\frac{1}{|\sin\theta_l|}U^{(l-1)}.$$

Hence, by property (i) of the sets $\{U\}$

$$\mu(U^{(l)}) \leq \frac{2}{|a^l||\sin\theta_l|^{l-1}}\mu(U^{(l-1)}). \tag{6.18}$$

We can now argue as in the case $l = 2$. We fix $\lambda_1, \dots, \lambda_{l-1}$ and integrate $J^l f$ over the cone $\Gamma(\lambda_1 a^1 + \cdots + \lambda_{l-1} a^{l-1}, \theta)$. Since the measure on the sphere of latitude in the unit sphere in $\mathbb{R}^n$ is $\sin^{n-2}\theta\, d\theta$ it follows from (6.18) that we obtain an answer bounded by $c\mu(U^{(l-1)})$ if $l - 1 \le n - 2$, i.e. $l \le n - 1$ (and $n \ge 2$). Induction now takes over and we have our result.

The extension to functions $f \in \mathcal{S}$ and the result for b derivatives proceed as before.

As for first a derivatives, we use property (ii) of the sets $U^{(l)}(\boxed{a}, b)$ which gives, in analogy to (6.18),

$$m(U^{(l)}) \le c \max_j \left(\frac{1}{|a^j||\sin\theta|} \right). \tag{6.19}$$

Combined with (6.18) this shows $\partial J^l(f)/\partial a^i_j$ is locally integrable (together with all b derivatives) for $n \ge l + 2$.

Again the same proof confirms the validity of the John equation (6.12) when $n \ge l + 2$.

This completes the proof of the sufficiency of the dimension conditions $n \ge l+1$ or $n \ge l+2$ for the local integrability of v derivatives of all orders of F and of $\nabla_a F$.

Proof of necessity. In case $l = 1$ the measure of the set of λ for which $|\lambda a| \le t$ is clearly $t/|a|$. Thus for $n = 1$ if $f \in \mathcal{D}$ is a positive function which is $\equiv 1$ near zero then

$$F(a, 0) = \int f(a\lambda)\, d\lambda \ge \frac{c}{|a|}$$

so F is not locally integrable.

For $l > 1$ we use the same iterative scheme that we used in the proof of the sufficiency. We fix $\lambda_1, \dots, \lambda_{l-1}, \theta_l$ as before and set $\alpha = \lambda_1 a^1 + \cdots + \lambda_{l-1} a^{l-1}$ and $\text{angle}(\alpha, a^l) = \theta_l$. If

$$|\alpha \sin\theta_l| \le \tfrac{1}{2}$$

then there are points of the form $\lambda_l a^l$ at a distance ≤ 1 from α. Moreover the set of such points has length $\ge \frac{1}{2}\sqrt{3} = c$. The total λ measure of such points (see Figure 6.1) is

$$\ge \frac{c}{|a^l|} \mu\left(\frac{1}{2\sin\theta_l} U^{(l-1)} \right).$$

Following our iteration scheme we find that if $f \in \mathcal{D}$ is a positive function with $f(0) \ne 0$ then $J^l f(\boxed{a}, 0)$ is not locally integrable.

This completes the proof of Theorem 6.2.

Our understanding of the parametric Radon transform will be enhanced by the computation of the parametric Radon transform of the δ function. To

accomplish this think of $\delta_{x=0}$ *à la* Dirac, namely

$$\delta_{x=0} = \lim_{N\to\infty} \delta^N_{x=0} = \begin{cases} N^n & \text{if } |x| \le \frac{1}{2N} \\ 0 & \text{otherwise.} \end{cases} \tag{6.20}$$

(We are using the norm $|x| = \max |x_j|$.) We have

$$\delta^N_{x=0}(\boxed{a}\lambda + b) = \begin{cases} N^n & \text{if } |\boxed{a}\lambda + b| \le \frac{1}{2N} \\ 0 & \text{otherwise.} \end{cases}$$

This means that

$$\begin{aligned} \operatorname{support} \mathbf{R_P}\delta_{x=0}(\boxed{a}, b) &= \{\boxed{a}, b \big| b \in \operatorname{span}(a^1, \dots, a^l)\} \\ &= S(\boxed{a}, b) \end{aligned}$$

say.

More precisely we want to know the measure on $S(\boxed{a}, b)$ which is defined by $\mathbf{R_P}\delta_{x=0}$. If $\boxed{a}\lambda^0 + b^0 = 0$ then the measure of the set of λ for which $|\boxed{a}\lambda + b^0| \le 1/N$ is the same as

$$\mu\{|\boxed{a}\lambda| \le 1/N\} = N^{-l}\mu(U^{(l)})$$

as we have seen above. Since $\delta^N_{x=0} = N^n$ on this set, $\mathbf{R_P}\delta_{x=0}$, which is the integral over such λ, is $N^{n-l}\mu(U^{(l)})$.

Note that the codimension of $S(\boxed{a}, b)$ is $n-l$ since the span of $\boxed{a}$ is l dimensional. The δ function of an affine plane L of codimension m is approximated by the function which is N^m on a neighborhood of L of width $1/N$. This suggests that $\mathbf{R_P}\delta_{x=0}$ is the δ function of $S(\boxed{a}, b)$ where we use the measure induced from $d\boxed{a}db$.

We leave the verification of this to the reader.

Let us return to equations (6.12). When $n = 2, l = 1$ the system (6.12) becomes determined, meaning there is only one equation for F. This equation has a fundamental solution e which is a solution outside $a = 0, b = 0$ which cannot be continued across the origin as a solution.

Problem 6.1 *For $n - l \ge 2$ does there exist a solution F of (6.12) on $\{\text{rank } \boxed{a} = l\}$ which cannot be continued to a solution on all of $\mathbb{R}^{nl}$?*

Problem 6.1 lies within the framework of Hartogs' extension. In [53] (see Section 9.1) we showed that the solutions of a truly overdetermined system of constant coefficient linear partial differential equations can be continued over suitable compact sets. But the set $\{\text{rank } \boxed{a} < l\}$ is not compact so we need a serious modification of the method of [53].

We have presented one aspect of the parametric Radon transform, namely its relation to the John and invariance equations. There is another aspect which leads to some generalizations: we can write F in the form

$$F(g, b) = \mu_g * f \tag{6.21}$$

where μ_g is the measure $\mu(\lambda)$ on the set $g\Lambda$ and convolution is in the additive variable. (Actually (6.21) is not exactly F because F involves $f(g\lambda + b)$ and the convolution should use $f(b - g\lambda)$. But this is a trivial point so we shall ignore it.)

Equation (6.21) suggests that we should study $F(g, b)$ by means of the Fourier transform in b. It also suggests that we can generalize the parametric Radon transform by using other types of convolution kernels.

There is another aspect of the John equations for J^l which is worthy of note. It is possible to write the equations in terms of the closure of differential 1 forms. Consider the forms

$$\omega^j = \frac{\partial F}{\partial a_1^j}\, db_1 + \cdots + \frac{\partial F}{\partial a_n^j}\, db_n \tag{6.22}$$

or

$$\omega_*^j = \frac{\partial F}{\partial b_1}\, da_1^j + \cdots + \frac{\partial F}{\partial b_n}\, da_n^j \tag{6.23}$$

or

$$\omega_*^{jk} = \frac{\partial F}{\partial a_1^j}\, da_1^k + \cdots + \frac{\partial F}{\partial a_n^j}\, da_n^k \tag{6.24}$$

in the respective variables b, a^j. The John equations are equivalent to $\{d\omega^j = 0\}$ or $\{d\omega_*^j = 0\}$. Since $\hat{a}^i$ is parallel to $\hat{b}$ by (6.11)ff. they imply $d\omega_*^{jk} = 0$.

Of course, these are forms on $\mathbb{R}^n$ and, as such, are exact. But we want more, namely, if F also satisfies the invariance equations, then there should be some analog of the invariance condition for the solutions h^j of $\omega^j = dh^j$. (We shall discuss only ω^j as the ω_*^j can be treated in a similar manner.)

Let $F = J^l f$ where f is small at infinity, say $f \in \mathcal{S}$. Then for each $\boxed{a}$ the invariance condition says that F is constant in b directions parallel to $\boxed{a}$, meaning for b in the linear space spanned by the columns of $\boxed{a}$. It is clear that $F(\boxed{a}, b)$ is small at infinity in b directions perpendicular to $\boxed{a}$ for generic $\boxed{a}$, meaning rank $\boxed{a} = l$.

Let us examine what this means for forms ω^j as in (6.22). We choose coordinates $b = (u, y)$ so that $u_1, \ldots, u_{n-l}$ are orthogonal to (the columns of) $\boxed{a}$ and $y_1, \ldots, y_l$ are parallel to $\boxed{a}$. The components of ω^j

$$\frac{\partial F}{\partial a_k^j}(\boxed{a}, b) = \int \lambda_j \frac{\partial f}{\partial x_k}(a^1\lambda_1 + \cdots + a^l\lambda_l + b)\, d\lambda \tag{6.25}$$

are constant in y and small at infinity in directions u. It makes sense to try to place the same requirements on h^j if $\omega^j = dh^j$. However, if a form ω whose components are constant in y is equal to dh then h is linear in y.

Proposition 6.3 *Suppose that $n - l \geq 2$. If ω is a smooth closed 1 form in b whose coefficients are small at infinity (e.g. fall off faster than $|b|^{-p}$ for all p) in directions perpendicular to a generic $\boxed{a}$ and constant in directions parallel to $\boxed{a}$ then $\omega = dh$ where h is a smooth function which is small at infinity in*

directions orthogonal to $\boxed{a}$ *and linear in directions parallel to* $\boxed{a}$. *If* ω *depends continuously on* $\boxed{a}$ *then* h *can be chosen to be continuous in* $\boxed{a}$.

The hypothesis $n - l \geq 2$ *is necessary.*

Proof of necessity. Thus we assume $n - l = 1$, meaning u is a single variable. We fix $\boxed{a}$ and let $h^0(u)$ be a smooth function which has different limits as $u \to \pm\infty$. Assume h^0 approaches its limit rapidly as $u \to \pm\infty$. We set $h(u, y) = h^0(u)$ for all y and let

$$\omega = dh = \frac{dh^0}{du}\, du.$$

Thus ω is constant in y, and is small at infinity in u directions, and is closed.

We claim that ω is not of the form $d\tilde{h}$ where $\tilde{h}$ is small in u directions and linear in y directions. For then we would have

$$d(h - \tilde{h}) = 0$$

so $h - \tilde{h}$ is constant. But this is impossible since $(h - \tilde{h})(u) \to h^0(\pm\infty)$ as $u \to \pm\infty$ and $h^0(\pm\infty)$ are not equal.

Proof of sufficiency. The distinction between $n - l = 1$ and $n - l \geq 2$ is that for $n = l + 1$ the points at infinity in the u axis are disconnected while for $n - l > 1$ they are connected.

Let us begin by fixing $\boxed{a}$. We pick a point p^0 at infinity in the u axis and define

$$h(u, y) = \int_{p^0}^{u,y} \omega.$$

The points at infinity form a connected set so the function $h(u, 0)$ is small at infinity since we can reach any $(u, 0)$ with u large by a path lying near infinity. Since $dh = \omega$ the function h is linear in y so $h(u, y)$ is small for any fixed y as $u \to \infty$.

If ω depends continuously on $\boxed{a}$ then we can construct h as above for each $\boxed{a}$ near a fixed $\boxed{a^0}$ with h depending continuously on $\boxed{a}$ because the splitting $b = (u, y)$ depends continuously (locally) on y. Then by using partitions of unity we can construct $h = h(\boxed{a})$ which depends continuously on $\boxed{a}$.

This completes the proof of Proposition 6.3.

6.2 Characterization by John equations

We now come to the crucial question:

Do the John equations and the invariance equations imply the existence of an f **satisfying** $F = J^l f$ **or** $F = J_1 f$**?**

The question is not yet precisely posed. We have to know what classes of functions f, F we consider.

Let us begin with J_1 as it is much simpler because $\Lambda = K$ is compact. For example, we could allow $F(a,b)$ and $f(b)$ to be any C^∞ functions.

Theorem 6.4 *The equations* $\mathbf{J}_1 F = \mathbf{I}_1 F = 0$ *for* $F \in \mathcal{E}$ *imply that* $F = J_1 f$ *for a unique* $f \in \mathcal{E}$.

Proof We use the John equations and the fundamental principle (see Section 1.4) to write

$$F(a,b) = \int_V e^{ia\cdot\hat{a}+ib\cdot\hat{b}}\, d\nu(\hat{a},\hat{b}). \tag{6.26}$$

ν is a measure on the variety $V = \{\hat{a}^2 = \hat{b}^2\}$ (which corresponds to the John equation) which is exponentially decreasing in imaginary directions and decreases faster than $(|\hat{a}|+|\hat{b}|+1)^{-N}$ for any N in real directions by the explicit description of the space $\hat{\mathcal{E}}'$.

For any $k \in K$

$$F(ka,b) = \int_V e^{ika\cdot\hat{a}+ib\cdot\hat{b}}\, d\nu(\hat{a},\hat{b}).$$

Since $F(ka,b) = F(a,b)$ we have

$$F(a,b) = \int dk \int e^{ika\cdot\hat{a}+ib\cdot\hat{b}}\, d\nu(\hat{a},\hat{b}). \tag{6.27}$$

We want to find an f such that

$$F(a,b) = \int f(ka+b)\, dk.$$

Suppose that for any $\hat{a},\hat{b}$ there is a κ with $\hat{a} = \kappa\hat{b}$. Then we can write

$$\begin{aligned} ka\cdot\hat{a} + b\cdot\hat{b} &= ka\cdot\kappa\hat{b} + b\cdot\hat{b} \\ &= (\kappa^{-1}ka + b)\cdot\hat{b}. \end{aligned} \tag{6.28}$$

In (6.27) we integrate first with respect to k for fixed $\hat{a},\hat{b}$. κ is absorbed in the measure so (6.27) becomes

$$\begin{aligned} F(a,b) &= \int dk \int e^{i(ka+b)\cdot\hat{b}}\, d\nu(\hat{a},\hat{b}) \\ &= \int f(ka+b)\, dk \\ &= J_1 f(a,b) \end{aligned} \tag{6.29}$$

where

$$f(b) = \int e^{ib\cdot\hat{b}}\, d\nu(\hat{a},\hat{b}). \tag{6.30}$$

Unfortunately there does not always exist a $\kappa \in K$ with $\hat{a} = \kappa\hat{b}$. κ does exist when $\hat{a},\hat{b}$ are real since $\hat{a}^2 = \hat{b}^2$, but for complex $\hat{b}$ we have to go to the

complexification $K^{\mathbb{C}}$ of K. Nevertheless

Claim. *The set* $\{k\hat{b}, \hat{b}\}_{\substack{\hat{b}\in\mathbb{C}^n \\ \kappa\in K}}$ *is sufficient for* V.

We proved in Theorem 4.12 that the union of real complex lines (complex lines with real direction numbers) $V^{\mathbb{CR}} \subset V$ is sufficient for $\mathcal{E}(V)$. Since any real point in V is of the form $(\kappa\hat{b}, \hat{b})$ it follows that

$$V^{\mathbb{CR}} \subset \{k\hat{b}, \hat{b}\}_{\substack{\hat{b}\in\mathbb{C}^n \\ \kappa\in K}}$$

so the latter is sufficient for V. This confirms our claim.

Theorem 6.4 now follows from our above remarks.

This calculation shows a totally new setting for the parametric Radon transform. According to (6.14) we start with the ultrahyperbolic equation $(\Delta_a - \Delta_b)F = 0$. We want to determine F by a function $f(b)$ by means of $F = J_1 f$. Note that $f(b) = F(0, b)$ as is clear from the definition of J_1. This leads us to examine the WP for the ultrahyperbolic equation with data given on the b axis. In analogy with the ideas of Chapter 4 the "normal derivatives" should consist of $h_j(\partial/\partial a)$ where h_j are the usual harmonic polynomials in a, because above each point $\hat{b}$ there lies an $m = n/2$ sphere in V with $(\text{radius})^2 = \hat{b}^2$. The invariance equation (6.13) for the parametric Radon transform has picked out the constant component of this WP. This corresponds to a solution F of $(\Delta_a - \Delta_b)F = 0$ which is K invariant in a and so has all its WD $= 0$ except for the datum f corresponding to the identity operator. If

$$[h_j\,(\partial/\partial a)\,F]\,(0, b) = \begin{cases} f(b) & \text{for } h_j \equiv 1 \\ 0 & \text{otherwise} \end{cases} \tag{6.31}$$

then $F = J_1 f$ since $J_1 f$ is a solution of the WP with

$$\text{WD}(J_1 f) = \text{WD}(F). \tag{6.32}$$

All this depends on the uniqueness theorem for the WP (see Section 4.2).

We now pass to the more difficult J^l. There are many WPs that we can associate to the John equation (6.12). For the purpose of inverting the parametric Radon transform the natural choice of WS is the b axis. Our proof of this inversion (Theorem 6.5) can be regarded as being within the framework of the WP with this WS.

Unlike the case of J_1 which we just presented, we have to restrict $J^l f$ to functions which are small at infinity. We have not examined optimal choices of regularity and growth at infinity; we restrict our considerations to $f \in \mathcal{S}$.

Theorem 6.5 *Suppose* $n - l \geq 2$. *The functions* F *of the form* $F = J^l f$ *for* $f \in \mathcal{S}$ *are characterized by the following properties:*

(i) $F \in \mathcal{S}'$.

(ii) *For any $n \times l$ matrix $\boxed{a} = a^1, \ldots, a^l$ of rank l, F belongs to $\mathcal{S}$ on the affine $n - l$ planes orthogonal to the l plane spanned by $a^1, \ldots, a^l$. The restriction of F to such planes depends smoothly in $\mathcal{S}$ on $\boxed{a}$ and on the affine (translation) parameter.*

(iii) *F satisfies the John equation $\mathbf{J}^l F = 0$ and the invariance equation $\mathbf{I}^l F = 0$.*

Proof All the essential difficulties already occur for $l = 1$ so we shall concentrate on that case. The variety V defined by the John equation consists of vectors $\hat{a}$ parallel to $\hat{b}$. It is important to observe that this condition is vacuous on $\hat{a}$ if $\hat{b} = 0$ and conversely.

If $F(a, b) = J^1 f(a, b)$ then it is clear from the definition (6.10) that F is constant for b directions parallel to a for $a \neq 0$; we have noted that this constancy is equivalent to the invariance equation. Thus F cannot be small in all directions at infinity. But if b stays in $\{a^\perp\}$ then F is small as $b \to \infty$. In fact it is clear in this case that for $a \neq 0$ and fixed we have

$$F(a, b) \in \mathcal{S}(a^\perp) \tag{6.33}$$

and $a \to F(a,)$ is a smooth map into $\mathcal{S}(b)$. In particular $F \in C^\infty$ on $a \neq 0$.

Theorem 6.2 asserts that F is locally integrable and hence defines a distribution in a, b. It is clear from the integral representation (6.10) of F that

$$F(a, b) = |a|^{-1} F(1, b).$$

This implies that $F \in \mathcal{S}'$.

This proves the necessity of (i), (ii), (iii).

We want to prove the converse. Besides the obvious technical problems in dealing with noncompact Λ we are faced with two special difficulties:

(1) We cannot define f in any obvious way from F since $F(0, b)$ is not well defined.

(2) The part of V where $\hat{b} = 0$ belongs to V^1 but not to V (see (6.4)).

We introduce two Fourier transforms of F:

$$\begin{aligned} \hat{F}(a, \hat{b}) &= \text{Fourier transform in } b \text{ for fixed } a \neq 0 \\ \hat{F}(\hat{a}, \hat{b}) &= \text{Fourier transform in } a, b. \end{aligned} \tag{6.34}$$

Our first task is to construct f from F. We cannot apply the idea of (6.29), (6.30) as is, because we cannot integrate the analogous measure $\int \mu(\lambda \hat{a}, \hat{b})\, d\lambda$ since μ is not small in $\hat{a}$. Rather we shall explain how to obtain f (or rather $\hat{f}$) as a sort of residue of F.

We use Fourier analysis to help us use the John and invariance equations since things become more transparent after Fourier transformation.

We shall work with $\hat{F}$. In contrast with our previous treatment of J_1 we start by examining the invariance equation rather than the John equation. We

fix $a \neq 0$ so $F(a,b)$ as a function of b is in $\mathcal{S}$ in the $a^\perp$ directions and is constant in directions parallel to a. This means that

$$\hat{}F(a,\hat{b}) = \hat{f}_a(\hat{b})\delta_{\hat{b}\in a^\perp} \tag{6.35}$$

where $\hat{f}_a \in \mathcal{S}(a^\perp)$. Moreover for $a \neq 0$ the map $a \to \hat{f}_a$ is smooth.

In order to see where we are going let us examine the nature of $\hat{h}_a$ when $F = H$ is of the form $J^1 h$.

We write $b = (x,y) = (x_a, y_a)$ with y parallel to a and $x \in a^\perp$. The dual decomposition is $\hat{b} = (\hat{x}, \hat{y})$. In these variables we have, formally,

$$\begin{aligned}\hat{h}_a(\hat{b})\delta_{\hat{y}=0} &= \iint e^{ix\cdot\hat{x}+iy\cdot\hat{y}}\,dx\,dy \int h(a\lambda + b)\,d\lambda \\ &= \delta_{\hat{y}=0}\iint e^{ix\cdot\hat{x}} h(a\lambda + x)\,dx\,d\lambda \\ &= \delta_{\hat{y}=0}\iint e^{ix\cdot\hat{x}+a\lambda\cdot\hat{0}} h(a\lambda + x)\,dx\,d\lambda. \end{aligned} \tag{6.36}$$

Suppose first that a is a unit vector; $\{a\lambda\}$ represents the line through a. We write $y = a\lambda$ so $dy = d\lambda$ and $dy\,dx = db$ so the last integral is $\hat{h}(\hat{x}, 0)$. We have thus shown

$$\hat{h}_a(\hat{b}) = \hat{h}(\hat{x}, 0). \tag{6.37}$$

In general, if $a \neq 0$, the same argument shows

$$\hat{h}_a(\hat{b}) = |a|^{-1}\hat{h}(\hat{x}, 0). \tag{6.38}$$

If we compare (6.35) with (6.38) we see that, in order to produce f satisfying $F = J^1 f$, we need

(α) *$\hat{f}_a$ is homogeneous in a of degree -1.*

This allows us to define, as in (6.37),

$$\hat{f}(\hat{x}_a, 0_a) = \hat{f}_a(\hat{b})$$

when $|a| = 1$. Note that, when $|a| = 1$,

$$H(a,x) = \mathbf{R}h(\mathbf{g}, x)$$

where $\mathbf{g}$ is the line $\{a\lambda\}$. In this case, by (6.37),

$$\begin{aligned}\hat{h}_a(\hat{b}) &= \hat{h}(\hat{x}, 0) \\ &= \text{restriction } \hat{h} \text{ to hyperplane } \{\hat{y} = 0\}.\end{aligned}$$

Thus if

$$\{\hat{f}_a(\hat{b})\} = \{\hat{f}(\hat{x}_a, 0)\}$$

($\hat{x}_a \in a^\perp$) is to define a function $\hat{f}(\hat{b})$ then we need

(β^*) *The $\hat{f}(\hat{x}, 0) = \hat{f}(\hat{x}_a, 0_a)$ defined for $\hat{x}_a \in a^\perp$ fit together as a varies with $|a| = 1$ to form a single function $\hat{f}(\hat{b})$ in $\mathcal{S}$, i.e. the various $\hat{f}(\hat{x}_a, 0_a)$ satisfy the moment conditions (see Section 2.1).*

Since $n - l \geq 2$ the moment conditions reduce to the identity

$$\hat{f}(\hat{x}_a, 0_a) = \hat{f}(\hat{x}_{a'}, 0_{a'}) \quad \{\hat{x}_a\} \cap \{\hat{x}_{a'}\}.$$

We can rewrite (β^*) in view of (α) for general a as

(β) $|a|\hat{f}_a(\hat{b})\delta_{\hat{b}\in a^\perp}$ *is independent of a.*

("Independent of a" is defined as in (β^*).)

Once we have completed the verification of (α) and (β) we can use the results of Section 2.1 to produce a function $f(b) \in \mathcal{S}$ whose Fourier transform is $\hat{f}(\hat{b})$. We have shown that $J^1 f(a, x_a) = \mathbf{R}f(\mathbf{g}, x_a)$ with $\mathbf{g} = a$ when $|a| = 1$. This means that

$$J^1 f(a, b) = \mathbf{R}^* \mathbf{R} f(\mathbf{g}, b)$$

since both sides agree for $b = (x_a, 0)$ and are constant in $a^\perp$. Thus

$$\hat{}J^1 f(a, \hat{b}) = \hat{f}(\hat{x}_a, 0)\delta_{\hat{b}\in a^\perp};$$

that is,

$$\hat{}F(a, \hat{b}) = \hat{}J^1 f(a, \hat{b})$$

when $|a| = 1$. Since both sides are homogeneous of degree -1 we obtain the desired result

$$F = J^1 f.$$

It remains to verify (α) and (β). The invariance property of

$$|a|\hat{f}_a(\hat{b})\delta_{\hat{b}\in a^\perp}$$

means that the distributions in $\hat{b}$ defined by fixing a are independent of a. We study

$$|a|\hat{f}_a(\hat{b})\delta_{\hat{b}\in a^\perp}$$

qua distributions in $\hat{b}$ and we examine the variation of these distributions as a varies.

It is in the verification of (α) and (β) that the first John equation comes into play. The Fourier transform in b of the John equation is

$$\left(\hat{b}_j \frac{\partial}{\partial a_i} - \hat{b}_i \frac{\partial}{\partial a_j}\right) \hat{}F(a, \hat{b}) = 0. \tag{6.39}$$

Now, $\hat{}F(a, \hat{b}) = \hat{f}_a(\hat{b})\delta_{\hat{b}\in a^\perp}$. We have assumed that the map $a \to \hat{}F(a, \hat{b})$ is a smooth map into $\mathcal{S}'$. Also $\hat{f}_a(\hat{b})$ isa smooth function of $\hat{b}$ so we can apply the

Leibnitz formula to compute the derivatives of $\hat{}F$ with respect to a. (Actually $\hat{f}_a(\hat{b})$ is defined only for $\hat{b} \in a^\perp$ but we can extend it smoothly in $a, \hat{b}$ for $a \neq 0$. Since we multiply $\hat{f}_a(\hat{b})$ by $\delta_{\hat{b}\in a^\perp}$ the extension is unimportant; its only purpose is to allow us to use the standard Leibnitz formula.)

We start with $a = (0, \alpha, 0, \dots, 0)$. We vary a and use (6.39). The main part of the proof consists in differentiating $\delta_{\hat{b}\in a^\perp}$ in a.

When we write a in the form $a = (0, \alpha, 0, \dots, 0)$ and vary $\alpha = a_2$, we do not change $\{a^\perp\}$ and hence $\partial\, \delta_{\hat{b}\in a^\perp} / \partial a_2 = 0$.

To understand the computation of $\partial\, \delta_{\hat{b}\in a^\perp} / \partial a_1$ examine Figure 6.2 depicting the $\hat{b}_1, \hat{b}_2$ plane and the variation $a \to a + \delta a_1$. We have drawn the figure for $\alpha > 0$; the figure for $\alpha < 0$ is analogous.

To compute $\partial\, \delta_{\hat{b}\in a^\perp} / \partial a_1$ we use the definition of derivative. Let $h(\hat{b})$ be a smooth function of compact support. We form

$$\begin{aligned} \frac{1}{\delta a_1} \int \left[h\left(\hat{b}_1, -\frac{\delta a_1}{\alpha}\hat{b}_1, \hat{b}_3, \dots, \hat{b}_n\right) - h\left(\hat{b}_1, 0, \hat{b}_3, \dots, \hat{b}_n\right) \right] d\hat{b}_1\, d\hat{b}_3 \dots d\hat{b}_n \\ \longrightarrow -\frac{\hat{b}_1}{\alpha} \int \frac{\partial h}{\partial \hat{b}_2}\left(\hat{b}_1, 0, \hat{b}_3, \dots, \hat{b}_n\right) d\hat{b}_1\, d\hat{b}_3 \dots d\hat{b}_n \\ = \left[\frac{\hat{b}_1}{\alpha} \frac{\partial}{\partial \hat{b}_2} \delta_{\hat{b}\in a^\perp} \right] \cdot h. \end{aligned} \tag{6.40}$$

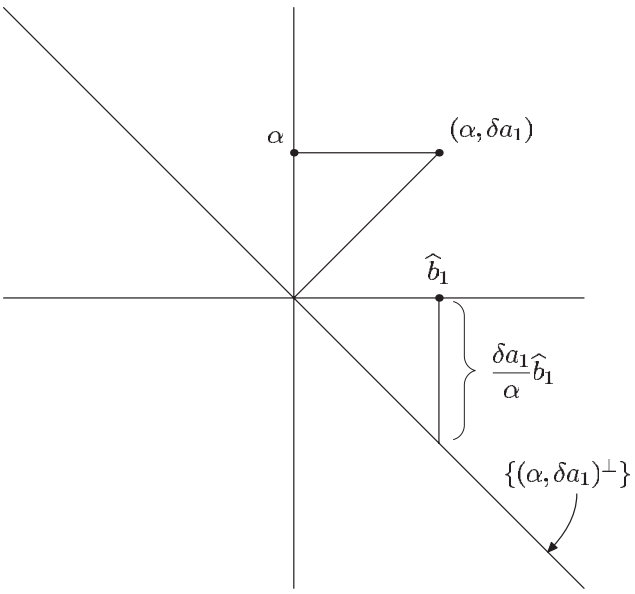

Figure 6.2

We can therefore rewrite (6.39) in the form

$$\hat{b}_2\left[\frac{\hat{b}_1}{\alpha}\hat{f}_a\frac{\partial}{\partial\hat{b}_2}+\frac{\partial}{\partial a_1}\hat{f}_a\right]\delta_{\hat{b}\in a^\perp}=\hat{b}_1\frac{\partial\hat{f}_a}{\partial a_2}\delta_{\hat{b}\in a^\perp}. \tag{6.41}$$

Observe that

$$\hat{b}_2\frac{\partial}{\partial\hat{b}_2}\delta_{\hat{b}\in a^\perp}=-\delta_{\hat{b}\in a^\perp} \tag{6.42}$$

because

$$\hat{b}_2\partial/\partial\hat{b}_2=(\partial/\partial\hat{b}_2)\hat{b}_2-\text{ identity}$$

and $\hat{b}_2$ vanishes on $\{(0,\alpha,0,\dots,0)^\perp\}$ since the second component $=0$. Hence (6.41) leads to

$$-\frac{\hat{b}_1}{\alpha}\hat{f}_a=\hat{b}_1\frac{\partial\hat{f}_a}{\partial a_2} \tag{6.43}$$

on $b\in a^\perp$ which is the only significant set for $\hat{f}_a(\hat{b})$.

Since $a=(0,\alpha,0,\dots,0)$ we can divide (6.43) by $\hat{b}_1$ when $\hat{b}_1\neq 0$ to deduce for such a and hence by a change of variables for all a

$$\sum a_j\frac{\partial\hat{f}_a}{\partial a_j}=-\hat{f}_a$$

which shows that $\hat{f}_a$ is homogeneous in a of degree -1 when $\alpha>0$ and $\hat{b}\neq 0$. A similar calculation applies when $\alpha<0$ so we have established the sought-for homogeneity of $\hat{f}_a$, except that we must examine what happens at $\hat{b}=0$.

Before passing to $\hat{b}=0$ we must strengthen the homogeneity result, which is (β). Homogeneity means that $|a|\hat{f}_a(\hat{b})$ is constant on a lines through the origin. We need to know that $|a|\hat{f}_a(\hat{b})$ is independent of a as long as $a\neq 0$ and $a\in\hat{b}^\perp$. (Recall that $\hat{b}\in a^\perp$ is the same as $a\in\hat{b}^\perp$.) For then $\hat{f}(\hat{b})$ is well defined. Thus it remains to show that, with the above notation, $\partial\hat{f}_a/\partial a_k=0$ for $k>2$ since $\partial(|a|)/\partial a_k=0$ because a is on the a_2 axis and hence is orthogonal to a_k.

We now make the same calculation as we did in (6.40)f. except that we vary a in the a_k direction. For $k\neq 2$ we need only change $i=1$ to $i=k$; (6.41) remains valid. If we divide this equation by $\hat{b}_k$ the right side is independent of $k\neq 2$. Thus

$$\hat{b}_1\left[\frac{\hat{b}_k}{\alpha}\hat{f}_a\frac{\partial}{\partial\hat{b}_2}+\frac{\partial}{\partial a_k}\hat{f}_a\right]\delta_{\hat{b}\in a^\perp}=\hat{b}_k\left[\frac{\hat{b}_1}{\alpha}\hat{f}_a\frac{\partial}{\partial\hat{b}_2}+\frac{\partial}{\partial a_1}\hat{f}_a\right]\delta_{\hat{b}\in a^\perp}. \tag{6.44}$$

We choose $\hat{b}=(1,0,\dots,0)$. Since $k>2$ the only nontrivial term in (6.44) is $\hat{b}_1\partial\hat{f}_a/\partial a_k$. We conclude that $\partial\hat{f}_a/\partial a_k=0$ for $k>2$ so that $|a|\hat{f}_a(\hat{b})$ is independent of a as long as $\hat{b}\neq 0$ and $a\in\hat{b}^\perp$.

For $a \neq 0$ the independence of $|a|\hat{f}_a(0)$ from a can be proven as follows. We know that $|a|\hat{f}_a(\hat{b})$ is smooth in $\hat{b}$ when $\hat{b} \in a^\perp$. This means that $|a|\hat{f}_a(0)$ is well defined as long as we approach $\hat{b} = 0$ in $a^\perp$. Moreover, if $\hat{b}_*$ is a line through the origin in $\{\hat{b}\}$ then we have shown that the value of $|a|\hat{f}_a(\hat{b})$ is independent of $a \in \hat{b}_*^\perp, \hat{b} \in \hat{b}_*, \hat{b} \neq 0$. Thus the limit $|a|\hat{f}_a(0)$ is independent of $a \in \hat{b}_*^\perp$ if the limit is taken for $\hat{b} \in \hat{b}_*$. But it might still be possible that if we start with a^1, a^2 and corresponding $\hat{b}_*^1, \hat{b}_*^2$ then the limits would be different.

However, this cannot happen when $n \geq 3$. For we can find a direction a^3 which is orthogonal to both $\hat{b}_*^1$ and $\hat{b}_*^2$. Hence in the plane $(a^3)^\perp$ the limits along $\hat{b}_*^1$ and $\hat{b}_*^2$ are the same. Since the limit along $\hat{b}_*^j$ is the same for all $a \in (\hat{b}_*^j)^\perp$ it is the same as for a^j.

Thus the value $|a|\hat{f}_a(0)$ is well defined and independent of a.

The proof of Proposition 2.5 allows us to conclude that the $|a|\hat{f}_a(\hat{b})$ fit together to form a smooth function $\hat{f}(\hat{b})$. According to our discussion following (6.38) this implies that $F = J^1 f$ for $a \neq 0$.

This completes the proof of Theorem 6.5 for $l = 1$.

When $l > 1$ we have to make several modifications. Equation (6.35) becomes

$$\hat{F}(\boxed{a}, \hat{b}) = \hat{f}_{\boxed{a}}(\hat{b})\delta_{\hat{b} \in \boxed{a}^\perp}. \tag{6.35^R}$$

(Recall that we are using the notation $\boxed{a}$ for the matrix and for the linear space spanned by its columns.) We have to verify the analogs of (α) and (β) for $l > 1$. The homogeneity is given by (6.10*).

We normalize our "initial" $\boxed{a}$ to be

$$\begin{aligned} a^1 &= (\alpha_1, 0, 0, \ldots, 0, 0, 0, \ldots, 0) \\ a^2 &= (0, \alpha_2, 0, \ldots, 0, 0, 0, \ldots, 0) \\ &\vdots \\ a^l &= (0, 0, 0, \ldots, 0, \alpha_l, 0, \ldots, 0). \end{aligned}$$

We have to differentiate $\delta_{\hat{b} \in \boxed{a}^\perp}$ considered as a distribution in $\hat{b}$. As in the case $l = 1$ when we vary $\boxed{a}$ within the l plane spanned by the a^j we do not change $\delta_{\hat{b} \in \boxed{a}^\perp}$ so

$$\frac{\partial}{\partial a_q^p}\delta_{\hat{b} \in \boxed{a}^\perp} = 0 \quad \text{if } p, q \leq l.$$

We want to vary $\boxed{a}$ outside this l plane. For simplicity of notation we compute $\partial/\partial a_k^1$ for $k > l$. The vectors $a^1, \ldots, a^l$ are fixed so the calculation of (6.40)f. remains valid as long as we ignore $a^2, \ldots, a^l$ and keep $\hat{b}_2, \ldots, \hat{b}_l$ fixed at 0.

We conclude as in the argument following (6.44) that

$$|\alpha_j| \hat{f}_{\boxed{a}}(\hat{b}) \qquad \text{is independent of } \alpha_j \text{ when } \alpha_j \neq 0.$$

So far we have studied one-dimensional variations in $\boxed{a}$ which preserve the orthogonality of the a_j. Let us now send $a^1 \to a^1 + (0, \epsilon, 0, \ldots, 0)$. As we mentioned $\delta_{\hat{b} \in \boxed{a}^\perp}$ is invariant so the John equation (6.39) becomes

$$\delta_{\hat{b} \in \boxed{a}^\perp} \left(\hat{b}_1 \frac{\partial f_{\boxed{a}}}{\partial a_2^1} \right) = 0$$

which implies

$$\frac{\partial f_{\boxed{a}}}{\partial a_2^1} = 0 \qquad \text{for } \hat{b}_1 \neq 0, \hat{b} \in \boxed{a}^\perp.$$

Note that when we make the variation $a^1 \to a^1 + (0, \epsilon, 0, \ldots, 0)$ we do not change $\det \boxed{a}^* \boxed{a}$. This is seen from the fact that

$$\det \begin{pmatrix} \alpha_1 & \epsilon & 0 & & \\ 0 & \alpha_2 & 0 & & \\ 0 & 0 & \alpha_3 & & \\ & & & \ddots & \\ & & & & \alpha_l \end{pmatrix} \begin{pmatrix} \alpha_1 & 0 & 0 & & \\ \epsilon & \alpha_2 & 0 & & \\ 0 & 0 & \alpha_3 & & \\ & & & \ddots & \\ & & & & \alpha_l \end{pmatrix} = (\alpha_1 \alpha_2 \ldots \alpha_l)^2.$$

Thus $\det(\boxed{a}^* \boxed{a})^{1/2} f_{\boxed{a}}(\hat{b})$ is preserved in accordance with (6.10*).

It is standard (and clear) that the $\boxed{a}$ variations that we have examined generate all variations so that $\det(\boxed{a}^* \boxed{a})^{1/2} f_{\boxed{a}}(\hat{b})$ is a well-defined function of $\hat{b}$ as long as no $\hat{b}_j = 0$. The analysis when $\hat{b}_j = 0$ is made as in the case $l = 1$.

This completes the proof of Theorem 6.5.

We have proven Theorem 6.5 using the Fourier transform in b. We now prove Theorem 6.5 using the Fourier transform in all variables. As usual the essential argument occurs for $l = 1$.

The variety V is the set of *real* $\hat{a}, \hat{b}$ with $\hat{a} \| \hat{b}$. Suppose $\hat{a} \neq 0$, e.g. $\hat{a}_1 \neq 0$. The equations

$$\hat{a}_1 \hat{b}_j = \hat{b}_1 \hat{a}_j$$

show that if $\hat{b}_1 = 0$ then $\hat{b}_j = 0$ for all j. We have

$$d(\hat{a}_1 \hat{b}_j - \hat{b}_1 \hat{a}_j) = \hat{a}_1 \, d\hat{b}_j + \hat{b}_j \, d\hat{a}_1 - \hat{a}_j \, d\hat{b}_1 - \hat{b}_1 \, d\hat{a}_j$$

where, as usual, d denotes exterior differentiation. Since $\hat{a}_1 \neq 0$ this differential form is not $\equiv 0$. In fact

$$d(\hat{a}_1 \hat{b}_2 - \hat{b}_1 \hat{a}_2) \wedge \cdots \wedge d(\hat{a}_1 \hat{b}_n - \hat{b}_1 \hat{a}_n) = \hat{a}_1^{n-1} \, d\hat{b}_2 \wedge \cdots \wedge d\hat{b}_n + \omega$$

where ω does not involve $d\hat{b}_2 \wedge \cdots \wedge d\hat{b}_n$. This shows that the exterior product of $d(\hat{a}_1\hat{b}_j - \hat{b}_1\hat{a}_j)$ for $j \neq 1$ does not vanish so the equations

$$\hat{a}_1\hat{b}_j - \hat{a}_j\hat{b}_1 = 0$$

define a variety which is a complete nonsingular intersection when $\hat{a}_1 \neq 0$.

The same property of V is true as long as $\hat{a} \neq 0$ or $\hat{b} \neq 0$.

Since $\hat{F} \in \mathcal{S}'$ satisfies $(\hat{a}_i\hat{b}_j - \hat{b}_i\hat{a}_j)\hat{F} = 0$ for all i, j it follows that $\hat{F}$ is a slowly increasing measure on V except possibly at the origin. For the local structure of V except at the origin becomes, after a simple change of variables, like that of the variety defined by $x_1 = 0, \ldots, x_{2n-1} = 0$ in $\mathbb{R}^{2n}$.

At the origin V develops multiplicities. Clearly any first-order operator annihilates the ideal $\mathcal{I}$ generated by $\hat{a}_i\hat{b}_j - \hat{b}_i\hat{a}_j$ at the origin as do operators depending on $\hat{a}$ or $\hat{b}$ only. The latter makes sense because the planes $\hat{a} = 0$ and $\hat{b} = 0$ are contained in V.

We have noted that the invariance equation becomes $(a \cdot \hat{b})\hat{}F = 0$. If we now reverse our viewpoint and think of $\hat{}F$ as a function of a for fixed $\hat{b}$ then, for $\hat{b} \neq 0$,

$$\hat{}F(a, \hat{b}) = g_{\hat{b}}(a)\delta_{a \in \hat{b}^\perp}. \tag{6.45}$$

It is not difficult to apply a Fubini argument to show that $\hat{F}(\hat{a}, \hat{b})$ is the Fourier transform of $\hat{}F$ in a for each fixed $\hat{b}$. Thus

$$\hat{F}(\hat{a}, \hat{b}) = \hat{g}_{\hat{b}}(\hat{a}) \tag{6.46}$$

where $\hat{g}_{\hat{b}}$ is constant in $\hat{a}$ directions parallel to $\hat{b}$ as long as $\hat{b} \neq 0$. (This is consistent with the fact that the Fourier transform of $a \cdot \nabla_b$ is $\hat{b} \cdot \nabla_{\hat{a}}$.) The fact that, in addition, support $\hat{g}_{\hat{b}}(\hat{a}) \subset \{\hat{a} \parallel \hat{b}\}$ (i.e. $\hat{a} = \lambda\hat{b}$ for some $\hat{b}$) means that $g_{\hat{b}}(a)$ is constant on its support which is $\{a \in \hat{b}^\perp\}$.

We have shown that

$$\hat{}F(a, \hat{b}) = c(\hat{b})\delta_{a \in \hat{b}^\perp}. \tag{6.47}$$

At first sight this equation might seem puzzling because $\hat{f}_a(\hat{b})$ (see (6.35), (6.38)) is *not* independent of a but rather it is $|a|\hat{f}_a(\hat{b})$ which is independent of a.

To show that (6.47) is really the same as the constancy of $|a|\hat{f}_a(\hat{b})$ in a we have to examine carefully the way we have identified functions with distributions.

We have regarded $V_I = \{a \cdot \hat{b} = 0\}$ in two ways:

(1) Fibered over $\{a\}$ with fibers $\hat{b} \in a^\perp$.
(2) Fibered over $\{\hat{b}\}$ with fibers $a \in \hat{b}^\perp$.

In case (1) we identify a function on V_I with a measure by multiplying by $da\delta_{\hat{b} \in a^\perp}$. Note that a is a variable in $\mathbb{R}^n$ and $\hat{b} \in a^\perp$ is a variable in $\mathbb{R}^{n-1}$. When we reverse things to case (2), $\hat{b}$ becomes a variable in $\mathbb{R}^n$ and a becomes a variable in $\mathbb{R}^{n-1}$.

The constancy of $\hat{F}(a,\hat{b})$ for $a \in \hat{b}^{\perp}$ which, as we have seen, is a consequence of the John equations, means that, as a distribution, $\hat{F}(a,\hat{b})$ is of the form $c(\hat{b})\delta_{a\in\hat{b}^{\perp}}$. When we reinterpret this in terms of (1) we have to multiply by $|a|^{-1}|\hat{b}|$ to compensate for the fact that the dimensionality of $(a,\hat{b})$ changes from $(n-1,n)$ in (2) to $(n,n-1)$ in (1). The measure on $(a,\hat{b})$ in (2) is thus $|a|^{n-2}|\hat{b}|^{n-1}d\theta_a d\theta_{\hat{b}}\, d|a|\, d|\hat{b}|$ while in (1) it is $|a|^{n-1}|\hat{b}|^{n-2}d\theta_a d\theta_{\hat{b}}\, d|a|\, d|\hat{b}|$. This accounts for the factor $|a|^{-1}|\hat{b}|$. (It would have been more precise to use variables $a,\hat{b}$ in (1) and $a',\hat{b}'$ in (2). Then $|a||\hat{b}|^{-1}$ represents the Jacobian of this change of coordinates.)

We conclude from this and (6.35) that

$$\hat{f}_a(\hat{b}) = c(\hat{b})|a|^{-1}|\hat{b}| \tag{6.48}$$

which is a second proof that $|a|\hat{f}_a(\hat{b})$ is independent of a.

The proof of Theorem 6.5 is concluded as before.

Remark 1 *The smoothness of $\hat{f}_a(\hat{b})$ in $\hat{b}$ for $a \neq 0$ means that the singularity of V^1 at $\hat{b}=0$ (for $a\neq 0$) plays no role. As seen in the conclusion of the proof of Theorem 6.5 the singularity at $a=0$ is irrelevant.*

The irrelevance of $\{\hat{b}=0\}$ can be seen from another point of view. If $\hat{h}\in\mathcal{S}'$ is a distribution on V supported on $\{\hat{b}=0\}$, which is the $\hat{a}$ axis, then it is readily verified that h cannot $\to 0$ in any b direction. Thus it plays no role in the nonparametric Radon transform on $\mathcal{S}$.

Remark 2 *Comparison of (6.35) with (6.38) or (6.47) and (6.48) shows that $\hat{f}$ is obtained from $\hat{F}$ via division by a δ function. This is analogous to forming a residue.*

We have noted that the John equations result from the geometry of parametrized linear spaces of dimension l. The invariance equation describes the measure $d\lambda$. We wish to examine some possibilities of varying the linearity of the spaces on which we integrate and also of varying the measure $d\lambda$. Let us begin with the invariance equation.

If we replace the measure $d\lambda$ used in the definition of $J^l f$ by the attenuated measure $\exp(i\hat{\lambda}\cdot\lambda)\,d\lambda$ then we see easily that the corresponding function $F_{\hat{\lambda}}(\boxed{a},b) = J^l_{\hat{\lambda}} f(\boxed{a},b)$ satisfies

$$F_{\hat{\lambda}}\left(\boxed{a}, b+\lambda_1 a^1+\cdots+\lambda_l a^l\right) = e^{-i\hat{\lambda}\cdot\lambda}F_{\hat{\lambda}}(a,b) \tag{6.49}$$

which, in infinitesimal form, becomes

$$(a^j\cdot\nabla_b)F_{\hat{\lambda}} = -i\hat{\lambda}_j F_{\hat{\lambda}}. \tag{6.50}$$

We can repeat our arguments used in the proof of Theorem 6.5 to show that the John equation and (6.50) and regularity imply $F_{\hat{\lambda}}$ is of the form $J^l_{\hat{\lambda}} f$.

Note that the Fourier transform of (6.50) is

$$(\hat{b}\cdot\nabla_{\hat{a}^j})\hat{F}_{\hat{\lambda}} = -i\hat{\lambda}_j\hat{F}_{\hat{\lambda}}. \qquad (6.50)\hat{}$$

Thus $\hat{F}_{\hat{\lambda}}$ is multiplied by $\exp(-i\hat{\lambda}_j\cdot\lambda_j)$ when we replace $\hat{a}^j$ by $\hat{a}^j+\lambda_j\hat{b}$. (This is the inverse of going from (6.49) to (6.50).) Since each $\hat{a}^j$ is parallel to $\hat{b}$, $F_{\hat{\lambda}}$ behaves like a character of the group Λ on V.

For a different example, we replace the measure $d\lambda$ by $\exp(-\sum\hat{\lambda}_j|\lambda_j|)\,d\lambda$; that is, use the *attenuation* $\exp(-\sum\hat{\lambda}_j|\lambda_j|)$. Since $\frac{1}{2}\exp(-|x|)$ is the fundamental solution for d^2/dx^2-1 the invariance equation becomes

$$\left[(a^j\cdot\nabla_b)^2-\hat{\lambda}_j^2\right]F(a,b) = 2\hat{\lambda}_j^2F(a^1,\dots,a^{j-1},0,a^{j+1},\dots,a^l,b)$$

as is easily seen by integration by parts.

Let us study the attenuation

$$\exp\left(-\int_\lambda^\infty \mu(a\tilde{\lambda}+b)\,d\tilde{\lambda}\right).$$

In our notation (with $b\sim x, a\sim\omega$) this corresponds to (A.11)ff. (where the attenuation is called μ). If we apply integration by parts to the corresponding Radon transform F_μ we find

$$\int\left\{[a\cdot\nabla_b+\mu(a\lambda+b)]\exp\left(-\int_\lambda^\infty\mu(a\tilde{\lambda}+b)\,d\tilde{\lambda}\right)\right\}d\lambda = 0.$$

This equation bears a striking similarity to the transport equation (A.10).

We want to replace the measure $\exp(i\hat{\lambda}\cdot\lambda)d\lambda$ by an arbitrary measure $d\mu(\lambda)$. As the analysis of this problem is complicated we restrict our considerations to $l=1$, $n=3$. We shall actually deal with a more general situation in which μ can be an arbitrary distribution in $\mathcal{S}'$ which is smooth at $\lambda=0$. (We use the notation $d\mu(\lambda)$ for this distribution.) This restriction is necessary in order that, for $f\in\mathcal{S}$,

$$F_\mu(a,b) = \int f(a\lambda+b)\,d\mu(\lambda)$$

is a locally integrable function which satisfies the John equations. The proof of this is contained in the proof of Theorem 6.2; it is the structure of μ only at $\lambda=0$ which is relevant to that argument.

We have already noted that F_μ satisfies the John equations. Moreover F_μ belongs to the space $\mathcal{S}$ in b directions orthogonal to a. *What replaces the invariance equation?*

The invariance equation for $d\mu(\lambda)=d\lambda$ was obtained from the translation invariance of $d\lambda$. Let us examine the behavior of F_μ under the action of translation and multiplication of λ. We have

$$F_\mu(a,b+a\lambda^0) = \int f(a\lambda+b)\,d\mu(\lambda-\lambda^0) \qquad (6.51)$$

$$F_\mu(\alpha a, b) = \int f(a\lambda + b)\, d\mu(\lambda/\alpha). \tag{6.52}$$

Equations (6.51) and (6.52) indicate that we should search for linear combinations of additive and multiplicative translates of μ which vanish, as such equations imply homogeneous equations on suitable b translates of F_μ. These equations are independent of f; they depend only on the "output" F_μ. For example, an additive convolution equation for μ

$$S * \mu = 0 \tag{6.53}$$

for a suitable distribution S implies, by (6.51),

$$(S * F_\mu)(a, b) = \int F_\mu(a, b + a\lambda^0) S(\lambda^0)\, d\lambda^0 = 0 \tag{6.54}$$

for all f, a, b.

Conversely, if (6.54) is satisfied for all f, a, b then $S * \mu$ vanishes on all suitable functions $f(a^0\lambda)$ where, for example, $a^0 = (1, 0, \ldots, 0)$ and f varies. These are arbitrary functions of λ so $S * \mu \equiv 0$.

Unfortunately the set of S such that $S * \mu = 0$ can only (under reasonable conditions on the space to which μ belongs) determine the support of the Fourier transform $\hat{\mu}$ of μ. This results formally from the fact that $\hat{S}\hat{\mu}$ is the Fourier transform of $S * \mu$.

The support of the Fourier transform is a subset of the dual of $\{\lambda\}$. This dual is "large" because $\mathbb{R}$ has many irreducible representations. Thus, in general, the convolution equations satisfied by μ do not restrict μ to lie in a well-defined finite dimensional space unless μ satisfies enough convolution equations so that the number of common zeros of the corresponding $\{\hat{S}\}$ is finite.

The same is true if we apply convolution on the multiplicative group $\mathbb{R}^\times$ based on (6.52).

Underlying these support properties is the theory of *spectral analysis* and *spectral synthesis* which is closely related to the Tauberian principle discussed in Section 1.6. We shall not enter into the details of the theory; we content ourselves with the remark that for the groups that we deal with in this problem the support of the Fourier transform can be determined by homogeneous convolution equations.

The Fourier transform of a function on a group is a function on the set of suitable representations. If the set of such representations is small then we might expect knowledge of the support to be close to knowledge of the function.

We call G the group generated by translations and multiplication of λ. G is generally referred to as the $A\lambda + B$ group. We denote elements of G by (A, B). Multiplication is defined by

$$(A, B)(A', B') = (AA', AB' + B). \tag{6.55}$$

One representation of G, which is called ρ_∞, is the representation on functions on the λ line whose additive Fourier transforms have no mass at the origin. We shall show presently that this representation is irreducible.

Other representations are ρ_m = representation on polynomials of degree m. These are not irreducible for $m \geq 1$ as {polynomials of degree m} $\supset$ {polynomials of degree $m-1$}, nor are they completely reducible.

There is no problem in understanding the representations ρ_m so let us pass to ρ_∞.

If ρ is a representation of G satisfying suitable regularity assumptions (which we shall not detail) then we can choose a basis for the representation space to consist of eigenvectors of the additive subgroup $\{(1, B)\}$ since it is abelian. Once we have one such eigenvector $v_{\hat{B}}$ such that

$$\rho(1, B)v_{\hat{B}} = e^{iB\hat{B}}v_{\hat{B}} \tag{6.56}$$

with $\hat{B} \neq 0$ then for any $A \neq 0$, by (6.55),

$$\begin{aligned}\rho(1, B)[\rho(A, 0)v_{\hat{B}}] &= \rho(A, 0)\rho(1, B/A)v_{\hat{B}} \\ &= e^{iB(\hat{B}/A)}\rho(A, 0)v_{\hat{B}}.\end{aligned} \tag{6.57}$$

This means that $\rho(A, 0)v_{\hat{B}}$ is an eigenvector of $\{(1, B)\}$ with eigenvalue $\exp(iB(\hat{B}/A))$.

We have shown that once we have one nonconstant eigenvector then we obtain a set of eigenvectors whose eigenvalues form the real line except for 0. These eigenvectors span a space which is the space of suitable (depending on our theory) functions on $\mathbb{R}^+ \setminus 0$. This space is therefore irreducible. It is the unique nontrivial irreducible representation of G. A model for this representation is

$$\rho_\infty(A, B)\mu(\lambda) = \mu(A\lambda + B). \tag{6.58}$$

It can be shown that the representations ρ_m, ρ_∞ provide a Fourier transform theory for G for which spectral synthesis holds for the space $\mathcal{S}'$. In particular we can determine the support of μ by the homogeneous left convolution equations it satisfies.

Actually more is true. For if we have a distribution $\nu(\lambda) \in \mathcal{S}'$ whose G Fourier transform belongs to the representation space for ρ_∞ then we can actually determine ν up to a constant multiple. This results from a general property of irreducible representations: the set of (left) convolutions consists of "all" operators on the representation space.

To understand how this determines ν, think of the finite dimensional analog. It is clear that any vector ν^0 in a finite dimensional vector space is determined up to scalar multiplication by the set of matrices U such that $U\nu^0 = 0$.

The same determination property is true for ρ_m even though it is not irreducible. To understand this point, let P be a polynomial. The degree of P is

determined by the smallest m such that P satisfies the additive convolution equation

$$\frac{d^{m+1}}{d\lambda^{m+1}}\delta_0 * P = \pm\frac{d^{m+1}P}{d\lambda^{m+1}} = 0.$$

We write

$$P = a_m\lambda^m + a_{m-1}\lambda^{m-1} + \cdots + a_0$$

with $a_m \neq 0$. We can determine a_{m-1}/a_m using G convolution as follows:

$$\frac{d^m P}{d\lambda^m} = m!a_m$$

(additive convolution) while for any γ

$$P(\gamma\lambda) - \gamma^m P(\lambda) = \gamma^{m-1}a_{m-1}(1-\gamma)\lambda^{m-1} + \dots$$

(multiplicative convolution) so

$$\frac{d^{m-1}}{d\lambda^{m-1}}[P(\gamma\lambda) - \gamma^m P(\lambda)] = (m-1)!\gamma^{m-1}(1-\gamma)a_{m-1}.$$

Thus we determine $\beta_{m-1} = a_{m-1}/a_m$ by the convolution equation

$$m\frac{d^{m-1}}{d\lambda^{m-1}}[P(\gamma\lambda) - \gamma^m P(\lambda)] - \gamma^{m-1}(1-\gamma)\beta_{m-1}\frac{d^m}{d\lambda^m}P(\lambda) = 0 \tag{6.59}$$

satisfied by P. By iteration we can determine all ratios a_{m-j}/a_m and hence we can determine P up to a constant.

From our discussion of the representations ρ_m, ρ_∞ we see that they are related to the structure of $\hat{\mu}(\hat{\lambda})$. If support $\hat{\mu} = \{0\}$ then $\hat{\mu}$ is a linear combination of derivatives of δ so μ is a polynomial and it can be determined up to a constant multiple by the above. Similarly if $\hat{\mu}$ is continuous at $\hat{\lambda} = 0$ then $\hat{\mu}$ belongs to the representation space of ρ_∞ and can be determined up to a constant multiple by convolutions on G as we have explained.

The problem arises when $\hat{\mu}$ is not smooth near the origin. If $\hat{\mu}$ is a distribution of order m near 0 then $\hat{\mu}_1 = \hat{\lambda}^{m+1}\hat{\mu}(\hat{\lambda})$ is continuous at 0. If we use some explicit form of division by $\hat{\lambda}^{m+1}$ then we obtain a distribution $\hat{\mu}_2$ which differs from $\hat{\mu}$ by a linear combination of $\delta^{(j)}$ with $j \leq m$.

In terms of μ or F_μ we can proceed as follows. Suppose we have a bound m on the order of $\hat{\mu}$; this is a bound on the growth of μ at infinity. We form $\mu_1 = d^{m+1}\mu/d\lambda^{m+1}$ to make $\hat{\mu}_1$ smooth at 0. We then use some integration process to construct μ_2 with

$$\frac{d^{m+1}\mu_2}{d\lambda^{m+1}} = \mu_1.$$

μ_2 is determined modulo a polynomial of degree m.

As we have noted, differentiation with respect to λ can be expressed by convolution of F_μ. Since μ_1 belongs to the representation space for ρ_∞ we can determine μ_1 up to a constant multiple by G convolutions.

We have proven

Theorem 6.6 *The homogeneous G convolution equations satisfied by μ determine μ up to a constant multiple if they force support $\hat{\mu} = \{0\}$. If support $\hat{\mu} \neq \{0\}$ then they determine μ up to a constant multiple and the addition of a polynomial.*

Our next task is to determine f from F_μ. If support $\hat{\mu} = \{0\}$ then by differentiating μ sufficiently we can make μ a nonzero constant and then apply Theorem 6.5.

Suppose support $\mu \neq \{0\}$. The theory of mean periodic functions [137] tells us that there is an exponential which is a limit of linear combinations of additive translates (convolutions) $S_t * \mu$ (spectral analysis). This means that there exists a real $\hat{\lambda}^0$ such that

$$\left(\frac{d}{d\lambda} - i\hat{\lambda}^0\right) \lim S_t * \mu = 0 \tag{6.60}$$

but $\lim S_t * \mu \neq 0$. Such $\hat{\lambda}^0$ can be defined in terms of additive convolutions of F_μ since $d/d\lambda$ is an additive convolution operator. This means that

$$\lim S_t * F_\mu = F_{ce^{i\lambda^0\lambda} d\lambda}.$$

We can now apply Theorem 6.5 and the succeeding remarks to determine f up to a constant multiple.

We have shown

Theorem 6.7 *f can be determined up to a constant multiple from F_μ by a process of*

(a) *Forming limits of additive convolutions.*
(b) *Using Theorem 6.5 as modified by replacing $d\lambda$ by $\exp(i\hat{\lambda}^0\lambda)\, d\lambda$ for a suitable $\hat{\lambda}^0$.*

Both (a) and (b) can be defined purely in terms of F_μ.

Problem 6.2 *Let μ be fixed. Is the range of $f \to F_\mu$, $f \in \mathcal{S}'$, the set of functions $H(a, b)$ which*

(i) *Belong to $\mathcal{S}'$*
(ii) *Belong to $\mathcal{S}$ in b variables orthogonal to a*
(iii) *Satisfy the John equations*
(iv) *Satisfy the same G convolution equations (in limit form) as μ?*

Conjecture. *The answer to Problem 6.2 is "yes" if these convolution equations determine μ up to a constant multiple.*

For another, more direct, approach to this problem we write

$$\begin{aligned}\hat{F}_\mu &= \iiint e^{ia\cdot\hat{a}+ib\cdot\hat{b}} f(at+b)\,da\,db\,d\mu(t)\\ &= \iiint e^{ia\cdot\hat{a}+i(y-at)\cdot\hat{b}} f(y)\,da\,dy\,d\mu(t)\\ &= \int e^{ia\cdot\hat{a}}\,da \int e^{iy\cdot\hat{b}} f(y)\,dy \int e^{-iat\cdot\hat{b}}\,d\mu(t)\\ &= \hat{f}(\hat{b}) \int e^{ia\cdot\hat{a}} \hat{\mu}(-a\cdot\hat{b})\,da.\end{aligned}$$

We decompose this a Fourier transform into a part perpendicular to $\hat{b}$ and a part parallel to $\hat{b}$. For each fixed $\hat{b}$, $\hat{\mu}(-a\cdot\hat{b})$ is constant in the a direction orthogonal to $\hat{b}$. Thus its Fourier transform in a is $\delta_{\hat{b}^\perp=0}$ (which is the δ function of the line consisting of those $\hat{a}$ whose $\hat{b}^\perp$ component is 0) times the (one-dimensional) Fourier transform of $\hat{\mu}(-a\cdot\hat{b})$ on the line $a \parallel \hat{b}$ which is $[\mu((\hat{a}\cdot\hat{b})/|\hat{b}|)]/|\hat{b}|$. Finally

$$\hat{F}_\mu(\hat{a}\cdot\hat{b}) = [\hat{f}(\hat{b})\mu((\hat{a}\cdot\hat{b})/|\hat{b}|)\delta_{\hat{b}^\perp}]/|\hat{b}|,$$

when $\hat{b}\neq 0$. It can be shown as in the proof of Theorem 6.5 that for $f\in\mathcal{S}$ and suitable μ there is no essential contribution from $\hat{b}=0$.

Since F_μ is given we can determine $\hat{F}_\mu$. The problem is the determination of $\hat{f}$ and μ from the products $\hat{f}(\hat{b})\mu(\hat{a}\cdot\hat{b}/|\hat{b}|)$. (Since $|\hat{b}|\neq 0$ we could multiply $\hat{F}_\mu$ by $|\hat{b}|$ and thereby remove the last factor $|\hat{b}|^{-1}$.)

The solution of this problem reduces to the case when $\hat{b}$ lies in a fixed line so we may as well assume at this point that $n=1$. The case $\hat{F}_\mu\equiv 0$ is uninteresting (and easily analyzed) so we assume $\hat{F}_\mu\not\equiv 0$. Let $\hat{a}_0,\hat{b}_0$ be fixed with $\hat{f}(\hat{b}_0)\mu((\hat{a}_0\cdot\hat{b}_0/|\hat{b}_0|)\neq 0$. In particular $\hat{f}(\hat{b}_0)=c\neq 0$. As $\hat{a}$ varies we examine

$$\hat{F}_\mu(\hat{a},\hat{b}_0) = c\mu((\hat{a}\cdot\hat{b}_0)/|\hat{b}_0|).$$

The set $\{(\hat{a}\cdot\hat{b}_0)/|\hat{b}_0|\}_a$ is the whole line. We have thus determined μ up to a constant multiple. Next, let $\hat{b}$ vary. Since we are assuming that $n=1$ we have

$$\mu((\hat{a}\cdot\hat{b})/|\hat{b}|) = \mu(\hat{a}\cdot\hat{b}_0/|\hat{b}_0|) = \mu(\hat{a}).$$

This leads to

$$\hat{F}_\mu(\hat{a},\hat{b}) = \hat{f}(\hat{b})\mu(\hat{a}).$$

Since $\hat{F}_\mu$ is known and $\mu(\hat{a})$ has been determined up to a constant multiple, $\hat{f}$ is determined up to the reciprocal constant multiple.

We can also study the case of general μ using the ideas discussed following Theorem 4.15 on decomposition of functions satisfying the first John equation.

We shall restrict ourselves to $l = 1$ as the case $l > 1$ follows the same general lines.

Since $V = \{\hat{a} \parallel \hat{b}\}$ we can parametrize $V^1 = V \setminus \{\hat{b} = 0\}$ by $\{(\hat{\lambda}, \hat{b})\}$ with $\hat{a} = \hat{\lambda}\hat{b}$. Whenever we have such a parametrization there are related decompositions of the space of solutions of the first John equation (defined by V^1) by the slices $\{\hat{\lambda} = \text{const.}\}$ and by the slices $\{\hat{b} = \text{const.}\}$. If $\mathcal{W}$ is a suitable function or distribution space then these slice decompositions are related to a sort of tensor product

$$\mathcal{W}(V) \cong \mathcal{W}(\hat{\Lambda}) \otimes \mathcal{W}(\{\hat{b}\}).$$

The slices tell us how to interpret this tensor product as a direct integral (see Section 1.8). When we use $\{\hat{b} = \text{const.}\}$ as slices then $\{\hat{b}\}$ is the domain of integration and the fibers are the isomorphic spaces $\mathcal{W}(\Lambda)$.

If we fix $\hat{\lambda}$ and let $\hat{b}$ vary then the contributions from this slice are of the form

$$\int e^{ia\cdot\hat{\lambda}\hat{b}+ib\cdot\hat{b}}\, d\nu_{\hat{\lambda}}(\hat{b}) = f_{\hat{\lambda}}(a\hat{\lambda} + b). \tag{6.61}$$

On the other hand if we fix $\hat{b}$ and let $\hat{\lambda}$ vary we obtain

$$\int e^{ia\cdot\hat{\lambda}\hat{b}+ib\cdot\hat{b}}\, d\nu_{\hat{b}}(\hat{\lambda}) = e^{ib\cdot\hat{b}} g_{\hat{b}}(a \cdot \hat{b}). \tag{6.62}$$

Functions of the form (6.61) satisfy the first-order equations (1.40) or (6.11) but functions of the form (6.62) do not satisfy first-order equations. (Of course they satisfy the first John equation.)

In our second proof of Theorem 6.5 we showed that any solution F of the first John equation is the Fourier transform of a measure on V plus a suitable polynomial. Hence, modulo functions which are Fourier transforms of distributions supported on the $\hat{a}$ axis, F can be written either as

$$F(a, b) = \int f_{\hat{\lambda}}(a\hat{\lambda} + b)\, d\mu(\hat{\lambda})$$

or as

$$F(a, b) = \int e^{ib\cdot\hat{b}} g_{\hat{b}}(a \cdot \hat{b})\, d\xi(\hat{b}).$$

Actually it would be more advantageous to replace the affine parameter $\hat{\lambda}$ by a projective parameter $\hat{\theta}$ so as to capture $\{\hat{b} = 0\}$.

According to the ideas formulated in Chapter 4 there should be a WP for solutions of the first John equation with data given on the b axis and "normal derivatives" corresponding to some basis for the measures $d\nu_{\hat{b}}(\hat{\lambda}) \in \mathcal{W}(\hat{\Lambda})$ used in (6.62).

A more interesting WP involves giving data on the a space and "normal derivatives" corresponding to "b derivatives." One practical problem of this genre is the *cone beam problem* in which we give data on a curve γ in $\{b\}$ for all a.

This problem is very important in tomography, in particular in magnetic resonance imaging (MRI). γ represents the set on which measurements of the line integrals are made.

Some conversations with Sarah Patch were useful in developing these ideas.

In standard MRI equipment γ is a helix, say of the form

$$b_1 = \cos t, \qquad b_2 = \sin t, \qquad b_3 = t.$$

The "normal derivatives" are defined by a basis on γ. We choose this basis to be $\{\exp(iut)\}$.

According to the Watergate method we want to parametrize "generically" suitable functions on the slices $V_{\hat{a}}$ of V, which are the points of V whose $\hat{a}$ coordinate is fixed, by the Fourier transform (in b) of $\{\exp(iut)\delta_\gamma\}$. This is the same as using the functions $\{\exp(ia\cdot\hat{a})\exp(iut)\delta_\gamma\}$ as a basis for solutions of the John equation.

At this point we meet a serious analytic problem. The basis functions on $V_{\hat{b}}$ are (formally) given by

$$\begin{aligned}\int e^{i\hat{b}_1\cos t + i\hat{b}_2\sin t + i\hat{b}_3 t + iut}\,dt &= \int e^{i\hat{\rho}\cos(t-\hat{\phi})+i(\hat{b}_3+u)t}\,dt \\ &= e^{i(\hat{b}_3+u)\hat{\phi}}\int e^{i\hat{\rho}\cos t}e^{i(\hat{b}_3+u)t}\,dt. \qquad (6.63)\end{aligned}$$

Here we have written $(\hat{b}_1,\hat{b}_2) = \hat{\rho}\exp(i\hat{\phi})$. We are using $\hat{\rho},\hat{\phi},\hat{b}_3$ as coordinates in $\{\hat{b}\}$.

The last integral has the appearance of a Bessel function. But the integral does not make sense as it stands. However, we can apply the ideas of the Poisson summation formula (Sections 1.7 and 8.2). The function $\exp(i\hat{\rho}\cos t)$ is periodic with period 2π. If $h(x)$ is any periodic function of period 2π then its Fourier (integral) transform is given by

$$\hat{h}(\hat{x}) = \sum \hat{h}_n\delta_n$$

where $\hat{h}_n$ are the Fourier series coefficients of h. (This results from the fact that the Fourier transform of $\exp(inx)$ is δ_n.)

By [11, vol. II, p. 81] the Fourier series coefficients of $\exp(i\hat{\rho}\cos t)$ are (up to a normalizing constant) $i^n J_n(\hat{\rho})$. This shows that

$$\int e^{i\hat{\rho}\cos t}e^{i(\hat{b}_3+u)t}\,dt = \sum i^n J_n(\hat{\rho})\delta_{\hat{b}_3+u=n}. \qquad (6.64)$$

Let $\hat{f}$ be a function on V and call $\hat{f}_{\hat{a}}$ the restriction of $\hat{f}$ to $V_{\hat{a}}$. Fixing $\hat{a}$ fixes $\hat{\phi}$ since $\hat{b} \parallel \hat{a}$. Thus $\hat{f}_{\hat{a}}$ can be regarded as a function of $\hat{\rho},\hat{b}_3$.

Equation (6.64) is a calculation of the basis functions; they are functions of $\hat{\rho},\hat{b}_3$ depending on the parameter u. The Watergate method (Section 1.4) involves

the search for a representation of $\hat{f}_{\hat{a}}$ as an integral of the basis functions. Using (6.63) and (6.64) this means the determination of a function $F(u)$ such that

$$e^{i\hat{b}_3\hat{\phi}} \sum i^n J_n(\hat{\rho}) \int e^{iu\hat{\phi}} \delta_{\hat{b}_3+u=n} F(u) = \hat{f}_{\hat{a}}(\hat{\rho}, \hat{b}_3). \tag{6.65}$$

As usual we solve equation (6.65) by fixing $\hat{b}_3$. Since the integral forces $u = n - \hat{b}_3$, equation (6.65) becomes

$$\sum i^n e^{in\hat{\phi}} G(\hat{b}_3 - n) J_n(\hat{\rho}) = \hat{f}_{\hat{a}}(\hat{\rho}, \hat{b}_3). \tag{6.66}$$

Equation (6.66) can be regarded as the expansion of $\hat{f}_{\hat{a}}(\hat{\rho}, \hat{b}_3)$, considered as a function of $\hat{\rho}$, in terms of the Bessel functions $J_n(\hat{\rho})$. Such expansions are known as Neumann series [11, vol. II, p. 63].

The possibility of such an expansion depends on $\hat{f}_{\hat{a}}(\hat{\rho}, \hat{b}_3)$ being an entire function (of $\hat{\rho}$) of exponential type <1.

If we trace the ideas of the Watergate method we observe that $\hat{f}$ arises from the restriction to V of the Fourier transform of a function $\mathbf{f}(b)$ which, in practice, represents the object being studied. In MRI studies this object is generally a small part of a person so $\hat{\mathbf{f}}$ and hence $\hat{f}$ have the desired exponential growth by the Paley–Wiener theorem.

In Section 1.2 we discussed a nonlinear version of the parametric Radon transform (see (1.56)ff.). It is defined by the map on $\mathbb{R}^1$

$$\lambda \to \left(\sum_0^m a_{1j}\lambda^j, \ldots, \sum_0^m a_{nj}\lambda^j \right).$$

We call $\overrightarrow{\lambda} = (1, \lambda, \ldots, \lambda^m)$ and $\overrightarrow{a}_j = (a_{1j}, \ldots, a_{nj})$ for $j = 0, 1, \ldots, m$. We define the nonlinear parametric Radon transform of degree m by

$$F\left(\boxed{a}\right) = J^{1,m} f \boxed{a} = \int f\left(\boxed{a}\overrightarrow{\lambda}\right) d\lambda.$$

In the notation $J^{1,m}$ the "1" signifies that we are dealing with maps from $\mathbb{R}^1$. It is possible to study $J^{l,m}$ in a similar manner. Here $\boxed{a}$ is the matrix whose columns are $\overrightarrow{a}_j$ and $\overrightarrow{\lambda}$ is a column vector. (The linear Radon transform corresponds to $m = 1$ and, in our previous notation, $b = \overrightarrow{a}_0$ and $a = \overrightarrow{a}_1$.)

Of course $F(\boxed{a})$ contains much more information when $m > 1$ than it does for $m = 1$. There is no problem in specializing the John equations to $\overrightarrow{a}_2 = \cdots = \overrightarrow{a}_m = 0$, hence arriving at the linear case which we have already analyzed. But certain aspects of $m > 1$ are noteworthy so we shall treat $m > 1$ directly. The first John equation is

$$\left(\frac{\partial^r}{\partial a_{i_1 j_1} \ldots \partial a_{i_r j_r}} - \frac{\partial^r}{\partial a_{i_1 k_1} \ldots \partial a_{i_r k_r}} \right) F = 0$$

when $\sum j_p = \sum k_p$. Thus the associated variety V (or V_m) is defined by

$$\hat{a}_{i_1 j_1} \dots \hat{a}_{i_r j_r} - \hat{a}_{i_1 k_1} \dots \hat{a}_{i_r k_r} = 0 \tag{6.67}$$

under the same conditions. To avoid trivial points we can assume $j_p \neq k_p$ for any p.

It is readily verified that (6.67) implies

$$\overrightarrow{\hat{a}}_j \parallel \overrightarrow{\hat{a}}_k \quad \text{for any } j, k. \tag{6.68}$$

But (6.68) is not equivalent to (6.67). For (6.68) says that the rank of the matrix $\boxed{\hat{a}}$ is 1. Suppose $\overrightarrow{\hat{a}}_0 \neq 0$ and $\overrightarrow{\hat{a}}_j = \hat{\lambda}_j \overrightarrow{\hat{a}}_0$ for $j \geq 1$. From (6.67) it follows that $\hat{\lambda}_{j_1} \hat{\lambda}_{j_2} \dots \hat{\lambda}_{j_r}$ depends only on $\sum j_p$ for any $j_1, \dots, j_r$ as long as $\overrightarrow{\hat{a}}_{j_1}, \dots, \overrightarrow{\hat{a}}_{j_r}$ are all $\neq 0$. We call $\hat{\lambda}_1 = \hat{\lambda}$. Upon setting $j_1 = j_2 = \dots = j_l = 1$ on the one hand and $j_1 = j_2 = \dots = j_{l-1} = 0$ and $j_l = l$ on the other we find from (6.67) that $\hat{\lambda}_j = \hat{\lambda}^j$ for all j. This means that we can parametrize V_m by $\hat{b} = \overrightarrow{\hat{a}}_0$ and $\hat{\lambda}$. We can now proceed as in the case $m = 1$ to show that any generic solution of the first John equation is an integral of solutions of the "isotropy equations" $\overrightarrow{J}_\lambda(D)$ which by (1.58) are

$$\left(\frac{\partial}{\partial a_{ij}} - \lambda^j \frac{\partial}{\partial a_{i0}} \right) G_\lambda = 0.$$

We have used the term "generic" because there are other solutions corresponding to $\{\overrightarrow{\hat{a}}_j = 0\}$ which do not play a significant role, as in the case $n = 1$.

To find the invariance equation let $\lambda \to \lambda + \lambda_0$. The integral defining $J^{1,m} f$ is invariant under this transformation which maps

$$\overrightarrow{a}_j \to \sum \binom{j+k}{k} \lambda_0^k \overrightarrow{a}_{j+k}.$$

For the infinitesimal version of invariance we set $\lambda_0 = \epsilon$ and ignore the terms involving ϵ^p for $p > 1$. We obtain

$$\sum (j+1) \left(\overrightarrow{a}_{j+1} \cdot \nabla_{\overrightarrow{a}_j} \right) F = 0,$$

i.e.

$$\left(\sum j \overrightarrow{\hat{a}}_{j-1} \cdot \nabla_{\overrightarrow{\hat{a}}_j} \right) \hat{F} = 0. \tag{6.69}$$

We can write $\hat{F}$ on the part V^1 on V where all $\overrightarrow{\hat{a}}_j \neq 0$ as $\hat{F}(\overrightarrow{\hat{a}}_0, \hat{\lambda} \overrightarrow{\hat{a}}_0, \hat{\lambda}^2 \overrightarrow{\hat{a}}_0, \dots, \hat{\lambda}^m \overrightarrow{\hat{a}}_0)$. V^1 and V are $n+1$ dimensional and $(\hat{\lambda}, \overrightarrow{\hat{a}}_0)$ are the variables on V^1. On V^1 equation (6.69) becomes

$$\frac{d}{d\hat{\lambda}} \hat{F} = 0$$

which implies (under suitable regularity) that

$$\hat{F}\left(\boxed{\hat{a}}\right) = \hat{g}\left(\overrightarrow{\hat{a}}_0\right) \tag{6.70}$$

on V^1.

Equation (6.70) gives the structure of the solutions of the John and invariance equations. The John equations for $m = 1$ are the same as the John equations for F on the subspace $\overrightarrow{a}_2 = \cdots = \overrightarrow{a}_m = 0$. Theorem 6.5 shows that if F satisfies them then $F(\overrightarrow{a}_0, \overrightarrow{a}_1, 0, \ldots, 0)$ is of the form $J^{1,1}f$ so that F is of the form $J^{1,m}f(\boxed{a})$ for general $\boxed{a}$.

All this means that f is determined from $J^{1,m}f$ on the subspace $\overrightarrow{a}_2 = \cdots = \overrightarrow{a}_m = 0$. We are interested in finding other subspaces of $\boxed{a}$ which are determining sets for f in terms of $J^{1,m}f$.

We can put things in a somewhat different framework. Regard V^1 as a fiber space over $\{\overrightarrow{\hat{a}}_0\}$ where the fiber is a complex line. According to the ideas we have set forth in Chapter 4 concerning exotic CPs we can state

Proposition 6.8 *Let us be given a linear subspace $L \subset \boxed{a}$ for which the projection of $\hat{L}$ on $\overrightarrow{\hat{a}}_2 = \cdots = \overrightarrow{\hat{a}}_m = 0$ is surjective. Let $F \in \mathcal{S}'(L)$ satisfy the restrictions of the John and invariance equations to L. Then F is of the form $J^{1,m}f$ on L.*

We can discuss the decompositions of such solutions $F(\boxed{a})$ and the corresponding Watergate problems in much the same way as in the linear case (see (6.52)ff.).

Remark. *We can regard our general parametric Radon transform as a special case of a construction involving families of hyperbolic operators. A function of the form $F_\lambda(a, b) = f(a\lambda + b)$ satisfies the first-order hyperbolic system*

$$\frac{\partial F_\lambda}{\partial a_j} = \lambda \frac{\partial F_\lambda}{\partial b_j} \qquad j = 1, \ldots, n. \tag{6.71}$$

A CS for this system is the b axis and the CD is $f(b)$. The parametric Radon transform relates f to an integral of solutions

$$J^1 f(a, b) = \int F_\lambda(a, b)\, d\lambda.$$

Thus we have formed an average of the solutions of (6.71) and we hope to be able to reconstruct f from this average. Since any one F_λ determines f it is reasonable that f could be determined from any (suitable) single average of the F_λ. A general average corresponds to replacing $d\lambda$ by another measure $d\mu(\lambda)$ (attenuation) and, indeed, our above analysis indicates that this is the case for several attenuations.

Problem 6.3 *For what other families of hyperbolic equations having a fixed CP can we obtain the CD from averages of the solutions?*

6.3 Non-Fourier analysis approach

We now return to the parametric Radon transform. One might be puzzled by the question of what happened to the moment conditions discussed in Chapter 2. How do they fit into the John differential equations for F?

Let us begin with $l = 1$ and $F = J^1 f$. By the invariance equation, $F(a, b)$ *qua* function of b is constant in directions parallel to a. When $|a| = 1$,

$$F(a, b) = (\mathbf{R}^1 f)(\mathbf{g}, b)$$

with $\mathbf{g} = a$ by (6.10). We call w the projection of b on $a^\perp$. The zeroth-order moment M^0 of $\mathbf{R}^1 f$ is given by

$$M^0(\mathbf{R}^1 f) = \int F(a, w)\, dw. \tag{6.72}$$

The zeroth-order moment condition asserts that this is independent of a.

To clarify our ideas we start with the case $n = 2$, $l = 1$. Precisely, we want to prove that if F is small at infinity in b in directions orthogonal to a, e.g. F satisfies the hypothesis of Theorem 6.5, then

$$\left(a_2 \frac{\partial}{\partial a_1} - a_1 \frac{\partial}{\partial a_2}\right) \int_{a^\perp} F(a, w)\, dw = 0. \tag{6.73}$$

Note that $a_2 \frac{\partial}{\partial a_1} - a_1 \frac{\partial}{\partial a_2}$ is the angular derivative in the a plane so (6.73) asserts the independence from a of the integral in (6.72).

Remark. *A word of caution: by Theorem 6.2, since $n = 2$, we cannot expect F to satisfy the first John equation when $a = 0$; for this reason we shall use the equations near $|a| = 1$.*

To prove (6.73) we write it in the form

$$\begin{aligned}\left(a_2 \frac{\partial}{\partial a_1} - a_1 \frac{\partial}{\partial a_2}\right) &\int F(a_1, a_2; -a_2 u, a_1 u)\, du \\ &= \int \left[a_2 F_{a_1} - a_1 F_{a_2} + u a_1 \frac{\partial F}{\partial b_1} + u a_2 \frac{\partial F}{\partial b_2}\right] du = \int -F_{a_1}\, db_1 - F_{a_2}\, db_2.\end{aligned} \tag{6.74}$$

The last equality uses $b_1 = -a_2 u$, $b_2 = a_1 u$ and the invariance condition (6.13) which is $a_1 F_{b_1} + a_2 F_{b_2} = 0$.

The 1 form $\omega = F_{a_1}\, db_1 + F_{a_2}\, db_2$ is closed as we have remarked above. But to conclude that the last integral is zero we have to know that $\omega = dh$ where h is small at infinity in directions orthogonal to a. Unfortunately Proposition 6.3 does not allow that conclusion because we are in dimension $n = 2$.

We might hope to be saved in dimension 2 for a different reason. Suppose our coordinates are adjusted so that $a = (a_1, 0)$. Then $db_1 = 0$ because we are

integrating along $a^\perp$. Thus we do not need the full closure of ω but we need only write $-\partial F/\partial a_2$ in the form $\partial h/\partial b_2$ where h is small in the b_2 direction.

To produce h it suffices to show that

$$\int \frac{\partial F}{\partial a_2}\, db_2 = 0.$$

This does not follow from the first John equation. For

$$\frac{\partial}{\partial b_1}\frac{\partial F}{\partial a_2} - \frac{\partial}{\partial b_2}\frac{\partial F}{\partial a_1} = 0$$

allows the possibility that $F = a_2\alpha(b_2)$ so that

$$F_{a_1} = 0$$
$$F_{a_2} = \alpha(b_2)$$

where $\alpha(b_2) \in \mathcal{S}(b_2)$ but $\int \alpha\, db_2 \neq 0$.

Note that F does not satisfy the hypotheses of Theorem 6.5. In fact

$$F(a_1, a_2; ta_1, ta_2) = a_2\alpha(ta_2)$$

is not constant in t if $a_2 \neq 0$ since $\alpha(ta_2) \to 0$ rapidly as $t \to \pm\infty$.

Problem 6.4 *Suppose F satisfies the hypothesis of Theorem 6.5 for $n = 2$ (except we do not assume that the first John equation holds at the origin). Are the moment conditions satisfied when $|a| = 1$? (Similarly for general l when $a^1, \ldots, a^l$ are orthonormal and $n = l + 1$.)*

When $n > 2$ and $l = 1$ we can proceed as before. We consider a lying in a plane and show that the angular derivative in the plane vanishes. For example, let $n = 3$ and choose coordinates so the plane is $a_3 = 0$. The computation is exactly as in (6.74) for $\int_{a^\perp} F$ which is now the two-dimensional integral

$$\int F(a, \omega)\, d\omega = \iint F(a_1, a_2, 0; -a_2u, a_1u, v)\, du\, dv.$$

The terms $a_1F_{b_1} + a_2F_{b_2} = a \cdot \nabla_b$ vanish as before. Now $\omega = F_{a_1}db_1 + F_{a_2}db_2 + F_{a_3}db_3$ is exact in the suitable sense; namely, by Proposition 6.3, $\omega = dh$ where h has suitable decrease at infinity.

We conclude that $\int_{a^\perp} F$ is constant, which is the zeroth-order moment condition.

For $n > 3$ the computation is the same.

When $n-l \geq 2$ subspaces $\hat{\beta}$ of dimension $< n-l$ are contained in various $\boxed{\hat{a}}^\perp$. (When $n - l = 1$, $\hat{\beta}$ is the origin.) Equation (6.10) asserts that when $a^1, \ldots, a^l$

form an orthonormal basis (in which case we write $\boxed{a}_0$ for $\boxed{a}$), $F(\boxed{a}_o, b) = \mathbf{R}f(\mathbf{g}, b)$ where $\mathbf{g} = \boxed{a}_0$. The moment conditions (Section 2.1) become

$$\hat{}F(\boxed{a}_0, \hat{b})\big|_{\hat{\beta}} \text{ is independent of } \boxed{a}_0 \text{ when } \boxed{a}_0^{\perp} \supset \hat{\beta}.$$

According to the projection slice theorem, restriction to a subspace is the Fourier transform of integration over its orthogonal component. This allows us to rephrase the moment conditions as

$$\int_{\beta^{\perp}} F(\boxed{a}_0, b)\, db \quad \text{is independent of } \boxed{a}_0 \text{ when } \boxed{a}_0 \subset \beta^{\perp}.$$

Actually this condition is meaningless as it stands because $F(\boxed{a}_0, b)$ is constant in directions parallel to $\boxed{a}_0$. We cannot use the usual form of the projection–slice theorem in this situation. For the correct translation of the moment condition we integrate $F(\boxed{a}_0, b)$ only over the orthogonal component of $\boxed{a}_0$ in $\beta^{\perp}$. (It is shown in Chapter 2 that higher moment conditions are unnecessary.) This condition is strongest when $\dim \beta^{\perp}$ is as small as possible, which means $\dim \beta^{\perp} = 1$ (so $\hat{\beta}$ is a hyperplane in $\boxed{a}_0$).

We give the computations for $l = 1, n = 3$ as the general computation for $l = 1$ and any $n \geq 3$ follows the same lines. We can assume that $a = (a_1, a_2, 0)$ and $\beta^{\perp} = \overrightarrow{a} + \{u(-a_2, a_1, 0)\}$. Our task is to show the independence of

$$\int F(a; -a_2 u, a_1 u, 0)\, du$$

from a_1, a_2. This is done exactly as in (6.74)f.

Finally there is the case of arbitrary $l \geq 1$. By choosing a new basis for the l plane we can restrict our considerations to the case in which $a^2 = (0, 0, 1, 0, \ldots, 0)$, $a^3 = (0, 0, 0, 1, 0, \ldots, 0), \ldots$, $a^l = (0, 0, \ldots, 1)$ and $a^1 = (a_1, a_2, 0, \ldots, 0)$ and $\beta^{\perp} = \boxed{a} + \{u(-a_2, a_1, 0, \ldots, 0)\}$. Our above discussion shows that we must prove that the integrals

$$\int F(a; -ua_2, ua_1, 0, \ldots, 0)\, du$$

are essentially independent of a_1, a_2. But this is exactly what is demonstrated in (6.74)f.; we need $n - l \geq 2$ to apply Proposition 6.3.

We have thus shown how the John equations imply the general zeroth-order moment conditions of Chapter 2 (see Proposition 2.5 and Theorem 2.6).

It is interesting to observe that the differential equations also give the homogeneity of F and of $\int_{a^{\perp}} F$ in the variable a. (Of course we no longer assume that $a, a^{\perp}$ form an orthonormal basis.) We again use a two-variable calculation although there is no difficulty in extending everything to $n > 2$. We want to

compute $a_1 \partial F/\partial a_1 + a_2 \partial F/\partial a_2$. We have

$$\left(\frac{\partial}{\partial b_1} - \frac{\partial}{\partial b_2}\right)\left(a_1 \frac{\partial F}{\partial a_1} + a_2 \frac{\partial F}{\partial a_2}\right) = a_1 \frac{\partial^2 F}{\partial a_1 \partial b_1} + a_2 \frac{\partial^2 F}{\partial a_2 \partial b_1} - a_2 \frac{\partial^2 F}{\partial a_2 \partial b_2} - a_1 \frac{\partial^2 F}{\partial a_1 \partial b_2}. \tag{6.75}$$

Now we apply the invariance equation $a_1 F_{b_1} + a_2 F_{b_2} = 0$ which, on differentiation with respect to a_1, yields

$$\frac{\partial F}{\partial b_1} + a_1 \frac{\partial^2 F}{\partial a_1 \partial b_1} + a_2 \frac{\partial^2 F}{\partial a_1 \partial b_2} = 0.$$

Using this to substitute in (6.75) for $a_1(\partial^2 F/\partial a_1 \partial b_1)$ and a similar equation for the derivative with respect to a_2 we find that (6.75) equals

$$\begin{aligned} &- \frac{\partial F}{\partial b_1} + \frac{\partial F}{\partial b_2} + a_2 \left(\frac{\partial^2 F}{\partial a_2 \partial b_1} - \frac{\partial^2 F}{\partial a_1 \partial b_2}\right) + a_1 \left(\frac{\partial^2 F}{\partial a_2 \partial b_1} - \frac{\partial^2 F}{\partial a_1 \partial b_2}\right) \\ &= - \left(\frac{\partial}{\partial b_1} - \frac{\partial}{\partial b_2}\right) F \end{aligned}$$

by (6.12). This means that

$$\left(\frac{\partial}{\partial b_1} - \frac{\partial}{\partial b_2}\right)\left(a_1 \frac{\partial F}{\partial a_1} + a_2 \frac{\partial F}{\partial a_2} + F\right) = 0.$$

Similarly,

$$\left(\frac{\partial}{\partial b_1} + \frac{\partial}{\partial b_2}\right)\left(a_1 \frac{\partial F}{\partial a_1} + a_2 \frac{\partial F}{\partial a_2} + F\right) = 0.$$

Thus $a_1 \partial F/\partial a_1 + a_2 F/\partial a_2 + F$ is constant in b. Since F and its derivatives are small at infinity in directions b orthogonal to a, this constant is 0.

We conclude that F is homogeneous in a of degree -1.

Next let us compute the homogeneity of $\int_{a^\perp} F$. We have

$$\begin{aligned} &\left(a_1 \frac{\partial}{\partial a_1} + a_2 \frac{\partial}{\partial a_2}\right) \int F(a_1, a_2, ; -a_2 u, a_1 u)\, du \\ &\quad = \int (a_1 F_{a_1} + a_2 F_{a_2})\, du + \int (a_1 F_{b_2} - a_2 F_{b_1}) u\, du \\ &\quad = - \int F\, du + \int \frac{dF}{du} u\, du = -2 \int F\, du \end{aligned} \tag{6.76}$$

so that $\int_{a^\perp} F$ is homogeneous of degree -2.

All these calculations are relevant to the zeroth moment. How about higher order moments? They are much more difficult to deal with and require a new technique which is of independent interest. As in the case of the zeroth-order

moment, the essential computation can be made for $n = 2$, $l = 1$ so we restrict ourselves to this situation.

Let us begin with the first-order moment $\int_{a^\perp} Fu\,du$. (We go back to $a, a^\perp$ being an orthonormal basis.) Differentiating with respect to the angle in a gives as before (see (6.74))

$$\left(a_2\frac{\partial}{\partial a_1} - a_1\frac{\partial}{\partial a_2}\right)\int F(a_1, a_2, ; -a_2u, a_1u)u\,du = \int [a_2F_{a_1} - a_1F_{a_2}]u\,du. \tag{6.77}$$

Differentiating again:

$$\begin{aligned}\left(a_2\frac{\partial}{\partial a_1} - a_1\frac{\partial}{\partial a_2}\right)^2 &\int Fu\,du \\ &= \int [a_2^2F_{a_1a_1} - 2a_1a_2F_{a_1a_2} + a_1^2F_{a_2a_2} - a_1F_{a_1} - a_2F_{a_2} \\ &\quad + ua_2^2F_{a_1b_2} - ua_1a_2F_{a_2b_2} + ua_1a_2F_{a_1b_1} - ua_1^2F_{a_2b_1}]u\,du. \end{aligned}\tag{6.78}$$

The term $-a_1F_{a_1} - a_2F_{a_2}$ is just F by the homogeneity of F described above. The terms involving b derivatives can be evaluated by differentiating the invariance equation $a_1F_{b_1} + a_2F_{b_2} = 0$:

$$\begin{aligned} 0 &= \left(a_2\frac{\partial}{\partial a_1} - a_1\frac{\partial}{\partial a_2}\right)(a_1F_{b_1} + a_2F_{b_2}) \\ &= a_1a_2F_{a_1b_1} + a_2^2F_{a_1b_2} - a_1^2F_{a_2b_1} - a_1a_2F_{a_2b_2} + a_2F_{b_1} - a_1F_{b_2} \\ &\quad + u(a_2^2F_{b_2b_2} + a_1^2F_{b_1b_1} + 2a_1a_2F_{b_1b_2}). \end{aligned}\tag{6.79}$$

The term involving second b derivatives is

$$a_1\frac{\partial}{\partial b_1}(a_1F_{b_1} + a_2F_{b_2}) + a_2\frac{\partial}{\partial b_2}(a_1F_{b_1} + a_2F_{b_2}) = 0$$

by invariance. We conclude

$$\begin{aligned}\left(a_2\frac{\partial}{\partial a_1} - a_1\frac{\partial}{\partial a_2}\right)^2 &\int Fu\,du \\ &= \int \left[a_2^2F_{a_1a_1} - 2a_1a_2F_{a_1a_2} + a_1^2F_{a_2a_2} + F - ua_2F_{b_1} + ua_1F_{b_2}\right]u\,du. \end{aligned}\tag{6.80}$$

As in (6.76) the term involving b derivatives is dF/du which leads to

$$\int \frac{dF}{du}u^2\,du = -2\int uF\,du. \tag{6.81}$$

This gives

$$\left(a_2\frac{\partial}{\partial a_1}-a_1\frac{\partial}{\partial a_2}\right)^2\int Fu\,du=-\int uF\,du + \int\left[a_2^2F_{a_1a_1}-2a_1a_2F_{a_1a_2}+a_1^2F_{a_2a_2}\right]u\,du. \tag{6.82}$$

It remains to evaluate the last term which can be written in the following interesting formalism:

$$\frac{1}{2}\int F_{a_1a_1}\,db_2\underset{s}{\otimes}db_2+2F_{a_1a_2}\,db_1\underset{s}{\otimes}db_2+F_{a_2a_2}\,db_1\underset{s}{\otimes}db_1$$

where $d\underset{s}{\otimes}d$ represents a "symmetric tensor product." The John differential equations imply

$$\begin{aligned}\frac{\partial}{\partial b_2}F_{a_1a_1}&=\frac{\partial}{\partial b_1}F_{a_1a_2}\\ \frac{\partial}{\partial b_1}F_{a_2a_2}&=\frac{\partial}{\partial b_2}F_{a_1a_2}\end{aligned} \tag{6.83}$$

from which

$$\frac{\partial^2}{\partial b_2^2}F_{a_1a_1}=\frac{\partial^2}{\partial b_1^2}F_{a_2a_2}$$

follows. These are the formal compatibility conditions for the existence of a function h such that

$$F_{a_1a_1}=h_{b_1b_1},F_{a_1a_2}=h_{b_1b_2},F_{a_2a_2}=h_{b_2b_2}. \tag{6.84}$$

We can write the integral in question formally as

$$\int(d\underset{s}{\otimes}d)h. \tag{6.85}$$

Along the line $\{b=(-a_2u,a_1u)\}$ in the b plane the integral is of the form

$$\int h''(u)u\,du=-\int h'(u)\,du=0 \tag{6.86}$$

provided that h is small at infinity.

Remark. *The equality of the last integral in (6.82) with the left side of (6.86) is a consequence of (6.84). The use of* $\underset{s}{\otimes}$ *is only suggestive.*

The proof of Proposition 6.2 can be extended to give

Proposition 6.9 *For $n \geq 3$ the compatibility conditions (6.83) are sufficient to guarantee that*

$$\sum F_{a_i a_j}\, db_i \underset{s}{\otimes} db_j = (d \underset{s}{\otimes} d)h$$

where for F small at infinity in directions perpendicular to a the same is true of h.

Collecting our results from (6.82)f. we have

$$\left(a_2 \frac{\partial}{\partial a_1} - a_1 \frac{\partial}{\partial a_2}\right)^2 \int_{a^\perp} Fu\, du = -\int_{a^\perp} Fu\, du. \tag{6.87}$$

This means that $\int_{a^\perp} uF$ is a linear combination of $\exp(i\theta_a)$ and $\exp(-i\theta_a)$ for fixed $|a|$ which is exactly the moment condition for the first moment. (We can analyze the homogeneity in a as before.)

There is no difficulty in extending the above to moments of any order. We have thus shown directly that the John equations imply the moment conditions for $n \geq 3$.

Remark 1 *We have shown that the John equations imply the moment conditions. Combined with the results of Section 2.1 this constitutes another proof of Theorem 6.5.*

Remark 2 *There is a great difference between the formalism related to (6.85) (and the analogous formalism for higher order moments) and the situation of zeroth-order moments, namely d. In the latter case we can deform the integral of a closed form to any reasonable curve, whereas in the former cases we must restrict ourselves to straight lines. This difference is founded in the transformation laws of the second derivative and of $d \underset{s}{\otimes} d$; they transform contragrediently only under linear transformation. On the other hand the first derivative and d transform contragrediently under all transformations.*

6.4 Some other parametric linear Radon transforms

Up to this point the Radon transform has involved integration over all affine subspaces of $\mathbb{R}^n$ of a fixed dimension l. Since this Radon transform is overdetermined when $l < n-1$ (i.e. the dimension of the space of affine l planes in $\mathbb{R}^n$ is then $> n$) one might search for suitable subsets Γ of the Grassmannian $\mathbf{G}(l,n)$ which, together with their translates, define a "good" Radon transform. The Tauberian principle enunciated in Chapter 1 says that

$$\text{Injectivity} \iff \bigcup_{\gamma \in \Gamma} \hat{\gamma}^\perp = \mathbb{R}^n. \tag{6.88}$$

(Actually, injectivity does not require equality in (6.88). Rather one needs density of $\cup \hat{\gamma}^\perp$. The meaning of "density"depends on the class of the functions considered.)

To make contact with the parametric Radon transform we shall deal with the "group" case. This means that Γ is defined by

$$\Gamma = \mathbf{G}\Lambda \tag{6.89}$$

where Λ is an "initial" l plane in $\mathbb{R}^n$ and $\mathbf{G}$ is a Lie subgroup of $GL(n)$. We normalize Λ to be the plane $\{\sum \lambda_j e_j\}$ where $e_1, \ldots, e_l$ are the first l unit vectors.

Let $\mathbf{A} = (\mathbf{G}, B)$ be the semi-direct product of $\mathbf{G}$ and B where B is the additive group of $\mathbb{R}^n$. We think of $\mathbf{A}$ as an "affine group." For any $\lambda \in \Lambda$ we denote by $\mathbf{A}_\lambda$ the isotropy group of λ in $\mathbf{A}$. We assume for the present that Λ is a line. The associated Radon transform is defined by (6.1).

In Section 6.1 above we studied the parametric Radon transform by examining functions F_λ on $\mathbf{A}$ which are right $\mathbf{A}_\lambda$ invariant for fixed λ, since this is the case if F_λ is of the form

$$F_\lambda(\mathbf{a}) = f(\mathbf{g}\lambda + b) = f(\mathbf{a}\lambda). \tag{6.90}$$

In the present case the isotropy group is

$$\mathbf{A}_\lambda = \{(\mathbf{g}', b') \text{ with } b' = \lambda - \mathbf{g}'\lambda\}. \tag{6.91}$$

Thus $\mathbf{A}_\lambda$ is parametrized by arbitrary $\mathbf{g}' \in G$; b' is determined by $\mathbf{g}', \lambda$.

A right $\mathbf{A}_\lambda$ invariant function $F_\lambda(\mathbf{g}, b)$ satisfies

$$F_\lambda(\mathbf{g}\mathbf{g}', \mathbf{g}\lambda - \mathbf{g}\mathbf{g}'\lambda + b) = F_\lambda(\mathbf{g}, b) \tag{6.92}$$

because the multiplication in $\mathbf{A}$ is $(\mathbf{g}, b)(\mathbf{g}_1, b_1) = (\mathbf{g}\mathbf{g}_1, \mathbf{g}b_1 + b)$.

As explained in Section 6.1 we obtain the John equations by differentiating (6.92) in $\mathbf{g}'$, meaning in the directions of the Lie algebra $\mathfrak{g}$ of $\mathbf{G}$ at $\mathbf{g} =$ identity. From a practical point of view this amounts to writing $\mathbf{g}' = 1 + t\epsilon$ with $\epsilon \in \mathfrak{g}$ and differentiating with respect to t and then setting $t = 0$. In the previous sections $\mathbf{G} = \mathbf{g}$ was the additive group M_{nl} of $n \times l$ matrices. We could choose as a basis of $\mathfrak{g}$ those $\epsilon = \epsilon_{ij}$ which are matrix units (i.e. the ij entry of the matrix is 1 while all others are 0). The computation of the derivative was easy.

Let us examine what happens if $\mathbf{G} = O(n)$ is the orthogonal group, or rather the connected component of the identity. Its Lie algebra consists of skew-symmetric matrices. A basis for this vector space is $\{(\varepsilon_{ij} - \varepsilon_{ji})\}$ for $i < j$.

For $\mathbf{g}' = 1 + t(\varepsilon_{ij} - \varepsilon_{ji})$ we have

$$(\mathbf{g}\mathbf{g}')_{km} = g_{km} + t \begin{cases} -g_{kj} & \text{if } m = i \\ g_{ki} & \text{if } m = j \\ 0 & \text{otherwise.} \end{cases} \tag{6.93}$$

Hence the infinitesimal form of (6.92) is

$$0 = \frac{\partial F_\lambda}{\partial \mathbf{g}'}(\mathbf{g}, \lambda)$$
$$= \left[\left(g_{ki} \frac{\partial}{\partial g_{kj}} - g_{kj} \frac{\partial}{\partial g_{ki}} \right) - \sum_k \left(\lambda_i g_{kj} \frac{\partial}{\partial b_k} - \lambda_j g_{ki} \frac{\partial}{\partial b_k} \right) \right] F_\lambda(\mathbf{g}, \lambda). \quad (6.94)$$

(Recall that $\lambda_p = 0$ if $p > l$.)

To find the explicit John equation we have to eliminate λ from the operators in (6.94) to obtain the enveloping equation. The John equation is not constant coefficient; the coefficients of the enveloping equation involve the g_{kl}. For this reason we cannot deal with it by the methods used in the proof of Theorem 6.5.

If $F(\mathbf{g}, b) = \int f(\mathbf{g}\lambda + b)\, d\lambda$ then the invariance of $d\lambda$ under translation leads to the invariance equation

$$F(\mathbf{g}, b) = F(\mathbf{g}, b + \mathbf{g}\lambda_0) \quad (6.95)$$

for any λ_0. In infinitesimal form (6.95) becomes

$$\sum_j g_{ji} \frac{\partial F}{\partial b_j}(\mathbf{g}, b) = 0 \quad i = 1, \ldots, l. \quad (6.96)$$

Problem 6.5 *Does the enveloping equation of (6.94) together with the invariance equation (6.96) imply (for suitable functions F) that*

$$F(\mathbf{g}, b) = \int f(\mathbf{g}\lambda + b)\, d\lambda? \quad (6.97)$$

For a simpler example we replace G by its Lie algebra $\mathfrak{g}$. This can be carried out by replacing $\mathbf{g} \in$ (6.90) or (6.97) by an element of $\mathfrak{g}$ and considering $\mathfrak{g}$ as an additive group. The isotropy group is still defined by (6.91) but the "product" $\mathbf{g}\mathbf{g}'$ in (6.92) is to be interpreted as $\mathbf{g} + \mathbf{g}'$. We set $\mathbf{g}' = t(\varepsilon_{ij} - \varepsilon_{ji})$ and let $t \to 0$. We find

$$\left(\frac{\partial}{\partial g_{ij}} - \frac{\partial}{\partial g_{ji}} - \lambda_j \frac{\partial}{\partial b_i} + \lambda_i \frac{\partial}{\partial b_j} \right) f(\mathbf{g}\lambda + b) = 0 \quad (6.98)$$

for $i < j$.

We want to find the enveloping equations of (6.98). This means we want to eliminate λ by taking linear combinations of the differential operators. Since the operators have constant coefficients they commute, so we can regard our task as that of finding the formal compatibility conditions for $\partial/\partial g_{ij} - \partial/\partial g_{ji}$ in order that they be expressible in the form $\lambda_j \partial/\partial b_i - \lambda_i \partial/\partial b_j$ for some λ.

The algebra goes as follows: Suppose $i < j < k \leq l$. Then, writing $\mathbf{g}_{ij}$ for $\partial/\partial g_{ij} - \partial/\partial g_{ji}$ $(= 2\partial/\partial g_{ij})$ we have, upon taking the Fourier transform in all

variables,

$$\begin{aligned}\hat{\mathbf{g}}_{ij} &= -\hat{b}_j\lambda_i + \hat{b}_i\lambda_j \\ \hat{\mathbf{g}}_{jk} &= -\hat{b}_k\lambda_j + \hat{b}_j\lambda_k \\ \hat{\mathbf{g}}_{ik} &= -\hat{b}_k\lambda_i + \hat{b}_i\lambda_k\end{aligned} \tag{6.99}$$

which yields

$$\hat{b}_k\hat{\mathbf{g}}_{ij} + \hat{b}_i\hat{\mathbf{g}}_{jk} - \hat{b}_j\hat{\mathbf{g}}_{ik} = 0 \tag{6.100}$$

as (generic) necessary conditions for the existence of λ.

When $i < j < k$ and $i \le l < j$ the equations simplify to

$$\hat{b}_k\hat{\mathbf{g}}_{ij} - \hat{b}_j\hat{\mathbf{g}}_{ik} = 0. \tag{6.100*}$$

Since $\partial/\partial g_{ij} = -\partial/\partial g_{ji}$ it is convenient to think of $\mathbf{g}_{ij}$ as $\partial/\partial g_{ij}$; that is, we replace the skew-symmetric matrices in $\mathfrak{g}$ by the upper triangular part. If we write (6.99) as a system of equations for λ then the $n(n-1)/2 \times n$ matrix of the system has rank 2. Equation (6.100) represents the compatibility defined by the generic row relations; this is the variety V^1 of (6.4).

These equations are the Lie algebra conditions (i.e. $\mathbf{g}$ close to identity) for F to be a parametric Radon transform when $\mathbf{G}$ is the orthogonal group. An essentially straightforward modification of the proof of Theorem 6.5 leads to

Theorem 6.10 *Let $n > 4$. Under the same hypotheses as in Theorem 6.5, a function $F(\mathbf{g}, b)$ is a parametric Radon transform on the Lie algebra of the orthogonal group if and only if F satisfies the invariance equation (6.96) and*

$$\frac{\partial^2 F}{\partial b_k \partial g_{ij}} + \frac{\partial^2 F}{\partial b_i \partial g_{jk}} - \frac{\partial^2 F}{\partial b_j \partial g_{ik}} = 0 \tag{6.101}$$

for any $i < j < k \le l$ and

$$\frac{\partial^2 F}{\partial b_k \partial g_{ij}} - \frac{\partial^2 F}{\partial b_j \partial g_{ik}} = 0 \tag{6.101*}$$

when $i \le l < j < k$.

The only remark we wish to make about the proof is that (6.100) and (6.100*) are the conditions for the existence of $\hat{\lambda}$ satisfying

$$\hat{\mathbf{g}}_{ij} - \hat{\lambda}_j\hat{b}_i + \hat{\lambda}_i\hat{b}_j = 0 \tag{6.102}$$

so the variety V defined by (6.100) and (6.100*) is fibered by lines of the form (6.102) for varying $\hat{\lambda}$. If we fix $\hat{\lambda}$ and let $\hat{b}$ vary then the Fourier transform of

a function on this part of V with $\hat\lambda$ fixed is of the form

$$\int \exp i \left[\sum_{i<j} g_{ij}(-\hat\lambda_j \hat b_i + \hat\lambda_i \hat b_j) + b\cdot\hat b \right] \hat H(\hat b)\, d\hat b$$
$$= \int \exp i \left[\left(\mathbf{g}\,\hat\lambda\right)\cdot \hat b + b\cdot\hat b \right] \hat H(\hat b)\, d\hat b = H\left(\mathbf{g}\,\hat\lambda + b\right) \tag{6.103}$$

for some measure $\hat H\, d\hat b$.

This gives the decomposition of solutions of (6.101) and (6.101*) into solutions of the first-order system (6.98).

Remark. *There is a difference in the sets* Γ *one obtains using the group* $\mathbf{G}$ *as contrasted to its Lie algebra* $\mathfrak{g}$. $\mathbf{G}\Lambda$ *(see (6.89)) consists of all linear spaces of dimension* l *while* $\mathfrak{g}\Lambda$ *is the set of planes of dimension* $\le l$.

It is reasonable to extend Theorem 6.10 to other Lie groups. Here is the computation for the symplectic group $\mathbf{G} = Sp(n,\mathbb{R})$. (Again we use the Lie algebra.) The symplectic group consists of 2×2 blocks of $n\times n$ matrices satisfying

$$\begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix} \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}' = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \tag{6.104}$$

The Lie algebra $\mathfrak{g}$ is thus the set of matrices satisfying

$$0 = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} + \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} \alpha' & \gamma' \\ \beta' & \delta' \end{pmatrix}$$
$$= \begin{pmatrix} -\beta+\beta' & \alpha+\delta' \\ -\delta-\alpha' & \gamma-\gamma' \end{pmatrix} \tag{6.105}$$

which means that β and γ are symmetric and $\delta = -\alpha'$.

A basis for $\mathfrak{g}$ consists of

$$\begin{aligned} &\{\beta_{ij}(\varepsilon_{i,j+n} + \varepsilon_{j,i+n})\}_{i<j} \\ &\{\beta_{ii}\varepsilon_{i,i+n}\} \\ &\{\gamma_{ij}(\varepsilon_{i+n,j} + \varepsilon_{j+n,i})\}_{i<j} \\ &\{\gamma_{ii}\varepsilon_{i+n,i}\} \\ &\{\alpha_{ij}(\varepsilon_{ij} - \varepsilon_{j+n,i+n})\}_{\text{all } i,j}. \end{aligned} \tag{6.106}$$

We set

$$F(\mathbf{g}, b) = \int f(\mathbf{g}\lambda + b)\, d\lambda \tag{6.107}$$

for $\mathbf{g} \in \mathfrak{g}$ and $b \in \mathbb{R}^{2n}$ and $\lambda \in \Lambda$. It is interesting to study Lagrangian planes, which are maximal isotropic planes of $\mathbb{R}^{2n}$. Isotropic means that the

inner product of any two points $(x, y), (z, w)$ (thought of as column vectors in $\mathbb{R}^n \oplus \mathbb{R}^n$) vanishes. The inner product is

$$(x' \quad y') \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} z \\ w \end{pmatrix} = x' \cdot w - y' \cdot z. \tag{6.108}$$

Lagrangians have dimension n. A prototype of a Lagrangian is $\Lambda = \{(x, 0)\}$. Thus in (6.107) we think of λ as the column vector $\binom{\lambda}{0}$ so

$$\mathbf{g}\lambda = \begin{pmatrix} \alpha\lambda \\ \gamma\lambda \end{pmatrix}. \tag{6.109}$$

To characterize Radon transforms of the form (6.107) we use (6.106). We find, for the derivatives of $f(\mathbf{g}\lambda + b)\, d\lambda$,

$$\begin{aligned} \frac{\partial}{\partial \gamma_{ij}} &= \lambda_j \frac{\partial}{\partial b_{i+n}} + \lambda_i \frac{\partial}{\partial b_{j+n}} \quad i < j \\ \frac{\partial}{\partial \gamma_{ii}} &= \lambda_i \frac{\partial}{\partial b_{i+n}} \\ \frac{\partial}{\partial \alpha_{ij}} &= \lambda_j \frac{\partial}{\partial b_i}. \end{aligned} \tag{6.110}$$

The α_{ij} derivatives lead to

$$\frac{\partial^2}{\partial b_k \partial \alpha_{ij}} = \frac{\partial^2}{\partial b_i \partial \alpha_{ij}}. \tag{6.111}$$

A typical relation involving γ_{ij} derivatives with $i < j$ is

$$\frac{\partial^3}{\partial b_1 \partial b_4 \partial \gamma_{23}} + \frac{\partial^3}{\partial b_3 \partial b_2 \partial \gamma_{14}} - \frac{\partial^3}{\partial b_1 \partial b_4 \partial \gamma_{13}} - \frac{\partial^3}{\partial b_1 \partial b_3 \partial \gamma_{24}} = 0. \tag{6.112}$$

Using the same methods as in the proof of Theorem 6.5 we can show

Theorem 6.11 *Let $n > 2$. Under the same hypotheses as in Theorem 6.5, a function $F(\mathbf{g}, b)$ is the parametric Radon transform "on affine Lagrangians," in the above sense, on $Sp(n, \mathbb{R})$ if and only if $F(\alpha, \gamma; b)$ satisfies the invariance equation and the compatibility conditions defined by (6.110).*

Problem 6.6 *Find a basis for the relations defined by (6.110) as well as for the analogous relations for interesting geometric actions of other simple Lie groups.*

7

RADON TRANSFORM ON GROUPS

Our previous work was centered on various Radon transforms of functions on $\mathbb{R}^n$. The Radon transform on homogeneous spaces of Lie groups G using leaves which are orbits of subgroups of G is the subject of this chapter. One method of analyzing this Radon transform is by embedding the orbit geometry in a euclidean space. In an abstract form this is the *method of affinization* introduced in Section 7.1.

In Section 7.2 the method is applied to the horocyclic Radon transform on the symmetric space of $SL(2,\mathbb{R})$ (upper half-plane). Affinization establishes a relation between the horocyclic Radon transform and the wave equation which leads to a complete understanding of the horocyclic Radon transform. A by-product is the Plancherel formula for this symmetric space.

In Section 7.3 a study is made of the geodesic Radon transform via affinization.

7.1 Affine and projection methods

In Section 1.6 we introduced the study of various Radon transforms on Lie groups G. These are often related to a double fibration

$$\begin{array}{ccc} & G/(K\cap H) & \\ {\scriptstyle \mathbf{R}_K}\swarrow & & \searrow{\scriptstyle \mathbf{R}_H} \\ G/K & \longleftrightarrow & G/H \end{array} \tag{7.1}$$

where K and H are closed subgroups of G. The related Radon transform can be described in two equivalent ways:

(1) *Intertwining.* We regard functions on G/K, G/H as right K or H invariant functions on G. A suitable function on G/K is mapped into the function on G/H

$$\mathbf{R}_{G/K\to G/H} f(g) = \int_H f(gh)\, dh. \tag{7.2}$$

This map is called intertwining because it commutes with the left action of G.

(2) *Fiber space.* We regard G as a fiber space over the base G/K with fiber K and as a fiber space over the base G/H with fiber H. We lift f from the

base G/K to G by making it constant on the fibers and then "push down" to the base G/H by integrating over the fibers.

Remark. *These ways emphasize invariance on K and H respectively. We could have a more general theory in which we identify functions on G/K (resp. G/H) with functions on G which transform according to a fixed irreducible representation of K (resp. H), not necessarily the trivial representation. (Such functions appear in Section 10.6.) The lift and push down have to be modified accordingly.*

We can interpret the Radon transform both geometrically and analytically:

(1) *Geometric.* We call $\mathfrak{o}$ the base point on the manifold G/K, i.e. the point fixed by K. We then regard $f(g) = f(g \cdot \mathfrak{o})$. Thus

$$\begin{aligned} \mathbf{R}_{G/K \to G/H} f(g) &= \int f(gh)\, dh \\ &= \int f(gh \cdot \mathfrak{o})\, dh \\ &= f(gH \cdot \mathfrak{o}), \end{aligned} \tag{7.3}$$

meaning the integral on the set $g(H \cdot \mathfrak{o})$ which is the transform of $H \cdot \mathfrak{o}$ by g.

(2) *Analytic.* We call δ_H the δ function of the subgroup H, thought of as a subset of G, with a suitably normalized Haar measure. We write the convolution on G in the form

$$(f_1 * f_2)(g) = \int f_1(gg_1^{-1}) f_2(g_1)\, dg_1.$$

Then

$$\mathbf{R}_{G/K \to G/H} f(g) = (f * \delta_H)(g) \tag{7.4}$$

(assuming H is unimodular, meaning the Haar measure is invariant under $h \to h^{-1}$).

For the usual Radon transform, G is the affine group A of $\mathbb{R}^n$, K is the rotation group in $\mathbb{R}^n$, and H is the additive group of some fixed hyperplane, e.g. $H = \{x_n = 0\}$. G/K can be identified with $\mathbb{R}^n$. $\mathbf{R}_{G/K \to G/H} f(g)$ is the integral of f over the g transform of the hyperplane H. Since g is a product of a translation s with a rotation $\mathbf{g}$, we obtain the usual Radon transform. $\mathbf{g}$ defines the spread and s is the spread parameter. Only the component of s orthogonal to $\mathbf{g}H$ is significant.

The natural analytic tool for the study of convolutions on a group G is the G analog of Fourier analysis. This involves the irreducible representations of G. When G is abelian we are back with ordinary Fourier analysis. For compact G, representation theory is quite well understood and is not very complicated. But for general noncompact Lie groups many complications arise.

If G is a semi-simple noncompact Lie group then much has been accomplished, especially through the work of Harish-Chandra. This work is a blend of analysis and algebra. As for the analysis, Harish-Chandra asserted

All analysis is eventually abelian.

We shall be directed towards Harish-Chandra's dictum with the modification that we allow analysis on compact groups—although that is essentially an interplay of algebra and abelian analysis.

In Chapter 2 we studied the ordinary Radon transform using abelian analysis, i.e. Fourier analysis on $\mathbb{R}^n$. What happened to the affine group A?

We recall the definition of a *semi-direct* product $G_1 \cdot G_2$ of groups G_1, G_2. Let

$$1 \to G_2 \to G_1 \cdot G_2 \to G_1 \to 0$$

be an exact sequence which splits. This means that G_2 is a normal subgroup and $g \in G_1 \cdot G_2$ has a unique representation

$$g = g_1 g_2.$$

$\{g_j\} = G_j$, the multiplication being as before. Also

$$\begin{aligned} g_2 g_1 &= (g_2 g_1 g_2^{-1}) g_2 \\ &= [\sigma(g_2) g_1] g_2. \end{aligned}$$

σ is a representation (homomorphism) of G_2 in the group of automorphisms of G_1 called an *action* of G_2 on G_1.

An *affine group* A is the semi-direct product of $\mathbb{R}^n$ with a group G acting on $\mathbb{R}^n$ by linear transformations $\sigma(g)$ satisfying

$$\sigma(g)x = gxg^{-1}$$

for $x \in \mathbb{R}^n$.

We want to explain the structure of representations ρ of A. Since $\mathbb{R}^n$ is abelian we should expect to be able to diagonalize the representation space U under the action of the subgroup $\mathbb{R}^n$. Thus a basis for U can be found of elements u which satisfy $\rho(x)u = \lambda(x)u$. ($\rho(x)$ is a linear transformation of U and $\lambda(x)$ is a scalar.) Since $\rho(x + x') = \rho(x)\rho(x')$, λ is a character of $\mathbb{R}^n$ and hence is of the form

$$\lambda(x) = e^{ix \cdot \hat{x}}$$

for some $\hat{x} \in \hat{\mathbb{R}}^n$; the inner product $x \cdot \hat{x}$ is defined by a convenient nondegenerate quadratic form.

For $g \in G$ and u an eigenfunction as above we call $u_g = \rho(g)u$. Then

$$\begin{aligned}\rho(x)u_g &= \rho(xg)u \\ &= \rho(g)\rho(g^{-1}xg)u \\ &= \rho(g)e^{ig^{-1}xg\cdot\hat{x}}u \\ &= \rho(g)e^{i\sigma(g^{-1})x\cdot\hat{x}}u \\ &= e^{ix\cdot\sigma'(g^{-1})\hat{x}}u_g.\end{aligned}$$

Putting things in other terms, once we have produced one eigenfunction u corresponding to an eigenvalue $\exp(ix\cdot\hat{x})$ of $\mathbb{R}^n$ the whole orbit $\exp(ix\cdot\sigma'(G)\hat{x})$ is a set of eigenvalues corresponding to the eigenfunctions $\{u_g\}$.

We can reverse matters somewhat. We start with a linear action σ' of G on $\hat{\mathbb{R}}^n$. Let V be (the closure of) an orbit. Then A acts on the linear combinations of the set of exponentials

$$E_{\mathcal{O}} = \{\exp ix\cdot\hat{x}\}_{\hat{x}\in V}$$

thought of as functions of x, namely

$$g\exp ix\cdot\hat{x} = \exp ix\cdot\sigma'(g)\hat{x}$$

and $\mathbb{R}^n$ acts by translation. We thus obtain a representation of A on functions of x which are limits of suitable linear combinations of the exponentials in $E_{\mathcal{O}}$, i.e. whose Fourier transforms are supported in V.

In many interesting cases the orbits V are algebraic varieties or Zariski open sets of algebraic varieties (algebraic varieties minus proper algebraic subvarieties). They provide a (slice) decomposition of $\hat{\mathbb{R}}^n$ of the form

$$\mathcal{O}_{\vec{c}} = \{\vec{P}(\hat{x}) = \vec{\hat{c}}\}. \tag{7.5}$$

$\vec{P} = (P_1, \ldots, P_r)$ where the P_j are fixed polynomials and $\vec{\hat{c}} = (\hat{c}_1, \ldots, \hat{c}_r)$ are constants; the orbits correspond to the constants $\vec{\hat{c}}$. By the fundamental principle (Section 1.4) the above representation spaces of A correspond to solutions of the equations

$$\left[\vec{P}\left(i\frac{\partial}{\partial x}\right) - \vec{\hat{c}}\right] = 0. \tag{7.6}$$

We are interested in studying function theory on homogeneous spaces G/K. We can start with G/K and construct a related affine group whenever we can find a finite dimensional representation σ of G on some $\mathbb{R}^n$ for which there exists

a point x_K which is fixed exactly by K. The affine group A is the semi-direct product of G with $\mathbb{R}^n$ using the action σ.

We can now concentrate on the orbit $\mathcal{O} = G \cdot x_K \approx G/K$. At this point we can apply our theory of harmonicity if, in fact, the closure of $G \cdot x_K$ is defined by equations of the form (7.5) and if there is an orbit $\hat{\mathcal{O}}_0$ in the action of σ' on $\hat{\mathbb{R}}^n$ which is homogeneous; that is, the polynomials P_j are homogeneous and $\vec{c} = 0$. We also need the strongly free property of the P_j.

We refer to this passage from G to the corresponding affine group A as *affinization*. It involves embedding G in a larger group which is easier to work with.

There is a process, which we call *projection*, which can sometimes be used to go from a known Radon transform to a new Radon transform. Suppose that we have an integral geometric transform on a manifold M with spreads $\{\mathbf{g}\}$ and leaves $\{(s, \mathbf{g})\}$ which we understand. Let $\tilde{M}$ be another manifold. Suppose we have a homeomorphism $\phi : \tilde{M} \to M$. Then we can "lift" the spreads from M to $\tilde{M}$ to obtain a Radon transform on $\tilde{M}$.

So far this process is trivial; but we can obtain nontrivial results by a slight modification. We start with an integral geometric transform on $\tilde{M}$ which we want to analyze. Suppose we can map the leaves on $\tilde{M}$ into leaves $\{(s, \mathbf{g})\}$ on a manifold M for which we understand the associated Radon transform. However, the Radon transform depends also on a measure $\mu(s, \mathbf{g})$ on each leaf. This measure may not coincide with the corresponding measure on the preimage of the leaf $(s, \mathbf{g})$ in $\tilde{M}$. Adjusting the measure on $(s, \mathbf{g})$ to agree with that in $\tilde{M}$ is the same as an *attenuation*, i.e. replacing $\mu(s, \mathbf{g})$ by $\lambda(s, \mathbf{g})\mu(s, \mathbf{g})$.

There is one case in which the attenuation causes no difficulty: there is a global nonvanishing function $h(m)$ on M for which $\lambda(s, \mathbf{g}) = h(m)$ on the leaf $(s, \mathbf{g})$. $h(m)$ depends only on the point $m \in M$. We call ϕ *conformal* when this is the case.

We use the term "conformal" because when the leaves are geodesics or, more generally, totally geodesic submanifolds for metrics $g, \tilde{g}$ on $M, \tilde{M}$ respectively then ϕ is a conformal map.

We can go one step further. Instead of ϕ being a homeomorphism, we allow ϕ to be a covering map. Thus $\tilde{M}$ is regarded as an l sheeted covering of M; the leaves in $\tilde{M}$ "project" onto leaves in M. Again the passage from the Radon transform on M to that on $\tilde{M}$ is easy when the map is conformal.

7.2 The nilpotent (horocyclic) Radon transform on G/K

In this section we shall show how to reduce the study of certain group Radon transforms to problems in Euclidean space using affinization.

For the simplest example, let $G = SL(2, \mathbb{R})$ be the group of 2×2 matrices $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ of determinant 1. (See Section 10.1 for more details on $SL(2, \mathbb{R})$.) We

refer to the notation of Chapter 10:

$$K = \text{maximal compact} = \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \right\} \tag{7.7}$$

$$A = \text{diagonal} = \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \right\} \tag{7.8}$$

$$N = \text{nilpotent} = \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \right\} \tag{7.9}$$

$$W = \text{Weyl} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right\} \tag{7.10}$$

$$M = \pm I. \tag{7.11}$$

We shall sometimes deal with G/M instead of G. For most of our consideration there is very little difference between G and G/M so we shall often write G for G/M when it does not cause any difficulty.

G (or really G/M) acts on $\mathbb{R}^3$ which can be identified with 2×2 real symmetric matrices $X = \left(\begin{smallmatrix} u & y \\ y & v \end{smallmatrix}\right)$

$$g(X) = gXg^t. \tag{7.12}$$

The isotropy group of I is clearly K. The isotropy group of $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ is seen to be MN. The isotropy group of $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ is MA.

Looked at from another viewpoint,

$$G/K = \{\det\, X = 1,\ \mathrm{tr}X > 0\}. \tag{7.13}$$

$$G/MN = \{\det\, X = 0,\ \mathrm{tr}X > 0\}. \tag{7.14}$$

Note that G/K can be identified with the upper half-plane H^+ if we think of G as the conformal group of H^+ acting by fractional linear transformation:

$$z \to gz = \frac{az+b}{cz+d}. \tag{7.15}$$

(This is really an action of G/M as M acts trivially.) The isotropy group of $z = i$ is clearly K. To make the explicit identification of H^+ with (7.13) we note that, group-theoretically,

$$z = \xi + i\eta = \begin{pmatrix} \eta^{1/2} & \xi\eta^{-1/2} \\ 0 & \eta^{-1/2} \end{pmatrix} \cdot i. \tag{7.16}$$

Thus $\xi + i\eta$ is identified with the point in (7.13)

$$\begin{pmatrix} \eta^{1/2} & \xi\eta^{-1/2} \\ 0 & \eta^{-1/2} \end{pmatrix} I \begin{pmatrix} \eta^{1/2} & 0 \\ \xi\eta^{-1/2} & \eta^{-1/2} \end{pmatrix} = \begin{pmatrix} \eta + \xi^2\eta^{-1} & \xi\eta^{-1} \\ \xi\eta^{-1} & \eta^{-1} \end{pmatrix}. \tag{7.17}$$

We can also exhibit a more convenient model for the (positive) light cone Γ^+ of (7.14). Note that, under the usual linear transformation of the plane, the isotropy group in G of the point $(1,0)$[1] is defined by $a = 1, c = 0$ and so is N. Thus the plane (minus the origin) $P = G/N$. This means that P is a two sheeted covering of Γ^+. The explicit two sheeted map can be realized in two ways:

(1) The Iwasawa decomposition $G = KA^+N$, which suggests that $P = KA^+(1,0)$ (which is clear). The KA^+ coordinates of (x, y) are exactly polar coordinates. If we apply ka to $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ in the three-dimensional representation we obtain

$$ka \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a^2 \cos^2\theta & -a^2 \cos\theta \sin\theta \\ -a^2 \cos\theta \sin\theta & a^2 \sin^2\theta \end{pmatrix}. \tag{7.18}$$

The two-to-one nature of the map is seen from the fact that $\theta \to \theta + \pi$ leaves the right side of (7.18) invariant. (The right side is, in fact, a function of 2θ.)

(2) The quadratic (albanese) map of the plane into $\mathbb{R}^3$

$$(\alpha, \beta) \to \begin{pmatrix} \alpha^2 & \alpha\beta \\ \alpha\beta & \beta^2 \end{pmatrix}. \tag{7.19}$$

It is readily verified, e.g. by comparison with (7.18), setting $\alpha = a\cos\theta$, $\beta = -a\sin\theta$, that this maps P in a two-to-one fashion onto Γ^+. From a group theoretical point of view, the three-dimensional representation is realized as the symmetric square of the representation on P.

Looked at from the three-dimensional viewpoint, G is the component of the identity of the orthogonal (Minkowski) group $SO(1,2)$. This results from the fact that

$$\det gX = \det X \tag{7.20}$$

and $\det X$ is a nondegenerate indefinite quadratic form. The corresponding Laplacian is the wave operator

$$\begin{aligned} \Box &= 4\frac{\partial^2}{\partial u \partial v} - \frac{\partial^2}{\partial y^2} \\ &= \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} \end{aligned} \tag{7.21}$$

with

$$t = \tfrac{1}{2}(u + v) \quad x = \tfrac{1}{2}(u - v).^2 \tag{7.22}$$

(The coefficient 4 in (7.21) is necessary to make $\Box$ invariant.)

[1] We generally think of vectors as column vectors. Due to manuscript technicalities, vectors such as (1,0) are written as though they are row vectors.

[2] The letter y is used both for the off-diagonal entry of X and for a euclidean coordinate. The transformation (7.22) implicitly identifies these two coordinates.

Now for the geometry. The group N leaves invariant the point $\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$ and, in fact, the whole ray $\rho_0 = \left(\begin{smallmatrix}a&0\\0&0\end{smallmatrix}\right)$. G acts linearly so acts on points, lines, planes, etc. Moreover G preserves the Minkowski metric and hence preserves Minkowski orthogonality. Since N leaves $\left(\begin{smallmatrix}a&0\\0&0\end{smallmatrix}\right)$ invariant it also leaves invariant (though not pointwise) the plane Q which is (Minkowski) orthogonal to the ray through this point. It is easily seen that the inner product is (up to an unimportant constant)

$$X \cdot X' = uv' + u'v - 2yy'. \tag{7.23}$$

Thus the plane Q, which is orthogonal to $(u', v', y') = (1, 0, 0)$, is defined by $v = 0$.

We use the plane Q to define the spread of planes

$$Q^s = Q + \begin{pmatrix}0&0\\0&s\end{pmatrix} = \{X|\ v = s\}$$

which means

$$Q^s = \left\{X\Big|\ X \cdot \begin{pmatrix}1&0\\0&0\end{pmatrix} = s\right\} = \left\{\begin{pmatrix}u&y\\y&s\end{pmatrix}\right\}. \tag{7.24}$$

Thus $\{Q^s\}$ is a spread of N invariant planes.

The other rays ρ_θ on Γ^+ can be obtained from ρ_0 by rotation by $k_\theta \in K/M$. The isotropy group of ρ_θ is the conjugate N^θ of N. In this way we obtain a family of spreads whose "Grassmann" parameter space is K/M. We denote the leaves of these spreads by Q^s_θ. From the euclidean point of view, Q is the plane which is tangent to the light cone and passes through the generator $v = y = 0$. In t, x, y coordinates the plane is defined by $t - x = 0$ and the generator is $\{t - x = 0, y = 0\}$.

A *horocycle* on a homogeneous manifold V of G is a set of the form gNO where O is the "origin" in V. In particular for $V = G/K$ the origin is the K fixed point. We use the Iwasawa decomposition and write $g = kan$. Since A normalizes N we have

$$\begin{aligned} kan\,NO &= ka\,NO \\ &= kaNa^{-1}aO \\ &= kNk^{-1}kaO \\ &= N^k\,ka\,O \end{aligned} \tag{7.25}$$

where N^k is the conjugate of N by k.

Since $G = KAK$, $\{kaO\} = G/K$ so the horocycles on G/K can also be described as the orbits of the conjugates of N on G/K.

The *nilpotent (horocyclic) Radon transform* $\mathbf{R}^N f$ is defined as the set of integrals of f over the horocycles (see (7.28)). We fix one Haar measure on N and that defines the measure on all horocycles. We shall discuss this from the Minkowski viewpoint presently.

We focus our attention on functions on G/K. Looked at from the three-dimensional (affinization) viewpoint we have

$$Q^1 \cap G/K = NI$$

since the identity I is the base point on G/K and

$$\begin{aligned} Q^1 &= \left\{ \begin{pmatrix} u & y \\ y & 1 \end{pmatrix} \right\} \\ Q^1 \cap G/K &= \left\{ \begin{pmatrix} 1+n^2 & n \\ n & 1 \end{pmatrix} \right\} \\ &= NI \end{aligned}$$

by (7.17) with $\eta = 1, \xi = n$. Hence the horocycles are of the form

$$gQ^1 \cap G/K = gNO.$$

Note that

$$aQ^1 = \left\{ \begin{pmatrix} a^2 u & y \\ y & a^{-2} \end{pmatrix} \right\}$$

so

$$\begin{aligned} aQ^1 \cap G/K &= \left\{ \begin{pmatrix} a^2(1+y^2) & y \\ y & a^{-2} \end{pmatrix} \right\} \\ &= Q^{a^{-2}} \cap G/K. \end{aligned} \tag{7.26}$$

This means that we can regard $\{aQ^1 \cap G/K\} = \{aNO\}$ as a horocyclic spread on G/K; this spread is the intersection of the spread

$$\{Q + \left(\begin{smallmatrix} 0 & 0 \\ 0 & s \end{smallmatrix}\right)\}$$

in Minkowski space with G/K.

By (7.25) the general horocycle on G/K is of the form $kaNO$. We call $Q_\theta = kQ$ so Q_θ is the orthogonal plane to

$$\begin{aligned} k \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} &= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \cos^2\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & \sin^2\theta \end{pmatrix} \end{aligned}$$

which is $\frac{1}{2}(1, \cos 2\theta, -\sin 2\theta)$ in (t, x, y) coordinates. Q_θ can also be described as the tangent plane to the light cone containing the generator $\{k\left(\begin{smallmatrix} u & 0 \\ 0 & 0 \end{smallmatrix}\right)\}$.

We call Q^s_θ the translate of Q_θ by $k\left(\begin{smallmatrix} 0 & 0 \\ 0 & s \end{smallmatrix}\right)$, so that $Q^s_\theta = kQ + k\left(\begin{smallmatrix} 0 & 0 \\ 0 & s \end{smallmatrix}\right) = kQ^s$. Since G preserves the Minkowski inner product we can write

$$Q^s_\theta = \left\{X \middle| X \cdot k\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right) = s\right\}.$$

This shows that the spread $\{Q^s_\theta\}_s$ is formed by parallel planes.

In analogy with (7.26) we have

$$\begin{aligned} Q^{a^{-2}}_\theta \cap G/K &= kQ^{a^{-2}} \cap G/K \\ &= kaQ^1 \cap G/K. \end{aligned} \tag{7.26*}$$

Hence for each k the spread $\{Q^s_\theta \cap G/K\}_s$ of horocycles on G/K can be described as the intersection with G/K of the spread $\{Q^s_\theta\}_s$ of parallel planes in $\mathbb{R}^3$.

In our spread terminology k (or θ) is the Grassmann parameter and s is the spread parameter.

We have thus realized a goal of affinization: the manifold G/K is realized as a submanifold of $\mathbb{R}^3$ and the horocyclic spreads are intersections with G/K of $\mathbb{R}^3$ (abelian) spreads. We expect that the analysis of the horocylic Radon transform should be abelian.

Since N fixes $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ it follows from (7.24) that $Q^s \cap G/K$ is N invariant. We have

$$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1+y^2 & y \\ y & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix} = \begin{pmatrix} 1+(y+n)^2 & y+n \\ y+n & 1 \end{pmatrix}$$

which shows that N acts simply transitively on $Q^1 \cap G/K$ and similarly on any $Q^s \cap G/K$. (This also follows from our above observation that $Q^1 \cap G/K = NI$.) The same is true for the action of N^k on $Q^s_\theta \cap G/K$. The Minkowski metric, being N invariant, induces a Haar measure on each $Q^s_\theta \cap G/K$. This measure can also be realized as the metric on $Q^s_\theta \cap G/K$ which is induced from the Minkowski metric on G/K or on Q^s_θ because G preserves the Minkowski metric.

Let us compute the explicit metric on $Q^s \cap G/K$ in terms of N. We can write

$$\begin{aligned} Q^s \cap G/K &= N \begin{pmatrix} s^{-1} & 0 \\ 0 & s \end{pmatrix} \\ &= \left\{ \begin{pmatrix} s^{-1} + n^2 s & ns \\ ns & s \end{pmatrix} \right\}. \end{aligned} \tag{7.27}$$

The Minkowski metric $d\tilde{s}^2 = -\frac{1}{4}du\,dv + dy^2$ becomes a constant multiple of

$$\begin{aligned} d\tilde{s}^2 &= [d(ns)]^2 \\ &= s^2\,dn^2 \end{aligned}$$

since s is a constant on Q^s.

We have shown that the Minkowski metric corresponds to the normalization

$$d\tilde{s} = s\,dn$$

of the Haar measure of N on $Q^s \cap G/K$ in the above coordinates. (s is positive on G/K; negative s corresponds to $t < 0$.)

We call δ_{Q_θ} the δ function on Q_θ with the Minkowski metric. Let f be a suitable function on G/K; we identify f with a measure on $\mathbb{R}^3$ by multiplying f with $\delta_{G/K}$. We have

$$\begin{aligned} \mathbf{R}^N f(s,\theta) &= \int_{Q^s_\theta \cap G/K} f \\ &= (\delta_{Q_\theta} * f)(s). \end{aligned} \tag{7.28}$$

Since δ_{Q_θ} is invariant under translations in the Q_θ plane the convolution $\delta_{Q_\theta} * f$ depends only on s (and θ).

In the Poincaré upper half-plane the origin is the K fixed point $z = i$ so NO is the line $\Im z = 1$. The corresponding spread is the set of lines $\Im z = \text{const}$. The spread corresponding to the Grassmann parameter θ is the set of circles tangent to the real axis at the point $z = k\infty = -\cot\theta$.

There is one more geometric construction which is fundamental for us. By Γ^+_X we denote the translate of the positive light cone Γ^+ of (7.14) by X. Since Γ^+ is G invariant, if X is invariant under a subgroup L of G then Γ^+_X is L invariant, so $\Gamma^+_X \cap G/K$ is also invariant under L.

In particular, let X_0 be a point on Γ^+ which is N invariant. Then $\Gamma^+_{X_0} \cap G/K$ is a union of horocycles; we claim it is a single horocycle.

It is actually more transparent to make the computation in (t,x,y) coordinates. Then for $X_0 = (r,r,0)$ we must solve

$$\begin{aligned} t^2 - x^2 - y^2 &= 1 \quad (\text{this is } G/K) \\ (t-r)^2 - (x-r)^2 - y^2 &= 0 \quad (\text{this is } \Gamma^+_{X_0} \text{ for } X_0 = (r,r,0)) \end{aligned} \tag{7.29}$$

which gives

$$v = t - x = \frac{1}{2r}. \tag{7.30}$$

This means that the intersection of $\Gamma^+_{(r,r,0)}$ with G/K is the same as the intersection of $v = 1/2r, t \geq 0$, with G/K which is exactly $Q^{1/2r} \cap G/K$. The plane

$t - x = 1/2r$ has the Minkowski normal $(1, 1, 0)$ which is, of course, consistent with the fact that the plane is invariant under the isotropy group of $(1, 1, 0)$.

We can make the same computation for points $k(r, r, 0)$ which make up the whole of Γ^+. In this way the light cone construction gives an explicit identification of the horocycles with points on Γ^+.

As we have noted many times in this book, the primary function of spreads is to give us a way of introducing analysis into geometry. Analysis is introduced by replacing the δ functions of the individual leaves of the spread by suitable linear combinations or limits of such (spread functions). In particular we shall concentrate on exponentials which are spread functions; this is Fourier analysis.

We now come to an important observation. Let us define the Fourier transform using the Minkowski inner product (7.12) instead of the usual definition using the euclidean inner product. Recall that the plane Q_θ is the orthogonal to the generator L_θ of the light cone. Thus by the projection–slice theorem

$$\delta_{Q^s_\theta} = \widehat{e^{i\lambda s}\delta_{\hat{L}_\theta}} \tag{7.31}$$

where λ is a parameter on $\hat{L}_\theta$. In particular, since support $\delta_{\hat{L}_\theta} \subset$ light cone,

$$\Box\delta_{Q^s_\theta} = 0; \tag{7.32}$$

that is, $\delta_{Q^s_\theta}$ *are solutions of the wave equation.* (This can readily be proven directly.)

This means that we are using the restrictions of solutions of $\Box$ to G/K to define $\mathbf{R}^N$ (see (7.28)). Thus the whole theory of the horocyclic Radon transform on G/K fits into the framework of harmonic function theory for $\Box$—this is what we gain from abelianization.

Equation (7.31) is a relation between $\mathbf{R}^N$ and the light cone $\hat{\Gamma}$. Our above observation (centering around (7.29) and (7.30)) that the horocycles can be interpreted as the intersections with G/K of the translates of Γ^+ by points in Γ^+ can be used, via $\mathbf{R}^N$, to relate functions on G/K to functions on Γ^+. Thus $\mathbf{R}^N f$ defines the function $\mathbf{R}^N_{\Gamma^+} f$ on Γ^+ by

$$\mathbf{R}^N_{\Gamma^+} f(X) = \int_{(\Gamma^+ + X) \cap G/K} f. \tag{7.33}$$

By the discussion following (7.30) the spreads $\{Q^s_\theta \cap G/K\}_s$ correspond to the points X on the ray θ in Γ^+.

The relation amongst $\mathbf{R}^N$, the light cone, and the wave operator is capable of generalization to nongroup situations. As in Chapter 3 let $\vec{\imath}$ be a set of strongly free homogeneous polynomials. Suppose that the variety V of common zeros[3] of the $i_j(\hat{x})$ is "nice," meaning that the $\vec{\imath}$ harmonic functions are spanned by the

[3]We are using a somewhat inconsistent notation by writing V although it is a variety in $\hat{x}$ space.

Fourier transforms of derivatives of δ functions along the generators $\hat{L}_\theta$ of V as in Theorem 3.5. We call Q_θ the orthogonal to L_θ in some suitable nondegenerate quadratic form. Then we can define a *general horocyclic Radon transform* $\mathbf{R}^N$ of suitable functions f on $V_{\vec{c}} : \{\vec{\imath}(x) = \vec{c}\}$ by

$$\mathbf{R}^N f(\theta, s) = (\delta_{Q_\theta} * f)(s). \tag{7.34}$$

Here s is a suitable parameter. f is identified with the measure $f\delta_{\vec{\imath}=\vec{c}}$ on $\mathbb{R}^n$. We assume that $\delta_{Q_\theta} * f$ is well defined, which entails that translates of Q_θ are not tangent to $V_{\vec{c}}$. Since $\delta_{Q_\theta} * f$ is constant in the directions of Q_θ it is a function of some suitable s.

It would be convenient to choose s as the natural parameter on L_θ but, as in the case of $L =$ light cone, it may happen that $L_\theta \subset Q_\theta$, in which case s has to be chosen as a parameter on the translates Q^s_θ of Q_θ to force $\mathbb{R}^n = \bigcup Q^s_\theta$.

We now return to our main concern which is the relation described above between G/K and $\hat{\Gamma}$. As the study of $\mathbf{R}^N$ is somewhat complicated we begin with an outline of our approach.

Equation (7.32) and the preceding discussion show that $\mathbf{R}^N f$ (say for $f \in \mathcal{D}(G/K)$) is defined by considering f as an element of the dual of the kernel of $\Box$. However, the solutions u of $\Box u = 0$ are determined by their CD on G/K—that is, two functions on G/K whereas f is only 1. We can remedy this situation by injecting $\mathcal{D}(G/K)$ into $\mathcal{D}(G/K) \oplus \mathcal{D}(G/K)$. Of course there are many such injections; we find it convenient to choose the *Kelvin odd* injection $f \to (0, f)$. (This will be clarified below.) It is natural to think of $(0, f)$ as the CD of a Kelvin odd solution F of $\Box F = 0$. We often identify f with F.

$\mathbf{R}^N$ is defined using the "basis" $\{\delta_{Q^s_\theta}\}$ for solutions of $\Box$. (The fundamental principle and (7.31) show that this is a basis.) As in the case of the ordinary Radon transform it is more convenient to use the Fourier basis $\{\exp[i(t\hat{t} - x\hat{x} - y\hat{y})]\}_{\hat{t}^2=\hat{x}^2+\hat{y}^2}$. This leads to

$$\begin{aligned}\hat{f}(\hat{t}, \hat{x}, \hat{y}) &= \int_{G/K} f(t, x, y) e^{it\hat{t} - ix\hat{x} - iy\hat{y}} \\ &= e^{it\hat{t} - ix\hat{x} - iy\hat{y}} \cdot f\delta_{G/K}\end{aligned}$$

where $\delta_{G/K}$ is the δ function of G/K with the G invariant measure. We consider $\hat{f}$ as a function on $\hat{\Gamma}$ because the fundamental principle tells us that exponentials with frequencies in $\hat{\Gamma}$ form a basis for solutions of $\Box$. We can regard $\hat{f}$ as

$$\hat{f} = \mathcal{F}_{G/K \to \hat{\Gamma}} f$$

which is the *Fourier transform of f from G/K to $\hat{\Gamma}$*.

We are now faced with the problem of inverting this Fourier transform. The Fourier transform of any function on $\hat{\Gamma}$ is a solution of $\Box$. We shall show that the inverse Fourier transform of $\hat{f} \operatorname{sgn} \hat{t}$ is the Kelvin odd solution F whose CD is

$(0, f)$. By writing $\hat{f}$ as an integral over $\hat{\theta}$ of functions on the rays of $\hat{\Gamma}$ this inverse Fourier transform also inverts the Radon transform $\mathbf{R}^N$ as in the analogous situation for $\mathbf{R}$ on $\mathbb{R}^n$.

It remains to characterize $\widehat{\mathcal{D}(G/K)}$. Certainly such functions $\hat{\alpha}$ are entire functions of exponential type on $\hat{\Gamma}$ (i.e. restrictions of entire functions of exponential type to $\hat{\Gamma}$). It is somewhat complicated to prove that they have suitable growth conditions in real directions because $f \in \mathcal{D}(G/K)$ is *not* smooth in $\mathbb{R}^3$ but only on G/K. Thus the Fourier transform of f is not small in all real directions on $\mathbb{R}^3$ and it requires a rather subtle argument to prove that $\hat{f}$ is small along the rays of $\hat{\Gamma}$. We derive conditions in (7.47)ff. which deal with the relation of the compact support of f to the behavior of $\hat{f}$ in $\Im t$. We provide another condition which characterizes the Kelvin odd nature of functions $\hat{\alpha}$ on $\hat{\Gamma}$. (We call $\hat{\alpha}$ Kelvin odd if its (inverse) Fourier transform is Kelvin odd.)

As in the case of the ordinary Radon transform we need moment conditions to pass from the $\{Q^s_\theta\}$ basis to the Fourier basis; these moment conditions allow us to pass from $\mathbf{R}^N$ to the Fourier transform.

We list the steps which we use to carry out this program:

(**A**) We use harmonic function theory (Chapter 3) to prove the injectivity of $\mathbf{R}^N$ (for general $\vec{\imath}$).

Henceforth we specialize $\vec{\imath}$ to be $\hat{t}^2 - \hat{x}^2 - \hat{y}^2$.

(**B**) We study the analog of the moment conditions of Chapter 2. This involves the question of determining compatibility conditions on a set of functions $\{\hat{f}_{\hat{\theta}}\}$ defined on $\{\hat{L}_{\hat{\theta}}\}$ in order that there exist a smooth function $\hat{f}$ on V whose restriction to each $\hat{L}_{\hat{\theta}}$ is $\hat{f}_{\hat{\theta}}$.

(**C**) Using an inner product on solutions h of $\Box h = 0$ we express the Fourier transform $\hat{f}$ of $f\delta_{G/K}$ as the inner product of the solution of the CP for $\Box$ with CD $= (0, f)$ with exponential solutions.

(**D**) This allows us to regard $f \longrightarrow \hat{f}$ as a *Fourier transform on varieties* which maps functions on the algebraic variety G/K into functions on the algebraic variety $\hat{\Gamma}$. Functions $\hat{f}$ in the image of this Fourier transform are odd with respect to a certain transformation called the Kelvin transformation.

(**E**) Using the CP we write the inversion formula for this Fourier transform when f is the δ function of a point on G/K. This leads to a general Fourier inversion formula.

For simplicity we restrict our considerations in (**F**) and (**G**) to K invariant functions on G/K (called functions of *type 0*).

(**F**) For any suitable K invariant function $\hat{u}$ on $\hat{\Gamma}$ we compute the Watergate data of the function U on the t axis; this is the inverse Fourier transform of $\hat{u} \operatorname{sgn} \hat{t}$. (Clearly $\Box U = 0$.) This allows us to give an intrinsic characterization in terms of $\hat{u}$ of Kelvin invariance of U.

(**G**) We relate the euclidean Fourier transform of $u\delta_{G/K}$ for $u \in \mathcal{D}(G/K)$ to the group Fourier transform on G/K. As a by-product we derive the Fourier inversion formula (Plancherel formula) for K invariant functions on

G/K. We also need this relation to prove that $\hat{f}$ is rapidly (polynomially) decreasing on the rays of $\hat{\Gamma}$.

(**H**) We extend the above to all functions in $\mathcal{D}(G/K)$ (not necessarily K invariant).

This completes the inversion formula for $\mathbf{R}^N$ on $\mathcal{D}(G/K)$. The same ideas apply to other spaces of functions on G/K.

Remark. *We shall prove* (**A**) *for general* $\vec{\imath}$ *for which* V *satisfies the regularity of Theorem 3.5. After that we shall restrict our considerations to the group* $SL(2,\mathbb{R})$ *and the corresponding light cone. We shall make some comments on the possibilities of dealing with more general* $\vec{\imath}$.

Now for the details.

(**A**) Let $\mathcal{W}(V_{\vec{c}})$ be a space of functions f on $V_{\vec{c}}$ which are exponentially decreasing. It is a standard fact in polynomial approximation that functions $u \in \mathcal{W}'$ can be approximated in the topology of $\mathcal{W}'$ by polynomials (see [72]). In Chapter 3 we showed that any polynomial p agrees with a harmonic polynomial on $V_{\vec{c}} = \{\vec{\imath} = \vec{c}\}$. Thus the harmonic polynomials are dense in $\mathcal{W}'$.

The regularity of V means (see Section 3.1) that a basis for harmonic polynomials can be constructed from spread harmonic polynomials. If $f \in \mathcal{W}$ and $\mathbf{R}^N f = 0$ then f is orthogonal to all spread harmonic polynomials, as is clear from (7.34) and Theorem 3.5. Thus $\mathbf{R}^N f = 0$ means f is orthogonal to $\mathcal{W}'$, i.e. $f \equiv 0$.

Theorem 7.1 *If* V *satisfies the regularity condition then* $\mathbf{R}^N$ *is injective on spaces* $\mathcal{W}$ *of functions which are (at least) exponentially decreasing. In fact, if* V *is connected and* $\mathbf{R}^N f$ *vanishes on an infinite number of spreads* $\{\theta\}$ *for which* $\bigcup \hat{L}_{\hat{\theta}}$ *is not contained in any analytic subvariety of* V *then* $f \equiv 0$.

Proof We have proven the injectivity of $\mathbf{R}^N$. Since f is exponentially decreasing $\hat{f}$ is holomorphic. The vanishing of $\mathbf{R}^N f$ on the spread θ means $f * \delta_{Q_\theta} = 0$. By Fourier transform $\hat{f} = 0$ on $\hat{L}_{\hat{\theta}}$. If $\hat{f} = 0$ on $\{\hat{L}_{\hat{\theta}}\}$ for an infinite number of θ such that $\bigcup \hat{L}_{\hat{\theta}}$ is not contained in any analytic subvariety of V and V is connected, then $\hat{f}$ vanishes on V so, by the above, $f = 0$.

In particular Theorem 7.1 applies to the space $\mathcal{D}(V_{\vec{c}})$.

(**B**) As we mentioned, our further considerations will be centered on $V =$ light cone; we shall make some remarks on the possibility of extension to more general V.

Claim. *For each* l *the dimension of the homogeneous harmonics of degree* l *is* $2l + 1$.

Proof $h(t, x, y)$, being a solution of the wave equation, is equivalent to the usual harmonicity of $h(it, x, y)$. Since the change from t to it does not affect homogeneity, our claim follows from the standard result for ordinary harmonic functions.

The number $2l+1$ can be understood from another point of view: The dimension of the space of homogeneous polynomials P of degree l is $\binom{l+2}{2}$. To be harmonic (or a solution of the wave equation) means all coefficients of $\Box P$ vanish. There are $\binom{l}{2}$ such coefficients. Thus the harmonics have dimension $\binom{l+2}{2} - \binom{l}{2} = 2l+1$.

We shall give a different explanation below.

For the ordinary Radon transform the compatibility conditions arose from the condition that smooth functions on lines fit together by means of COMP ORIGIN to form a function which is smooth at the origin. Functions on lines appear because they are the Fourier transforms of the Radon transform on spreads (see Chapter 2). In the case of $\mathbf{R}^N$, the Radon transform of f on the spread θ is given by convolution with δ_{Q_θ}. Using (7.31) we deduce that the Fourier transform of $\mathbf{R}^N f(\theta)$ is the restriction of $\hat{f}$ to the generator of the light cone defined by θ. Thus we are in a situation which is similar to that in Chapter 2.

To go further we make a general

Definition. *A function $\hat{F}$ defined on a set σ is said to be smooth (holomorphic) at a point $p \in \sigma$ if $\hat{F}$ extends to be smooth (holomorphic) in a neighborhood of p in $\mathbb{R}^n(\mathbb{C}^n)$.*

Note that if $\hat{F}$ is a smooth function on $\hat{\Gamma}$ then $\hat{F}$ can be uniquely written on $\hat{\Gamma}$ in the form

$$\begin{aligned}\hat{F}(\hat{t},\hat{x},\hat{y}) &= \frac{1}{2}\left[\hat{F}(\sqrt{\hat{x}^2+\hat{y}^2},\hat{x},\hat{y}) + \hat{F}(-\sqrt{\hat{x}^2+\hat{y}^2},\hat{x},\hat{y})\right] \\ &\quad + \hat{t}\,\frac{\hat{F}(\sqrt{\hat{x}^2+\hat{y}^2},\hat{x},\hat{y}) - \hat{F}(-\sqrt{\hat{x}^2+\hat{y}^2},\hat{x},\hat{y})}{2\sqrt{\hat{x}^2+\hat{y}^2}} \\ &= \hat{F}_0(\hat{x},\hat{y}) + \hat{t}\hat{F}_1(\hat{x},\hat{y}). \end{aligned} \tag{7.35}$$

Equation (7.35) is crucial for understanding smoothness at the origin of the light cone.

In Chapter II of FA we showed that if $\hat{F}$ is holomorphic on $\hat{\Gamma}$ (or on $\hat{\Gamma}$ near 0) then $\hat{F}_0$ and $\hat{F}_1$ are holomorphic in $\mathbb{C}^n$ (resp. locally). (The converse is obvious.) The same argument applies with one minor change, namely the replacement of convergence of power series by Taylor's formula with remainder, to C^∞ functions. The essential idea of the proof is to use the equation $\hat{t}^2 = \hat{x}^2 + \hat{y}^2$ of the light cone to reduce all powers of $\hat{t}$ to $\hat{t}^0$ and $\hat{t}^1$.

In studying the compatibility conditions at the origin for functions which are smooth on lines it suffices, as in Chapter 2, to restrict our considerations to real generators of $\hat{\Gamma}$. (Generators lying in $\hat{t}=0$ present an additional complication.)

Our first observation is that if we choose any $2l+1$ generators $\{\lambda\theta_i\}$ of $\hat{\Gamma}$ which are distinct from each other and none is equal to the reflection of another in the $\hat{x}, \hat{y}$ plane then

$$\{[\theta_i \cdot (t,x,y)]^l\}$$

is a basis for homogeneous harmonics of degree l.

To prove our contention we invoke the CP for $\Box$: the CD is given by

$$\mathcal{C}h = \left[h(0,x,y), \frac{\partial h}{\partial t}(0,x,y)\right].$$

We write $\theta_i = (\theta_i^1, \theta_i^2, \theta_i^3)$ where the θ_i^j are real and $\theta_i^1 \neq 0$; we have

$$\mathcal{C}[\theta_i \cdot (t,x,y)]^l = [(\theta_i^2 x + \theta_i^3 y)^l, l\theta_i^1(\theta_i^2 x + \theta_i^3 y)^{l-1}].$$

According to our construction $\{(\theta_i^2, \theta_i^3)\}$ define $2l+1$ distinct lines in $\mathbb{R}^2$. Thus, by Lemma 1.1, $[(\theta_i^2, \theta_i^3) \cdot (x,y)]^l, [(\theta_i^2, \theta_i^3) \cdot (x,y)]^{l-1}$ span all pairs of polynomials which are homogeneous of respective degrees $l, l-1$.

This verifies

Proposition 7.2 *If $\theta_i = (\theta_i^1, \theta_i^2, \theta_i^3)$ are $2l+1$ real vectors on $\hat{\Gamma}$ such that $\{(\theta_i^2, \theta_i^3), (-\theta_i^2, -\theta_i^3)\}$ are distinct and $\theta_i^1 \neq 0$ then*

$$\{[\theta_i \cdot (t,x,y)]^l\}$$

is a basis for homogeneous harmonics (solutions of the wave equation) of degree l. In particular this space has dimension $2l+1$.

Now suppose we have two such bases. Changing from one basis to another leads to relations

$$[\theta_i \cdot (t,x,y)]^l = a_{ij} \sum [\theta'_j \cdot (t,x,y)]^l. \tag{7.36}$$

Thus for any $f \in \mathcal{W}$

$$\int f[\theta_i \cdot (t,x,y)]^l = \sum \int a_{ij} f[\theta'_j \cdot (t,x,y)]^l. \tag{7.37}$$

These relations are *moment conditions* for f in analogy with the moment conditions discussed in Chapter 2. As in Chapter 2 we shall study these moment conditions by means of the Fourier transform.

Equation (7.36) can be stated in terms of $\mathbf{R}^N f$. For, $[\theta \cdot (t,x,y)]^l$ is a spread function for the spread defined by θ (see the discussion preceding (7.34)). Hence (7.36) represents relations amongst the $\mathbf{R}^N f(\theta_i)$ and $\mathbf{R}^N f(\theta'_i)$.

Suppose that, conversely, we are given functions $g(\theta, s)$ for each θ which satisfy, in the notation of (7.36) and (7.37),

$$\int g(\theta_i, s)\{[\theta_i \cdot (t,x,y)](s)\}^l \, ds = \sum a_{ij} \int g(\theta'_i, s)\{[\theta'_i \cdot (t,x,y)](s)\}^l \, ds \tag{7.38}$$

for all relations (7.36). (Since $\theta_i \cdot (t,x,y)$ is a spread function for θ_i the expression $[\theta_i \cdot (t,x,y)](s)$ is defined.)

We want to prove that there is an f on G/K with $\mathbf{R}^N f = g$.

We shall prove this later. For the present we want to prove a preliminary result. If such an f exists then our above discussion of the relation of the Fourier transform to $\mathbf{R}^N$ shows that

$$\hat{f} = \hat{\mathbf{g}} \quad \text{on } \hat{\Gamma}.$$

We have written $\hat{\mathbf{g}}$ for the function on the generators of $\hat{\Gamma}$ which is defined by the Fourier transform in s of $g(\theta, s)$.

We are left with the problem of determining conditions for holomorphic functions defined on generators of $\hat{\Gamma}$ to fit together to form a holomorphic function on $\hat{\Gamma}$. In case $\hat{\Gamma}$ is replaced by $\mathbb{R}^n$ this problem is solved in Section 2.1 using COMP ORIGIN. We shall reduce the problem from $\hat{\Gamma}$ to $\mathbb{R}^n$.

We have shown above that the CD of $\{[\theta_i \cdot (t,x,y)]^l\}_i$ spans all pairs of polynomials which are homogeneous of degrees $(l, l-1)$. Thus, given any $P^0(x,y)$ which is homogeneous of degree l we can find $\{\alpha_i\}, \{{}^0\theta_j\}$ such that

$$\mathrm{CD}\left\{\sum \alpha_i [{}^0\theta_i \cdot (t,x,y)]^l\right\} = (P^0, 0). \tag{7.39}$$

Hence if we have a relation

$$\hat{h}^0 \equiv \sum \beta_j P_j^0 = 0 \tag{7.40}$$

it leads to

$$\sum \beta_j \alpha_{ij} [{}^0\theta_i \cdot (t,x,y)]^l = 0 \tag{7.41}$$

since $\mathrm{CD}(A) = 0$ is equivalent to $A = 0$.

Let $\hat{F}$ be a "nice" function on $\hat{\Gamma}$. By (7.35) we can write

$$\hat{F} = \hat{F}^0 + i\hat{t}\hat{F}^1.$$

Note that (7.35) does not require that $\hat{F}$ be a holomorphic function on $\hat{\Gamma}$; it suffices that $\hat{F}$ is holomorphic on each generator of $\hat{\Gamma}$ and then $\hat{F}^0, \hat{F}^1$ are holomorphic on the lines through the origin in $\{\hat{t} = 0\}$.

We want to verify that $\hat{F}^0$ and $\hat{F}^1$ satisfy COMP ORIGIN. We can restrict our considerations to homogeneous $\hat{F}$. Suppose, in the notation of (7.41), that

$$\sum \beta_j \alpha_{ij} \left[{}^0\theta_i \cdot \left(\frac{\partial}{\partial \hat{t}}, \frac{\partial}{\partial \hat{x}}, \frac{\partial}{\partial \hat{y}}\right)\right]^l \hat{F} = 0 \qquad \text{on } \hat{\Gamma}. \tag{7.42}$$

(This is defined because $\hat{F}$ is a function on the generators of $\hat{\Gamma}$ and $\theta_{ij} \cdot \left(\partial/\partial\hat{t}, \partial/\partial\hat{x}, \partial/\partial\hat{y}\right)$ are differentiations at the origin along generators of $\hat{\Gamma}$.) Then by (7.35), (7.39), and (7.40)

$$\sum \beta_j P_j^0 \left(\frac{\partial}{\partial \hat{x}}, \frac{\partial}{\partial \hat{y}}\right) \hat{F}^0 = 0.$$

We have shown that the compatibility (7.42) implies that $\hat{F}^0$ satisfies COMP ORIGIN.

A similar result is valid for $\hat{F}^1$.

The result for $\mathbb{R}^2$ in place of $\hat{\Gamma}$ (Section 2.1) gives

Theorem 7.3 *Let $\hat{F}$ be a holomorphic function on the generators of $\hat{\Gamma}$ which depends analytically on θ and satisfies*

$$\text{COMP ORIGIN}(\hat{\Gamma}),$$

meaning

$$\sum \gamma_j [\phi_j \cdot (t,x,y)]^l = 0 \tag{7.43}$$

implies

$$\sum \gamma_j \left[\phi_j \cdot \left(\frac{\partial}{\partial \hat{t}}, \frac{\partial}{\partial \hat{x}}, \frac{\partial}{\partial \hat{y}}\right)\right]^l \hat{F}(0,0,0) = 0. \tag{7.44}$$

Then $\hat{F}$ is holomorphic on $\hat{\Gamma}$.

Remark. *In Chapter IX of FA we constructed general formulas of the genre (7.35). Since (7.35) is the essential ingredient in the proof of Proposition 7.2 and of Theorem 7.3 we can extend our arguments to prove analogs of these results for $\vec{\imath}$ which are Cauchy hyperbolic, which means there is a CP like that for $\Box$.*

(**C**) We now pass to a study of the CP for the wave operator $\Box$ on G/K. This will enable us to relate the Fourier transform $\hat{f}$ of a suitable function f on G/K to the CP.

Hyperbolic spherical coordinates are defined by

$$\begin{aligned} t &= r\cosh\zeta \\ x &= r\sinh\zeta\cos\theta \\ y &= r\sinh\zeta\sin\theta \end{aligned} \tag{7.45}$$

For $r > 0$ ($r < 0$) these form a coordinate system in the interior of the upper (lower) light cone. Points on the light cone are represented in the form $(t, t\cos\theta, t\sin\theta)$ with $t \neq 0$.

The euclidean measure becomes

$$dt\,dx\,dy = r^2 \sinh\zeta\, dr\, d\zeta\, d\theta. \tag{7.46}$$

If f is a smooth function of compact support on G/K (which we identify with $r = 1, t > 0$) then for $(\hat{t}, \hat{x}, \hat{y}) = (\hat{t}, \hat{t}\cos\hat{\theta}, \hat{t}\sin\hat{\theta}) \in \hat{\Gamma}$

$$\begin{aligned} &\hat{f}(\hat{t},\hat{x},\hat{y}) \\ &\quad = \iint f(\cosh\zeta, \sinh\zeta\cos\theta, \sinh\zeta\sin\theta) e^{i\hat{t}[\cosh\zeta + \sinh\zeta\cos(\theta-\hat{\theta})]} \sinh\zeta\, d\zeta\, d\theta. \end{aligned} \tag{7.47}$$

By expanding a general f in terms of characters of K we can reduce to the situation where f transforms like a character in θ, i.e. $f(k_\theta X) = \exp(im\theta)f(X)$. (See (**H**) for more details.) Transformation under a nontrivial character does not differ much from transformation under the trivial character, so we begin by concentrating on K invariant functions.

f and hence also $\hat{f}$ are now functions of $t, \hat{t}$ respectively since K is the group of rotations about the t axis. We can replace $\cos(\theta - \hat{\theta})$ by $\cos\theta$ so that

$$\hat{f}(\hat{t}) = \iint f(t) \exp\left\{ i\hat{t}[t + (t^2-1)^{1/2}\cos\theta] \right\} d\theta\, dt. \tag{7.48}$$

Suppose support $f(\zeta) \subset \{-\log A \le \zeta \le \log A\}$. Since $t = \cosh\zeta$

$$A^{-1} \le t + (t^2-1)^{\frac{1}{2}}\cos\theta \le A \tag{7.49}$$

so that $\hat{f}$ is an entire function of exponential type and for $\hat{t}$ real

$$\left|\hat{f}(i\hat{t})\right| \le c \begin{cases} e^{A|\hat{t}|} & \hat{t} \to -\infty \\ e^{-|\hat{t}|/A} & \hat{t} \to +\infty. \end{cases} \tag{7.50}$$

Similar estimates are valid in the entire half-planes as $\Re\hat{t} \to -\infty$ (resp. $\Re\hat{t} \to +\infty$.)

If we apply the Paley–Wiener theorem to $\hat{f}(\hat{t})$ we find that $\hat{f}$ is the one-dimensional Fourier transform of $f_1(t)$ where

$$\text{support } f_1 \subset \{A^{-1} \le t \le A\}. \tag{7.51}$$

However, the smoothness of f does not translate readily into conditions on f_1 because f (more precisely $f\delta_{G/K}$) is smooth only in the directions of G/K so $\hat{f}$ is certainly not small in all directions. We postpone a discussion of the decrease of f_1.

We can think of $\hat{f}$ in another way, one which provides us with a closer relation to the wave equation.

It is a standard fact that the CP for the wave operator $\Box$ with data on $r = 1$ is well posed for C^∞ solutions u inside the forward light cone. The CD of u can be described by

$$\mathrm{CD}(u) = \left(u\big|_{G/K}, r\,\frac{\partial u}{\partial r}\Big|_{G/K} \right). \tag{7.52}$$

We can define a nondegenerate skew-symmetric product of two solutions u, v by

$$(u,v)(r) = \iint_r \left[u\frac{\partial v}{\partial r} - v\frac{\partial u}{\partial r} \right] d\Omega \tag{7.53}$$

where $d\Omega = r^2 \sinh\zeta\, d\zeta\, d\theta$ is the normalized G invariant measure on $r = \text{const.}$ (see (7.46)). We assume u, v satisfy conditions of smoothness and smallness at infinity which allow our formal calculations to be valid.

In hyperbolic spherical coordinates

$$\Box = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} - \frac{1}{r^2}\Delta \tag{7.54}$$

where Δ is the group invariant Laplacian on G/K. Δ is the unique (up to a constant) second-order differential operator on G/K which is G invariant.

Let us make the important observation

$$\begin{aligned}
&\frac{\partial}{\partial r}\iint_{r=\text{const.}}\left[v\frac{\partial u}{\partial r} - u\frac{\partial v}{\partial r}\right] d\Omega \\
&\quad = \iint\left[v\frac{\partial^2 u}{\partial r^2}r^2 + 2v\frac{\partial u}{\partial r}r - u\frac{\partial^2 v}{\partial r^2}r^2 - 2u\frac{\partial v}{\partial r}r\right]\sinh\zeta\, d\zeta\, d\theta \\
&\quad = \iint\left[v\left(\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r}\right) - u\left(\frac{\partial^2 v}{\partial r^2} + \frac{2}{r}\frac{\partial v}{\partial r}\right)\right] r^2\sinh\zeta\, d\zeta\, d\theta \\
&\quad = \iint\left[v\Delta u - u\Delta v\right] r^2\sinh\zeta\, d\zeta\, d\theta \\
&\quad = 0.
\end{aligned} \tag{7.55}$$

The third equality follows from $\Box u = \Box v = 0$ while the last is valid because Δ is formally self-adjoint on $r = \text{const.}$ and u, v satisfy conditions at infinity which allow the formal self-adjointness to become an actual self adjointness.

Equation (7.55) shows the invariance of the skew-symmetric inner product defined by (7.53) under variation of r.

Remark. *The skew symmetry of the inner product is not important; the crucial points are the nondegeneracy and invariance of the inner product.*

We can understand the inner product and its invariance properties in terms of integration by parts. Let us integrate $u\Box v - v\Box u$ in the region between $r = r_1$ and r_2. The contribution to the integral which comes from the G/K Laplacian Δ vanishes because Δ is self-adjoint on the sets $r = \text{const}$. We are left with the r part $\partial^2/\partial r^2 + (2/r)\partial/\partial r$ of $\Box$. Integration by parts on the r lines leads to (if $r_1 > r_2$)

$$\begin{aligned}
\iiint_{r_1}^{r_2}(u\Box v - v\Box u)\, d\Omega &= \iint_{r_1}\left(u\frac{\partial v}{\partial r} - v\frac{\partial u}{\partial r}\right) d\Omega - \iint_{r_2}\left(u\frac{\partial v}{\partial r} - v\frac{\partial u}{\partial r}\right) d\Omega \\
&= (u, v)(r_1) - (u, v)(r_2).
\end{aligned} \tag{7.56}$$

Since the left side of (7.56) vanishes we have again established the invariance of the inner product (7.53).

Let F be the solution of $\Box F = 0$ with $\mathrm{CD}(F) = (0, f)$ on G/K and let v be the solution

$$\begin{aligned} v &= \exp(it\hat{t} - ix\hat{x} - iy\hat{y}) \\ &= \exp\left[ir\hat{t}(\cosh\zeta - \sinh\zeta\cos(\theta - \hat{\theta}))\right] \end{aligned}$$

for $(\hat{t}, \hat{x}, \hat{y}) = (\hat{t}, \hat{t}\cos\hat{\theta}, \hat{t}\sin\hat{\theta}) \in \hat{\Gamma}$. The CD of v on G/K is

$$\mathrm{CD}(v)_{r=1} = \left\{ v\big|_{r=1}, \left[i\hat{t}(\cosh\zeta - \sinh\zeta\cos(\theta - \hat{\theta}))\right] v\big|_{r=1} \right\}. \tag{7.57}$$

Thus

$$(F, v) = -\hat{f}(\hat{t}, \hat{x}, \hat{y}). \tag{7.58}$$

Here and in the following we identify a function f on G/K with the measure $f\delta_{G/K}$ on $\mathbb{R}^3$; $\hat{f}$ is the Fourier transform on $\mathbb{R}^3$ of this measure.

Remark. *A comparison of (7.58) with (7.31), (7.34) shows that the Fourier transform of $\mathbf{R}^N f$ in the spread parameter is related to the solution of the CP with data $(0, f)$. This relation is made more precise in Theorem 7.8 below.*

Since $\mathrm{CD}(F) = (0, f)$ the set of solutions $\{u\}$ is really only "half" the solutions of $\Box$. Such solutions are characterized by vanishing on $r = 1$.

Remark. *In terms of general systems $\vec{\imath}$ the inner product (u, v) of solutions is termed energy. It would take us too far afield to discuss energy in any detail.*

(**D**) We introduce the *Kelvin transformation*

$$w(r, \zeta, \theta) \to \mathcal{K}w(r, \zeta, \theta) = r^{-1}w(r^{-1}, \zeta, \theta). \tag{7.59}$$

A simple direct computation shows that

$$\Box\mathcal{K}f = \phi(r^5\Box f)$$

where ϕ is the map $r \to r^{-1}$. This equation shows that $\mathcal{K}$ preserves the kernel of $\Box$. Since $\mathcal{K}f = \phi(rf)$ we can rewrite the equation in the form (since $\phi = \phi^{-1}$)

$$\phi\Box\phi = r^5\Box r^{-1}.$$

In the terminology of [58] the map ϕ is *conformal* for $\Box$; that is, conjugation of $\Box$ by ϕ is equivalent to left and right multiplication by factors called *conformal weights*.

We shall discuss the deeper meaning of $\mathcal{K}$ below.

Instead of defining the CD by the operators $(\mathrm{id}, r\partial/\partial r)$ we replace $r\partial/\partial r$ by $r\partial/\partial r + 1/2$. This is advantageous since $r\partial/\partial r + 1/2$ is odd under $\mathcal{K}$, meaning

$$\mathcal{K} \circ (r\partial/\partial r + 1/2) = -(r\partial/\partial r + 1/2) \circ \mathcal{K}. \tag{7.60}$$

Since $r = 1$ we can ignore powers of r. Thus $\mathcal{K}(F)$ is the solution of the CP with

$$\mathrm{CD}(\mathcal{K}(F)) = \left(0, -\left(r\partial/\partial r + 1/2\right) F\big|_{r=1}\right) \tag{7.61}$$

$$= (0, -f). \tag{7.62}$$

We have shown that

$$\mathcal{K}(F) = -F. \tag{7.63}$$

Conversely if $\mathcal{K}(F) = -F$ then $F\big|_{r=1} = 0$.

This means that

The set of solutions F of $\Box F = 0$ whose CD on $r = 1$ is of the form $(0, f)$ is characterized as being the set of solutions of $\Box F = 0$ which are odd under the Kelvin transformation.

A consequence of our study of the CP is that the Fourier transforms $\hat{f}$ of functions f on $r = 1$ when restricted to $\hat{\Gamma}$, which we term the *Fourier transform from G/K to $\hat{\Gamma}$*, constitute only "half" of the functions on $\hat{\Gamma}$ since, by the fundamental principle, solutions of $\Box$ correspond to "arbitrary" functions or measures on $\hat{\Gamma}$ while the CD of solutions coming from $\{(0, f)\}$ is only half the possible CD. In order to determine this "half" we first characterize the solutions F which are odd under $\mathcal{K}$ in a manner which lends itself to Fourier analysis.

Remark. *The Kelvin transform selects (by oddness) CD of the form $(0, f)$. As we have seen in Chapter 4 the CP is closely related to harmonicity in the direction orthogonal to the Cauchy surface. In the case where harmonicity is determined by a reflection group W we can use representation theory of W to pick out special subspaces of CD. In the present situation W is the two-element group and $\mathcal{K}$ can be considered as a generator of this group.*

(**E**) We want to show that the Fourier transform of the Kelvin condition (7.62) together with suitable regularity conditions determine the range of the Fourier transform on G/K, and, consequently, the range of the nilpotent Radon transform on G/K. In so doing we shall also construct an inversion for the Fourier transform from G/K to $\hat{\Gamma}$. In order to accomplish this goal we have to understand how to find the range and the inverse of a map. One technique involves finding certain objects in the range and then expressing all potential elements in terms of these special objects. For example, one proof of the ordinary Fourier inversion formula starts with the explicit calculation of the Fourier transforms of $\exp(-ax^2)$ for varying a and then showing how to express "arbitrary" functions as linear combinations of $\{\exp(-ax^2)\}_a$.

The important observation in this connection is that we can compute the Fourier transform $\hat{f}$ on $\hat{\Gamma}$ when f is the δ function of $(1, 0, 0)$ on G/K. Such $\hat{f}$ and some simple modifications form our "special objects." In fact

Theorem 7.4 *The solution of the CP on G/K with CD $= (0, \delta_{(1,0,0)})$ is the same as the solution of the CP on $T =$ tangent plane to G/K at $(1, 0, 0)$ with the same CD.*

The value of our assertion is that it is easy to compute the solution of the CP on T since T is linear.

Proof We call $\Delta = \Delta_{(1,0,0)}$ the solution of the CP for $\Box$ with data $(0, \delta_{(1,0,0)})$ on T. The propagation speed of solutions of the wave equation is 1. This means that

$$\text{support}\,\Delta \subset \text{closure interior}\,\Gamma + (1,0,0).$$

In particular

$$(\text{support}\,\Delta) \cap G/K = \{(1,0,0)\}$$

so that $\text{CD}_{G/K}(\Delta)$ has support at $(1,0,0)$.

$\text{CD}_T(\Delta) = (0, \delta_{(1,0,0)})$ means that

(a) $\lim \int_{T+\epsilon(1,0,0)} \psi\Delta = 0$

(b) $\lim \frac{2}{\epsilon} \left[\int_{T+\epsilon(1,0,0)} \psi\Delta - \int_{T-\epsilon(1,0,0)} \psi\Delta\right] = \psi(1,0,0)$

for smooth functions ψ: (a) says that $\lim \Delta\big|_{T+\epsilon(1,0,0)} = 0$ and (b) computes the normal derivative of Δ, i.e. $\partial/\partial t$, on T.

$\text{CD}_{G/K}(\Delta)$ can be characterized in a similar fashion.

Since $\text{support}\,\Delta \subset \Gamma + (1,0,0)$ it is clear that

$$(\text{support}\,\Delta) \cap [T + \epsilon(1,0,0)] \subset B[(1,0,0);\epsilon] \cap [T + \epsilon(1,0,0)] \tag{7.64}$$

where $B(a;\lambda)$ is the ball centered at a, radius λ.

We can express Δ as a Fourier transform of a measure $\boldsymbol{\Delta}$ on $\hat{\Gamma}$. We claim that

$$\boldsymbol{\Delta} = ie^{i\hat{t}} \frac{|\hat{t}|}{2\hat{t}}\, d\hat{t}\, d\hat{\theta}. \tag{7.65}$$

To verify (7.64) we take the inverse Fourier transform $\tilde{\boldsymbol{\Delta}}$ of $\boldsymbol{\Delta}$

$$\tilde{\boldsymbol{\Delta}}(t,x,y) = \frac{i}{2} \iint e^{-i\hat{t}(t - x\cos\hat{\theta} - y\sin\hat{\theta}) + i\hat{t}} \frac{|\hat{t}|}{\hat{t}}\, d\hat{t}\, d\hat{\theta}.$$

This vanishes on $t = 1$ because $|\hat{t}|/\hat{t}$ is odd. (The factor $e^{i\hat{t}}$ changes $t = 0$ into $t = 1$.)

On the other hand

$$\begin{aligned}
\frac{\partial}{\partial t}\tilde{\boldsymbol{\Delta}}(t,x,y) &= \frac{1}{2} \iint e^{-i\hat{t}[(t-1) - x\cos\hat{\theta} - y\sin\hat{\theta}]} |\hat{t}|\, d\hat{t}\, d\hat{\theta} \\
&= \iint e^{-i\hat{t}(t-1) + ix\hat{x} + iy\hat{y}}\, d\hat{x}\, d\hat{y} \\
&= \delta_{x=y=0} \quad \text{on } t = 1
\end{aligned}$$

which verifies our claim that $\tilde{\boldsymbol{\Delta}} = \Delta$.

We digress to show how to derive (7.64) from the Lagrange interpolation formula (7.35); this is the Watergate method introduced in Section 1.4. Actually a slight variation of (7.35) is needed because the harmonic basis $(1, \hat{t})$ used for (7.35) relates to the CP with CD on $\{t = 0\}$. In our situation the CS is $\{t = 1\}$ so the harmonic basis is replaced by $(e^{i\hat{t}}, \hat{t}e^{i\hat{t}})$.

The modified harmonic expansion is

$$\hat{H}(\hat{t}, \hat{x}, \hat{y}) = \hat{H}_0(\hat{x}, \hat{y})e^{i\hat{t}} + \hat{H}_1(\hat{x}, \hat{y})\hat{t}e^{i\hat{t}} \quad \text{on } \hat{\Gamma}.$$

This means that $H \in \mathcal{E}'$ acts on CD by

$$H \cdot (f_0, f_1) = \hat{H}_0 \cdot \hat{f}_0 + \hat{H}_1 \cdot \hat{f}_1.$$

Upon multiplication by $\exp(-i\hat{t})$ we obtain the modified (7.35):

$$\begin{aligned}\hat{H}(\hat{t}, \hat{x}\hat{y}) = {} & \frac{e^{i\hat{t}}}{2}\left[\hat{H}(\sqrt{\hat{x}^2+\hat{y}^2}, \hat{x}, \hat{y})e^{-i\sqrt{\hat{x}^2+\hat{y}^2}} + \hat{H}(-\sqrt{\hat{x}^2+\hat{y}^2}, \hat{x}, \hat{y})e^{i\sqrt{\hat{x}^2+\hat{y}^2}}\right] \\ & + \frac{\hat{t}e^{i\hat{t}}}{2}\left[\frac{\hat{H}(\sqrt{\hat{x}^2+\hat{y}^2}, \hat{x}, \hat{y})e^{-i\sqrt{\hat{x}^2+\hat{y}^2}} - \hat{H}(-\sqrt{\hat{x}^2+\hat{y}^2}, \hat{x}, \hat{y})e^{i\sqrt{\hat{x}^2+\hat{y}^2}}}{\sqrt{\hat{x}^2+\hat{y}^2}}\right].\end{aligned} \tag{7.37*}$$

If we start with CD of the form $(0, f)$ on $t = 1$ then the solution F of the CP is given by

$$\begin{aligned}F(t, x, y) &= \delta_{txy} \cdot F \\ &= i(\hat{\delta}_{txy})_1 \cdot \hat{f} \\ &= \frac{i}{2}\left\{\frac{e^{i[(t-1)\sqrt{\hat{x}^2+\hat{y}^2} - x\hat{x} - y\hat{y}]} - e^{i[(-t+1)\sqrt{\hat{x}^2+\hat{y}^2} - x\hat{x} - y\hat{y}]}}{\sqrt{\hat{x}^2+\hat{y}^2}} \cdot \hat{f}(\hat{x}, \hat{y})\right\}.\end{aligned}$$

The measure used on $\hat{x}, \hat{y}$ is

$$\hat{r}\, d\hat{r}\, d\hat{\theta} = |\hat{t}|\, d\hat{t}\, d\hat{\theta} \quad \text{on } \hat{\Gamma}.$$

Putting things together,

$$F(t, x, y) = \frac{i}{2}\int_{\hat{\Gamma}} e^{i[t(\hat{t}-1) - x\hat{x} - y\hat{y}]}\hat{f}(\hat{x}, \hat{y})\frac{|\hat{t}|}{\hat{t}}\, d\hat{t}\, d\hat{\theta}$$

because integration over all of $\hat{\Gamma}$ encompasses both square roots of $\hat{x}^2 + \hat{y}^2$. The factor i appears because the Fourier transform of $\partial/\partial t$ is $i\hat{t}$ while we use $(1, \hat{t})$ as the harmonic basis. We have completed the derivation of (7.64) by means of harmonic expansions.

Let us return to the proof of Theorem 7.4. Note that the invariant measure on $\hat{\Gamma}$ is (a constant times) $d\hat{t}\, d\hat{\theta}$. This is a simple consequence of the fact that the action of G on $\hat{\Gamma}$ is the quadratic transform of the linear action on $\mathbb{R}^2 - \{0\}$. The linear action preserves the euclidean measure $\rho\, d\rho\, d\phi$. Since $\hat{t} = \rho^2$ and $\hat{\theta} = 2\phi$ the invariant measure is $d\hat{t}\, d\hat{\theta}$. (Another computation can be made using coordinates $\hat{t}, \hat{\theta}, \hat{\Box} = \hat{t}^2 - \hat{r}^2$ near $\hat{\Gamma}$, where $\hat{r}^2 = \hat{x}^2 + \hat{y}^2$. The Jacobian of the transformation from $\hat{t}, \hat{\theta}, \hat{\Box}$ to the euclidean coordinates $\hat{t}, \hat{\theta}, \hat{r}$ whose G invariant measure is $\hat{r}\, d\hat{t}\, d\hat{r}\, d\hat{\theta}$ is

$$\frac{\partial(\hat{t}, \hat{\theta}, \hat{\Box})}{\partial(\hat{t}, \hat{\theta}, \hat{r})} = -2\hat{r}$$

which gives a second verification since $|\hat{t}| = \hat{r}$ on $\hat{\Gamma}$.)

Note that

$$|\hat{t}|/\hat{t} = \operatorname{sgn} \hat{t} = \text{Heaviside function}$$

is invariant under the group G. By (7.64) this means that $\Delta(t-1, x, y)$ is G invariant and so depends only on the Minkowski distance. The one-dimensional Fourier transform of $\operatorname{sgn} \hat{t}$ (which is the same as $\Delta(t-1, 0, 0))$ is t^{-1}; this is the foundation of the study of the Hilbert transform. We have

$$\Delta(t, x, y) = \begin{cases} [\operatorname{sgn}(t-1)]\{\text{distance}[(t-1, x, y) \text{ to } \Gamma_{1,0,0}]\}^{-1} & \text{for } (t-1)^2 \geq x^2 + y^2 \\ 0 & \text{otherwise.} \end{cases} \tag{7.66}$$

For, (7.65) holds on the t axis by our remark concerning the Hilbert transform. For other values we use G invariance.

Comparing (7.65) with the well-known formula for the fundamental solution of the wave equation yields

$$\Delta = e^+(t-1, x, y) - e^-(t-1, x, y) \tag{7.67}$$

where $e^{\pm}$ is the fundamental solution for $\Box$ with support in the forward (backward) light cone.

We are now in a position to complete the proof of Theorem 7.4; we must compute the CD of Δ on G/K and show it is the same as the CD on T.

Let us examine Δ on $G/K + (\epsilon, 0, 0)$. Now $G/K + (\epsilon, 0, 0)$ is defined by

$$(t - \epsilon)^2 - x^2 - y^2 = 1.$$

Thus $G/K + (\epsilon, 0, 0)$ meets the forward light cone translated to $(1, 0, 0)$ in

$$(t-1)^2 = (t - \epsilon)^2 - 1,$$

i.e.

$$t = \frac{1}{2(1-\epsilon)} + \frac{1+\epsilon}{2} = 1 + \epsilon + \frac{1}{2}\epsilon^2 + \mathcal{O}(\epsilon^3).$$

Remark. *The underlying reason for the equality of the CD of Δ on T and G/K is that T is tangent to G/K at $(1,0,0)$ so for values of t between 1 and ϵ the tangent plane T differs locally by $\mathcal{O}(\epsilon^2)$ from G/K. Moreover, $\Box$ is second order so $\mathcal{O}(\epsilon^2)$ is good enough to establish equality of normal derivatives. (Actually $\mathcal{O}(\epsilon^{1+\eta})$ would suffice.)*

We write points on $G/K+(\epsilon,0,0)$ in the form $[\epsilon+\cosh\zeta, \sinh\zeta\cos\theta, \sinh\zeta\sin\theta]$. The distance of such a point to the light cone translated by $(1,0,0)$ is given by

$$\begin{aligned}\text{dist}^2 &= (\epsilon + \cosh\zeta - 1)^2 - \sinh^2\zeta \\ &= \epsilon^2 + 2 + 2(\epsilon-1)\cosh\zeta - 2\epsilon.\end{aligned}$$

In particular dist $= 0$ when

$$\begin{aligned}\cosh\zeta &= 1 + \frac{\epsilon^2}{2(1-\epsilon)} \\ &= 1 + \frac{\epsilon^2}{2} + \mathcal{O}(\epsilon^3).\end{aligned}$$

For small ζ this means

$$\zeta = \epsilon + \mathcal{O}(\epsilon^2).$$

We have shown (see (7.65)) that

$$\begin{aligned}&\Delta(\epsilon + \cosh\zeta, \sinh\zeta\cos\theta, \sinh\zeta\sin\theta) \\ &\quad = \begin{cases} \{2(\epsilon-1)(\cosh\zeta - 1) + \epsilon^2\}^{-1/2} & \text{for dist} > 0 \\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

In particular the nonvanishing set of Δ is given by

$$|\zeta| \le \epsilon + \mathcal{O}(\epsilon^2).$$

We wish to compare this with the corresponding value for $T + (\epsilon,0,0)$. Using polar coordinates (ρ,θ) on T we have

$$\Delta(1+\epsilon, \rho\cos\theta, \rho\sin\theta) = \begin{cases} (\epsilon^2 - \rho^2)^{-1/2} & \rho \le \epsilon \\ 0 & \text{otherwise.}\end{cases}$$

To compare the two expressions we call $\tilde\rho = \sinh\zeta$. By (7.46) the measure on G/K is

$$\begin{aligned}\sinh\zeta\, d\zeta\, d\theta &= (\cosh\zeta)^{-1}\tilde\rho\, d\tilde\rho\, d\theta \\ &= \tilde\rho\, d\tilde\rho\, d\theta[1 + \mathcal{O}(\zeta^2)].\end{aligned}$$

Since $|\zeta| \le c\epsilon$ by the above, the $\mathcal{O}(\zeta^2)$ term is not important.

Since $\tilde{\rho} = \sinh \zeta = \zeta + \mathcal{O}(\zeta^3)$ the set on $(\epsilon, 0, 0) + G/K$ on which Δ does not vanish is of the form

$$\begin{aligned} \tilde{\rho} &= \zeta + \mathcal{O}(\zeta^3) \\ &= \epsilon + \mathcal{O}(\epsilon^2) \end{aligned}$$

by the above. Thus if we map $\epsilon + G/K$ onto $\epsilon + T$ locally by $(\tilde{\rho}, \theta) \to (\rho, \theta)$ then the nonvanishing sets of Δ correspond up to order ϵ^2. Moreover the value of Δ^{-2} on $(\epsilon, 0, 0) + G/K$ is

$$2(\epsilon - 1)\frac{\tilde{\rho}^2}{2} + \epsilon^2 + \mathcal{O}(\tilde{\rho}^3) = \epsilon^2 - \tilde{\rho}^2 + \mathcal{O}(\epsilon\tilde{\rho}) + \mathcal{O}(\tilde{\rho}^3).$$

Since $\tilde{\rho} \leq c\epsilon$ this agrees with the value of Δ^{-2} on $(\epsilon, 0, 0) + T$ up to $\mathcal{O}(\epsilon^3)$.

We have shown that the values of Δ on $(\epsilon, 0, 0) + G/K$ and on $(\epsilon, 0, 0) + T$ agree up to $\mathcal{O}(\epsilon^2)$. Moreover the measures agree to order ϵ^2. This implies that the CD agrees. This completes the proof of Theorem 7.4.

Theorem 7.5 *For $f \in \mathcal{D}(G/K)$, the space of C^∞ functions of compact support on G/K, we can represent the solution F of the CP with* $\mathrm{CD}(F) = (0, f)$ *as the (inverse) Fourier transform of the measure*

$$\mathbf{F} = \frac{i|\hat{t}|}{2t}\hat{f}\delta_{\hat{\Gamma}}.$$

Proof We have verified Theorem 7.5 for $f = \delta_{(1,0,0)}$. (Strictly speaking $\delta_{(1,0,0)} \notin \mathcal{D}(G/K)$ but the same result is valid for $\mathcal{E}'(G/K)$.) We apply $g \in G$ to the formula for $\delta_{(1,0,0)}$. Since the Fourier transform is defined via the Minkowski inner product,

$$\widehat{\delta_{g(1,0,0)}} = g\hat{\delta}_{(1,0,0)}.$$

The measure $\delta_{\hat{\Gamma}}$ is G invariant and, since G is the proper Lorentz group, it preserves $\operatorname{sgn} \hat{t} = |\hat{t}|/\hat{t}$. Thus

$$\text{Inverse FT}\left[\frac{i|\hat{t}|}{2\hat{t}}\hat{\delta}_{g(1,0,0)}\delta_{\hat{\Gamma}}\right] = g\,\text{Inverse FT}\left[\frac{i|\hat{t}|}{2\hat{t}}\hat{\delta}_{(1,0,0)}\delta_{\hat{\Gamma}}\right]. \tag{7.68}$$

We have already verified that the right side of (7.68) is $\delta_{g(1,\,0,\,0)}$.

Thus we have proven the validity of Theorem 7.5 for the "special objects" $\delta_{g(1,0,0)}$. The general result follows by taking limits of linear combinations (integrals).

(**F**) Theorem 7.5 shows us how to express the solution F of the CP whose CD on G/K is $(0, f)$ in terms of the euclidean Fourier transform of $f\delta_{G/K}$ restricted to $\hat{\Gamma}$. To obtain a more complete picture we study the Watergate problem (WP) associated to $\Box$ (see Chapter 4).

If $f \in \mathcal{D}(G/K)$, then $\hat{f}$ is an entire function of exponential type which is bounded in real directions and of exponential growth in imaginary directions. Moreover $\hat{f} = \mathcal{O}(\exp(-c|\Im \hat{t}|))$ on $\hat{\Gamma}$ as $\Im \hat{t} \to +\infty$ and $\hat{f}$ is exponentially bounded as $\Im \hat{t} \to -\infty$ and exponentially decreasing as $\Im \hat{t} \to +\infty$ by (7.50).

We shall show below that f is rapidly decreasing on $\hat{\Gamma}$ for real $\hat{t}$.

For the remainder of (**F**) and for (**G**) we shall assume that the functions f we deal with are of *type 0*, meaning they are invariant under the group K of rotations around the t axis. Then $\hat{f}$ is also K invariant so $\hat{f}\big|_{\hat{\Gamma}}$ is a function of the single variable $\hat{t}$.

For K invariant functions f we can express $\hat{f}\big|_{\hat{\Gamma}}$ in an interesting way. f depends only on $\cosh \zeta$ and $\hat{f}\big|_{\hat{\Gamma}}$ on $\hat{t}$ so, by (7.47),

$$\hat{f}(\hat{t}) = \iint f(\cosh \zeta) e^{i\hat{t}[\cosh \zeta + \sinh \zeta \cos \theta]} \sinh \zeta \, d\zeta \, d\theta.$$

We change variables from ζ, θ to ζ and $u = \cosh \zeta + \sinh \zeta \cos \theta$. This leads to

$$\hat{f}(\hat{t}) = \int e^{i\hat{t}u} \, du \int f(\cosh \zeta) \sqrt{2u \cosh \zeta - u^2 - 1} \, d\zeta. \tag{7.69}$$

This representation of $\hat{f}$ is related to the Abel transform (see [29, vol. II] and Section 8.2).

Let $\tilde{h}(\hat{t})$ be an entire function of exponential type which is rapidly (polynomially) decreasing for real $\hat{t}$ and $\mathcal{O}(\exp(-c|\Im \hat{t}|))$ as $\Im \hat{t} \to +\infty$ for some $c > 0$. We want to determine when $\tilde{h} = \hat{f}$ for some $f \in \mathcal{D}(K \backslash G/K)$. By the Paley–Wiener theorem for functions of one real variable we can write

$$\tilde{h}(\hat{t}) = \int e^{it\hat{t}} \underset{\sim}{h}(t) \, dt \tag{7.70}$$

where $\underset{\sim}{h} \in \mathcal{D}$ and support $\underset{\sim}{h} \subset [c, \infty)$.

Think of the integral as a limit of linear combinations of $\exp(it_0\hat{t})\underset{\sim}{h}(t_0)$. $\exp(it_0\hat{t})\underset{\sim}{h}(t_0)$ is the value on the light cone of the Fourier transform in Minkowski space of $\underset{\sim}{h}(t_0)\delta_{(t_0,0,0)}$ thought of as a measure on $r = t_0$. Hence by Theorem 7.5 (with the obvious modification of replacing $r = 1$ by $r = t_0$)

$$\underset{\sim}{h}(t_0)\Delta_{(t_0,0,0)} = \frac{i}{2} \iint \underset{\sim}{h}(t_0) e^{-i\hat{t}[(t-t_0) - x\cos\hat{\theta} - y \sin\hat{\theta}]} \frac{|\hat{t}|}{\hat{t}} \, d\hat{t} \, d\hat{\theta} \tag{7.71}$$

is the solution of the CP for $\Box$ whose CD on $t = t_0$, hence also on $r = t_0$, is $(0, \underset{\sim}{h}(t_0)\delta_{(t_0,0,0)})$.

We call $\mathcal{H}$ the integral of (7.71) with respect to t_0. Using (7.71) we deduce

$$\begin{aligned}\mathcal{H}(t,x,y) &= \frac{i}{2}\iiint \underset{\sim}{h}(t_0)e^{-i\hat{t}[(t-t_0)-x\cos\hat{\theta}-y\sin\hat{\theta}]}\frac{|\hat{t}|}{\hat{t}}\,d\hat{t}\,d\hat{\theta}\,dt_0 \\ &= \int \underset{\sim}{h}(t_0)\Delta_{(t_0,0,0)}(t,x,y)\,dt_0 \\ &= \int \underset{\sim}{h}(t_0)\Delta_{(0,0,0)}(t-t_0,x,y)\,dt_0 \qquad (7.72)\end{aligned}$$

because $\Box$ is translation invariant.

The calculation in (7.72) depends on integration first with respect to $\hat{t},\hat{\theta}$ and then with respect to t_0. We now reverse the order of integration. By (7.70) the t_0 integral replaces $\underset{\sim}{h}(t_0)$ by $\tilde{h}(\hat{t})$. We are left with

$$\mathcal{H}(t,x,y) = \iint \tilde{h}(\hat{t})e^{-i\hat{t}[t-x\cos\hat{\theta}-y\sin\hat{\theta}]}\frac{|\hat{t}|}{\hat{t}}\,d\hat{t}\,d\hat{\theta}. \qquad (7.73)$$

Since $\tilde{h}$ and $\underset{\sim}{h}$ are "nice" functions the change of order of integration is not hard to justify.

$\mathcal{H}$ is the solution of the wave equation which is the Fourier transform of

$$\tilde{\mathbf{H}}(\hat{t}) = \tilde{h}(\hat{t})\frac{|\hat{t}|}{\hat{t}}$$

on the light cone.

On the other hand, since the propagation speed of light is 1

$$\text{support}\,\Delta_{(t_0,0,0)} \subset \Gamma + (t_0,0,0).$$

Since $\underset{\sim}{h}$ has compact support in $t_0 > 0$,

$$\text{support}\,\mathcal{H}(t,x,y) \cap G/K \text{ is compact.}$$

We can describe $\mathcal{H}$ in terms of its WD on the t axis. Since all our functions are K invariant this WD is just the restriction of $\mathcal{H}$ to the t axis. It is a standard fact that the Fourier transform of $|\hat{t}|/\hat{t}$ is t^{-1} (suitably defined). Since $\Delta_{(0,0,0)}$ is the Fourier transform of $c|\hat{t}|/\hat{t}$ on the light cone its WD is t^{-1} by the above (up to a constant which we shall ignore). Our formula (7.73) for $\mathcal{H}$ represents

$$\mathcal{H}(t,0,0) = \underset{\sim}{h} * t^{-1}$$

(convolution in t). This is exactly the Hilbert transform of $\underset{\sim}{h}$.

We have shown

Proposition 7.6 *The WD of the inverse Fourier transform of $\tilde{h}(\hat{t})(|\hat{t}|/\hat{t})$ on $\hat{\Gamma}$ is the Hilbert transform of $\underset{\sim}{h}$.*

If $\mathcal{H}$ is odd under the Kelvin transform then it vanishes on $r = 1$ so its CD is of the form $(0, f)$. We have explained why

$$\mathcal{H} = \int \underset{\sim}{h}(t_0)\Delta_{(t_0,0,0)}\, dt_0$$

has compact support on G/K. The regularity of its CD is easy to establish.

We say that $\underset{\sim}{h}$ or $\tilde{h}$ is *Kelvin odd* if $\mathcal{H}$ is Kelvin odd.

The solutions of $\Box$ are determined by their WD. We claim that the Kelvin oddness of $\mathcal{H}$ is equivalent to the Kelvin oddness of the Hilbert transform of $\underset{\sim}{h}$. For it is clear that if $\mathcal{H}$ is Kelvin odd then so is its WD. Conversely, since the Kelvin transform commutes with $\Box$ the oddness of the WD of a solution H of $\Box$ implies the oddness of H.

Theorem 7.7 *The image $\hat{\mathcal{D}}(K\backslash G/K)$ of K invariant functions on G/K under Fourier transformation restricted to $\hat{\Gamma}$ consists of all functions $\tilde{h}(\hat{t})$ satisfying*

(i) *$\tilde{h}$ is an entire function of exponential type.*
(ii) *$\tilde{h}$ is rapidly decreasing for real $\hat{t}$.*
(iii) *$\tilde{h}(\hat{t}) = \mathcal{O}(\exp(-c|\Im\hat{t}|))$ for $\Im\hat{t} \to +\infty$, for some $c > 0$.*
(iv) *$\tilde{h}$ is Kelvin odd.*

Proof We have shown the necessity of (i), (iii), and (iv). We now verify the necessity of (ii). Since the group Laplacian Δ on G/K is elliptic, a continuous function h of compact support on G/K is in $\mathcal{D}(G/K)$ if and only if $\Delta^j h$ is continuous for any $j \geq 0$. We now use separation of variables. By (7.54) we can write

$$r^2\Box = \left(r\frac{\partial}{\partial r}\right)^2 + r\frac{\partial}{\partial r} - \Delta.$$

When we compute $\tilde{h}$ on the light cone we are only interested in exponentials whose frequencies lie on the light cone; such exponentials belong to the kernel of $\Box$. Thus, for $\hat{x} \in \hat{\Gamma}$,

$$\begin{aligned}\int \Delta[f\delta_{G/K}]e^{ix\cdot\hat{x}} &= \int \left[\Box - \left(r\frac{\partial}{\partial r}\right)^2 - r\frac{\partial}{\partial r}\right](f\delta_{G/K})e^{ix\cdot\hat{x}} \\ &= -\int f\delta_{G/K}\left[\left(r\frac{\partial}{\partial r}\right)^2 + r\frac{\partial}{\partial r}\right]e^{ix\cdot\hat{x}}\end{aligned}$$

because $r = 1$ and Δ commutes with multiplication by r.

Now x is of the form $r(\cosh\zeta, \sinh\zeta\cos\theta, \sinh\zeta\sin\theta)$ while $\hat{x} = \hat{r}(1, \cos\hat{\theta}, \sin\hat{\theta})$ where $\hat{r}$ is the euclidean distance to the origin on $\hat{\Gamma}$. Thus

$$x\cdot\hat{x} = r\hat{r}[\cosh\zeta - \sinh\zeta\cos(\theta - \hat{\theta})].$$

This means that

$$r\frac{\partial}{\partial r}e^{ix\cdot\hat{x}} = \hat{r}\frac{\partial}{\partial\hat{r}}e^{ix\cdot\hat{x}}.$$

Since we are integrating in x the integral commutes with $\hat{r}$ differentiation.

We have shown

$$\int \Delta[f\delta_{G/K}]e^{ix\cdot\hat{x}} = -\left[\left(\hat{r}\frac{\partial}{\partial\hat{r}}\right)^2 + \hat{r}\frac{\partial}{\partial\hat{r}}\right]\hat{f}(\hat{x}) \quad \text{on } \hat{\Gamma}.$$

In general this means that

$$\int \Delta^j[f\delta_{G/K}]e^{ix\cdot\hat{x}} = (-1)^j\left[\left(\hat{r}\frac{\partial}{\partial\hat{r}}\right)^2 + \hat{r}\frac{\partial}{\partial\hat{r}}\right]^j\hat{f}(\hat{x}) \quad \text{on } \hat{\Gamma}.$$

Since $\Delta^j[f\delta_{G/K}]$ is a measure it follows that $(\hat{r}\partial/\partial\hat{r})^j\hat{f}$ is a bounded function on $\hat{\Gamma}$ for every j.

If we define $\hat{r}$ to be negative on the negative light cone the Fourier transform $\underset{\wedge}{f}(\underset{\wedge}{r})$ of $\hat{f}(\hat{r})$ in $\hat{r}$ is a distribution whose support is in $\underset{\wedge}{r} \geq c$ for some $c > 0$ since $\hat{f}(\hat{r})$ is an entire function of exponential type which is exponentially decreasing in $\Im\hat{r} < 0$. (The change from $\hat{r}$ to $-\hat{r}$ clearly does not affect $\hat{r}\partial/\partial\hat{r}$.)

The Fourier transform of $\hat{r}\partial/\partial\hat{r}$ is

$$\frac{\partial}{\partial\underset{\wedge}{r}}\underset{\wedge}{r} = 1 + \underset{\wedge}{r}\frac{\partial}{\partial\underset{\wedge}{r}}.$$

Since $\{(\hat{r}\partial/\partial\hat{r})^j\hat{f}\}$ are uniformly bounded functions on the real $\hat{r}$ axis the Fourier transforms $\{[(\partial/\partial\underset{\wedge}{r})\underset{\wedge}{r}]^j\underset{\wedge}{f}\}$ are distributions of bounded order; by the commutation relation of $\hat{r}$ and $\hat{\partial}/\partial\hat{r}$ the same is true of $\{(\partial/\partial\underset{\wedge}{r})^j\underset{\wedge}{f}\}$. (Support $\underset{\wedge}{f}$ does not meet the origin so the factor $\underset{\wedge}{r}$ has no essential effect.) This means that $\hat{\underset{\wedge}{f}} \in \mathcal{D}(\underset{\wedge}{r})$ so its (inverse) Fourier transform $\hat{f}$ is rapidly decreasing for $\hat{r}$ real.

The necessity of (ii) is established.

The discussion preceding Proposition 7.6 shows that any $\tilde{h}$ satisfying (i), (ii), (iii), (iv) defines a Kelvin odd solution $\mathcal{H}$ of $\Box\mathcal{H} = 0$ whose CD on G/K belongs to $\mathcal{D}(K\backslash G/K)$. $\mathcal{H}$ is the Fourier transform of the function

$$\tilde{\mathbf{H}}(\hat{t}) = \hat{h}(\hat{t})\frac{|\hat{t}|}{\hat{t}}$$

on $\hat{\Gamma}$.

By (7.71) $\mathcal{H}$ is the integral in t_0 of the solutions $\underset{\sim}{h}(t_0)\Delta_{(t_0,0,0)}$ of $\Box$ whose CD on $r = t_0$ is $(0, \underset{\sim}{h}(t_0)\delta_{(t_0,0,0)})$. The CD of $\mathcal{H}$ on $r = 1$ is of the form $(0, h)$. We can evaluate $\widehat{h\delta_{G/K}}$ on $\hat{\Gamma}$ using the invariance of the inner product (7.53). In the notation of (7.58)

$$(\mathcal{H}, v) = -\hat{h}(\hat{t})$$

where we have written $\hat{t}$ for $(\hat{t}, \hat{x}, \hat{y})$ since $\mathcal{H}$ is K invariant.

Instead of evaluating the inner product on $r = 1$, we write $\mathcal{H}$ as an integral of $\underset{\sim}{h}(t_0)\Delta_{(t_0,0,0)}$ and evaluate the inner product on $r = t_0$ for each t_0. Since

$$\mathrm{CD}_{t_0}(\Delta_{(t_0,0,0)}) = (0, \delta_{(t_0,0,0)})$$

we find

$$\left(\Delta_{(t_0,0,0)}, e^{it\hat{t}+ix\hat{x}+iy\hat{y}}\right) = e^{it_0\hat{t}}.$$

Thus

$$(\mathcal{H}, v) = \int \underset{\sim}{h}(t_0)e^{it_0\hat{t}}\, dt_0 = \tilde{h}(\hat{t}).$$

This proves that $\tilde{h} \in \mathcal{D}(K\backslash G/K)$ which establishes Theorem 7.7.

We can use Theorem 7.7 to characterize the horocyclic Radon transform on G/K.

Theorem 7.8 *Let $f^N(s)$ be a smooth function of compact support in the horocyclic spread parameter s vanishing for $s \leq c$ for some $c > 0$ with f^N independent of the spread. A necessary and sufficient condition that*

$$f^N = \mathbf{R}^N f$$

for some $f \in \mathcal{D}(K\backslash G/K)$ is that the Hilbert transform of f^N is Kelvin odd.

Proof We start with (7.34) which shows how to express $\mathbf{R}^N f$ as $\{\delta_{Q_\theta} * f\}$ where f is identified with $f\delta_{G/K}$. We have

$$\widehat{\delta_{Q_\theta} * f} = \hat{f}\big|_{L_{\hat{\theta}}}$$

where $L_{\hat{\theta}}$ is the ray on $\hat{\Gamma}$ corresponding to the angle θ.

If $f^N(s)$ is of the form $\mathbf{R}^N f(s)$ for some $f \in \mathcal{D}(K\backslash G/K)$ then $f^N = \delta_{Q_\theta} * f$ as functions of s.

We call $\tilde{f}^N$ the Fourier transform of f^N in the variable s, so

$$\tilde{f}^N(\hat{t}) = \hat{f}\big|_{L_{\hat{\theta}}}(\hat{t})$$

where the right side is independent of $\hat{\theta}$.

We now define $\hat{f}$ by this formula. Theorem 7.8 now follows from Theorem 7.7.

What happened to the moment conditions?

The point is that for functions on $\hat{\Gamma}$ which depend only on $\hat{t}$ there are no moment conditions. To verify this recall that for the hyperplane Radon transform on $\mathbb{R}^n$

$$\mathbf{R}^{n-1}f(s,\mathbf{g}) = \mathbf{R}^{n-1}f(-s,-\mathbf{g}).$$

Since we can go from $\mathbf{g}$ to $-\mathbf{g}$ by a rotation, if f is rotationally invariant

$$\mathbf{R}^{n-1}f(s,\mathbf{g}) = \mathbf{R}^{n-1}f(-s,\mathbf{g});$$

that is, $\hat{f}(\hat{s},\hat{\mathbf{g}}^{\perp})$ is an even function on the line $\hat{\mathbf{g}}^{\perp}$.

We claim that COMP ORIGIN holds (see Section 2.1). A rotationally invariant formal power series is a power series in $\hat{r}^2$. Since $\hat{f}(\hat{s},\hat{\mathbf{g}}^{\perp})$ is an even function of $\hat{s}$ which is independent of $\hat{\mathbf{g}}^{\perp}$ its power series at the origin depends only on $\hat{r}^2$. This confirms the claim.

It follows that there are no moment conditions for $\mathbf{R}^{n-1}$ for rotationally invariant functions.

The argument leading to the proof of Theorem 7.3 shows how rotationally invariant functions on lines on $\hat{\Gamma}$ correspond to pairs of rotationally invariant functions on lines on $\{\hat{t}=0\}$. The correspondence is given by the following. The function $\hat{h}(\hat{t},\hat{L}^{\perp})$ of the generators of $\hat{\Gamma}$ corresponds to the pair

$$\frac{1}{2}[\hat{h}(\hat{t},\hat{L}^{\perp}) + \hat{h}(-\hat{t},\hat{L}^{\perp})] = \hat{h}^{+}$$
$$\frac{1}{2\hat{t}}[\hat{h}(\hat{t},\hat{L}^{\perp}) - \hat{h}(-\hat{t},\hat{L}^{\perp})] = \frac{1}{\hat{t}}\hat{h}^{-}.$$

Since by the above remarks concerning $\mathbf{R}^{n-1}$, $\hat{h}^{+}$ and $\hat{h}^{-}$ are arbitrary (except for regularity and growth at infinity) rotationally invariant functions of $(\hat{x},\hat{y})$, $\hat{h}$ is an arbitrary rotationally invariant function on $\hat{\Gamma}$.

Remark. *The same argument shows that there are no moment conditions for functions on the generators of $\hat{\Gamma}$ which transform under a fixed character of K.*

(**G**) Theorem 7.7 can be regarded as the Paley–Wiener theorem for the euclidean Fourier transform on G/K. We are now going to relate this Fourier transform to the Fourier transform in the sense of the group G. For functions of type 0 (K invariant) on G/K the kernel of the G Fourier transform is the Legendre function

$$P_s^0(\cosh\zeta) = \int (\cosh\zeta - \sinh\zeta\cos\theta)^s \, d\theta$$

(see [68], [11] vol. I, p. 157 (15)]).

We have discussed the relation between euclidean Fourier analysis on G/K and $\mathbf{R}^N$, which is related to N and hence G/N. We have identified G/N with Γ^+ by $g \leftrightarrow g\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$. Under this identification

$$\begin{aligned} ga &\leftrightarrow ga\begin{pmatrix}1&0\\0&0\end{pmatrix} \\ &= g\begin{pmatrix}a^2&0\\0&0\end{pmatrix} \\ &= g\begin{pmatrix}a^2&0\\0&a^2\end{pmatrix}\begin{pmatrix}1&0\\0&0\end{pmatrix} \\ &= \begin{pmatrix}a^2&0\\0&a^2\end{pmatrix} g\begin{pmatrix}1&0\\0&0\end{pmatrix}. \end{aligned}$$

(The notation may be slightly confusing. On the first line a is an element of A. In the subsequent lines a is the matrix entry $\left(\begin{smallmatrix}a&0\\0&a^{-1}\end{smallmatrix}\right)$ of that element of A.) Thus right action of A on G/N (the first line), which commutes with the left action of G, is equivalent to scalar multiplication. This suggests that the Mellin transform in scalar multiplication on $\Gamma = G/N$ should be a useful tool for the study of analysis on G/N.

The relation between G Fourier analysis and euclidean Fourier analysis starts from the formal identity

$$\begin{aligned} \iint_{\hat{t}\geq 0} \hat{t}^s e^{i\hat{t}(\cosh\zeta - \sinh\zeta\cos\hat{\theta})}\frac{d\hat{t}}{\hat{t}}\,d\hat{\theta} &= \Gamma(s)(-i)^{-s}\int(\cosh\zeta - \sinh\zeta\cos\hat{\theta})^{-s}\,d\hat{\theta} \\ &= \Gamma(s)(-i)^{-s}P^0_{-s}(\cosh\zeta). \end{aligned} \tag{7.74}$$

It is important that we can regard the left side of (7.74) as either the additive Fourier transform on $\hat{\Gamma}$ of $\hat{t}^s$ or the Fourier transform in scalar multiplication (Mellin transform) on $\hat{\Gamma}$ of the exponential function which defines the Fourier transform from $\hat{\Gamma} = G/N$ to G/K.

There are many ways of providing a rigorous treatment of (7.74). One approach is via the theory of distributions, but we prefer a method which incorporates analytic continuation.

The Γ function[4] is defined by

$$\Gamma(s) = \int_0^\infty e^{-x}x^s\frac{dx}{x}$$

for $\Re s > 0$. We are primarily interested in what happens when x is replaced by $\pm ix$. Let z be a complex number. If $\Im z < 0$ then a change of variables (which

[4]The reader should be careful to distinguish between the Γ function and the light cone which is also denoted by Γ.

is valid by Cauchy's theorem) gives

$$(iz)^{-s}\Gamma(s) = \int_0^\infty e^{-izx} x^s \frac{dx}{x}. \tag{7.75}$$

We cannot let z be real in the integral but we can use the left side of (7.75) to define the integral as long as we are in the complex z plane cut along the positive imaginary axis.

The integrand in (7.75) is a holomorphic function of x in the complex plane cut along the negative real x axis. If $\arg z = \theta$ with $-\pi < \theta \leq -\pi/2$ then $-\pi/2 < \arg iz \leq 0$. In this case we can shift the contour in the complex x plane to $\arg x = \epsilon > 0$. This means that even if z moves to a point on the negative real axis

$$\arg iz + \epsilon > 0$$

the integral converges.

An analogous process holds if $0 \leq \arg iz < \pi/2$.

Thus the integral in (7.75) makes sense for z real if we interpret it in terms of an integral in x along $\arg x = \epsilon$ where $\epsilon > 0$ is suitably chosen if z is in the negative real axis and $\epsilon < 0$ if z is in the positive real axis.

Actually we do not need to shift the contour to $\arg x = \epsilon$. It suffices to shift to a contour of the form

$$\Im x = \pm c(1 + \log(1 + |x|)) \tag{7.76}$$

for $c > \Re s$. In this case we can interpret the integral in the sense of distributions. For if ϕ is a test function (C^∞ function of compact support) of a single real variable $\hat{x}$ then it is a standard result of Paley–Wiener–Schwartz theory that

$$|\hat{\phi}(x)| \leq c(1 + |x|)^{-N}$$

on

$$|\Im x| \leq N(1 + \log|x|)$$

for any N (c depends on N). Thus the integral (7.75) converges in the space $\mathcal{D}'$ for z real, meaning that we modify the integrand on the contour (7.76) by multiplying by $\hat{\phi}(x)$ and regarding the integral as a linear function of $\hat{\phi}$. (We can shift the so-modified integral from (7.76) to the real x axis.)

A more delicate analysis such as that developed in Section 9.3 shows that the integral converges for z real in the space $\mathcal{S}'$ of Schwartz. Thus the identity (7.75) has meaning for z real since both sides are well defined.

The integral (7.75) involves $\exp(-izx)$. We shall also meet a corresponding integral

$$\int_0^\infty e^{ixz} x^s \frac{dx}{x}.$$

Now z belongs to the complex plane cut along the negative imaginary axis and in the above sense of analytic continuation

$$\int_0^\infty e^{ixz} x^s \frac{dx}{x} = (-iz)^{-s}\Gamma(s). \tag{7.77}$$

In (7.75) $-\pi > \arg z > -2\pi$ while in (7.77) $0 < \arg z < \pi$. In these formulas

$$i = e^{i\pi/2} \quad -i = e^{-i\pi/2} \quad -1 = e^{-i\pi}. \tag{7.78}$$

Let us clarify the branch of $(\pm iz)^{-s}$ that we are using. (7.75) is the standard definition of the Γ function when $z = -i$ and we identify $[i(-i)]^{-s} = 1$. Thus z in the lower half-plane and $z \to \pm 1$ corresponds to a rotation from $z = -i$ by $\pi/2$ in the positive (negative) direction. As $z \to \pm 1$ we have

$$-iz \to \mp i, \quad iz \to \pm i.$$

This leads to the interpretation

$$\int_0^\infty e^{-i(\pm x)} x^s \frac{dx}{x} = e^{\mp i\pi s/2}\Gamma(s). \tag{7.79}$$

On the other hand, the "base point" in (7.77) is $z = i$ and we identify $[-i(i)]^{-s} = 1$. $z \to \pm 1$ constitutes a rotation in the negative (positive) direction from the base point by $\pi/2$. As $z \to \pm 1$ we have

$$(-iz) \to \mp i, \quad iz \to \pm i$$

so that (7.77) becomes, in this limit sense,

$$\int_0^\infty e^{i(\pm x)} x^s \frac{dx}{x} = e^{\pm i\pi s/2}\Gamma(s). \tag{7.79*}$$

The two analytic continuations are consistent.

Since $\cosh\zeta - \sinh\zeta\cos\hat{\theta} > 0$ we can use these ideas to give a rigorous treatment of (7.74).

Equation (7.74) contains another ingredient, the interplay of Fourier and Mellin transforms. The Fourier transform involves the characters of the additive group $\mathbb{R}^+$ while the Mellin transform depends on the characters of the multiplicative group $\mathbb{R}^\times$ (see [63] and Section 10.3). The interrelation is governed by the Γ function which, on the one hand, is the Mellin transform of the exponential and, on the other hand, is the Fourier transform (at i) of the multiplicative characters.

We can now apply (7.74) to relate the G Fourier transform to the euclidean Fourier transform.

Let $f \in \mathcal{D}(K\backslash G/K)$ so

$$\hat{f}(\hat{t}) = \int f(\cosh\zeta) e^{i\hat{t}(\cosh\zeta - \sinh\zeta\cos\theta)} \sinh\zeta \, d\zeta \, d\theta. \tag{7.80}$$

We have justified the process of taking the Mellin transform of the exponential for $\hat{t} \geq 0$. Since f is of compact support and smooth we can take the Mellin transform $\overset{\vee}{\hat{f}}{}^{\pm}$ of $\hat{f}^+$ or $\hat{f}^-$ which are the respective restrictions of $\hat{f}$ to $\hat{t} \geq 0$ and $\hat{t} \leq 0$. This leads to

$$\begin{aligned}
\overset{\vee}{\hat{f}}{}^{+}(s) &= \Gamma(s)(-i)^{-s} \int P^0_{-s}(\cosh\zeta) f(\cosh\zeta) \sinh\zeta \, d\zeta \\
&= \Gamma(s)(-i)^{-s} \check{f}(s) \\
&= \Gamma(s) e^{i\pi s/2} \check{f}(s)
\end{aligned} \tag{7.81}$$

where $\check{f}$ is the G Fourier transform of f.

Similarly

$$\overset{\vee}{\hat{f}}{}^{-}(s) = \Gamma(s) e^{-i\pi s/2} \check{f}(s). \tag{7.82}$$

We have seen in (7.64)ff. that the Fourier transform F of $\frac{i}{2}\hat{f}(\hat{t})\operatorname{sgn}\hat{t}$ on $\hat{\Gamma}$ has CD $= (0, f)$ on $r = 1$. We call $F^{\pm}$ the Fourier transform of $\frac{i}{2}\hat{f}^{\pm}(\hat{t})\operatorname{sgn}\hat{t}$ on $\hat{\Gamma}^{\pm}$. We apply the techniques of analytic continuation and distribution theory that we have developed in (7.74)ff. Using the inverse Mellin transform of (7.81), (7.82),

$$\begin{aligned}
F^{\pm}(r\cosh\zeta) &= \frac{i}{2} \iint \hat{f}^{\pm}(\hat{t}) e^{-i\hat{t}r(\cosh\zeta - \sinh\zeta\cos\hat{\theta})} \operatorname{sgn}\hat{t} \, d\hat{t} \, d\hat{\theta} \\
&= \frac{i}{2} \int \Gamma(s) e^{\pm i\pi s/2} \check{f}(s) \int e^{-i\hat{t}r(\cosh\zeta - \sinh\zeta\cos\hat{\theta})} \hat{t}^{-s} \operatorname{sgn}\hat{t} \, d\hat{t} \, ds \\
&= \pm\frac{i}{2} \int \Gamma(s) e^{\pm i\pi s/2} (\pm i)^{s-1} r^{s-1} \check{f}(s) \Gamma(1-s) P^0_{s-1}(\cosh\zeta) \, ds \\
&= \frac{\pi}{2} \int r^{s-1} e^{\pm i\pi s} \check{f}(s) P^0_{s-1}(\cosh\zeta) \frac{ds}{\sin\pi s}
\end{aligned} \tag{7.83}$$

because $\Gamma(s)\Gamma(1-s) = \pi \csc\pi s$. Since $F = F^+ + F^-$

$$F(r\cosh\zeta) = \pi \int r^{s-1} \check{f}(s) P^0_{s-1}(\cosh\zeta) \cot\pi s \, ds. \tag{7.84}$$

The CD of F on $r = 1$ is $(0, f)$. By (7.60) f is the restriction of $(r\partial/\partial r + 1/2)F$ to $r = 1$, i.e.

$$f(\cosh\zeta) = \pi \int \check{f}(s) P^0_{s-1}(\cosh\zeta) \left(s - \frac{1}{2}\right) \cot\pi s \, ds. \tag{7.85}$$

(Recall that the "normal derivative" is $r\partial/\partial r + 1/2$ by (7.59)ff.)

Formula (7.85) is known as the *Plancherel formula* for functions on G of type $(0,0)$, meaning K bi-invariant functions on G.

It is a standard fact that the line $s = \frac{1}{2} + it$ corresponds to unitary representations of G (see (7.106)ff. for a different view). Note that on $\Re s = \frac{1}{2}$ the function $\cot \pi s = \tanh \pi t$ is odd, as is $s - \frac{1}{2} = it$.

The fact that F vanishes on $r = 1$ is a consequence of the invariance of P_{s-1} and hence $\check{f}(s)$ under $s \to 1 - s$. We wish to give a direct verification of this invariance called *Weyl invariance*.

Note that, by (7.58), $\int \exp[i\hat{t}(\cosh\zeta - \sinh\zeta\cos\theta)]\, d\theta$ is the Fourier transform from G/K to $\hat{\Gamma}$ of δ_ζ; here δ_ζ is the δ function of the circle defined by ζ on $r = 1$. By (7.74) we can interpret $\Gamma(s)(-i)^{-s}P^0_{-s}(\cosh\zeta)$ as the Mellin transform on $\hat{\Gamma}$ of this Fourier transform. (In this case we replace $\hat{\theta}$ in (7.74) by θ.) Since any function is an integral of δ functions the Weyl invariance of $\check{f}(s)$ and P^0_{s-1} are equivalent.

The oddness of F under the Kelvin transformation means that if we replace r^{s-1} by r^{-s} in the right side of (7.84) then F changes sign since this transformation is equivalent to $s \to 1 - s$. Now if f is essentially $\delta_{(1,0,0)}$ then $\check{f} \sim 1$ since clearly $P^0_s(0) = 1$ (when measures are suitably normalized). Since $\cot \pi s$ changes sign under $s \to 1 - s$ the oddness of F under $\mathcal{K}$ for $\check{f} = 1$ in (7.84) implies that

$$P^0_{s-1}(\cosh\zeta) = P^0_{-s}(\cosh\zeta)$$

by the uniqueness of the Mellin transform. Hence for any "reasonable" f

$$\check{f}(s) = \check{f}(1 - s) \tag{7.86}$$

in accordance with our assertion.

We can use the relation between the Fourier transform $\mathcal{F}_{G/K \to \hat{\Gamma}}$ and the group Fourier transform on G/K to give another proof (see the proof of Theorem 7.7) that for $f \in \mathcal{D}(K\backslash G/K)$ the euclidean Fourier transform $\hat{f}(\hat{t})$ is rapidly decreasing on $\hat{\Gamma}$.

We write

$$\hat{f}(\hat{t}) = \iint e^{i\hat{t}(\cosh\zeta - \sinh\zeta\cos\theta)} f(\zeta) \sinh\zeta\, d\zeta\, d\theta.$$

We use our discussion of analytic continuation in (7.74)ff. to write the exponential in the integrand as the inverse Mellin transform of its Mellin transform:

$$e^{i\hat{t}(\cosh\zeta + \sinh\zeta\cos\theta)} = \int \Gamma(s)(-i)^{-s}(\cosh\zeta - \sinh\zeta\cos\theta)^{-s}\hat{t}^{-s}\, ds.$$

We can now justify the change of variables to write

$$\begin{aligned}\hat{f}(\hat{t}) &= \int \Gamma(s)(-i)^{-s}\hat{t}^{-s}\int f(\zeta)(\cosh\zeta-\sinh\zeta\cos\theta)^{-s}\sinh\zeta\,d\theta\,d\zeta\\ &= \int \Gamma(s)(-i)^{-s}\hat{t}^{-s}\int f(\zeta)P^0_{-s}(\zeta)\sinh\zeta\,d\zeta\\ &= \int \check{f}(s)\Gamma(s)(-i)^{-s}\hat{t}^{-s}\,ds. \end{aligned}\tag{7.87}$$

We claim that $\check{f}(s)$ is an entire function of exponential type which is rapidly decreasing for $\Im s\to\pm\infty$ uniformly for $\Re s$ in any compact interval. The exponential type is a consequence of the fact that the same is true of P^0_{-s}, as is clear from the integral representation for P^0_{-s} which also shows that P^0_{-s} is bounded as $\Im s\to\pm\infty$ uniformly for ζ and $\Re s$ in any compact intervals.

We now use the fact that P^0_{-s} is an eigenfunction of the Laplacian $\Delta_{G/K}$ with eigenvalue $s(1-s)$. This means that

$$\begin{aligned}[s(1-s)]^k\check{f}(s) &= \int f(\cosh\zeta)\Delta^k_{G/K}P^0_{-s}(\cosh\zeta)\sinh\zeta\,d\zeta\\ &= \int(\Delta^k_{G/K}f)(\cosh\zeta)P^0_{-s}(\cosh\zeta)\sinh\zeta\,d\zeta\end{aligned}$$

which is bounded as $\Im s\to\pm\infty$ when $f\in\mathcal{D}(G/K)$ by our above remarks concerning P^0_{-s}. Thus $\check{f}(s)$ is rapidly decreasing. (Recall that $\sinh\zeta\,d\zeta$ is the Haar measure and $\Delta_{G/K}$ is self-adjoint for this measure.)

We conclude that $\check{f}(s)\in\hat{\mathcal{D}}(\mathbb{R}^1)$. We claim that this implies $\hat{f}\in\mathcal{D}(\mathbb{R}^1)$. To understand this point, suppose that $h\in\mathcal{D}(\mathbb{R}^1)$ and support $h\subset(0,\infty)$. Then, by the same notion of analytic continuation we have used,

$$\begin{aligned}\hat{h}(\hat{t}) &= \int e^{ix\hat{t}}h(x)\,dx\\ &= \iint\Gamma(s)(-i)^{-s}\hat{t}^{-s}x^{-s}h(x)\,dx\,ds\\ &= \int\Gamma(s)(-i)^{-s}\hat{t}^{-s}\underset{\vee}{h}(s)\,ds.\end{aligned}$$

Thus the function $\hat{h}(\hat{t})$ is represented as a Mellin transform of $\Gamma(s)(-i)^{-s}\underset{\vee}{h}(s)$ where $\underset{\vee}{h}$ is an arbitrary function of $\hat{\mathcal{D}}(\mathbb{R}^1)$.

Conversely such a representation implies $\hat{h} \in \hat{\mathcal{D}}(\mathbb{R}^1)$ by reversing the steps:

$$\begin{aligned}\int \Gamma(s)(-i)^{-s}\hat{t}^{-s}\check{h}(s)\,ds &= \iint \Gamma(s)(-i)^{-s}\hat{t}^{-s}h(x)x^{-s}\,ds\,dx \\ &= \int h(x)\,dx \int x^{-s}(-i)^{-s}\Gamma(s)\hat{t}^{-s}\,ds \\ &= \int h(x)e^{ix\hat{t}}\,dx.\end{aligned}$$

We conclude that (7.87) defines an automorphism of $\hat{\mathcal{D}}(\mathbb{R}^1)$. In particular $f \in \mathcal{D}(G/K)$ implies $\hat{f} \in \hat{\mathcal{D}}(\hat{\Gamma})$. The rapid decrease of $\hat{f}$ is the sought-after property.

Remark 1 *The Weyl invariance under $s \to 1-s$ of $\check{f}$ is the counterpart of the Kelvin oddness of $\hat{f}$.*

Remark 2 *Our study of wave front sets as given in Chapter 5 indicates that if a function h is smooth in a given codirection then its Fourier transform should be small (in a suitable sense) in that codirection. "Codirection" is defined by the linear structure of $\mathbb{R}^n$.*

Our above discussion can be considered as an analogous idea except that "linear codirections" are replaced by the hyperboloid G/K.

Problem 7.1 *Develop a general theory of such "nonlinear wave front sets."*

(**H**) We shall now explain how to pass from K invariant functions on G/K to general functions $f \in \mathcal{D}(G/K)$. By decomposing functions on $\mathbf{g}/K$ under left action of the circle group K, i.e. using coordinates ζ, θ on G/K and expressing functions in terms of Fourier series in θ, we can see that any such f admits a convergent decomposition

$$f = \sum f_n$$

where

$$f_n(\zeta, \theta + \theta_0) = e^{in\theta_0} f_n(\zeta, \theta).$$

f_n is called a function of type n.

It is clear that the Fourier transform and the Radon transform commute with K so it suffices to restrict our considerations to functions of fixed type. However, the analog of Theorem 7.4 is somewhat more complicated. The differential operator $T_1 = \{t = 1\}$ of lowest order which transforms according to the character $\exp(in\theta)$ is $\partial^n/\partial\bar{z}^n$ if $n > 0$ and $\partial^n/\partial z^n$ if $n < 0$. The role in Theorem 7.4 of $\delta_{(1,0,0)}$ on T_1 is replaced by $(\partial^n/\partial\bar{z}^n)\delta_{(1,0,0)}$ or $(\partial^n/\partial z^n)\delta_{(1,0,0)}$. For simplicity of notation we shall assume $n \geq 0$ in what follows.

We call $\Delta^n_{(1,0,0)}$ the solution of the CP with

$$\mathrm{CD} = [0, (\partial^n/\partial\bar{z}^n)\delta_{(1,0,0)}] \quad \text{on } T_1.$$

We have to compute the CD of Δ^n on G/K. Since T_1 is tangent to $r = 1$ at $(1, 0, 0)$ the fact that the propagation speed of solutions of $\Box$ is 1 leads to

$$(\text{support } \Delta^n_{(1,0,0)}) \cap \{r = 1\} = \{(1, 0, 0)\}.$$

$\Delta^n_{(1,0,0)}$ is odd under $t \to 1-t$, but this does not imply directly that $\Delta^n_{(1,0,0)}$ is odd under the Kelvin transform $\mathcal{K}$ when $n \neq 0$. Since $\Box$ commutes with rotations around the t axis $\mathrm{CD}_{r=1}(\Delta^n_{(1,0,0)})$ transforms like $\exp(in\theta)$ under such rotations.

In polar coordinates on T_1

$$\frac{\partial}{\partial \bar{z}} = e^{i\theta}\frac{\partial}{\partial r} + i\frac{e^{i\theta}}{r}\frac{\partial}{\partial \theta}$$

so that

$$\frac{\partial^n}{\partial \bar{z}^n} = e^{in\theta}\sum \binom{n}{j} i^j r^{-j}\frac{\partial^n}{\partial r^{n-j}\partial \theta^j}.$$

We can compute the CD of $\Delta^n_{(1,0,0)}$ on $r = 1$ as follows. Find the formal power series in t for $\Delta^n_{(1,0,0)}$ on $T_1 = \{t = 1\}$ near $(1, 0, 0)$. This can be computed in the usual manner (*à la* Cauchy–Kowalewski) using the expression $\Box = \partial^2/\partial t^2 - \Delta$ where Δ is the Laplacian[5] in x, y. The coefficients of the power series are distributions on T_1 supported at $(1, 0, 0)$. Then make a change of variables to realize $\Delta^n_{(1,0,0)}$ as a formal power series on $r = 1$ with coefficients which are distributions on $r = 1$ supported at $(1, 0, 0)$. These formal power series do not converge; the CD is the first two terms of the power series. The computation shows that, in fact, $\Delta^n_{(1,0,0)}$ is Kelvin invariant.

We shall not go through the details of this complicated, though straightforward, calculation. They lead to the analog of the Hilbert transform for functions of type n.

This method can be carried out to derive the extension to $n \neq 0$ of all of our above results for $n = 0$. However, the method is somewhat complicated because $\partial^n/\partial \bar{z}^n$ is not of first order. A simpler approach starts with the solution of Δ_p of $\Box\Delta_p = 0$ whose CD at the point $p \in G/K$ is $(0, \delta)$. Δ_p is clearly invariant under K_p which is the Lorentz rotation group around the T_p axis which is the line through the origin passing through p.

Since G commutes with $\Box$ and with the Kelvin transformation, Δ_p is obtained from $\Delta_{(1,0,0)}$ by applying a Lorentz transformation g_p which takes $(1, 0, 0)$ to p. The Fourier transform $\hat{\Delta}_p$ is obtained from $\hat{\Delta}_{(1,0,0)}$ by applying g_p. By multiplying by $\exp(-in\theta)$ and integrating over K we can compute the Fourier transform of the solution ${}_n\Delta_{Kp}$ of the wave equation with

$$\mathrm{CD}({}_n\Delta_{Kp}) = (0, e^{in\theta}\delta_{Kp}).$$

We can use these ideas to determine the image of $\mathbf{R}^N$ on all of $\mathcal{D}(G/K)$. Suppose we are given functions $f^N(\theta, s)$ for each spread θ which depend smoothly

[5]The reader should distinguish between the Laplacian Δ and the solution $\Delta^n_{(1,0,0)}$.

on θ and, as functions of s, are in $\mathcal{D}(s)$ and vanish for $s \leq c$ for some $c > 0$. Suppose f^N satisfies the moment conditions as in Theorem 7.3. Then the Fourier transforms $\hat{f}^N(\theta, \hat{s})$ in the spread variable s fit together to define a function $\hat{f}^N \in \hat{\mathcal{D}}(\hat{\Gamma})$. By decomposition in Fourier series in θ (or $\hat{\theta}$) we obtain functions $\hat{f}_n^N \in \hat{\mathcal{D}}(\hat{\Gamma})$ of type n.

Using essentially the same technique as in the case $n = 0$ we can then determine that f^N is of the form $\mathbf{R}^N f$ for some $f \in \mathcal{D}(G/K)$ if and only if each f_n^N satisfies the analog of Kelvin invariance.

The G Fourier inversion formula and the G Paley–Wiener theorem for G/K can also be derived along the same lines. The Kelvin invariance, in this form, is referred to as *Weyl invariance*. $P_{-s}^0 = P_{-(1-s)}^0$ is replaced by

$$P_{-s}^n = P_{s-1}^n$$

(see [11], vol. I, p. 140 (1)). We can derive this for arbitrary n as we did for $n = 0$. Following the method developed above for $n = 0$ we find that a function f_n of type n must satisfy

$$\check{f}_n(s)\frac{\Gamma(s)}{\Gamma(s+m)} = \check{f}_n(1-s)\frac{\Gamma(1-s)}{\Gamma(1-s+m)}. \tag{7.88}$$

This implies that

$$\check{f}_n(s) = 0 \quad \text{for } s = 0, -1, \ldots, -n+1 \ (n > 0) \tag{7.89}$$

with similar vanishing when $m < 0$.

The Paley–Wiener theorem for G is due to [68–70].

Our presentation is different from the original one.

The approach we have given is based on the WP on the t axis and on the CP on the plane $t = 1$. The reason we can deal directly with these parametrization problems is the linear nature of the parametrization surfaces. In particular we used the CP on the plane $t = 1$ to explicitly invert the Fourier transform of $\delta_{(1,0,0)}$ on G/K.

There is another PS which can be dealt with in an essentially direct way; this is Γ (or, more precisely, Γ^+). The idea of parametrizing solutions of the wave equation by data on Γ is due to d'Adhémar (see [134]).

The integration by parts method is very useful in the present situation, but we wish to digress and explain a different approach—one which can be used when the real field is replaced by other fields, e.g. finite or p-adic (see Section 10.7).

By duality we want to express, e.g. $\delta_{1,0,0}$, in the form (see Section 1.4)

$$\delta_{1,0,0} = \int \alpha(t)\delta_{r=t}\, dt + \Box T$$

where $\delta_{1,0,0}$ is the unit measure on the circle $r^2 = x^2 + y^2 = t^2$. (Such a formula can be expected because $\delta_{1,0,0}$ is invariant under rotations around the t axis.)

We want to solve this equation for α. As usual we take the (three-dimensional) Fourier transform; this leads to

$$e^{i\hat{t}} = \iint \alpha(t) e^{it\hat{t}+ir\hat{r}\cos\theta}\, dt\, d\theta \quad \text{on } \hat{\Gamma}.$$

We have $t = r$ and $\hat{t} = \hat{r}$ since we are on $\hat{\Gamma}$. Thus

$$e^{i\hat{t}} = \int \alpha(t)\, dt \int e^{it\hat{t}(1+\cos\theta)}\, d\theta.$$

The right side of this equation is essentially a multiplicative convolution (in t). We can solve for α by taking the Mellin transform in $\hat{t}$. Except for powers of i (which cancel) the equation becomes

$$\begin{aligned}\Gamma(s) &= \Gamma(s) \int \alpha(t) t^{-s}\, dt \int (1+\cos\theta)^{-s}\, d\theta \\ &= \Gamma(s)\beta(1/2, 1/2-s)\underset{\vee}{\alpha}(1-s)\end{aligned}$$

by standard β function formulas [11, vol. I, p. 10]. Here $\underset{\vee}{\alpha}$ is the Mellin transform of α.

We have shown that, up to simple constants,

$$\begin{aligned}\underset{\vee}{\alpha}(1-s) &= [\beta(1/2, 1/2-s)]^{-1} \\ &= \beta(-1/2, 1-s) \\ &= \int t^{-s}(1-t)^{-3/2}\, dt.\end{aligned}$$

Thus

$$\alpha(t) = \begin{cases} (1-t)^{-3/2} & t \le 1 \\ 0 & t > 1 \end{cases}$$

which is d'Adhémar's result.

We return to our main theme.

One might wonder why Γ is easier to deal with than G/K. Although G/K and Γ both have the structure of KA, in the case of Γ the components K, A commute because A acts by scalar multiplication. In fact the A action commutes with all of G. For G/K the components K, A do not commute.

From the point of view of general harmonicity we have seen in Chapter 3 that $\vec{\imath} = 0$ is easier in several ways than $\vec{\imath} = \vec{c}, \vec{c} \neq 0$.

There is an "intertwining" between G/K and Γ^+ which is accomplished by means of integration by parts as in (7.56), except that the coordinate system (r, ζ, θ) degenerates on the light cone and so must be replaced by another coordinate system near the light cone.

We can use the following idea to deal with the limit on Γ (which is treated in detail in [134]). Note that scalar multiplication, hence also the infinitesimal generator $r\partial/\partial r$, preserve the kernel of $\Box$. If F is the solution of the CP with $\mathrm{CD}(F) = (0, f)$ then

$$G = r\frac{\partial \bar{F}}{\partial r} + \frac{1}{2}\bar{F}$$

satisfies $\Box G = 0$ and

$$\mathrm{CD}(G) = (\bar{f}, g)$$

for some function g. g is of no significance to us because the inner product

$$\begin{aligned}\iint_{G/K}\left[F\frac{\partial G}{\partial r} - G\frac{\partial F}{\partial r}\right] &= \iint_{G/K}\left[F\left(r\frac{\partial}{\partial r}+\frac{1}{2}\right)G - G\left(r\frac{\partial}{\partial r}+\frac{1}{2}\right)F\right]\\ &= -\iint_{G/K}|f|^2.\end{aligned}$$

Since $G = (r\partial/\partial r + 1/2)\bar{F}$ the invariance of the inner product under scalar multiplication gives

$$-\iint_{G/K}|f|^2 = \iint_{r=\text{const.}} F\left(r\frac{\partial}{\partial r}+\frac{1}{2}\right)^2\bar{F} - \left|\left(r\frac{\partial}{\partial r}+\frac{1}{2}\right)F\right|^2.$$

This expresses the L_2 norm of f in terms of integrals of F on the half-hyperboloids $r = \text{const.} > 0$, $t > 0$. To pass to the light cone $r = 0$ we need to examine a suitable coordinate system near the light cone.

One coordinate system that makes sense near Γ^+ is cylindrical coordinates t, ρ, θ where

$$\begin{aligned}x &= \rho\cos\theta\\ y &= \rho\sin\theta.\end{aligned} \tag{7.90}$$

The measure is

$$dt\,dx\,dy = \rho\,dt\,d\rho\,d\theta$$

and

$$\Box = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial \rho^2} - \frac{1}{\rho}\frac{\partial}{\partial \rho} - \frac{1}{\rho^2}\frac{\partial^2}{\partial \theta^2}. \tag{7.91}$$

Actually, it is better to use a slight variation of cylindrical coordinates, i.e. we set

$$\begin{aligned}\rho &= r_\Gamma + \sigma\\ t &= r_\Gamma - \sigma.\end{aligned} \tag{7.92}$$

Thus

$$\begin{aligned}\frac{\partial}{\partial t} &= \frac{1}{2}\left(\frac{\partial}{\partial r_\Gamma} - \frac{\partial}{\partial \sigma}\right)\\ \frac{\partial}{\partial \rho} &= \frac{1}{2}\left(\frac{\partial}{\partial r_\Gamma} + \frac{\partial}{\partial \sigma}\right).\end{aligned} \tag{7.93}$$

The wave operator becomes

$$\Box = -\frac{\partial^2}{\partial r_\Gamma \partial \sigma} - \frac{1}{2(r_\Gamma + \sigma)}\left(\frac{\partial}{\partial r_\Gamma} + \frac{\partial}{\partial \sigma}\right) - \frac{1}{4(r_\Gamma+\sigma)^2}\frac{\partial^2}{\partial \theta^2}. \tag{7.94}$$

The measure is

$$2(r_\Gamma + \sigma)\, d\theta\, dr_\Gamma\, d\sigma.$$

It is now easy to integrate by parts in σ and compute the contribution from the light cone which is $\{\sigma = 0\}$. (r_Γ is the natural parameter on the generators of Γ.) We find

$$\begin{aligned}\iiint u\Box v = \iiint v\Box u + \iint_{G/K}\left[v\frac{\partial u}{\partial r} - u\frac{\partial v}{\partial r}\right]\\ + 2\iint_{\Gamma^+}\left[-vr_\Gamma\frac{\partial u}{\partial r_\Gamma} - \frac{1}{2}uv\right] dr_\Gamma\, d\theta.\end{aligned} \tag{7.95}$$

Remark 1 *There is no problem in integrating by parts using different coordinate systems in different regions. For the boundary terms are intrinsically defined by $\Box$ and hence the interior boundary terms cancel.*

Remark 2 *One has to take into account the singular point which is the origin. It does not play any role for functions on G/K. It is significant for G/A.*

We shall return to this later.

In our above treatment based on the CP much use was made of the Γ function and its analytic continuation (see (7.74)ff.). The Γ function arose from the Mellin transform of $\exp it\hat{t}[\cosh\zeta - \sinh\zeta\cos(\theta - \hat{\theta})]$ which is the exponential of the inner product of points on G/K and on the light cone $\hat{\Gamma}$.

We must now deal with the exponential of the inner product $t\hat{t}[1-\cos(\theta-\hat{\theta})]$ between points $(t, t\cos\theta, t\sin\theta)$ and $(\hat{t}, \hat{t}\cos\hat{\theta}, \hat{t}\sin\hat{\theta})$ on the light cones Γ and $\hat{\Gamma}$. There is an additional complication which comes from the vanishing of $1 - \cos(\theta - \hat{\theta})$ for $\theta - \hat{\theta} = 0$; our previous methods do not apply.

To remedy this situation we use the analytic continuation we have described above on each hyperboloid $r = c > 0$. We then define operations on Γ in terms of limits from $r = c > 0$. Some care must be taken in passing to the limit; the calculations are similar to the calculations made in Chapters 3 and 4 regarding passage in the plane from hyperbolas to the light cone. The main point is that

the limits are taken for fixed t as a point x, y with $x^2 + y^2 < t^2$ approaches the light cone.

We shall not give the details of the limit process as they are tedious but essentially straightforward.

As in our previous treatment (see (7.56)ff.) we set

$$v(r, \zeta, \theta) = e^{it\hat{t} - ix\hat{x} - iy\hat{y}}$$

with $(\hat{t}, \hat{x}, \hat{y}) = (\hat{t}, \hat{t}\cos\hat{\theta}, \hat{t}\sin\hat{\theta})$ and $u = F =$ solution of CP with CD $= (0, f)$ on G/K. Equation (7.95) becomes

$$\hat{f}(\hat{t}, \hat{x}, \hat{y}) = -\iint_{\Gamma^+} e^{it\hat{t}[1+\cos(\theta-\hat{\theta})]} \left(r_\Gamma \frac{\partial F}{\partial r_\Gamma} + \frac{1}{2} F \right) dr_\Gamma \, d\theta \tag{7.96}$$

because $r\cosh\zeta = t$ and $r\sinh\zeta = \rho \to t$. Moreover $r_\Gamma = t\sqrt{2}$ so the difference between t and r_Γ plays no essential role.

We assume that f and hence F are K invariant; the general case, as before, is an easy modification of this one combined with expansion in Fourier series on K.

Here is our strategy:

(α) f is determined by $\hat{f}$. ($\hat{f}$ is defined only on $\hat{\Gamma}$.)
(β) F is determined by Kelvin invariance and $\hat{f}$.
(γ) $\hat{f}$ can be expressed as a "Fourier transform from Γ^+ to $\hat{\Gamma}$" of $(r_\Gamma \partial/\partial r_\Gamma + 1/2)F\big|_{\Gamma^+}$.
(δ) Starting with $\hat{f}$ we can construct F such that $\hat{f}$ is the Fourier transform of $(r_\Gamma \partial/\partial r_\Gamma + 1/2)F\big|_{\Gamma^+}$ from Γ^+ to $\hat{\Gamma}$. This inverts (γ).
(ϵ) We express CD(F) on G/K in terms of the group Fourier transform $\check{f}(s)$ of f. This establishes the Paley–Wiener theorem and the Plancherel inversion formula for G/K.

A is the injectivity of $\mathbf{R}^N$ which we have seen is (α). (β) is the uniqueness of the CP on $\{r = 1\}$. (γ) is (7.96). We are left with (δ) and (ϵ).

To invert (7.96) we think of Γ^+ as $K \times \mathbb{R}^\times$. Since F is independent of θ we can replace $\cos(\theta - \hat{\theta})$ by $\cos\theta$. We evaluate the integral by using the Mellin transform in $\hat{t}$. By our previous discussion of analytic continuation related to the Γ function we can replace the exponential by the inverse Mellin transform of its Mellin transform in $\hat{t}$. Thus

$$\begin{aligned}
\hat{f}(\hat{t}) &= \iiint_{\Gamma^+ \times \{s\}} (-it\hat{t})^{-s} (1 + \cos\theta)^{-s} \Gamma(s) \left(t\frac{\partial}{\partial t} + \frac{1}{2} \right) F(t) \, dt \, d\theta \, ds \\
&= \iint \Gamma(s)(-i\hat{t})^{-s} (1 + \cos\theta)^{-s} \left(s - \frac{1}{2} \right) \underset{\vee}{F}(-s+1) \, ds \, d\theta \\
&= \int \Gamma(s)(-i\hat{t})^{-s} c(-s) \left(s - \frac{1}{2} \right) \underset{\vee}{F}(-s+1) \, ds.
\end{aligned} \tag{7.97}$$

$c(s)$ is Harish–Chandra's function [86–88]

$$\begin{aligned} c(s) &= \int_0^{2\pi} (1+\cos\theta)^s \, d\theta \\ &= 2^s \int_0^{\pi} \cos^{2s}\theta \, d\theta \\ &= 2^{2+1}\beta\left(\tfrac{1}{2}, \tfrac{1}{2}+s\right) \end{aligned} \tag{7.98}$$

as is seen by making the change of variable $u = \cos^2\theta$ [11, vol. I, p. 10 (19)].

To carry out (δ) we want to construct F from $\hat{f}$ as given by (7.97). Since F is a solution of $\Box F = 0$ we should express F as a Fourier transform over $\hat{\Gamma}$. Equation (7.97) suggests that everything depends on $\hat{t}^s$, so we compute the Fourier transforms

(a) $\mathcal{F}_{\hat{\Gamma}^+ \to G/K}(\hat{t}^s)$,
(b) $\mathcal{F}_{\hat{\Gamma}^+ \to \Gamma^+}(\hat{t}^s)$.

We shall make the computations formally. The discussion in (7.74)ff. can be used to put these calculations on a rigorous basis.

To compute the Fourier transform (a) we have as in (7.74)

$$\begin{aligned} &\iint_{\Gamma^+} e^{i\hat{t}(\cosh\zeta - \sinh\zeta\cos\hat{\theta})} \hat{t}^s \, d\hat{t} \, d\hat{\theta} \\ &\quad = \Gamma(s+1)(-i)^{-s-1} \int (\cosh\zeta - \sinh\zeta\cos\hat{\theta})^{-s-1} \, d\hat{\theta} \\ &\quad = \Gamma(s+1)(-i)^{-s-1} P^0_{-s-1}(\cosh\zeta). \end{aligned} \tag{7.99}$$

This is the Fourier transform of $\hat{t}^s$ from Γ^+ to $G/K = \{r = 1\}$. For general r we have clearly

$$\iint_{\Gamma^+} e^{ir\hat{t}(\cosh\zeta - \sinh\zeta\cos\hat{\theta})} \hat{t}^s \, d\hat{t} \, d\hat{\theta} = \Gamma(s+1)(-ir)^{-s-1} P^0_{-s-1}(\cosh\zeta). \tag{7.100}$$

For (b)

$$\begin{aligned} \iint_{\Gamma^+} e^{it\hat{t}(1-\cos\hat{\theta})} \hat{t}^s \, d\hat{t} \, d\hat{\theta} &= \Gamma(s+1)(-it)^{-s-1} \int (1-\cos\hat{\theta})^{-s-1} \, d\hat{\theta} \\ &= \Gamma(s+1)(-it)^{-s-1} c(-s-1). \end{aligned} \tag{7.101}$$

We now compare (7.97) with (7.87). We find

$$\check{f}(s) = \left(-s+\tfrac{1}{2}\right) c(-s) \underset{\vee}{F}(-s+1). \tag{7.102}$$

We use (7.102) to verify that F is the Fourier transform of $\frac{i}{2}\hat{f}\operatorname{sgn}\hat{t}$, which is the "d'Adhémar proof" of Theorem 7.5. This is the same as saying that the Fourier transform of $\hat{f}\operatorname{sgn}\hat{t}$ is Kelvin odd and that $F\big|_{\Gamma} = (i/2)\mathcal{F}_{\hat{\Gamma}\to\Gamma^+}(\hat{f}\operatorname{sgn}\hat{t})$. By (7.87), (7.102), and (b) above (taking into account $\hat{\Gamma} = \hat{\Gamma}^+ \cup \hat{\Gamma}^-$)

$$\begin{aligned}
&\int_{\hat{\Gamma}} \hat{f}(\hat{t})\operatorname{sgn}\hat{t}e^{it\hat{t}(1-\cos\hat{\theta})}\,d\hat{t}\,d\hat{\theta}\\
&\quad= \iiint \check{f}(s)\Gamma(s)(-i)^{-s}\hat{t}^{-s}\operatorname{sgn}\hat{t}e^{it\hat{t}(1-\cos\hat{\theta})}\,d\hat{t}\,d\hat{\theta}\,ds\\
&\quad= \iint \check{f}(s)\Gamma(s)(-i)^{-s}\Gamma(-s+1)[(-i)^{s-1}-(i)^{s-1}](1-\cos\hat{\theta})^{s-1}t^{s-1}\,ds\,d\theta\\
&\quad= -i\int \check{f}(s)\Gamma(s)\Gamma(1-s)[(-1)^{-s}+(-1)^{s}]c(s-1)t^{s-1}\,ds\\
&\quad= -2i\int \left(s-\frac{1}{2}\right)c(-s)c(s-1)[\sin\pi s]^{-1}\cos\pi s\underset{\vee}{F}(-s+1)t^{s-1}\,ds. \qquad (7.103)
\end{aligned}$$

From (7.98), up to a simple constant multiple,

$$\begin{aligned}
c(-s)c(s-1) &= \frac{\Gamma(\frac{1}{2}-s)\Gamma(s-\frac{1}{2})}{\Gamma(1-s)\Gamma(s)}\\
&= \frac{\Gamma(\frac{1}{2}-s)\Gamma(\frac{1}{2}+s)}{(s-\frac{1}{2})\Gamma(s)\Gamma(1-s)}\\
&= \frac{\sin\pi s}{(s-\frac{1}{2})\cos\pi s}. \qquad (7.104)
\end{aligned}$$

Applying the Mellin inversion formula to the last integral in (7.103) yields

$$\mathcal{F}_{\hat{\Gamma}\to\Gamma^+}(\hat{f}\operatorname{sgn}\hat{t}) = -2iF(t) \qquad (7.105)$$

which is the sought-after result.

For (ϵ) we use (a), i.e. (7.100), in place of (b). We replace $t(1-\cos\hat{\theta})$ on the left side of (7.103) by $r(\cosh\zeta - \sinh\zeta\cos\hat{\theta})$.

$$\begin{aligned}
&\int_{\hat{\Gamma}} \hat{f}(\hat{t})\operatorname{sgn}\hat{t}e^{ir\hat{t}(\cosh\zeta-\sinh\zeta\cos\hat{\theta})}\,d\hat{t}\,d\hat{\theta}\\
&\quad= \iiint \check{f}(s)\Gamma(s)(-i)^{-s}\hat{t}^{-s}\operatorname{sgn}\hat{t}e^{ir\hat{t}(\cosh\zeta-\sinh\zeta\cos\hat{\theta})}\,d\hat{t}\,d\hat{\theta}\,ds\\
&\quad= \int \check{f}(s)\Gamma(s)(-i)^{-s}\Gamma(-s+1)r^{s-1}[(-i)^{s-1}-(i)^{s-1}]P^0_{s-1}(\cosh\zeta)\,ds\\
&\quad= -2i\int \check{f}(s)\cot\pi s P^0_{s-1}(\cosh\zeta)r^{s-1}\,ds. \qquad (7.103^*)
\end{aligned}$$

By (7.105) the function defined by (7.103*) is $-2iF$. The CD on $r = 1$ is $(0, f)$. Applying the operator $r\partial/\partial r + 1/2$ (the normal derivative) to (7.103*) leads to

$$f(\cosh\zeta) = -2i \int \check{f}(s) \left(s - \frac{1}{2}\right) \cot \pi s P^0_{s-1}(\cosh\zeta)\, ds$$

which is the Plancherel formula for K invariant functions on G/K.

The same proof can be used to derive the Plancherel formula for functions on G/K which transform under a fixed character of K and hence for all functions on G/K. The same technique can be used to establish the Paley–Wiener theorem for G/K.

There is another aspect of this d'Adhémar PS which is noteworthy. If we start with (7.100) and, as usual, let $r \to 0$ holding t fixed, we find that

$$\lim_{\substack{r\cosh\zeta = t \\ (t,x,y)\to\Gamma}} r^{-s} P^0_{-s}(\cosh\zeta) = t^{-s} c(-s) \tag{7.106}$$

because the point $(r\cosh\zeta, r\sinh\zeta\cos\theta, r\sinh\zeta\sin\theta) \to (t, t\cos\theta, t\sin\theta)$ so the left side of (7.100) $\to$ left side of (7.101). Since $t = r\cosh\zeta$ this is the same as

$$P^0_{-s}(\cosh\zeta) \sim c(-s)\cosh^{-s}\zeta. \tag{7.107}$$

This is the standard asymptotic formula for the Legendre function (see [11, vol. I, p. 164 (19),(20)]); formula (7.107) differs from Bateman's by a constant.

The standard [153, 154] theory of eigenfunction expansions suggests that the Plancherel measure for the P^0_{-s} expansion is

$$|c(-s)|^2 = c(-s)c(s-1) \tag{7.108}$$

because $s = \frac{1}{2} + i\tau$ on the unitary line for P^0_{-s}.

The fact that $\Im s = 1/2$ is the unitary line for P^0_{-s} is seen as follows. The asymptotic formula for $P^0_{-s}(\cosh\zeta)$ is $\exp(-s|\zeta|)$ while the ζ measure is $\sinh\zeta\, d\zeta \sim \exp|\zeta|\, d\zeta$. Thus, wave packets (see Section 1.1) formed by $P^0_{-s}(\cosh\zeta)$ are close to L_2 when

$$\begin{aligned} |P^0_{-s}(\cosh\zeta)|^2 \sinh\zeta &\sim e^{|\zeta|(-2s+1)} \\ &\sim \in L_2 \end{aligned}$$

which means

$$\Re s = \tfrac{1}{2}.$$

7.3 Geodesic Radon transform

Our treatment of the horocyclic Radon transform was based on the fact that any N orbit on G/K is a horocycle. Such is no longer the case for A orbits.

To understand this point note that an N orbit on G/K is of the form NgK. On the other hand, a horocycle is a translate of NK that is g_1NK. Thus we want to know if for every g there is a g_1 with

$$g_1NK = NgK. \tag{7.109}$$

Since $G = NAK = KAN$ this equation becomes

$$k_1a_1NK = NaK. \tag{7.110}$$

The solution of (7.110) is $k_1 =$ identity, $a_1 = a$. The crucial point is that A normalizes N so g_1 depends only on g.

On the other hand, if we reverse the roles of A and N then equation (7.110) becomes

$$k_1n_1AK = AnK \tag{7.111}$$

or

$$k_1n_1a_1 = ank. \tag{7.112}$$

To solve (7.112) for k_1, n_1, a_1, we regard G as the group of fractional linear transformations of the Poincaré upper half-plane as in (7.15)ff. We write $n = \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right), a = \left(\begin{smallmatrix} a & 0 \\ 0 & a^{-1} \end{smallmatrix}\right)$. Then (7.112) is

$$k_1(a_1^2i + n_1) = a^2i + a^2n. \tag{7.113}$$

Now, n_1 and a_1 are uniquely determined by (7.112). For if $k_1n_1a_1 = k_2n_2a_2$ then $n_1a_1a_2^{-1}n_2^{-1} \in K$. But $n_1a_1a_2^{-1}n_2^{-1}$ is a triangular matrix and hence cannot be in K unless it is the identity matrix, which means $n_1 = n_2, a_1 = a_2$. One solution of (7.113) is $k_1 =$ identity, $a_1 = a$, and $n_1 = a^2n$. Hence n_1 (which is unique) depends on a so AnK is not of the form g_1AK.

We present two approaches to the geodesic Radon transform:

(a) *Upper half-plane approach.* A geodesic is defined as gAK or, in terms of coordinates in the upper half-plane, gAi. For $g =$ identity the geodesic is the imaginary axis. Since g defines a conformal (angle-preserving) map and g preserves the real axis, it follows that gAi is a (semi)circle orthogonal to the real axis. Since G acts transitively on such semicircles, $\{gAi\}$ is exactly the set of semicircles centered on the real axis (see Section 10.1 for details on what follows).

To compute the isotropy group of a geodesic we use the polar decomposition $G = AKA$; we compute the isotropy group of Ai. For $akAi$ to equal Ai it must be the case that $kAi = Ai$. Since k is a rotation about the point i and Ai is the imaginary axis, it follows that $k =$ identity or $k =$ rotation by π, i.e. $k = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, so the space of geodesics can be identified with $G/\tilde{A}$ where $\tilde{A}$ is the normalizer of A in G. $\tilde{A}$ is generated by A and the Weyl group

$$W = \left\{\text{identity}, w = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\right\}. \tag{7.114}$$

It is easily seen that in the corresponding situation for horocyclic flow the space of horocycles is G/N.

Let us examine the action of W on Minkowski space:

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u & z \\ z & v \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} v & -z \\ -z & u \end{pmatrix}. \tag{7.115}$$

In terms of t, x, y coordinates

$$w : (t, x, y) \mapsto (t, -x, -y). \tag{7.116}$$

This means that w preserves t coordinate and acts as the antipodal map in the (x, y) variables. Moreover A is the subgroup leaving $(0, 0, 1)$ invariant. In particular if we factor the hyperboloid $G/A : t^2 - x^2 - y^2 = -1$ by W then we replace the circles $x^2 + y^2 = t^2 + 1$ (t fixed) by the projectivized circles. Hence we can identify the space of geodesics with this projectivized hyperboloid.

What about the spread formulation? A reasonable idea for spreads is the set of all circles orthogonal to the real axis with fixed center. The spread parameter is the radius of the circle and the Grassmann parameter is the center. For any fixed Grassmann parameter c the corresponding spread functions consist of all functions of $(x - c)^2 + y^2 = \rho_c^2$.

We must compute the Fubini (spread) measure. We use coordinates θ, ρ_c where

$$\begin{aligned} x - c &= \rho_c \cos\theta \\ y &= \rho_c \sin\theta. \end{aligned} \tag{7.117}$$

Unlike our previous examples the coordinate system depends on the specific spread.

The Poincaré measure in the upper half-plane is

$$\frac{dx\, dy}{y^2} = \frac{d\rho_c\, d\theta}{\rho_c \sin^2\theta}. \tag{7.118}$$

This means that the spread measure is $d\theta/(\rho_c \sin^2\theta)$ or $d\theta/\sin^2\theta$ (i.e. we can combine the factor ρ_c with $d\rho_c$). The line element is

$$\begin{aligned} ds^2 &= \frac{dx^2 + dy^2}{y^2} \\ &= \frac{d\rho_c^2 + \rho_c^2\, d\theta^2}{\rho_c^2 \sin^2\theta}. \end{aligned} \tag{7.119}$$

Hence on $\rho_c = \text{const.}$ the measure is

$$ds = \frac{d\theta}{\sin\theta}. \tag{7.120}$$

Since ds^2 is G invariant (7.155) is the A invariant measure on $\rho_c = \text{const}$. Thus the spread measure (as defined by (7.32)ff.) is *not* the (A invariant) measure we use to define $\mathbf{R}^A$.

Suppose $\mathbf{R}^A f$ vanishes on the spread c. Let h be a function of ρ_c only. Then for "nice" functions $h(\rho_c)$

$$\begin{aligned}\iint yh(\rho_c)f(x,y)\frac{dx\,dy}{y^2} &= \int h(\rho_c)\,d\rho_c \int f(\rho_c,\theta)\,\frac{d\theta}{\sin\theta}\\ &= \int h(\rho_c)\mathbf{R}^A f(\rho_c)\,d\rho_c\\ &= 0. \end{aligned} \tag{7.121}$$

This means that $\mathbf{R}^A f \equiv 0$ is equivalent to yf being orthogonal to all (Fubini) spread functions $h(\rho_c)$. (We assume that f vanishes sufficiently at $y = 0$.)

We are left with the problem of determining whether the linear combinations of $\{h(\rho_c)\}$ for varying c are dense in the dual of the space of f.

Remark. *We do not need to use the Poincaré area element $y^{-2}\,dx\,dy$ to define the spread measure. Rather we need a measure whose Fubini measure for the given spreads is simply related to the given (A invariant) measure on the geodesics. The only property of the measure on G/K that is needed is that it is positive and smooth and that the resultant Fubini measure on the geodesics (leaves of the spreads) is, on each leaf, a simple multiple of the given (A invariant) measure on that geodesic. For we want to know that if the linear combinations of spread functions are dense then $\mathbf{R}^A f = 0$ implies $f = 0$. In fact, the above shows it is better to use $y^{-1}\,dx\,dy$.*

Our first approach to the geodesic Radon transform is

(a) *Polynomial approach based on Lemma 1.1.* A reasonable analog of Lemma 1.1 would involve writing an arbitrary polynomial which is even in y in terms of polynomials which are constant on the leaves of spreads. A basis for such polynomials is $\{[(x+a)^2 + y^2]^n\}$. We have

Lemma 7.9 *Any monomial $x^\alpha y^{2\beta}$ can be written as a linear combination of $\{[(x+a)^2 + y^2]^n\}$ with $n \le 2\beta+\alpha+1$ and at most $(1/2)(2\beta+\alpha+3)(2\beta+\alpha+2)$ values of a.*

Proof We use the binomial theorem to expand $[(x+a)^2 + y^2]^n$ in powers of a:

$$\begin{aligned}[(x+a)^2 + y^2]^n &= \sum_j \binom{n}{j}(x+a)^{2n-2j}y^{2j}\\ &= \sum_k a^k \sum_j \binom{n}{j}\binom{2n-2j}{k}x^{2n-2j-k}y^{2j}. \end{aligned} \tag{7.122}$$

Note that $k \leq 2n$. Thus by the nonvanishing of the Vandermonde determinant, if we use $2n$ different values a_l of a we can express all the coefficients of the a^k, i.e. the inner sums in the second equation of (7.122), as linear combinations of the $[(x+a_l)^2+y^2]^n$. (We do not need the coefficient of a^{2n} as that is 1.)

We want to express the monomials $x^\alpha y^{2\beta}$ in terms of linear combinations of the $[(x+a_l)^2+y^2]^n$. The problem is that when n and k are fixed there may be several j in the sum in (7.122) so we have only expressed a linear combination of the $x^\alpha y^{2\beta}$ (rather than the individual terms) in terms of $\{[(x+a_l)^2+y^2]^n\}$. If we allow n and k to vary but fix $2n-k$ then the set of possible j is fixed. Hence the monomials $x^\alpha y^{2\beta}$ that appear in the coefficient of a^k for those k (which depend on n) are fixed.

Fixing n and k fixes the degree of the terms $x^{2n-2j-k}y^{2j}$ in (7.122) at $2n-k$. But the coefficients $\binom{n}{j}\binom{2n-2j}{k}$ depend on n,k. It is for this reason that we can solve for fixed $x^\alpha y^{2\beta}$.

To understand this point, we set $2n-k=4$. Then the monomials that occur in the sum in (7.122) are

$$x^4, x^2y^2, y^4. \tag{7.123}$$

We have the problem of determining these monomials from

$$\binom{n}{0}\binom{2n}{2n-4}x^4 + \binom{n}{1}\binom{2n-2}{2n-4}x^2y^2 + \binom{n}{2}\binom{2n-4}{2n-4}y^4. \tag{7.124}$$

It is easy to see that the vectors of coefficients span a three-dimensional space as n varies. This allows us to express the individual terms (7.123) in terms of the (known) sums (7.124).

The same result is valid when $2n-k$ is any fixed number. The proof involves some intricate (but elementary) combinatorics of binomial coefficients which we shall omit as the main consequence of Lemma 7.9 is the known result of Helgason [95]:

Theorem 7.10 *The geodesic Radon transform is injective on* $\mathcal{D}(G/K)$.

Proof Given any $f \in \mathcal{D}(G/K)$ we extend f to the whole complex plane by making it even in y. Then Lemma 7.9 together with (7.121) show that if $\mathbf{R}^A f \equiv 0$ then $|y|f$ is orthogonal in the Poincaré metric to all polynomials which are even in y. (When using the whole plane instead of the half-plane we replace y in (7.121) by $|y|$, or what amounts to the same thing, we use the measure $|y|^{-1}\,dx\,dy$ instead of the Poincaré measure.) Weierstrass's theorem on polynomial approximation implies that $|y|^{-1}f \equiv 0$ and hence $f \equiv 0$.

Lemma 7.9 can be used to study the analog of the moment conditions. These are derived for $\mathbf{R}^A$ just as the usual moment conditions are derived from Lemma 1.1. Thus Lemma 7.9 gives many relations amongst linear combinations of spread polynomials, since as seen from the proof of Lemma 7.9 there is great freedom in choosing spread polynomials to represent a given monomial $x^\alpha y^{2\beta}$.

Problem 7.2 *Do the moment conditions derived from such polynomial identities give the complete set of compatibility conditions for functions $H_x(r)$ to be of the form*

$$H_c(\rho_c) = \int_{(x-c)^2+y^2=\rho_c^2} f$$

for $f \in \mathcal{D}(G/K)$)?

The Iwasawa decomposition $G = NAK = KNA$ indicates that, in terms of the G action on G/K, N acts as a "complement" of A. This suggests that there is another "reasonable" way to define geodesic spreads: $\{nAi\}$ is the spread of geodesics which are parallel to the imaginary axis, meaning that they all pass through ∞. The invariant polynomials are polynomials in x. A general spread is $\{knAi\}_{n\in N}$ which consists of all geodesics passing through a point fixed real x_0. For example, when $k = \left(\begin{smallmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{smallmatrix}\right)$ we have $x_0 = -\cot\theta$. Such a spread can be described as the set of circles of the form (varying c)

$$(x-c)^2 + y^2 = (x_0 - c)^2.$$

A spread function is a function of the center c. Since

$$c = \frac{x^2 + y^2 - x_0^2}{2(x - x_0)}$$

there are no polynomials which are spread functions.

It seems difficult to deal with such functions purely from an algebraic–combinatorial viewpoint.

The two types of spreads we have introduced, namely circles with a fixed center on the real axis and geodesics passing through a given real point, can be described in a somewhat common manner. Circles with a fixed real center x_0 are the geodesics orthogonal to the geodesic $\Re z = x_0$. Thus the set of such spreads can be thought of as follows. We start with a fixed geodesic Ai. The associated spread consists of all geodesics orthogonal to Ai. Then we form the family of "parallel" geodesics $\{nAi\}$; for each we have an associated spread of orthogonal geodesics. Of course we have the freedom of choice as to the meaning of "parallel."

On the other hand, the spread $\{nAi\}$ can be considered as the set of geodesics orthogonal to the horocycle Ni (or any Nai). We now form the set of horocycles $\{kNi\}$ and for each of them the spread $\{knAi\}_{n\in N}$ of geodesics orthogonal to kNi; these are the geodesics passing through the real point $c = -\cot\theta$ where $k = \left(\begin{smallmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{smallmatrix}\right)$.

(b) *Three-dimensional approach.* Recall that, just as in euclidean geometry, geodesics on $G/K =$ hyperboloid $\{t^2 - x^2 - y^2 = 1\}$ are the intersections of G/K with planes through the origin. From this point of view a natural geometric

spread consists of all 2 planes through the origin which intersect the x, y plane in a fixed line. (This is a spread in Minkowski space; we shall relate it to G/K below.)

We write such a plane in the form

$$\begin{aligned} L_{\alpha\theta} : t &= \alpha(x\cos\theta + y\sin\theta) \\ &= ax + by. \end{aligned} \tag{7.125}$$

(The spread is defined by θ fixed, α varying.) Using the expression (7.17) for the points in $\mathbb{R}^3$ which correspond to points in the upper half-plane we deduce from (7.125)

$$\eta + \xi^2\eta^{-1} + \eta^{-1} = a(\eta + \xi^2\eta^{-1} - \eta^{-1}) + 2b\xi\eta^{-1} \tag{7.126}$$

or

$$(1-a)(\xi^2+\eta^2) - 2b\xi = -1 - a,$$

i.e.

$$\begin{aligned} \eta^2 + \left(\xi - \frac{b}{1-a}\right)^2 &= -\frac{(1+a)}{(1-a)} + \frac{b^2}{(1-a)^2} \\ \eta^2 + \left(\xi - \frac{\alpha\sin\theta}{1-\alpha\cos\theta}\right)^2 &= \frac{\alpha^2-1}{(1-\alpha\cos\theta)^2}. \end{aligned} \tag{7.127}$$

This is a circle centered on the real (ξ) axis at $\xi = \alpha\sin\theta/(1-\alpha\cos\theta)$ with radius $= (\alpha^2-1)^{1/2}/(1-\alpha\cos\theta)$.

Problem 7.3 *Give a natural interpretation of this spread on G/K.*

From (7.127) we see that we must have $\alpha^2 \geq 1$ for there to be an intersection of $L_{\alpha\theta}$ with G/K. When $\alpha = 1$, $L_{\alpha\theta}$ is given by

$$t - x\cos\theta - y\sin\theta = 0$$

whose Minkowski normal is $(1, \cos\theta, \sin\theta)$. Note that $L_{\alpha\theta}$ contains the line $\{r(1, \cos\theta, \sin\theta)\}$ which is the generator of the light cone through $(1, \cos\theta, \sin\theta)$. Thus $L_{\alpha\theta}$ is the tangent to the light cone containing this generator.

The only interesting part of this spread is the set of such planes between the tangent plane to the light cone at $(1, \cos\theta, \sin\theta)$ and the tangent plane at $(1, -\cos\theta, -\sin\theta)$ which corresponds to $\alpha \in [-1, 1]$. We call $\mathcal{D}_\theta$ the union of these planes in $t > 0$.

We can characterize spread functions f_θ for the spread θ [as in (7.125)] corresponding to the angle θ by means of differential equations. Such functions are supported by $\mathcal{D}_\theta$. In the first place the derivative d_θ in the direction of $x\cos\theta + y\sin\theta = 0$ annihilates all spread functions, i.e.

$$d_\theta f_\theta \equiv \left(\sin\theta\frac{\partial}{\partial x} - \cos\theta\frac{\partial}{\partial y}\right) f_\theta \equiv 0. \tag{7.128}$$

In addition the δ functions of the leaves of the geometric spread are constant in radial directions.

Thus we can add to (7.128)

$$x\frac{\partial f_\theta}{\partial x} + y\frac{\partial f_\theta}{\partial y} + t\frac{\partial f_\theta}{\partial t} = 0. \tag{7.129}$$

Proposition 7.11 *Equations (7.128) and (7.129) characterize the spread functions corresponding to θ. The δ functions of the leaves of the spreads form a basis for the space of spread functions.*

Proof We have shown that spread functions satisfy (7.128) and (7.129) so let us prove the converse.

Let L^0_θ be any line parallel to $x\cos\theta + y\sin\theta = 0$ not lying in $\{t = 0\}$. If h_θ satisfies (7.128) then h_θ is constant on L^0_θ. Equation (7.129) implies that h_θ is constant on all of $L_{\alpha\theta}$ where α is chosen uniquely by the condition $L^0_\theta \subset L_{\alpha\theta}$. Conversely every $L_{\alpha\theta}$ is fibred by lines L^0_θ. Thus h_θ is a spread function for θ.

It follows from the definition that the δ functions of $\{L_{\alpha\theta}\}_\alpha$ form a basis for the spread functions.

This completes the proof of Proposition 7.11.

We want to find the enveloping equation $\mathbf{E}f = 0$ of the systems E_θ defined by (7.128), (7.129) for varying θ (see Section 6.1). Recall that $\mathbf{E}f = 0$ is a system of equations whose solutions are the integrals of solutions of the equations $E_\theta f_\theta = 0$. (Integral is defined in a generalized sense so we allow the "measure" used in the integral to be a distribution.)

Let us clarify the concept of enveloping equations.

The fundamental principle (Section 1.5) associates to a system of constant coefficient partial differential equations $P(D)h = 0$ an algebraic variety V such that solutions of the equation are represented as Fourier integrals over V. Fourier integrals over the finite union of varieties $V^1 \cup \cdots \cup V^l$ are sums of solutions of $P^1(D)h_1 = 0, \ldots, P^l(D)h_l = 0$. Hence the finite union of the varieties V^j corresponds to the enveloping equation of $P^1, \ldots, P^l$ and we may imagine that the same is true of infinite unions.

When the enveloping equation of the constant coefficient systems $P^j(D)h = 0$ corresponds to the union of the V^j we say that $\{P^j\}$ is an *envelopable system.*

We shall show that the systems (7.128), (7.129) for varying θ form an envelopable system. For each θ the variety V_θ associated to (7.128) is given by

$$V_\theta = \{\hat{x}\sin\theta - \hat{y}\cos\theta = 0\}. \tag{7.130}$$

It is clear that $\bigcup_\theta V_\theta$ is the whole space. Thus one might suspect that every function is an integral over θ of functions satisfying (7.130). Indeed this is the case. For, if L^0_θ is a line parallel to the x, y plane then it is parallel to some line through the origin in the x, y plane, say corresponding to θ, so $\delta_{L^0_\theta}$ satisfies (7.128). Every 2 plane L in $\mathbb{R}^3$ through the origin is clearly a union of such lines

so δ_L is an integral of solutions of (7.128) for a fixed θ. On the other hand, the results of Section 2.1 assert that, for the spaces $\mathcal{W}$ we studied, any $f \in \mathcal{W}$ can be represented in terms of $\mathbf{R}^*\mathbf{R}f$ which is an integral (in a generalized sense, i.e. allowing differential and pseudo-differential operators as the "measure") of $\{\delta_L\}$.

In the present case we have added the homogeneity condition (7.129) to the equations (7.128). We should expect (7.129) to become the enveloping equation. Certainly any linear combination of spread functions satisfies (7.129).

Theorem 7.12 *Equation (7.129) is the enveloping equation for the systems (7.128) and (7.129).*

We give a slightly sharpened version of Theorem 7.12: we consider functions whose supports lie in the forward light cone.

Proof We want to show that any "nice" function h defined in the interior of the forward light cone which is homogeneous is an integral of spread functions. (The interior of the light cone is the intersection of the domains $\mathcal{D}_\theta$ where the spread functions for θ are defined as we have explained preceding Proposition 7.11.) Any such h is determined by its values on any cross-section of the rays through the origin (CS for the enveloping equation).

One choice for CS is the unit disk in the tangent plane T to G/K at I, which is the plane $\{t = 1\}$. Clearly any $L_{\alpha\theta} \cap T$ is a straight line; in fact the spread defined by θ intersects T exactly in a spread of parallel lines. We can apply the standard Radon transform theory as explained in Chapter 2 to the spreads $\{L_{\alpha\theta} \cap T\}_\alpha$. The restriction of h to T is a generalized integral of standard spread functions on T; extending by homogeneity gives the representation of h as an integral of spread functions.

This completes the proof of Theorem 7.12.

Theorem 7.12 shows that if functions on G/K are extended to be homogeneous then they are determined by the restrictions to any CS for (7.129).

To obtain explicit formulas for the geodesic Radon transform on G/K we transform the Poincaré measure on G/K to T via radial projection (i.e. homogeneity of degree 0). We obtain the *Klein model of hyperbolic geometry* for which geodesics (intersections of $L_{\alpha\theta}$ with T) are straight lines. We need only check that the Poincaré metric on D defines a Fubini measure for the spreads of parallel line segments. This is an essentially straightforward computation.

The idea of using the Klein model to study the geodesic Radon transform is due to Todd Quinto and to Carlos Berenstein and Casadio Tarabusi [14].

Another choice for CS is S^2 = *sphere centered at origin.*

Actually this works only for functions f on G/K of compact support. We choose S^2 large enough so that support $f \subset$ interior S^2. It is clear that geodesics on G/K correspond to geodesics on S^2. The latter can be analyzed directly using the representation theory of the *compact* group $SO(3)$.

This can be used to give another treatment of the geodesic Radon transform on G/K.

8

RADON TRANSFORM AS THE INTERRELATION OF GEOMETRY AND ANALYSIS

The previous chapters dealt with the Radon transform of functions. The analog for forms related to certain systems $\vec{P}$ of partial differential equations is the subject matter of Section 8.1. In particular we develop a relation between $\vec{P}$ closed, $\vec{P}$ exact, and an integral geometric condition which generalizes the integral of forms over cycles. We also introduce a different method which is a take-off on the classical notion of splines.

In general the Poisson summation formula (PSF) depends on a relation between two bases for a spread. If the spread functions f satisfy $\vec{P}f = 0$ then the δ functions of points on a parametrization surface provide one basis. Section 8.2 deals with iteration of the PSF which is possible when one parametrization surface is contained in the parametrization surface for another operator. Selberg's trace formula is formulated in these terms.

The Euler–Maclaurin sum formula is a "cut-off" of the usual PSF. More subtle cut-offs are presented in Section 8.3. They can be used to derive the approximate functional equation for the ζ function. Analogs are given for partial differential equations and for discrete groups.

Section 8.4 gives a method for studying infinite dimensional representations of groups in terms of finite dimensional representations; this extends Weyl's work on finite dimensional representations (the "compact trick"). More generally, it is a study of the relation of different eigenfunction expansions for a fixed differential operator.

8.1 Integral geometry and differential equations

One impetus I had for studying the Radon transform was provided by Hubert Goldsmidt who asked me whether a differential 1 form ω on $\mathbb{R}^n$ (small at infinity), whose integral over every straight line is zero, is exact. If we write

$$\omega = \vec{f} = \sum f_j \, dx_j \tag{8.1}$$

then does the geometric condition

$$\int_L \omega = 0 \tag{8.2}$$

for every line L imply (and is equivalent to) the analytic condition of the existence of a function g which is zero at infinity such that $dg = \vec{f}$, i.e.

$$\frac{\partial g}{\partial x_j} = f_j? \tag{8.3}$$

To put this in spread formulation, let L_0 be a fixed line through the origin. If $\tau(L_0)$ denotes the tangent vector to L_0 (which is independent of the point) then (8.2) can be written in the form

$$\int_L \tau(L_0) \cdot \vec{f} = 0 \tag{8.4}$$

for all L in the spread determined by L_0. Here $\vec{f}$ is the vector function $\vec{f} = (f_1, \ldots, f_n)$.

As usual this means

$$\delta_{L_0} * [\tau(L_0)] \cdot \vec{f} = 0 \tag{8.5}$$

or, by Fourier transform,

$$\tau(L_0) \cdot \vec{\hat{f}}\Big|_{\hat{L}_0^\perp} = 0. \tag{8.6}$$

Note that $\tau(L_0)$ is the normal to $\hat{L}_0^\perp$. Thus if $\hat{x} \neq 0$ is given and $\hat{L}_0^\perp$ is any hyperplane containing $\hat{x}$ then $\vec{\hat{f}}(\hat{x}) \cdot \tau(L_0) = 0$ so that $\vec{\hat{f}}(\hat{x}) \in \hat{L}_0^\perp$. Hence $\vec{f}(\hat{x})$ belongs to any hyperplane containing $\hat{x}$ so $\vec{\hat{f}}(\hat{x})$ is in the direction of $\hat{x}$. This implies that

$$\hat{x}_i \hat{f}_j(\hat{x}) - \hat{x}_j \hat{f}_i(\hat{x}) = 0 \tag{8.7}$$

for all i, j; (8.7) means that ω is closed.

We rewrite (8.7) in the form

$$\frac{\hat{f}_i(\hat{x})}{\hat{x}_i} = \frac{\hat{f}_j(\hat{x})}{\hat{x}_j} = \hat{\mathbf{g}}(\hat{x}) \tag{8.8}$$

for any i, j. The possible singularities of $\hat{\mathbf{g}}$ are contained in the intersection of the zeros of the denominators, i.e. $\hat{x} = 0$.

Now, $\vec{\hat{f}}(0) \cdot \tau(L_0)$ is the integral of $\vec{f} \cdot \tau(L_0)$ over $\mathbb{R}^n$ which can be thought of as the union of lines parallel to L_0. Condition (8.2) implies that $\vec{\hat{f}}(0) \cdot \tau(L_0) = 0$ for each L_0, which means that $\vec{\hat{f}}(0) = 0$. Hence $\hat{\mathbf{g}}$ has no singularities.

We conclude that, in fact, $\hat{\mathbf{g}}$ has no singularities so $\vec{f}$ is exact.

The equivalence of (8.2) and (8.3) represents a relation between geometry and analysis. How did the differential operators $\partial/\partial x_j$ get "absorbed" in the line integral?

Before examining this question in detail let us give some other examples of this "geometrization" of analysis.

We have mentioned Delsarte's theorem (see [30]) that f is harmonic if it satisfies the mean value theorem for all circles of radii r_1, r_2 where r_1/r_2 is not the zero of a certain Bessel function. The mean value property is

$$(\delta_{S^0(r_j)} - \delta) * f = 0 \quad j = 1, 2.$$

Here $S^x(r)$ is the sphere of radius r centered at x.

Delsarte's theorem can be reformulated as follows: the linear combinations of the sums

$$T_1 * (\delta_{S^0(r_1)} - \delta) + T_2 * (\delta_{S^0(r_2)} - \delta)$$

for $T_1, T_2 \in \mathcal{W}'$ are dense in $\mathcal{W}'$.

It is interesting to observe that the density of such linear combinations cannot be replaced by surjectivity of the sums themselves unless r_1/r_2 has some Diophantine properties. (This is best seen when $n = 1$ and "harmonic" is replaced by "linear.")

Mean value theorems arise from *splines*. Classical splines are piecewise linear continuous functions. In practice they are used to approximate solutions of various equations.

For an example of our ideas let $\boxed{\sigma}$ be an $n \times n$ matrix of classical splines of compact support. We call

$$d^2 = \{\partial^2/\partial x_i \partial x_j\} = \text{the symmetric second derivative.}$$

Let f be a linear function. Then

$$\begin{aligned} 0 &= \boxed{\sigma} \cdot d^2 f \\ &= (d^2)' \boxed{\sigma} \\ &= \sum \mu_{ijk} \cdot f. \end{aligned}$$

The μ_{ijk} are measures on the boundaries of the regions on which the σ_{ij} are linear.

The μ_{ijk} are measures because d^2 is second order, so the derivative of a piecewise linear continuous function is a measure. If we used d in place of d^2 then $\boxed{\sigma}$ would be replaced by $\vec{\sigma} = \{\sigma_j\}$ where the σ_j are characteristic functions of regions. This can be regarded as the source of the fact that the closure of an ordinary $n-1$ form f can be described by the vanishing of the integral of f over the boundaries of "sufficiently many" regions.

In general we define a $\vec{P}$ *spline* for the system $\vec{P} = \vec{P}(x, D)$ as a vector $\vec{\sigma}$ of functions such that $\vec{P}' \cdot \vec{\sigma}$ vanishes except on lower dimensional sets where it is

a measure. When P is a second-order differential operator then one choice for σ is the Green's function for P for the point x and the region Ω, meaning σ is a fundamental solution for P at x which vanishes on the boundary of Ω (hence is continuous except possibly at x) so that

$$P'\sigma = \delta_x + \mu$$

where μ is a measure on $\operatorname{bd}\Omega$. When applied to $\log|f|$, where f is a meromorphic function, this leads to deep results in complex function theory starting with the Poisson–Jensen formula.

The study of $\vec{P}$ splines clearly falls within the framework of the Radon ansatz. Although they provide nice geometric necessary conditions for a function f to be in the kernel of $\vec{P}$, it is often difficult to use them directly to obtain sufficient conditions.

Problem 8.1 *Give a general method to use splines to determine the kernel of* $\vec{P}$.

An "infinitesimal" solution to this problem when P is the Laplacian on a Riemannian manifold is given by writing

$$Pf(x) = \lim \frac{\int_{S^x(2\epsilon)} f - 2\int_{S^x(\epsilon)} f + f}{\epsilon^2}.$$

Analogous formulas exist for linear combinations of powers of the Laplacian.

Is there a global result *à la* Delsarte?

A question which is in the same spirit as Problem 8.1 is

Problem 8.2 *For which* $\vec{P}$ *does there exist a* $\vec{P}$ *Green's function?*

Definition *A* $\vec{P}$ *Green's function for a point* x *is a* $\vec{P}$ *spline* $\vec{G}$ *for which*

$$\vec{P}\cdot\vec{G} = \delta_x + \sum \mu_j$$

where the μ_j *are measures supported on lower dimensional sets and* $x \notin \operatorname{support}\mu_j$ *for any* j.

We shall not deal further with splines. Rather we use an approach which is in keeping with the spirit of this work.

The exactness (8.3) of ω implies that the f_j satisfy the same operator relations as the $\partial/\partial x_j$; that is, ω is closed. By Fourier transform

$$\hat{x}_j\hat{f}_i(\hat{x}) = \hat{x}_i\hat{f}_j(\hat{x}). \tag{8.9}$$

Put in other terms, the polynomial vectors ${}_{ij}\vec{Q}$ defined by

$$\left({}_{ij}\vec{Q}\right)_k = \begin{cases} \hat{x}_j & \text{if } k = i \\ -\hat{x}_i & \text{if } k = j \\ 0 & \text{otherwise} \end{cases} \tag{8.10}$$

generate the (polynomial) module of relations $\{\vec{Q}\}$ of $\hat{d} = (\hat{x}_1, \ldots, \hat{x}_n)$. (The polynomial module of relations of $\hat{d}$ is the set of vectors $\vec{Q}$, whose components are polynomials, satisfying $\vec{Q} \cdot \hat{d} = 0$.) It is crucial that we can find a basis for the module of relations which is homogeneous. This results from the homogeneity of $\hat{d}$.

Our point of departure is the replacement of the elements $\vec{Q}$ by the constant vectors $\vec{e} = \vec{e}(\hat{x}^0) = \vec{Q}(\hat{x}^0)$ for suitable $\hat{x}^0$.

We shall distinguish two approaches to this interplay of geometry and analysis:

(1) Consider the restriction of $\hat{x}$ to a line through the origin, say $\hat{L} = \hat{L}(\hat{x}^0)$, which is the line through $\hat{x}^0 \neq 0$. Since the components of $\vec{\hat{d}}$ and $\vec{Q}$ are homogeneous of degree 1, for $\hat{x} \in \hat{L}(\hat{x}^0)$ the orthogonality of $\vec{\hat{f}}(\hat{x})$ to $\vec{Q}(\hat{x})$ is equivalent to the orthogonality to $\vec{e}(\hat{x}^0) = \vec{Q}(\hat{x}^0)$. (This is not true at $\hat{x} = 0$ but if $\vec{f}$ is small at infinity, which we assume, $\vec{\hat{f}}$ is smooth so $\hat{x} = 0$ gives no trouble.) It is easily seen, e.g. by examining the case $\hat{L} = \hat{x}_1$ axis, that the $\vec{Q}(\hat{x}^0)$ span $\hat{L}^\perp(\hat{x}^0)$.

By the projection–slice theorem the Fourier transform of $\vec{e}(\hat{x}^0) \cdot \vec{\hat{f}}\big|_{\hat{L}(\hat{x}^0)}$ is the Radon transform in the spread defined by the hyperplane $H = L^\perp(\hat{x}^0)$. We have shown

$$\mathbf{R}_H(\vec{f} \cdot \vec{e}) = 0 \tag{8.11}$$

for any hyperplane H and any $\vec{e} = \vec{e}(H)$ which is the value of an element of the module of relations of $\hat{d}$ evaluated at some point $\hat{x}^0 \in \hat{H}^\perp$.

Equation (8.11) is, at least superficially, somewhat weaker than the line integral condition (8.4) because H can be represented as the union (in many ways) of translates of lines L whose tangent is an $\vec{e}$.

(2) For this approach we need the *ellipticity* of $\hat{d}$, meaning that if $\vec{Q} \cdot \vec{\hat{f}} = 0$ for all Q (which we now assume) then, for any $\hat{x}^1 \neq 0$, $\vec{\hat{f}}(\hat{x}^1) = \hat{\mathbf{g}}(\hat{x}^1)\hat{d}(\hat{x}^1)$. Put in other terms, the complex defined by $\hat{d}$ is exact at the first stage at each point $\hat{x}^1 \neq 0$. We do not assume actual exactness which would entail some "regularity" of $\hat{\mathbf{g}}$.

We fix a vector $\vec{e} \neq 0$ which is of the form $\vec{Q}(\hat{x}^0)$ for some $\vec{Q}$ in the module of relations of $\hat{d}$. We define $\hat{\Lambda}(\vec{e})$ by

$$\hat{\Lambda}(\vec{e}) = \{\hat{x} \big| \vec{e} \cdot \hat{d}(\hat{x}) = 0\}. \tag{8.12}$$

By ellipticity $\vec{\hat{f}}$ is orthogonal to $\vec{e}$ on $\hat{\Lambda}(\vec{e})$.

Note that $\hat{\Lambda}(\vec{e}) \supset \hat{L}(\hat{x}^0)$.

Conversely suppose $\vec{e}\cdot\vec{\hat{f}}(\hat{x})$ vanishes on $\hat{\Lambda}(\vec{e})$. Since $\vec{Q}(\hat{x}^0)\cdot\hat{d}(\hat{x}^0)=0$ the homogeneity of $\vec{Q}$ and $\hat{d}$ implies that $\vec{Q}\cdot\vec{\hat{f}}=0$ on $\hat{L}(\hat{x}^0)$. We conclude that the condition $\vec{e}\cdot\vec{\hat{f}}=0$ on $\hat{\Lambda}(\vec{e})$ for all $\vec{e}$ of the form $\vec{Q}(\hat{x}^0)$ is equivalent to $\vec{Q}\cdot\vec{\hat{f}}\equiv 0$, i.e. $\vec{\hat{f}}$ is closed.

By the projection–slice theorem the vanishing of $\vec{e}\cdot\vec{f}$ on $\hat{\Lambda}(\vec{e})$ is equivalent to the Radon transform condition

$$\mathbf{R}^1_{\hat{\Lambda}(\vec{e})^\perp}(\vec{e}\cdot\hat{f})=0. \tag{8.13}$$

This is the line integral condition for closure.

Before passing to the general case it is instructive to analyze the symmetric second derivative d^2 defined by

$$\vec{f}_{kl}=(d^2f)_{kl}=\frac{\partial^2 f}{\partial x_k\partial x_l}. \tag{8.14}$$

Clearly $\vec{f}$ is symmetric. The analog of closed is:

$$\text{for any fixed } k_0 \text{ the 1 form } \vec{f}_{k_0}=(f_{k_0 l}) \text{ is closed.} \tag{8.15a}$$

$$\text{For any } iji'j' \quad \frac{\partial^2 f_{i'j'}}{\partial x_i\partial x_j}=\frac{\partial^2 f_{ij}}{\partial x_{i'}\partial x_{j'}}. \tag{8.15b}$$

In case one of k,l equals i or j (8.15b) is a consequence of (8.15a). As in the case of d the module of relations of $\hat{d}^2$ is generated by the usual module for the $\vec{f}_{k_0}$ and by

$$\left({}_{iji'j'}\vec{Q}\right)_{kl}=\begin{cases}\hat{x}_i\hat{x}_j & \text{if } kl=i'j'\\ -\hat{x}_{i'}\hat{x}_{j'} & \text{if } kl=-ij\\ 0 & \text{otherwise.}\end{cases} \tag{8.16}$$

These generators are homogeneous and we have ellipticity.

We now follow our two methods for introducing geometry:

(1) Fix $\hat{x}^0\neq 0$.

On the line $\hat{L}(\hat{x}^0)$ orthogonality of $\vec{\hat{f}}$ to $\vec{Q}$ is equivalent (for smooth $\vec{\hat{f}}$) to orthogonality to $\vec{e}=\vec{Q}(\hat{x}^0)$. We arrive at

$$\mathbf{R}_{L^\perp(\hat{x}^0)}(\vec{e}\cdot\vec{f})=0 \tag{8.17}$$

for any $\vec{e}$ of the form $\vec{Q}(\hat{x}^0)$ as the necessary and sufficient condition for $\vec{f}$ to be closed (in the d^2 sense).

(2) Fix $\vec{e}$ of the form $\vec{Q}(\hat{x}^0)$ for some $\vec{Q}$ and some $\hat{x}^0\neq 0$.

We call

$$\hat{\Lambda}(\vec{e})=\{\hat{x}\big|\vec{e}\cdot\hat{d}^2(\hat{x})=0\}. \tag{8.18}$$

$\hat{\Lambda}(\vec{e})$ is now a quadric. By ellipticity the vanishing of $\vec{Q}\cdot\vec{\hat{f}}$ implies the vanishing of $\vec{e}\cdot\vec{\hat{f}}$ on $\hat{\Lambda}(\vec{e})$. Conversely, $\vec{e}\cdot\vec{\hat{f}}=0$ on Λ implies, by homogeneity, that $\vec{Q}\cdot\vec{\hat{f}}$ vanishes on $\hat{L}(\hat{x}^0)$. Thus the vanishing of $\vec{e}\cdot\vec{f}$ on $\hat{\Lambda}(\vec{e})$ for all $\vec{e}=\vec{Q}(\hat{x}^0)$ is equivalent to the d^2 closure of $\vec{f}$.

We want to formulate this vanishing in terms of the Radon transform. We call $\hat{\chi}_{\hat{\Lambda}(\vec{e})}$ the characteristic function of $\hat{\Lambda}(\vec{e})$ (δ function). Our condition can be stated as

$$\hat{\chi}_{\hat{\Lambda}(\vec{e})}(\vec{e}\cdot\vec{\hat{f}})\equiv 0. \tag{8.19}$$

By Fourier transform

$$\chi_{\hat{\Lambda}(\vec{e})} * (\vec{e}\cdot\vec{f})=0 \tag{8.20}$$

for all $\vec{e}$. $\chi_{\hat{\Lambda}(\vec{e})}$ is the Fourier transform of $\hat{\chi}_{\hat{\Lambda}(\vec{e})}$; it is not the characteristic function of $\hat{\Lambda}(\vec{e})$.

Equation (8.19) is the assertion that $\vec{e}\cdot\vec{\hat{f}}$ is orthogonal to all measures supported on $\hat{\Lambda}(\vec{e})$. It can be shown that the fundamental principle is valid for the real cones $\hat{\Lambda}(\vec{e})$ with the Schwartz space $\mathcal{S}$ playing a role analogous to an AU space. In this sense (8.20) asserts that $\vec{e}\cdot\vec{f}$ is orthogonal to the solutions of the system of equations $\vec{\mathbf{P}}(\vec{e})$ defined by the variety $\hat{\Lambda}(\vec{e})$.

In Chapter 1 we defined the Radon transform $\mathbf{R}_{\vec{\mathbf{P}}(\vec{e})}(\vec{e}\cdot\vec{f})$ as the restriction of $\vec{e}\cdot\vec{f}$ (as a linear functional) to the kernel of $\vec{\mathbf{P}}(\vec{e})$. Thus our result is the equivalence of

$$\mathbf{R}_{\vec{\mathbf{P}}(\vec{e})}\vec{e}\cdot\vec{f}=0 \tag{8.21}$$

for all $\vec{e}$ together with the closure of the $\vec{f}_{k_0}$ to the d^2 closure[1] of $\vec{f}$.

We can now pass to the general theory. We start with a system $\vec{P}(D)$ of homogeneous operators of fixed degree. The module of relations of $\vec{\hat{P}}$ is generated by homogeneous $\vec{Q}$. We want to determine integral conditions characterizing the image $\{\vec{f}\}$ of $\vec{\mathbf{P}}$.

(1) Fix $\hat{x}^0\neq 0$. For any $\vec{Q}$ the orthogonality of $\vec{e}=\vec{Q}(\hat{x}^0)$ to $\vec{\hat{f}}$ on $\hat{L}(\hat{x}^0)$ is equivalent to $\vec{Q}\cdot\vec{f}=0$ on $\hat{L}(\hat{x}^0)$. This leads, as in the examples, to

Theorem 8.1 *$\vec{f}$ is in the image of $\vec{P}$ if and only if*

$$\mathbf{R}_{L^\perp(\hat{x}^0)}(\vec{e}\cdot\vec{f})=0 \tag{8.22}$$

for all $\hat{x}^0$ and all $\vec{e}$ of the form $\vec{Q}(\hat{x}^0)$.

[1]The image of d^2 was discussed for the first time by Kähler [101] using a different method.

Remark. *We can regard (8.22) as a multiplicity hyperplane Radon transform, the multiplicities being* $\{\vec{e}\}$.

(2) We now assume ellipticity. Fix an $\vec{e}$ of the form $\vec{Q}(\hat{x}^0)$ and define

$$\hat{\Lambda}(\vec{e}) = \{\hat{x} \big| \vec{e} \cdot \vec{P}(\hat{x}) = 0\}. \tag{8.23}$$

Ellipticity implies that for any $\vec{f}$ in the image of $\vec{P}$ we have $\vec{e} \cdot \vec{\hat{f}} = 0$ on $\hat{\Lambda}(\vec{e})$. Moreover if $\vec{e} \cdot \vec{\hat{f}}$ vanishes on $\hat{\Lambda}(\vec{e})$ then $\vec{Q} \cdot \vec{\hat{f}}$ vanishes on $L(\hat{x}^0)$ because $\vec{Q}$ is homogeneous.

We conclude that the vanishing of $\vec{e} \cdot \vec{\hat{f}}$ on $\hat{\Lambda}(\vec{e})$ for all such $\vec{e}$ is equivalent to $\vec{f}$ being $\vec{P}$ closed (i.e. in the kernel of all $\vec{Q}$). That is,

Theorem 8.2 *The* $\vec{P}$ *exactness of* $\vec{f}$ *is equivalent to*

$$\mathbf{R}_{\mathbf{P}(\vec{e})} \vec{f} = 0 \tag{8.24}$$

for all $\vec{e}$ *of the form* $\vec{Q}(\hat{x}^0)$.

Lie groups. The most natural generalization of a straight line is a curve in a Lie group G or a homogeneous manifold $M = G/K$ of the form gL (or gLm_0) where L is a one-parameter subgroup of G. The analog of the euclidean equation $\vec{f} = (\partial h/\partial x_j)$ is

$$\vec{f} = \big(\partial(\vec{\alpha})h\big) \tag{8.25}$$

where $\partial(\vec{\alpha}) = (\partial(\alpha_j))$ is a basis for the Lie algebra $\mathcal{G}$ of G thought of as left invariant differential operators.

Problem 8.3 *Is the condition (say for* f, h *of compact support)*

$$\int_{gL} \vec{f} = 0 \tag{8.26}$$

for all g, L *sufficient for the solvability of (8.25)?*

The integral in (8.26) is defined by thinking of $\vec{f}$ as a linear function (one cochain) on $\mathcal{G}$. In the language of cohomology, Problem 8.3 asks whether (8.26) is the condition for a one cochain on $\mathcal{G}$ with coefficients in $\mathcal{D}(G)$ to be a one coboundary.

In order for Problem 8.3 to be meaningful we must first make some normalizations. Even for $\mathbb{R}^n$ there is a relation between the basis $\{\partial/\partial x_j\}$ for the Lie algebra and the measure used on geodesics. Thus if we replaced $\{\partial/\partial x_j\}$ by $\{c_j \partial/\partial x_j\}$ with $c_j \neq 0$ depending on j, the relation between line integrals and exact (or closed) 1 forms would not be valid if we used the usual euclidean measure to define line integrals.

Suppose that G is semi-simple so there is a nondegenerate invariant bilinear form $B(X,Y)$ on the Lie algebra $\mathcal{G}$. Let $\{\alpha_j\}$ be an orthonormal basis for $\mathcal{G}$ and let $\partial(\alpha_j)$ be the left invariant differential operator on G corresponding to α_j. Since everything is G invariant the α_j define an orthonormal basis for the tangent space at any point of G.

By definition

$$\partial(\alpha_j)f(g) = \frac{d}{dt}f(g(\exp(t\alpha_j)))\big|_{t=0}. \tag{8.27}$$

We choose $\{\omega_j\}$ as a dual basis to $\{\alpha_j\}$ so that $\{\omega_j\}$ define the Haar measure of G and they define the measure dt on the $\{g\exp(t\alpha_j)\}_{-\infty}^{\infty}$. With this normalization, if $\vec{f}$ is of the form (8.25) then (8.26) is valid.

We do not know the answer to Problem 8.3 but we conjecture it is true at least for real semi-simple Lie groups G.

One can proceed in a somewhat different direction. In Chapter 7 we discussed the horocyclic Radon transform $\mathbf{R}^N$ on G/K where $G = SL(2,\mathbb{R})$. In particular we showed that $\mathbf{R}^N$ is injective on $\mathcal{D}(G/K)$ and we computed its image.

Now, let $\vec{f}$ be a three-component vector function on G/K; We want to know whether $\vec{f} = (\partial h/\partial x_j)$ on G/K; that is, $\vec{f}$ is the restriction to G/K of $(\partial h/\partial x_j)$ where h is a function on $\mathbb{R}^3$. The first approach would seem to be to integrate $\vec{f}$ over suitable curves in G/K. But we shall follow a somewhat different idea which is based on the spirit of Theorems 8.1 and 8.2 but which yields a different type of result.

We integrate $\vec{f}\cdot\vec{e}$ over the horocycles which are the intersections $Q_\theta^s \cap G/K$ where Q_θ^s is parallel to the tangent plane to the light cone containing the line $L(\theta)$ and $\vec{e}$ represents the Minkowski direction cosines of a line lying in such a tangent plane.

Theorem 8.3

$$\int_{Q_\theta^s \cap G/K} \vec{f}\cdot\vec{e} = 0 \tag{8.28}$$

for all such $Q_\theta^s, \vec{e}$ is necessary for $\vec{f}$ to be the restriction to G/K of $(\partial h/\partial x_j)$.

Proof As in Chapter 7, the vanishing of the integrals (8.28) means that for any generator $\hat{\theta}$ of the light cone $\hat{\Gamma}$ we have

$$\vec{\hat{f}}\cdot\vec{e} = 0 \quad \text{on } L(\hat{\theta}) \quad \text{for all } \vec{e}\in L(\hat{\theta})^{\perp}. \tag{8.29}$$

(Orthogonality refers to the Minkowski metric.) Equation (8.29) means that $\vec{\hat{f}}|_{L(\hat{\theta})}$ is parallel to $\hat{\theta}$. We conclude, as in our discussion following (8.6), that $\vec{\hat{f}}$ is of the form $\hat{x}\hat{h}$ on $\hat{\Gamma}$. By Fourier transform $\vec{f} = dh$ on the kernel of the wave operator $\Box$. We have assumed that support $\vec{f}\subset G/K$. By hyperbolicity or the

harmonic function theory of Chapter 3 the restrictions of solutions of $\Box g = 0$ to G/K form all functions on G/K. Thus $\vec{f} = dh$ on G/K.

Remark. *We can actually take h as a solution of the wave equation.*

Problem 8.4 *What is the analog of Theorem 8.3 for the geodesic Radon transform?*

We wish to conclude by giving a general framework for the interrelation between analysis and integral geometry.

We start with a manifold M and a system of differential equations

$$\boxed{P}(x, D)\vec{g} = \vec{f}. \tag{8.30}$$

Here $\vec{f}$ and $\vec{g}$ are cross-sections of vector bundles and $\boxed{P}$ is a matrix of differential operators with appropriate transformation properties. This is analysis.

For the geometry, we are given a collection of submanifolds V_α of M. (The same submanifold may correspond to many α.) For each α we are given a cross-section $e^{\vec{\alpha}}$ of an appropriate jet bundle on V_α. $e^{\vec{\alpha}}$ has the same number of components as $\vec{f}$.

For integral geometry we are provided with a measure μ on M. We call μ_α the induced measure on V_α. μ_α can be defined by the Fubini method or some other convenient method.

Problem 8.5 *For which $\boxed{P}$ can we determine $\{V_\alpha, e^{\vec{\alpha}}\}$ so that*

$$\int_{V_\alpha} \vec{f} \cdot e^{\vec{\alpha}} = 0 \tag{8.31}$$

for all α is necessary and sufficient for the existence of $\vec{g}$ satisfying (8.30)?

In the spirit of (8.20) one may wish to replace $V_\alpha, e^{\vec{\alpha}}$ by kernels $\vec{K_\alpha}$ and replace integration over V_α by integration of $\vec{K_\alpha} \cdot \vec{f}$ over M. (Equation (8.31) corresponds to $\vec{K_\alpha} = e^{\vec{\alpha}} \delta_{V_\alpha}$.)

8.2 The Poisson summation formula and exotic intertwining

We have introduced the Poisson summation formula (PSF) at several places in this work. It was seen (see Section 1.7) to depend on hyperbolic equations and two bases for functions on a single Cauchy Surface S, one geometric (usually $\{\delta_s\}$) and one analytic.

We wish to introduce an allied construction which involves two Cauchy or parametrization surfaces. This falls into the realm of intertwining which is a comparison of solutions on two different Cauchy (or parametrization) surfaces. One of the most common forms of intertwining is integration by parts. This is

usually most successful when the Cauchy surfaces are hypersurfaces. We shall, however, study cases in which at least one of the parametrization problems is a Watergate problem.

Let us begin with $\mathbb{R}^{n+1}$ with coordinates (t, x). We study the CP for the wave operator $\Box$ with CD on $t = 0$ and also the WP with WD on the t axis. We could intertwine the CD on $t = 0$ with the WD on $x = 0$ by applying the fundamental solution to the CD. However, this works only because the CD is given on $t = 0$ which is a hypersurface; we cannot reverse the process. We shall, therefore, use a method based on the Fourier transform, which works in more general situations. In particular it is easy to reverse our procedure and go from WD to CD. Our computations are mostly formal. It is easy to justify them; we leave such details to the reader.

We start with CD of the form $(0, h)$. If h has a (formal) Fourier representation

$$h(x) = \int \hat{h}(\hat{x}) e^{ix\cdot\hat{x}}\, d\hat{x}$$

then the solution H of the CP is given by

$$H(t,x) = \int \hat{h}(\hat{x}) e^{ix\cdot\hat{x}} \frac{\sin t\sqrt{\hat{x}^2}}{\sqrt{\hat{x}^2}}\, d\hat{x} \tag{8.32}$$

since the right side of (8.32) is a solution of the wave equation whose CD is clearly $(0, h)$.

Put in other terms, H has the Fourier representation

$$H(t,x) = \int_V \frac{\hat{h}(\hat{x})}{2i\hat{t}} e^{ix\cdot\hat{x}+it\hat{t}}\, d\hat{x} \tag{8.33}$$

where V is the variety $\hat{t}^{\,2} = \hat{x}^2$.

We want to intertwine this CD with the WD of H. We first express $H(t, x)$ in terms of h; for this we introduce the Fourier representation for $\hat{h}$ in (8.32). Thus

$$H(t,x) = \int e^{ix\cdot\hat{x}} \frac{\sin t\sqrt{\hat{x}^2}}{\sqrt{\hat{x}^2}}\, d\hat{x} \int h(u) e^{-iu\cdot\hat{x}}\, du.$$

To evaluate this integral suppose for simplicity that $n = 2$. We write h in a Fourier series using polar coordinates $\tilde{r}, \tilde{\theta}$

$$h(u) = \sum h_j(\tilde{r}\,) e^{ij\tilde{\theta}}$$

so that

$$\begin{aligned} H(t,x) &= \int e^{ix\cdot\hat{x}} \frac{\sin t\sqrt{\hat{x}^2}}{\sqrt{\hat{x}^2}}\, d\hat{x} \int \sum h_j(\tilde{r}\,) e^{ij\tilde{\theta}} e^{i\tilde{r}\hat{r}\cos(\tilde{\theta}-\hat{\theta})} \tilde{r}\, d\tilde{r}\, d\tilde{\theta} \\ &= \int e^{itx\cdot\hat{x}} \frac{\sin t\hat{r}}{\hat{r}} \hat{r}\, d\hat{r}\, d\hat{\theta} \sum e^{ij\hat{\theta}} \int h_j(\tilde{r}\,) e^{ij\tilde{\theta}} e^{i\tilde{r}\hat{r}\cos\tilde{\theta}} \tilde{r}\, d\tilde{r}\, d\tilde{\theta}. \end{aligned}$$

Thus

$$H(t,0) = \int \sin t\hat{r}\, d\hat{r} \int h_0(\tilde{r}\,) e^{i\tilde{r}\hat{r}\cos\tilde{\theta}} \tilde{r}\, d\tilde{r}\, d\tilde{\theta}.$$

(Only the term $j = 0$ in $\exp(ij\hat{\theta})$ occurs.)

If we formally interchange the order of operations, we obtain

$$\begin{aligned} H(t,0) &= \int h_0(\tilde{r}\,)\tilde{r}\, d\tilde{r} \int \sin t\hat{r}\, d\hat{r} \int e^{i\tilde{r}\hat{r}\cos\tilde{\theta}}\, d\tilde{\theta} \\ &= \int h_0(\tilde{r}\,)\tilde{r}\, d\tilde{r} \int \sin t\hat{r} J_0(\tilde{r}\hat{r})\, d\hat{r} \end{aligned} \tag{8.34}$$

where J_0 is the standard Bessel function (see [11, vol. II])

$$J_0(r) = \int e^{ir\cos\theta}\, d\theta.$$

We now use the Mehler–Sonine formula for the Fourier transform of J_j [11, vol. II, p. 81 (12)] which reads

$$J_0(\hat{t}\,) = \int_1^\infty \left(t^2 - 1\right)^{-1/2} \sin t\hat{t}\, dt \tag{8.35}$$

for $\hat{t} > 0$. (Our computations are correct up to normalizing constants.) $(t^2 - 1)^{-1/2}$ and $J_0(\hat{t})$ are even functions. If we define $(t^2 - 1)^{-1/2} = 0$ when $|t| \le 1$ and multiply by sgn t then we can write the right side of (8.35) as a Fourier transform over $(-\infty, \infty)$. The Fourier inversion formula reads

$$\int_{-\infty}^\infty J_0(\hat{t}) \operatorname{sgn}\hat{t} e^{it\hat{t}}\, d\hat{t} = \begin{cases} \operatorname{sgn} t \left(t^2 - 1\right)^{-1/2} & |t| \ge 1 \\ 0 & |t| < 1. \end{cases}$$

Hence

$$\int_0^\infty \sin t\hat{r} J_0(r\hat{r})\, d\hat{r} = \begin{cases} c \operatorname{sgn} t \left(t^2 - r^2\right)^{-1/2} & |t| \ge r \\ 0 & |t| < r \end{cases} \tag{8.36}$$

which is the usual formula for the fundamental solution.

We now use this intertwining as a secondary PSF. The first step is the usual PSF on $\mathbb{R}^2$. We set

$$\hat{h}(\hat{x}) = \sum \delta_m = \sum e^{il\cdot\hat{x}}.$$

From (8.32) we find from the expression $\hat{h}(\hat{x}) = \sum \delta_m$

$$H(t,x) = \sum e^{im\cdot x} \frac{\sin|m|t}{|m|}.$$

On the other hand, if we use (8.34) and (8.36) for $h(x) = \sum \delta_m$ we find that (up to a normalizing constant)

$$H(t,0) = \sum_{|m|<t} \operatorname{sgn} t(t^2 - m^2)^{-1/2} \tag{8.37}$$

since $h_0(r)$ is the number of lattice points $\{m\}$ on the circle of radius r times δ_r. The measure $r\,dr$ is the natural one relating radial functions to distributions, meaning that the formalism gives $(\delta_m r\,dr) \cdot f = f(m)$ for any radial function.

If we equate the two expressions for $H(t,0)$ we find a new version of the PSF:

$$\sum \frac{\sin |m|t}{|m|} = \sum_{|m|<t} \operatorname{sgn} t \left(t^2 - m^2\right)^{-1/2}. \tag{8.38}$$

Remark 1 *The reader might be somewhat confused by the Mehler–Sonine formula (8.35) for J_0 since the usual expression as a Fourier transform is*

$$\begin{aligned} J_0(\hat{t}) &= \int e^{i\hat{t}\cos\theta}\, d\theta \\ &= \int_{-1}^{1} e^{i\hat{t}t} \left(1 - t^2\right)^{-1/2} dt. \end{aligned} \tag{8.39}$$

Equations (8.35) and (8.39) seem to be two different expressions for J_0 as a Fourier transform! The point is that (8.39) is valid for $J_0(\hat{t})$ for all $\hat{t}$ while (8.35) represents $J_0(\hat{t})$ for $\hat{t} > 0$ and $-J_0(\hat{t})$ for $\hat{t} < 0$. The relation between the integrals from $[-1,1]$ and $[1,\infty]$ can be explained by contour integration.

Remark 2 *The Mehler–Sonine formula shows that the Fourier transform of $J_0(\hat{t})$ sgn $\hat{t}$ vanishes on $|t| < 1$. It would be interesting to verify this directly using the criteria developed in Chapter 5.*

Remark 3 *The PSF (8.38) comes from the zeroth WD; that is, the datum corresponding to the identity operator. The same method allows us to produce a PSF for each operator $(\partial/\partial x \pm i\partial/\partial y)^j$.*

The example just given can be regarded as a combination of the PSF for the plane with the intertwining of the CP and WP for the wave equation. From the viewpoint of the operator $\mathbf{R}^*$ it can also be regarded as an iteration of the general (abstract) PSF as formulated in Section 1.7. We now present another example of this iteration which introduces a new ingredient. To that end we examine the simplest case which is the iteration of the PSF for the line with intertwining given by the hyperbolic operator $\partial/\partial t - \partial/\partial x \equiv \partial$.

If we start with

$$h(x) = \sum_l e^{ilx}$$

then it is clear that the analytic representation for the solution of $\partial H = 0$, $\mathrm{CD}(H) = h$ is

$$H(t,x) = \sum e^{il(t+x)}. \tag{8.40}$$

If we use the PSF to write $h(x) = \sum \delta_m$ then the natural (geometric) representation is

$$H(t,x) = \sum_m \delta_{x-t=m}. \tag{8.41}$$

The equality of (8.40) and (8.41) on the t axis is just the PSF on the t axis.

In this case the intertwining via the operator $\partial/\partial t - \partial/\partial x$ produces nothing new. This operator is too simple.

There is another procedure of passing from one PSF to a second PSF. This is based on the trace instead of on intertwining. In the present situation it uses the operator $\tau = \tau(t^0)$ which sends $h \to H(t^0, x)$ for $h(x)$ periodic of period 1.

The trace of an operator $\mathcal{O}$ is invariant under the choice of basis. There are two types of bases which are usually used to compute the trace formally:

(1) *Eigenfunctions of* $\mathcal{O}$. If $\{\phi(\lambda)\}$ are a basis of eigenfunctions (eigenvectors) of $\mathcal{O}$ with eigenvalues $\{\lambda\}$ then

$$\mathrm{Tr}\,\mathcal{O} = \sum \lambda. \tag{8.42}$$

(2) δ *function basis.* If $\mathcal{O}$ acts on a suitable space of functions on a space S then, in the sense of Section 1.2, $\{\delta_x\}_{x\in S}$ is a basis for the function space. $\mathcal{O}$ maps δ_x into a function k_x whose expression in the δ basis is

$$\mathcal{O}(\delta_x) = k_x = \int k_x(y)\delta_y\, dy. \tag{8.43}$$

Hence for any function $g = \int g(x)\delta_x\, dx$ we have

$$\mathcal{O}g = \int g(x)k_x\, dx$$

so

$$(\mathcal{O}g)(y) = \int g(x)k_x(y)\, dx. \tag{8.44}$$

This means that $k_x(y)$ is the integral kernel representing the operator $\mathcal{O}$. From this we have

$$\mathrm{Tr}\,\mathcal{O} = \int k_x(x)\, dx. \tag{8.45}$$

By (8.43) we can regard $k_x(x)$ as the *fixed point index* of $\mathcal{O}$ since it represents that "proportion" of $\mathcal{O}(\delta_x)$ in the δ function basis which is δ_x.

The identity of (8.42) and (8.45) is often of considerable interest.

In the case of the map $\tau = \tau(t^0)$ we use the basis $\{\exp(ilx)\}$ of eigenfunctions and the δ function basis $\{\delta_x\}_{x\in[0,1]}$. The exponential basis clearly leads to

$$\operatorname{Tr}\tau(t^0) = \sum e^{ilt^0} \text{ (eigenbasis)} \tag{8.46}$$

which we regard as a distribution in t^0.

As for the δ basis we use the fact that if $\mathrm{CD}(H) = \delta_x$ then $H(t^0, x) = \delta_{t^0+x}$. This contributes to $\operatorname{Tr}\tau(t^0)$ when $t^0 + x \equiv x \mod \mathbb{Z}$ since we are dealing with periodic CD. Thus $\operatorname{Tr}\tau(t^0) = 0$ if $t^0 \notin \mathbb{Z}$. When $t^0 \in \mathbb{Z}$ each δ_x is an eigenvector of $\tau(t^0)$ with eigenvalue 1. A distribution calculation leads easily to

$$\operatorname{Tr}\tau(t^0) = \sum \delta_l. \tag{8.47}$$

The identity of (8.46) and (8.47) is a formal restatement of the PSF. Thus, as above, we gain nothing.

Let us contrast the intertwining and trace methods. We focus on the eigenfunction basis (analytic aspect).

The eigenfunction $\exp(ilx)$ as CD yields the solution $H(t, x) = \exp[il(x + t)]$. We have

$$H(t^0, 0) = e^{ilt^0}$$

which is exactly the eigenvalue of $\exp(ilx)$ for the operator $\tau(t^0)$.

Suppose we had a CP for some hyperbolic or parabolic differential equation of the form

$$\frac{\partial^2}{\partial t^2} + P\left(x, \frac{\partial}{\partial x}\right).$$

We call $\mathcal{O}(t)$ the operator representing the motion of CD in t (like $\tau(t)$). Let $\{\phi_l(x)\}$ be a basis for the space of CD we consider consisting of normalized eigenfunctions of $P(x, \partial/\partial x)$ with eigenvalues $\lambda(l)$. Thus

$$\mathcal{O}(t)\phi_l(x) = e^{it\sqrt{\lambda(l)}}\phi_l(x).$$

We call $H_l(t, x)$ the solution of the CP with $\mathrm{CD} = [\phi_l(x), i\sqrt{\lambda(l)}\phi_l(x)]$ so

$$H_l(t, x) = e^{it\sqrt{\lambda(l)}}\phi_l(x).$$

The contribution of ϕ_l to $\operatorname{Tr}[\mathcal{O}(t)]$ is $\exp(it\sqrt{\lambda(l)})$. On the other hand intertwining leads to

$$H_l(t, 0) = e^{it\sqrt{\lambda(l)}}\phi_l(0).$$

The trace and intertwining methods give the same result only when $\phi_l(0) = 1$, which is the case of the exponential basis used in (8.46)f.

The values $\{\phi_l(0)\}$ associated to eigenfunction expansions are generally not well understood.

Problem 8.6 *Use the intertwining method to study* $\{\phi_l(0)\}$ *for classical eigenfunction expansions.*

One might be tempted to rectify the situation by using the orthonormality of the ϕ_l to replace $\phi_l(0) = 1$. In this case

$$\int |H_l(t,x)|^2\,dx = \exp(-2\Im t\sqrt{\lambda(l)}).$$

For hyperbolic and parabolic equations λ_l is real so we do not obtain anything of significance. We would have to change our viewpoint and study elliptic equations; the CD becomes entire functions, or functions whose wave front set is restricted.

Problem 8.7 *Study trace formulas related to elliptic equations.*

We now come to our main example which is related to the Selberg trace formula discussed in Section 1.7. We start with the wave operator $\Box$ in three-dimensional Minkowski space and the CS $= G/K =$ half-hyperboloid, $r^2 = t^2 - x^2 - y^2 = 1, t > 0$. ($G = SL\,(2,\mathbb{R})/\pm I$. We use the notation and ideas of Chapter 7.) On G/K we study Γ invariant functions; Γ is a discrete subgroup of G with compact fundamental domain and no elliptic elements. (Actually finite volume $\Gamma \setminus G$ and Γ with elliptic elements could be treated by our methods with some additional technical difficulties.)

As usual the geometric basis is $\{\delta_x\}_{x\in\Gamma\setminus G/K}$. The analytic basis is $\{\phi(x,\lambda_j)\}$ where $\phi(x,\lambda_j)$ are the Γ invariant eigenfunctions of the noneuclidean Laplacian Δ on G/K. The eigenvalue of $\phi(x,\lambda_j)$ is

$$\lambda_j = s_j(1+s_j).$$

Recall that

$$\Box = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} - \frac{1}{r^2}\Delta$$

by (7.54) and, as in (7.60), the "normal derivative" is $\frac{1}{2} + r\partial/\partial r$. The solution of the CP for $\Box$ with CD $= [\phi(x,\lambda_j), 0]$ is

$$\tfrac{1}{2}\phi(x,\lambda_j)\left[r^{s_j} + r^{-1-s_j}\right] \tag{8.48}$$

where r is the Minkowski distance. (We could also use CD $= (0,\phi)$ which would involve a minor change.)

The Fourier–eigenfunction expansion on $\Gamma \setminus G/K$ is

$$\delta_{x^0} = \sum \bar{\phi}(x^0,\lambda_j)\phi(x,\lambda_j).$$

Here x^0 is a point in a fundamental domain for Γ on $G\backslash K$. As in the case of the usual PSF for $\mathbb{R}^1$ (Section 1.7) we apply $\mathbf{R}_0^*$ to this identity, where $\mathbf{R}_0$ is the Radon transform defined by summation over Γ; the CS is this fundamental

domain. $\mathbf{R}_0^*$ is the operator which extends functions on the fundamental domain to Γ invariant functions on G/K. This leads to the identity on G/K

$$\sum_{\gamma\in\Gamma}\delta_{\gamma x^0} = \sum \bar{\phi}(x^0,\lambda_j)\phi(x,\lambda_j). \tag{8.49}$$

By (8.48) the solution of the CP for $\Box$ with data $(\delta_{x^0},0)$ on G/K is

$$\mathbf{R}^*\left(\sum \delta_{\gamma x^0},0\right)(x,r) = \tfrac{1}{2}\sum \bar{\phi}(x^0,\lambda_j)\phi(x,\lambda_j)\left[r^{s_j}+r^{-1-s_j}\right]. \tag{8.50}$$

$\mathbf{R}^*$ is the usual (dual) Radon transform associated to a CP, namely the map of the CD into the solution. In particular, if $x=0$ then

$$\mathbf{R}^*\left(\sum \delta_{\gamma x^0},0\right)(0,r) = \tfrac{1}{2}\sum \bar{\phi}(x^0,\lambda_j)\phi(0,\lambda_j)\left[r^{s_j}+r^{-1-s_j}\right]. \tag{8.51}$$

As mentioned above, this formula is complicated because $\phi(0,\lambda_j)\not\equiv 1$. If we had $\phi(0,\lambda_j)\equiv 1$ then we could set $x^0=0$ in (8.49) and then the right side of (8.51) would reduce to

$$\tfrac{1}{2}\sum r^{s_j}+r^{-1-s_j} \tag{8.52}$$

on the t axis. This sum is of the same form as we described above for the operator $\partial/\partial t-\partial/\partial x$.

To remove the burdensome term $\phi(0,\lambda_j)$ we return to the idea of trace. Instead of setting $x^0=x=0$ in (8.50) we set $x^0=x$ and integrate over $x\in\Gamma\setminus G/K$. We arrive at (8.52).

This is the analytic computation. The geometric computation depends on applying the fundamental solution of the wave equation to the δ function basis $\{\delta_{x^0}\}$ to compute the solution at (x^0,r) and then integrating over x^0. It is the fundamental solution which is the kernel $k_x(y)$ of (8.44).

In Selberg's work the geometric form of the trace is expressed in terms of the eigenvalues of the matrices $\gamma\in\Gamma$. We wish to explain why this is so from our vantage point.

The fundamental solution e_Δ for the Laplacian Δ in $\mathbb{R}^3$ is r^{-1}; for the wave equation the formula for $e_\Box$ is essentially the same except that r becomes the Minkowski distance to the origin and support $e_\Box$ lies in the forward or backward light cone.

We are interested in evaluating the solution of $\Box$ on the hyperboloid $r=t^0$ in terms of its values on $r=1$. By Lorentz invariance we can reduce the computation to that of the solution at the point $(t^0,0,0)$ on the t axis.

As described above we are interested in the values of $e_\Box(t^0,0,0;1,\zeta,\theta)$ for those points $(1,\zeta,\theta)$ on $r=1$ which are of the form $\gamma(1,0,0)$ since these represent the "fixed point indices" of such points.

In terms of the representation of Minkowski 3 space as 2×2 symmetric matrices as in Chapters 7 and 10, $\gamma(1,0,0)$ is the matrix $\gamma\gamma'$ and $(t^0,0,0)$ is $t^0 I$. If we conjugate γ by an orthogonal matrix k then

$$\gamma\gamma' \to k\gamma k' k\gamma' k' = k\gamma\gamma' k'$$

while $t^0 I$ is invariant under such a conjugation. The Lorentz distance between X and Y is clearly invariant under conjugation by k. This shows that the fixed point index of $\gamma \cdot I$ relative to $t^0 \cdot I$ is invariant under conjugation by $k \in K$. Thus this fixed point index depends only on $\operatorname{Tr}\gamma$ since $\det\gamma = 1$.

To compute the fixed point index we may assume $\gamma = \begin{pmatrix} e^{\zeta} & 0 \\ 0 & e^{-\zeta} \end{pmatrix}$. Then

$$\begin{aligned}[2\ \text{distance}(tI, \gamma\gamma')]^2 &= (t - \operatorname{Tr}\gamma)^2 - (e^{\zeta} - e^{-\zeta})^2 \\ &= t^2 - 2t \operatorname{Tr}\gamma + 4.\end{aligned}$$

Because the support of the fundamental solution lies in the light cone, only those γ for which $\text{distance}(tI, \gamma\gamma') \geq 0$ enter the trace formula.

This gives the fixed point index for tI. Other points on the hyperboloid $G \cdot tI$ are treated exactly the same way. The trace is then obtained by integration over a fundamental domain.

The explicit formulas and the method of rigorizing (i.e. changing from the above formal computation to a rigorous one) can be made in many ways. The usual rigorization involves replacing $\operatorname{Tr}\tau(t)$ by $\int f(t) \operatorname{Tr}\tau(t)\, dt$ for suitable f.

We refer the reader to [139] and many subsequent works for such treatments.

We return to the analytic side of the trace formula. We mentioned in Section 1.7 that Selberg's trace formula is a form of Frobenius' formula for the induced character. We want to show how the eigenfunction relation (8.49) leads to Frobenius' formula.

We mentioned in Section 1.7 that one could essentially substitute the spherical function for the character. We now make this more precise. The point is that $\phi(x,\lambda)$, being an eigenfunction of Δ, belongs to the principal series representation corresponding to s. Thus $\phi(x,\lambda)$ can be expanded in the basis of the eigenfunctions $u_p(x,s)$ of Δ on G/K which satisfy

$$u_p(k_\theta g, s) = e^{ip\theta} u(g,s)$$

for $k_\theta \in K$ (see [68–70] for details). (The u_p are matrix coefficients of the principal series representation.) In particular u_0 is the spherical function (bi-K-invariant).

u_0 is very closely related to the character $\chi(g,s)$ of the representation corresponding to s. We have pointed out this relation in Section 1.7. In fact,

$$\int \chi(gk,s)dk = cu_0(g,s). \tag{8.53}$$

Now, by applying left integration over K to the eigenfunction relation (8.49) we obtain an analytic–geometric identity

$$\sum \delta_{K\gamma x^0} = \sum c_j \bar{\phi}(x^0, \lambda_j) u_0(x, \lambda_j). \tag{8.54}$$

To show the relation between (8.54) and Frobenius' formula for the induced character we must invert (8.53). To perform this inversion we introduce another aspect of the spherical function. This uses some of the structure of G detailed in Section 10.1.

The spherical function can be defined as the bi-K-invariant function in the principal series of $G = SL(2, \mathbb{R})$. These representations are induced from the representation

$$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} = \begin{pmatrix} a & a^{-1}n \\ 0 & a^{-1} \end{pmatrix} \to a^{2s}$$

of the subgroup NA. The Iwasawa decomposition $G = NAK$ allows us to identify the representation space with the space of functions which are homogeneous of degree $2s$ on $N \setminus G =$ plane minus origin. Precisely the functions $f(g)$ belonging to the representation space satisfy $f(nag) = a^{2s} f(g)$.f is identified with the function $\tilde{f}$ on the plane by

$$\begin{aligned} f(nak) &= \tilde{f}[(0 \quad 1) \cdot ak] \\ &= \tilde{f}\left((0 \quad a^{-1})k\right) \\ &= a^{2s}\tilde{f}\left((0 \quad 1)k\right). \end{aligned}$$

The representation space also contains a unique right K invariant function $\tilde{f}_0$ which, being left N and right K invariant, can be identified with the function $h_0(z) = \eta^s$ on the upper half-plane $G/K = \{z = \xi + i\eta\}$. The precise identification is given by

$$\begin{aligned} \eta^s &= h_0 \left[\begin{pmatrix} \eta^{1/2} & \xi\eta^{-1/2} \\ 0 & \eta^{-1/2} \end{pmatrix} \cdot i \right] && h_0 \text{ as function on } \{z\} \\ &= h_0 \left[\begin{pmatrix} 1 & \xi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta^{1/2} & 0 \\ 0 & \eta^{-1/2} \end{pmatrix} k \cdot i \right] && h_0 \text{ lifted to } G \\ &= \tilde{f}_0 \left[(0 \quad 1) \begin{pmatrix} 1 & \xi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta^{1/2} & 0 \\ 0 & \eta^{-1/2} \end{pmatrix} k \right] && \tilde{f}_0 = h_0 \text{left } N \text{ invariant} \\ &= \tilde{f}_0 \left[(0 \quad \eta^{-1/2}) \, k \right] && \tilde{f}_0 \text{ as function on plane.} \end{aligned} \tag{8.55}$$

From the function $\eta^s = (\Im z)^s$ we can construct the unique function on the upper half-plane which is bi-K-invariant and lies in the representation space. This is

$$\int \Im(kz)^s \, dk = y^s \int |z \sin\theta + \cos\theta|^{-2s} \, d\theta. \tag{8.56}$$

Since this function is K invariant, it is constant on every hyperbolic circle centered at i. Thus we can evaluate it at any point on such a circle, in particular the point $i\eta$ on the imaginary axis. We obtain

$$\int \left[\left(\frac{\eta + \eta^{-1}}{2} \right) - \left(\frac{\eta - \eta^{-1}}{2} \right) \cos 2\theta \right]^{-s} d\theta = P^0_{-s}(\cosh \zeta) \tag{8.57}$$

for $\eta = \exp(\zeta)$. P^0_{-s} is the Legendre function [11, vol. I, Chapter 3].

From this point of view P^0_{-s} appears as a Radon transform, namely the Radon transform of the function η^s on the spread of hyperbolic circles centered at i.

In order to clarify matters we wish to digress and analyze the Legendre function and the Legendre transform, i.e. the integral transform with kernel $P^0_s(\cosh \zeta)$, and measure $\sinh 2\zeta \, d\zeta$ which is the Haar measure on G restricted to bi-K-invariant functions. By definition

$$P^0_s(\cosh \zeta) = \int (\cosh \zeta + \sinh \zeta \cos \theta)^s \, d\theta.$$

If we make the change of variables

$$\cosh \zeta + \sinh \zeta \cos \theta = e^u$$

we find, up to normalizing constants (see [11, Vol. I, p. 159]),

$$\begin{aligned} P^0_s(\cosh \zeta) &= \int_{-\zeta}^{\zeta} \frac{e^{u(s+1/2)}}{(\cosh \zeta - \cosh u)^{1/2}} du \\ &= \int_0^{\zeta} \frac{\cosh u(s + 1/2)}{(\cosh \zeta - \cosh u)^{1/2}} \frac{d \cosh u}{\sinh u}. \end{aligned} \tag{8.58}$$

For any function $h(\cosh \zeta)$ we can express its Legendre transform as

$$\begin{aligned} \check{h}(s) &= \int P^0_s(\cosh \zeta) h(\cosh \zeta) \sinh 2\zeta \, d\zeta \\ &= \int_0^{\infty} \cosh u(s + \tfrac{1}{2}) du \int_u^{\infty} \frac{h(\cosh \zeta)}{(\cosh \zeta - \cosh u)^{\frac{1}{2}}} \sinh 2\zeta \, d\zeta. \end{aligned} \tag{8.59}$$

Thus $\check{h}(s)$ is the composition of the (adjoint) Abel transform of $h(\cosh \zeta)$ with the Fourier transform (see e.g. [29, vol. I, p. 158]). (There is a change of variable from 2ζ to ζ which we are ignoring. Moreover the kernel of the "Fourier transform" is cosh rather than cos; this means that $\Re s = -\frac{1}{2}$.) In particular, by (8.58), P^0_s is the Abel transform of $[\cosh u(s + \frac{1}{2})]/\sinh u$ considered as a function of $\cosh u$.

There is a standard formula for the inversion of the Abel transform: if

$$f(x) = \int_0^x \frac{\phi(t)}{\sqrt{x - t}} \, dt \tag{8.60}$$

then

$$\phi(x) = \int_0^x \frac{f'(t)}{\sqrt{x-t}}\, dt. \tag{8.61}$$

The same formulas are valid if we replace the interval $[0, x]$ by $[x, \infty)$ provided that we change $\sqrt{x-t}$ to $\sqrt{t-x}$ (also there are constant factors which depend on normalizations).

The inversion formulas (8.60), (8.61) were found by Abel. Since the Legendre transform is a composition of the Abel transform and the Fourier (Mellin) transform, the inversion formula leads easily to the inversion of the Legendre transform and hence to the Plancherel measure for the Legendre transform (see Chapter 7). This method has been used by many authors (see e.g. [68, 69]).

The inversion formula for the Abel transform can also be found from the explicit formula for the fundamental solution

$$e(t, x, y) = (t^2 - x^2 - y^2)^{-1/2} = r^{-1}. \tag{8.62}$$

This approach uses the fact that, according to the ideas of Chapter 7, the Fourier transform v of the function $\hat{r}^{is}$ on the light cone $\hat{\Gamma}$ is of the form $\gamma(s) r^{-is} P^0_{-is}(\cosh \zeta)$ in the forward light cone. Here $\gamma(s)$ is a quotient of products of Γ functions. v is of the form $\gamma(s) r^{-is} = \gamma(s) t^{-is}$ on the t axis. We can use the fundamental solution of the wave equation to evaluate $v(t, 0, 0)$ in terms of $v(1, r, \theta)$. This evaluates t^{-is} in terms of $P^0_{-is}(\cosh \zeta)$. The inversion formula for the Mellin transform then yields the inversion for the Legendre transform and hence, by (8.58), for the Abel transform.

After this digression to Legendre functions and the Abel transform we can understand the passage from the spherical functions to characters. We have seen that the spherical functions can be obtained from characters by right integration over K. The characters as functions on the Cartan group A are essentially Weyl sums of a^{2s}; that is, they are obtained from a^{2s} by summation over the Weyl group W and dividing by the Weyl denominator. (In this case, W consists of identity and $a \to a^{-1}$ and the denominator is $a - a^{-1}$.) To compute $\int \chi(gk)\, dk$ we have to find the conjugacy class of gk. Since this integral is spherical, i.e. bi-K-invariant, we can assume that $g \in A$ since $G = KAK$.

Let $g = \alpha \in A$. Then

$$\begin{aligned} gk &= \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \alpha\cos\theta & \alpha\sin\theta \\ -\alpha^{-1}\sin\theta & \alpha^{-1}\cos\theta \end{pmatrix}. \end{aligned}$$

The conjugacy class of gk is determined by

$$\operatorname{tr} gk = (\alpha + \alpha^{-1}) \cos\theta.$$

To find the element of A which is conjugate to gk we have to solve

$$a + a^{-1} = (\alpha + \alpha^{-1}) \cos\theta.$$

We denote this a by $a(\alpha, k)$. Then

$$\int \chi(\alpha k)\, dk = \int \chi(a(\alpha, k))\, dk.$$

A direct computation shows that this integral is just the Abel transform of $\chi(\alpha)$. (In any case this is intrinsically clear since by (8.51) $\int \chi(gk)\, dk$ is the spherical function $P^0_{-s}(\cosh \log \alpha)$ which, as we have noted in (8.58), is the Abel transform of $(\alpha^{s+\frac{1}{2}} + \alpha^{-(s+1/2)})/(\alpha - \alpha^{-1})$.)

We can now start with (8.54) and apply the inverse Abel transform. Integration over $x^0 \in \Gamma \setminus G/K$ yields the Frobenius reciprocity theorem.

Remark. *The above discussion shows that after conceptual analysis we need one of three explicit computations to derive Selberg's trace formula:*

(a) *The explicit formula for the fundamental solution. (This is needed for the geometry of the trace or intertwining methods.)*
(b) *The explicit formula for the characters of those irreducible representations that contain* Γ *invariant vectors. (This is needed to make the Frobenius reciprocity formula explicit.)*
(c) *The inversion of the Abel transform. (This is needed for the approach using spherical functions.)*

Another ingredient of the Selberg trace formula involves Hecke operators. These exist when Γ is the modular group and for some other discrete groups. In such cases $\Gamma \setminus G$ may be noncompact so there is a continuous series of eigenfunctions $\phi_\lambda(\tau)$ in addition to the discrete series. Although the noncompactness of $\Gamma \setminus G$ gives rise to technical difficulties the formal aspects remain the same. We shall deal only with the latter.

The Hecke operators T_n are formed from finite sets $\{M_{nl}\}$ of point transformations of G/K such that

$$T_n f(z) = \sum_l f(M_{nl} z)$$

is Γ invariant whenever f is [91]. For the modular group Hecke chose M_{nl} as fractional linear transformations defined by matrices of the form

$$M_{nl} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

where a, b, c, d are integers satisfying $ad - bc = n$. Precisely $\left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right) \right\}$ is a set of representatives for integer matrices of determinant n modulo Γ. We can choose

the $\{M_{nl}\}$ to be

$$\left\{\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}\right\}_{\substack{ad=n \\ d>0 \\ b \mod d}}.$$

Thus

$$T_n f(z) = \sum_{\substack{ad=n \\ d>0 \\ b \mod d}} f\left(\frac{az+b}{d}\right).$$

The Hecke operators commute with each other and with the Laplacian Δ and are self-adjoint on $L_2(\Gamma\backslash G/K)$. Thus we can choose our basis $\{\phi_m(z)\}$ on $L_2(\Gamma\backslash G/K)$ to consist of simultaneous eigenfunctions for Δ and all T_n. In particular if we fix n then we can study the "orbits" of Γ and T_n. These are the sets $\{M_{nl}\gamma \cdot x\}$ for x in a fundamental domain for Γ.

Since T_n commutes with the Laplacian it also commutes with the solution of the CP. Hence we can apply T_n to the solution $F(t^0, z)$ of the CP for $\Box$ with $\mathrm{CD} = (f, 0)$ for fixed t^0 and obtain the solution of the CP with $\mathrm{CD} = (T_n f, 0)$. In this way we can compute $T_n \tau_{t^0}$. The analytic computation leads to the trace of T_n on each eigenspace of Δ. The geometric computation involves the fixed points of $T_n \tau_{t^0}$. (All these ideas appear in a different form in Selberg's article [139].)

We conclude this section with another type of PSF.

Consider the light cone Γ^2 in the plane. The lattice points on Γ^2 are $\{(m, m)\}$ and $\{(m, -m)\}$. Using the one-dimensional PSF on each of the lines $y = \pm x$ we find, for the two-dimensional Fourier transform,

$$\widehat{\sum \delta_{m,m}} + \widehat{\sum \delta_{m,-m}} - \widehat{\delta_{0,0}} = \sum \delta_{x+y=m} + \sum \delta_{x-y=m}. \tag{8.63}$$

It is clear that, except for $m = 0$, the δ function of $x \pm y = m$ has the restriction $\delta_{m/2,\pm m/2}$ to Γ^2. Hence, modulo normalization and the simple terms $\delta_{0,0}, \delta_{x\pm y=0}$ we can say $\sum_{p\in\Gamma^2} \delta_p$ *is essentially its own Fourier transform on* Γ^2.

What happens in higher dimensions? The answer is that such a result is false in $\mathbb{R}^n$. One has to go to the adelization to find the correct formulation. It is closely related to the Minkowski–Siegel theorem on quadratic forms. But this subject is far from the ideas of the present work and will appear elsewhere.

8.3 The Euler–Maclaurin summation formula

The PSF is an identity between infinite sums. The Euler–Maclaurin summation formula (EMSF) can be regarded as a "finitistic" version of the PSF.

We call

$$T_N = \sum_{-N}^{N} \delta_j \tag{8.64}$$

so T_N satisfies

$$(\tau - I)T_N = \delta_{N+1} - \delta_{-N} \tag{8.65}$$

where τ represents translation by 1.

We can solve (8.65) by Fourier transform:

$$\hat{T}_N(\hat{x}) = \frac{e^{i(N+1)\hat{x}} - e^{-iN\hat{x}}}{e^{i\hat{x}} - 1}. \tag{8.66}$$

Of course, (8.65) has many solutions but, by Paley–Wiener, the holomorphic function defined by (8.66) is the only solution whose Fourier transform has support in $[-N, N]$.

All the above is "reasonable" because τ is defined in terms of the additive group of $\mathbb{R}$ and we have used the additive Fourier transform. But now we "change gears" and switch to multiplicative Fourier analysis. In the present case this amounts to writing $(\exp(i\hat{x}) - 1)^{-1}$ as a Laurent series (with remainder). Note that $(\exp(i\hat{x}) - 1)^{-1}$ has a simple pole at $\hat{x} = 0$. In fact, the Laurent series was calculated by Bernoulli

$$\frac{1}{e^{i\hat{x}} - 1} = \frac{1}{i\hat{x}} - \frac{1}{i} + \frac{B_1}{2!} i\hat{x} - \frac{B_2}{4!}(i\hat{x})^3 \pm \dots \tag{8.67}$$

where B_j are the Bernoulli numbers. Upon breaking the series (8.67) after the term involving $\hat{x}^M$ (8.66) becomes

$$\begin{aligned}\hat{T}_N(\hat{x}) = {} & \frac{e^{i(N+1)\hat{x}} - e^{-iN\hat{x}}}{i\hat{x}} - Q_M^N(\hat{x}) \left[e^{i(N+1)\hat{x}} - e^{-iN\hat{x}}\right] \\ & + R_M^N(\hat{x}) \left[e^{i(N+1)\hat{x}} - e^{-iN\hat{x}}\right].\end{aligned} \tag{8.68}$$

Q_M^N is a polynomial of degree M and R_M^N is the remainder, which has simple poles at $\{j\}_{j\neq 0}$. (In our normalization $\pi = 1$.) Note that $[\exp i(N + 1)\hat{x} - \exp(-iN\hat{x})]/i\hat{x}$ is the Fourier transform of the characteristic function $\chi[-N, N + 1]$. Moreover the Fourier transforms of $Q_M^N(\hat{x}) \exp i(N + 1)\hat{x}$ and $Q_M^N \exp(-iN\hat{x})$ have support at the endpoints of the interval $[-N, N + 1]$. Thus the Fourier transform of (8.68) takes the form

$$T_N = \chi(-N, N + 1) + B_M^+(N + 1) + B_M^-(-N) + E_M^N. \tag{8.69}$$

$B_M^{\pm}$ are the punctual distributions equal to $Q_M^N(i\,d/dx)\delta_{N+1}$ (resp. $Q_M^N(i\,d/dx)\delta_{-N}$). They are distributions of order $\leq M$ supported at $N + 1$ and $-N$ respectively, whose coefficients are expressible in terms of Bernoulli numbers.

E_M^N has support in $[-N, N+1]$ and is "small." χ represents the characteristic function.

Equation (8.69) is the Euler–Maclaurin sum formula.

Instead of using finite intervals we could have used a semi-infinite interval, e.g. $[0, \infty]$. Applying the Mellin transform to the resulting formula yields one of Riemann's proofs of the functional equation of $\zeta(s)$.

We can formulate a general analog of the EMSF. Let S be a distribution of compact support and let h be a solution of $S * h = 0$. (In the classical case $S = \tau - 1, h = \sum \delta_n$, and $b = \delta_{N+1} - \delta_{-N}$.) If $\chi = \chi_\Omega$ is the characteristic function of a compact domain Ω we have

$$S * \chi h = b$$

where support b is near bd Ω.

Fourier transformation allows us to write

$$(\widehat{\chi h})(\hat{x}) = \frac{\hat{b}(\hat{x})}{\hat{S}(\hat{x})}. \tag{8.70}$$

Since the left side of (8.70) is entire, $\hat{S}$ divides $\hat{b}$.

Suppose we can write $\hat{S} = \hat{S}_0 \hat{S}_1$ where $\hat{S}_0$ is a polynomial which vanishes as the origin and $\hat{S}_1(0) \neq 0$. Then (8.70) becomes

$$(\widehat{\chi h})(\hat{x}) = \frac{\hat{b}(\hat{x})}{\hat{S}_0(\hat{x})} \cdot \frac{1}{\hat{S}_1(\hat{x})}. \tag{8.71}$$

The EMSF for S depends on finding a "nice" expansion for $1/\hat{S}_1$. A particular "nice" expansion would be

$$\frac{1}{\hat{S}_1(\hat{x})} = Q(\hat{x}) + \hat{S}_0(\hat{x})\hat{B}(\hat{x}) \tag{8.72}$$

where Q is a polynomial and B is an entire function. (In the classical case $\hat{S}_0 = i\hat{x}, \hat{S}_1 = (\exp(i\hat{x}) - 1)/i\hat{x}$, and $Q = 1$. Thus (8.72) corresponds to (8.67) multiplied by $i\hat{x}$.) Inserting this value of $1/\hat{S}_1$ into (8.71) leads to

$$(\widehat{\chi h})(\hat{x}) = \frac{\hat{b}(\hat{x})Q(\hat{x})}{\hat{S}_0(\hat{x})} + \hat{b}(\hat{x})\hat{B}(\hat{x}) \tag{8.73}$$

which we regard as a general EMSF.

The inverse Fourier transform of $\hat{x}^{-1}$ is the Heaviside function which is the fundamental solution for d/dx and defines integration. For this reason the classical EMSF is regarded as an "integration formula"; it gives the approximate value of integrals of suitable functions. Thus our general EMSF (8.73) should be regarded as an approximation to the fundamental solution for $S_0(D)$.

Problem 8.8 *Find a suitable setting for this approximation.*

This discussion is relevant to a single operator ∂. To treat the Euler–Maclaurin formula in $\mathbb{R}^n, n > 1$, we must deal with overdetermined systems $\vec{\partial} = (\partial_1, \dots, \partial_j)$ since, for example, the usual lattice satisfies $(\tau_{x_k} - 1)\sum \delta_j = 0$ for $k = 1, \dots, n$. Our method works reasonably well when Ω is a rational convex polygon since we can iterate results for half-planes. As this would lead us too far afield we defer it to a future publication.

If we examine the relation of the PSF to the EMSF we notice a glaring asymmetry. The PSF involves all the integers j and all the characters $\exp(ijx)$. Though the EMSF involves $2N+1$ points it involves only the identity character which, when cut off, becomes the characteristic function $\chi(-N, N+1)$ in (8.69).

It makes more sense to obtain a more symmetric formula. We write[2]

$$\frac{1}{e^{i\hat{x}} - 1} = \frac{e^{-i\hat{x}/2}}{2i \sin \hat{x}/2}.$$

Instead of using only the pole at $\hat{x} = 0$ we use the poles at $\hat{x} = n$ for $n \in [-J, J]$. The residue is 1. Thus

$$\frac{1}{e^{i\hat{x}} - 1} = \frac{1}{i} \sum_{-J}^{J} \frac{1}{\hat{x} - n} + P_J \tag{8.74}$$

where the remainder P_J can be expanded in a power series which converges up to $|\hat{x}| = J + 1$.

This allows us to write, in analogy to (8.68),

$$\begin{aligned} \hat{T}_N(\hat{x}) = \sum_{-J}^{J} \frac{e^{i(N+1)\hat{x}} - e^{-iN\hat{x}}}{i(\hat{x} - n)} - Q^N_{MJ}(\hat{x}) \left[e^{i(N+1)\hat{x}} - e^{-iN\hat{x}} \right] \\ + R^N_{MJ}(\hat{x}) \left[e^{i(N+1)\hat{x}} - e^{-iN\hat{x}} \right]. \end{aligned} \tag{8.75}$$

Q^N_{MJ} is a polynomial of degree M and R^N_{MJ} has simple poles at $\hat{x} = n, |n| > J$.

Note that the inverse Fourier transform of

$$\frac{e^{i(N+1)\hat{x}} - e^{iN\hat{x}}}{i(\hat{x} - n)}$$

is $\exp(inx)\chi[-N, N+1]$. We arrive at the formula

$$T_N = \sum_{-J}^{J} e^{inx} \chi[-N, N+1] + B^+_{MJ}(N+1) + B^-_{MJ}(-N) + E^N_{MJ} \tag{8.76}$$

with the same notation as in (8.69).

[2]Recall our notation convention that we have "normalized" $2\pi = 1$. Thus the zeros of $\sin \hat{x}/2$ are at the integers and the above symbol n is 2π times an integer. (But the x integers in $[-N, N+1]$ are actual integers.)

We refer to (8.76) as the (finite) *approximate PSF.*

The EMSF deals with $\sum_0^N \delta_j$ or $\sum_{-N}^N \delta_j$. We can give an analogous treatment of $\sum_0^\infty \delta_j$. We call

$$T_N^\infty = \sum_N^\infty \delta_j \tag{8.64*}$$

so

$$(\tau - 1)T_N^\infty = \delta_N \tag{8.65*}$$

or

$$\hat{T}_N^\infty(\hat{x}) = \frac{e^{iN\hat{x}}}{e^{i\hat{x}} - 1}. \tag{8.66*}$$

When thinking of $\hat{T}_N^\infty$ as a distribution we associate it to a contour in the complex $\hat{x}$ plane which lies above the poles of $\exp(i\hat{x}) - 1$ as this is the path of integration needed to make the inverse Fourier transform of $\hat{T}_N^\infty$ vanish for $x < N$.

We multiply (8.74) by $\exp(iN\hat{x})$ and derive the analog of (8.75) for $\hat{T}_N^\infty$. Since the inverse Fourier transform of

$$\frac{e^{iN\hat{x}}}{i(\hat{x} - \pi n)}$$

(using our contour) is $\exp(inx)\chi[N, \infty)$ we obtain an analog of (8.76) for T_N^∞ in which the term $\chi[-N, N+1)$ is replaced by $\chi[N, \infty)$.

We refer to this formula as the infinite approximate PSF.

If we are given a Dirichlet series

$$A(s) = \sum_0^\infty a_n n^{-s}$$

then we can break up the sum into $\sum_0^N$ and $\sum_{N+1}^\infty$ and apply the infinite approximate PSF to $\sum_{N+1}^\infty$. The ordinary PSF is used to prove the functional equation for the Riemann ζ function. We can apply the finite and infinite approximate PSF to derive the approximate functional equation for the ζ function [150, Chapter 4].

Some special forms of (8.76) appear in various works on trigonometric sums.

For example, the work of Vander Corput can be formulated in terms of the approximate EMSF.

Vinogradoff's method (see [150]) involves an analogous idea related to the nonlinear Fourier transform (Chapter 5). For, Vinogradoff's method involves the study of sums

$$\sum_a^{a=q} e^{2\pi i(\alpha_0 + \alpha_1 n + \cdots + \alpha_k n^k)} = \text{nonlinear FT of } \chi[a, a+q] \sum \delta_n$$

By examining equation (8.66) in a different way we can gain a new insight into the EMSF.

We write (8.66) in the form

$$\frac{\sin(N+1/2)\hat{x}}{\hat{x}} = \hat{T}_N(\hat{x})\frac{\sin\hat{x}/2}{\hat{x}} \tag{8.66**}$$

Note that $[\sin(N+1/2)\hat{x}]/\hat{x}$ is the Fourier transform of the characteristic function of $[-(N+1/2),\ (N+1/2)]$. Since $\hat{T}_N(\hat{x})$ is the Fourier transform of $\sum_{-N}^{N}\delta_j$ we can interpret formula (8.66**) as saying that the integral of any function f over $[-N+l/2),\ (N+1/2)]$ is the sum of the integrals of f over the translates of $[-\frac{1}{2},\frac{1}{2}]$ by $j\in[-N,N]$ (which is obvious).

The usual EMSF approximates the integral of f over $[-(N+1/2),(N+1/2)]$ by a sum of values of f taken from these translates. There are many other settings in which such integrals of f are approximated by sums of contributions from the intervals $[j-1/2,j+1/2]$. One notable example occurs in the study of *almost periodic (AP) functions* (see [20]). In this case N is large and the subintervals are of the form $[j-l,j+l]$ where l is fixed and j may not be an integer.

Periodic functions are defined by the equations

$$(\delta_a-\delta_0)*f=0$$

and AP functions by

$$\left|(\delta_{a_j}-\delta_0)*f\right|<\epsilon$$

for suitable $\{a_j\}$ depending on ϵ.

Mean periodic functions (see [137]) satisfy more general convolution equations

$$S*f=0.$$

The above idea leads to the notion of *almost mean periodic functions* which are defined by

$$\left|S_{a_j}*f\right|<\epsilon.$$

This theory is closely related to the Hardy–Littlewood circle method (see [113]); we shall study it elsewhere.

In order to put the EMSF in a more general setting, we want to reformulate equation (8.65). Notice that $\tau=\exp(d/dx)$. (This is just Taylor's formula.) Instead of taking Fourier transforms we can work directly in x space. We solve (8.65) by writing

$$e^{d/dx}-1=\frac{d}{dx}\left[\frac{e^{\frac{d}{dx}}-1}{\frac{d}{dx}}\right] \tag{8.77}$$

where everything is defined by formal power series in d/dx.

We now solve (8.65) by inverting the right side of (8.77). The inverse of d/dx is integration and the coefficients of the power series of

$$\frac{\frac{d}{dx}}{e^{\frac{d}{dx}}-1}$$

are the Bernoulli numbers. We again arrive at (8.69).

The same method applies to the approximate PSF.

This latter method can sometimes be used in situations in which Fourier analysis seems to fail. Consider, for example, the hyperbolic plane. Let Γ_0 and Γ be two discrete subgroups of $SL(2,\mathbb{R})$ with fundamental domains of finite area and $\Gamma_0 \subset \Gamma$. We choose a fundamental domain Δ for Γ_0 which is composed of several fundamental domains for Γ. We pick one point $p_0 \in \text{interior}\,\Delta$. The EMSF should give an approximate formula for

$$\sum f(p_j)$$

for suitable functions f. The sum is over all points $p_j \in \Delta$ which are congruent to p_0 modulo Γ.

Suppose there is a $\gamma \in \Gamma$ such that

$$\{\gamma p_j\} \subset [\{p_j\} \cup \{q_k\}] \tag{8.78}$$

where "most" of the p_l occur in $\{\gamma p_j\}$. We write

$$(\gamma - 1)\sum \delta_{p_j} = \sum \delta_{q_k} - \sum \delta_{r_l} \tag{8.79}$$

where $\{r_l\}$ is a small subset of $\{p_j\}$. (This is the analog of (8.65).)

We also write

$$\gamma = \exp(X) \tag{8.80}$$

where X lies in the Lie algebra of $SL(2,\mathbb{R})$ and hence can be thought of as a first-order differential operator.

We now repeat the argument (8.77)f. Note that $\exp(X)$ is differentiation along the one-parameter group $\exp(tX)$. The reason the argument (8.77)f. works is because

$$\left(\frac{d}{dx}\right)^{-1}(\delta_{N+1} - \delta_{-N}) = \chi(-N, N+1). \tag{8.81}$$

Thus we need to have

$$\sum \delta_{q_k} - \sum \delta_{r_l} = \sum \left(\delta_{q_j} - \delta_{r_j}\right) \tag{8.82}$$

where (r_j, q_j) lies on the translate of the one-parameter group $\exp(tX)$ by r_j.

We have proven

Theorem 8.4 *Under the assumption (8.82)*

$$\sum \delta_{p_j} = \sum \chi(r_j, q_j) + B_M^+ + B_M^- + E_M^\Gamma \tag{8.83}$$

where $\chi(r_j, q_j)$ is the characteristic function of the part of the curve $r_j \exp(tX)$ lying between r_j and q_j, $B_M^\pm$ are distributions of order $\leq M$ supported at $\{q_j\}, \{r_j\}$ respectively, and E_M^Γ is "small."

Example Γ = modular group and Γ_0 = principal congruence subgroup modulo N, meaning the set of integral matrices a of determinant 1 satisfying

$$a \equiv I \,(\text{mod } N). \tag{8.84}$$

We can choose $\gamma =$ translation by 1.

Remark 1 *All we need for Theorem 8.4 is a set of points $\{p_j\}$ and a hyperbolic transformation γ for which (8.82) is satisfied. We formulated the theorem in terms of the discrete groups $\Gamma_0 \subset \Gamma$ because this formulation is parallel to the usual EMSF.*

Remark 2 *The above formulation of the EMSF involves points and curves. We do not have area integrals as in the case of the euclidean plane because $SL(2, \mathbb{R})$ has rank 1.*

Problem 8.9 *Is there an analog of the EMSF involving conjugacy classes, the Selberg zeta function, and area integrals? (Such an EMSF would lead to an approximate functional equation for the Selberg zeta function.)*

8.4 The compact trick

The origin of the ideas of this section is H. Weyl's compact trick (see [159]). Weyl's idea is that finite-dimensional representations of semi-simple noncompact Lie groups can be completely analyzed in terms of compact forms of the given groups.

We envisage a broader setting for this process; in particular we apply our method to some infinite dimensional representations. In its full generality we deal with a differential operator (or system of differential operators) with two eigenfunction expansions, one of which we understand (analog of the compact form) and which we use to shed light on the other expansion.

As we shall see from the examples, the compact trick can be thought of as some variation of the PSF.

The method can be described in the following terms. We have two prospective bases $\{\phi(s, x)\}$ and $\{\psi(\lambda, y)\}$. We are primarily interested in the ϕ expansion. The ψ expansion, which is understood, is a tool, e.g. for the computation of the inversion formula for the ϕ expansion.

We call

$$f_\phi(s) = \int f(x)\bar{\phi}(s, x)\, dx \tag{8.85}$$

$$f_\psi(\lambda) = \int f(y)\bar{\psi}(\lambda, y)\, dy. \tag{8.86}$$

It is not assumed that $\{x\}$ and $\{y\}$ are the same. For our ideas to work we need a way of passing from functions of x to functions of y.

We present several examples of increasing complexity.

For our first example we show how to derive the Fourier integral inversion formula for the line from the inversion for Fourier series. In his original work on Fourier series and integrals Fourier showed how to accomplish this by making the period $\to \infty$. It is important for our methodology that we use a *single period.*

In conformity with the above notation, y is a parameter on the circle and x is a parameter on the line. $\phi(s,x) = \exp(isx)$ and $\psi(\lambda, y) = \exp(iny)$ with $\{\lambda\} = \{n\}$.

Let $f(x)$ be a function with support $f \subset [-\epsilon, \epsilon]$; $f(y)$ is the periodic function defined by $f(x)$. In the present case $f_\phi(n) = f_\psi(n)$ by the PSF. By the inversion formula for Fourier series

$$f(y) = \sum f_\psi(n) e^{-iny}. \tag{8.87}$$

To obtain $f(x)$ from $f(y)$ we can multiply $f(y)$ by χ where χ is the characteristic function of $[-\frac{1}{2}, \frac{1}{2}]$.

We want to show that, for small x,

$$\chi f(y) = f(x) = \int f_\phi(s) e^{-isx}\, ds. \tag{8.88}$$

Since f_ϕ is an entire function of small exponential type we can employ the interpolation formula

$$f_\phi(s) = \sum f_\phi(n) \frac{\sin s}{s-n}. \tag{8.89}$$

(The reader may be confused because we have omitted a factor of 2π in $\sin s$. This is due to our normalization convention.)

We are left with the evaluation of

$$\int \left[\sum f_\phi(n) \frac{\sin s}{s-n} \right] e^{-isx}\, ds. \tag{8.90}$$

It is easy to regularize the interpolation formula to enable us to interchange the order of summation and integration. Since

$$\int \frac{\sin s}{s-n} e^{-isx}\, ds = e^{-isn} \int \frac{\sin s}{s} e^{-isx}\, ds \tag{8.91}$$

we are left with the task of showing

$$\int \frac{\sin s}{s} e^{-isx}\, ds = \chi \tag{8.92}$$

which can be accomplished by writing $\sin s = [\exp(is) - \exp(-is)]/2i$ and applying contour integration.

This verifies the Plancherel formula for functions of small support; the extension to general functions is standard.

The above passage from the circle to the line depended on the identification of functions on the circle with periodic functions on the line. For our next example of the compact trick the compact and noncompact groups meet only at one point (the group identity).

As before, the compact group = $\{y\}$ is the circle in the plane but now the noncompact group = $\{x\}$ is the scalar group, i.e. the multiplicative group of positive real numbers. Fourier analysis on the noncompact group is the Mellin transform so that $\phi(x) = x^s$.

Let $f(x)$ be an entire function which is small at infinity on $\mathbb{R}^+$, say, $f(x) = \mathcal{O}(x^{-N})$ for all N as $x \to \infty$. Then the Mellin transform

$$\underset{\vee}{f}(s) = f_\phi(s) = \int f(x) x^s \frac{dx}{x} \tag{8.93}$$

is holomorphic in the half-plane $\{\Re s > 0\}$ and is bounded in any strip

$$0 < \epsilon \leq \Re s \leq A.$$

The growth as $A \to \infty$ is a "Legendre conjugate" of the decrease of f as $x \to \infty$ (see Chapter V of FA).

We can also study $\underset{\vee}{f}$ in the left half-plane. If we write, for x small,

$$f(x) = \sum_{n=0}^{N} f_n x^n + \mathcal{O}(x^{N+1}) \tag{8.94}$$

then, if $a > 1$,

$$\begin{aligned} \int_0^a f(x) x^s \frac{dx}{x} &= \sum_{n=0}^{N} f_n \int_0^a x^{n+s} \frac{dx}{x} + \mathcal{O}\left[\int_0^a x^{N+1+s} \frac{dx}{x}\right] \\ &= \sum_{n=0}^{N} f_n \frac{a^{n+s}}{n+s} + \mathcal{O}\left(\frac{1}{N+1+s}\right) \end{aligned} \tag{8.95}$$

as long as $\Re(N+1+s) > 0$, i.e. $\Re s > -N-1$.

The integral on (a, ∞) converges so we conclude that $f_\phi(s)$ *has poles at the negative integers with residues* $\{f_n\}$.

For example, if $f(x) = \exp(-x)$ then $f_\phi(s) = \Gamma(s)$.

For any f which is small at infinity it is clear that the part of the integral (8.93) coming from $[a, \infty)$ is $\mathcal{O}(a^{\Re s})$ as $\Re s \to -\infty$. Moreover for any a the series $\sum f_n \frac{a^n}{n+s}$ converges uniformly on the s plane if we remove a circle of fixed radius around each of the poles since f is entire. It follows that for f entire the sum which represents the part of f_ϕ coming from $[0, a]$ is

$$\begin{aligned} \sum_0^\infty f_n \frac{a^{n+s}}{n+s} &= a^s \sum f_n \frac{a^n}{n+s} \\ &= \mathcal{O}(a^{\Re s}) \end{aligned} \tag{8.96}$$

as $\Re s \to -\infty$ outside of circles of fixed radius (independent of n) around the poles.

We can shift the contour in the (proposed) inverse Mellin transform

$$h(x) = \int_{\Re s = 0^+} f_\phi(s) x^{-s}\, ds \tag{8.97}$$

to $\Re s = -\infty$ for $x < a$ by (8.93), (8.94), and (8.96). Since a is arbitrary we can shift for all $x > 0$. This leads to

$$h(x) = \int_{\Re s = 0^+} f_\phi(s) x^{-s}\, ds = \sum f_n x^n. \tag{8.98}$$

Since the last series is $f(x)$ we have established the Mellin inversion formula for f.

All this works for entire functions f satisfying the above conditions. It is clear that such functions, e.g. $\{x \exp(-ax)\}$, are dense in L_2.

We have used power series in our derivation of the Mellin inversion formula. We now show how the power series representation, and hence also Cauchy's formula, are consequences of the inversion formula for Fourier series so our derivation of the Mellin inversion formula can be regarded as a compact trick.

If f is an entire function we can write the Fourier series expansion of f on the circle of radius r

$$f(z) = f(re^{i\theta}) = \sum a_n(r) e^{in\theta}.$$

The Cauchy–Riemann operator $\partial/\partial\bar{z}$ in polar coordinates is

$$\frac{\partial}{\partial \bar{z}} = \frac{e^{i\theta}}{2}\left(\frac{\partial}{\partial r} + \frac{i}{r}\frac{\partial}{\partial \theta}\right). \tag{8.99}$$

Since $\partial f/\partial \bar{z} = 0$ we find for $r \neq 0$

$$a_n'(r) - \frac{n a_n(r)}{r} = 0$$

so that

$$a_n(r) = c_n r^n.$$

From the continuity of f at the origin

$$a_n(r) = 0 \qquad \text{if } n < 0.$$

This is the power series expansion of f.

It is convenient to regard this use of the compact trick in the realm of harmonic functions (as in Chapter 3). The compact group gives the Fourier series

expansion of f on the circle $\{r = 1\}$. Since $\partial f/\partial \bar{z} = 0, f$ agrees with the harmonic extension of $\sum a_n(1)e^{in\theta}$ to the whole plane by the uniqueness for the CP for $\partial/\partial\bar{z}$ on the CS which is $\{r = 1\}$. Thus

$$f(z) = \sum_{n\geq 0} a_n(1)r^n e^{in\theta} + \sum_{n<0} a_n(1)e^{in\theta}r^{-n}.$$

But the second sum is antiholomorphic and hence is $\equiv 0$.

Thus $f(z) = \sum a_n(1)z^n$.

The multiplicative group G of positive reals meets the circle group $G^{\mathbb{C}}$ in the complex plane at the group identity. One might expect that, for suitable analytic functions f, we could shift the contour in

$$\int f(x)x^s \frac{dx}{x}$$

from $x \in [0, \infty)$ to the unit circle. It seems reasonable that the resultant integral would wrap infinitely often around the unit circle, accounting for the fact that $f_\psi(n)$ is the residue of the pole of $f_\phi(s)$ at $s = n$.

Problem 8.10 *Carry out this program.*

Remark. *The two forms of the compact trick that we have given parallel the two forms of the PSF described at the beginning of Section 1.8.*

We now pass to a higher level of sophistication in preparation for our treatment of the compact trick for $G = SL(2, \mathbb{R})$.

Let x be the standard parameter on the (half) unit hyperbola

$$\xi^2 - \eta^2 = 1 \qquad \xi > 0, \quad \xi = \cosh x, \ \ \eta = \sinh x. \tag{8.100}$$

As usual $\phi(s, x) = \exp(sx)$. We are going to derive the Plancherel formula for Fourier integrals using the harmonicity idea of the previous example.

Let F be a solution of the wave equation

$$\Box F \equiv \frac{\partial^2 F}{\partial \xi^2} - \frac{\partial^2 F}{\partial \eta^2} = 0 \tag{8.101}$$

such that F is exponentially small at infinity in the region between (8.100) and the (half) light cone Γ^+

$$\eta = \pm\xi \qquad \xi \geq 0. \tag{8.102}$$

Assume also that F is an entire function of ξ, η.

We call f the restriction of F to (8.100) and set

$$\begin{aligned} f_\phi(s) &= \int f(\cosh x, \sinh x)e^{-sx}\, dx \\ &= \int_{r=1} F(r\cosh x, r\sinh x)r^s e^{-sx}\, dx. \end{aligned} \tag{8.103}$$

We decompose f into odd and even parts under $x \to -x$. For simplicity we can assume that f is even. Thus the exponential factor $\exp(-sx)$ can be replaced by $\cosh sx$.

According to the ideas of the previous example we want to compute the compact harmonic coefficients, i.e. the coefficients in the expansion in the $\Box$ harmonic polynomials $\{(\xi \pm \eta)^n\}$ in terms of singularities of $f_\phi(s)$. In order to accomplish this we have to shift the contour in the last integral in (8.103) to $r = 0$ which is the half light cone (8.102).

This shifting is possible because F is a solution of the wave equation. We explained how to do this shifting in Chapter 7. Since $r\exp(\pm x) = \xi \pm \eta$ we have

$$r^s \cosh sx = \tfrac{1}{2}[(\xi + \eta)^s + (\xi - \eta)^s]. \tag{8.104}$$

Our usual method of defining the limits on Γ is via analytic continuation in s from $\Re s \gg 0$. The limit on $\xi = \pm\eta$ is $2^{s-1}\xi^s$ which is (up to an unimportant factor) r_Γ^s where r_Γ is the euclidean distance along the generators of the light cone Γ. We are thus left with the usual Mellin transform of $F\big|_\Gamma$ along the generators of Γ.

Our next observation is that the power series coefficients of F can be calculated using the r_Γ derivatives of $F\big|_\Gamma$. Although this is clear in the present situation it is important to note that this is part of the general theory of harmonicity discussed in Chapter 3. Since F is entire these derivatives can also be expressed, by (8.99)f., in terms of the Fourier series coefficients of f on the unit circle in the complex $z = \xi + i\eta$ plane.

Using the idea of (8.98) we can express $F\big|_\Gamma$ in terms of the inverse r_Γ Mellin transform of its r_Γ Mellin transform. This takes the form (up to normalizing constants)

$$F\big|_\Gamma(r_\Gamma) = \int_{\Re s = 0^+} f_\phi(s) r_\Gamma^{-s}\, ds = \sum F_n r_\Gamma^n. \tag{8.105}$$

F_n are the Taylor coefficients of F on the generators. (Since F is even $F\big|_\Gamma(r_\Gamma)$ is independent of the generator.) $f_\phi = \overset{\vee}{F|_\Gamma}$ is the Mellin transform of $F\big|_\Gamma$.

If we express r_Γ^s as in (8.104)f., then by (8.105) (up to normalizing constants)

$$\left\{\sum F_n\left[(\xi + \eta)^n + (\xi - \eta)^n\right]\right\}\Big|_\Gamma = \sum F_n r_\Gamma^n = F\big|_\Gamma(r_\Gamma). \tag{8.106}$$

Since a $\Box$ harmonic function is determined by its restriction to Γ, (8.105) and (8.106) imply

$$F(\xi, \eta) = \int_{\Re s = 0^+} f_\phi(s) r^{-s} \cosh sx\, ds \tag{8.107}$$

since both sides agree on Γ. Upon restriction to $r = 1$ we find that this is the Plancherel formula for the hyperbola.

Remark 1 *We can express the integral in (8.105) as the restriction to* Γ *of a Fourier transform on* $\hat{\Gamma}$. *Since such a Fourier transform defines a solution of the wave equation this gives another way of extending the right sides of (8.105) to a (unique) solution of the wave equation.*

Remark 2 *We have tacitly assumed the density in* L_2 *of the set of restrictions of entire solutions* F *of* $\Box$ *to the unit hyperbola which have the proper decrease at infinity. This depends on the hyperbolicity of* $\Box$ *and the space-like nature of (8.100).*

Let us examine how to use the above ideas to derive the Plancherel formula for bi-K-invariant functions on $G = SL(2, \mathbb{R})$. The same method works for functions on G/K which transform according to a character of K under left action.

The ideas of Chapter 7 show that the situation for G is similar to that of the previous example. The half-hyperbola is replaced by the half-hyperboloid

$$t^2 - \xi^2 - \eta^2 = 1 \qquad t > 0. \tag{8.108}$$

The exponential function is replaced by the Legendre function $P_s(\cosh x)$. As in Chapter 7 we use the invariance of energy to shift the contour in

$$f_\phi(s) = \int_{r=1} F(t, \xi, \eta) r^s P_s(\cosh x)\, dx \tag{8.109}$$

to the positive half of the light cone Γ.

We have shown in Chapter 7 that the limit of $r^s P_s(\cosh x)$ on the light cone $r = 0$ is given by

$$r^s P_s(\cosh x)\big|_\Gamma = \alpha(s) c(s) r_\Gamma^s. \tag{8.110}$$

$\alpha(s)$ is a simple factor and $c(s)$ is Harish-Chandra's c function.

Thus, except for simple factors,

$$f_\phi(s) \cong c(s) \underset{\vee}{F}\big|_\Gamma(s). \tag{8.111}$$

As usual $\underset{\vee}{F}\big|_\Gamma$ is meromorphic with poles at the negative integers. Moreover by Mellin's inversion formula

$$F\big|_\Gamma(r_\Gamma) = \int \underset{\vee}{F}\big|_\Gamma(s) r_\Gamma^{-s}\, ds. \tag{8.112}$$

Again using (8.110) and the uniqueness property of solutions of the wave equation on the light cone we find that, for F a suitable entire function,

$$\begin{aligned} F(t, \xi, \eta) &= \int \underset{\vee}{F}\big|_\Gamma(s) P_{-s}(\cosh \xi) [c(-s)]^{-1} r^{-s}\, ds \\ &= \int f_\phi(s) P_{-s}(\cosh \xi) [c(s)c(-s)]^{-1} r^{-s}\, ds. \end{aligned} \tag{8.113}$$

The compact trick involves shifting the contour in (8.113) to verify that both sides have the same expansion in spherical harmonic polynomials. This gives the Plancherel formula for the Legendre transform.

We cannot shift the contour directly because the integral is on $\Re s = \text{const.}$ and we have to shift to $\Re s \to \pm\infty$ where $P_s(\cosh\zeta)$ is large; in fact, it is asymptotic (essentially) to $\exp(|\Re s|\zeta)$. To understand how to proceed consider the analogous integral where $P_{-s}(\cosh\zeta)$ is replaced by $\cosh s\zeta$. In order to shift contours we need Euler's formula

$$\cosh s\zeta = \tfrac{1}{2}(e^{s\zeta} + e^{-s\zeta}). \tag{8.114}$$

We thus shift the term involving $\exp(\pm s\zeta)$ to $\Re s \to \mp\infty$ (we can assume $\zeta > 0$).

The corresponding formula for P_{-s} is given in [11, vol. I, p. 140 (3)]:

$$\pi P_{-s}(\cosh\zeta) = \tan\pi s[Q_{-s}(\cosh\zeta) - Q_{s-1}(\cosh\zeta)]. \tag{8.115}$$

Here Q is the associated Legendre function. Its asymptotics are essentially the same as those of the exponential function [11, vol. I, p. 164 (21)].

The factor $\tan\pi s$ cancels the factor $[c(s)c(-s)]^{-1}$ in (8.113) except for a factor $s - \frac{1}{2}$ so we can shift the contour to $\Re s \to -\infty$ as in (8.114)f. for the term Q_{-s}. The term Q_{s-1} can be treated by symmetry.

There remains the question of the meaning of (8.115) within our framework. Equation (8.115) can be derived from the null solutions η for the wave equation. We choose η so that its CD on $t = 1$ has support at $(1, 0, 0)$. Thus support η is contained in the light cone translated to $(1, 0, 0)$. We take the Mellin transform of η in scalar multiplication. By the separation of variables for the wave operator (see Section 7.2) the Mellin coefficients are eigenfunctions of Δ_G which is the G Laplacian on each hyperboloid. If we choose $\mathrm{CD}(\eta)$ to be invariant under rotations around the t axis these Mellin coefficients are of the form $\alpha(s)P_s(\cosh\zeta)$.

Note that the rays from the origin cut support η in two parts, which are bounded away from $r = 1$ (except for the ray through $(1, 0, 0)$). The Mellin transform of a function supported in $[0, a]$ with $a < 1$ is bounded by ca^s and hence is exponentially decreasing for $\Re s \to +\infty$ (similarly if support in $[a, \infty]$ with $a > 1$).

By checking the value of $\alpha(s)$ we derive (8.115). (It follows that Q_s is a fundamental solution for the Legendre operator but this is not important to us at this point.)

This completes our usage of the Weyl trick to compute the Plancherel formula for (left) K invariant functions on G/K.

Problem 8.11 *Modify this method to apply to G/A and to higher dimensional groups.*

9

EXTENSION OF SOLUTIONS OF DIFFERENTIAL EQUATIONS

Let f^0 be a function defined on a set Ω which satisfies the system of partial differential equations $\vec{P}(x, D)f = 0$ on Ω in some sense. In this chapter we study the question of whether f^0 can be extended to a solution f of $\vec{P}f = 0$ on a larger set Δ. We also examine the structure, e.g. wave front set, of the restriction of f^0 to certain subsets of Δ.

The problems studied in this chapter are described in Section 9.1.

In Section 9.2 we examine Hartogs' type extension, meaning Ω is part of the boundary of a domain Δ. In this case f^0 is a solution of $\vec{P}f^0 = 0$ in the jet sense.

In Section 9.3 we assume that Ω is the common boundary of several domains Δ^p. On each Δ^p we have a system of linear partial differential equations $\vec{P}^p f = 0$. f^0 is Cauchy data for each of these systems. We examine the wave front set of this common CD.

In Section 9.4 we study wave front sets when the CD on Ω is taken only in the limit sense for solutions on a domain Δ containing Ω in its boundary. We also study the wave front set of the CD of asymptotic solutions f of $\vec{P}$, meaning $\vec{P}f(x) \to 0$ rapidly as $x \in \Delta, x \to \Omega$. We prove an edge-of-the-wedge theorem for such asymptotic solutions.

A new illustration of the Radon ansatz is given in Section 9.5. We start with a function f defined in a region Δ and an elliptic operator P. We let $\{S_\alpha\}$ be the set of spheres of fixed radius (say 1) contained in Δ. f defines Dirichlet data for P on each S_α. We call f_α the solution of $Pf_\alpha = 0$ in the ball bounded by S_α whose Dirichlet data for P coincides with the DD defined by f. If f is "tangent" to all the f_α then we prove that $Pf = 0$ in Δ. These ideas are related to an analog for partial differential equations of contact manifolds.

9.1 Formulation of the problem

The Radon transformation is defined by integration. We have seen that, via the Fourier transform, it becomes restriction (projection–slice theorem). The adjoint of restriction is extension. In this chapter we shall analyze various extension problems related to partial differential equations.

We begin by stating some standard extension theorems for solutions of systems of linear partial differential equations with constant coefficients. This chapter is devoted to sharpening these theorems and putting some structure on the general extension question.

(1) *Hartogs' extension theorem.* Let Δ be a bounded domain in $\mathbb{C}^n$ $(n > 1)$ with smooth connected boundary Ω. Suppose f^0 is defined and holomorphic in Δ near Ω. Then f^0 extends to a holomorphic function f on all of Δ.

In [53] and FA we extended this theorem from holomorphic functions to solutions of

$$\vec{P}(D)f = 0 \tag{9.1}$$

where $\vec{P}(D) = (P_1(D), \ldots, P_r(D))$. The $P_j(D)$ are linear partial differential operators with constant coefficients such that the system (9.1) is truly overdetermined, meaning $r > 1$ and the P_j have no common factor.

Lewy [116] extended Hartogs' result in two different directions. In the first place he showed that f^0 need not be a solution of $\bar{\partial} f^0 = 0$ in a neighborhood of Ω in Δ. It suffices that f^0 is defined on Ω and satisfies the "tangential Cauchy–Riemann equations", meaning the equations induced by $\bar{\partial}$ on Ω.

Lewy also showed that we do not need f^0 to be defined on all of Ω. If Ω is strictly convex at a point $p^0 \in \Omega$ then there is a local extension of f^0 which is defined on the convex hull of Ω near p^0 and satisfies the Cauchy–Riemann equations.

In Section 9.2 we extend Lewy's theorems to solutions of general systems of the form (9.1).

(2) *Edge-of-the-wedge theorem.* Let f^0 be a function on $\mathbb{R}^n$, which is a Cauchy surface for the Cauchy–Riemann system $\bar{\partial}$. Suppose f^0 is common Cauchy Data on $\mathbb{R}^n$ for solutions f_j of the Cauchy–Riemann equation $\bar{\partial} f_j = 0$ on domains $\mathbb{R}^n + \Gamma_j$ where Γ_j is a proper convex cone in imaginary space. Then f^0 extends to (is the Cauchy data for) a solution f of $\bar{\partial} f = 0$ on

$$\mathbb{R}^n + \{\text{convex hull of } \bigcup \Gamma_j\}.$$

This result, as well as a local counterpart were discussed in Section 5.2.

The separate analyticity theorem of Hartogs and Bernstein discussed in Section 5.2 is of the same genre. Now f^0 is a function on the cube $\{|x_j| < 1\}$. For each $x_1^0, \ldots, x_{j-1}^0, x_{j+1}^0, \ldots, x_n^0$ fixed f^0 extends to a holomorphic function in the variable $x_j + iy_j$ for $0 < y_j < \delta$ where δ is independent of $x_1^0, \ldots, x_{j-1}^0, x_{j+1}^0, \ldots, x_n^0$. Then f^0 extends to a holomorphic function f in a set of the form: x arbitrary, y in the positive orthant (locally).

For the purposes of generalization to systems (9.1) let us make the following observation: the separate analyticity hypothesis is that f^0 extends to a function $\tilde{f}^0$ (locally) on

$$[\mathbb{R}^n + i\{y_1\}] \cup [\mathbb{R}^n + i\{y_2\}] \cup \cdots \cup [\mathbb{R}^n + i\{y_n\}] \tag{9.2}$$

where $i\{y_j\}$ is the (positive) j-th imaginary axis. Moreover $\tilde{f}^0$ is holomorphic on (9.2); that is, it is as holomorphic as it can be (satisfies the induced Cauchy–Riemann equations).

Note that the set (9.2) is obtained from the Cauchy surface $\mathbb{R}^n$ by forming the union of several one-dimensional extensions $\mathbb{R}^n + i\{y_j\}$ of $\mathbb{R}^n$. The conclusion is that f^0 extends to a solution on an open set. This puts the separate analyticity theorem in the framework of general systems (9.1). Let f^0 be "potential Cauchy data" on a (linear) Cauchy surface X for (9.1). Then there are suitable half-lines L_j^+ so that if f^0 is the Cauchy data for a solution of (9.1) on $\bigcup(X + L_j^+)$ then f^0 is the Cauchy data of a solution f of (9.1) on all of $X + \sum L_j^+$. (A solution on a lower dimensional manifold M is an element of the kernel of all operators P lying in the left ideal in the ring of differential operators with variable coefficients generated by the P_j such that P is also in the left ideal generated by the tangential derivatives to M.) This is generalized in Proposition 9.14ff.

There is another way of regarding the separate analyticity theorem. We call $\tilde{f}_j^0$ the restriction of $\tilde{f}^0$ to $[\mathbb{R}^n + i\{y_j\}]$. We extend $\tilde{f}_j^0$ to $\mathbb{C}^n$ (locally) by making it constant in the variables $y_1, y_2, \ldots, y_{j-1}, y_{j+1}, \ldots, y_n$. We call the resulting function f_j. It satisfies the system of equations

$$\frac{\partial f_j}{\partial y_1} = \cdots = \frac{\partial f_j}{\partial y_{j-1}} = \left[\frac{\partial}{\partial x_j} + i\frac{\partial}{\partial y_j}\right] f_j = \frac{\partial f_j}{\partial y_{j+1}} = \cdots = \frac{\partial f_j}{\partial y_n} = 0. \quad (9.3)$$

We denote this system of equations by

$$\vec{P}^j f_j = 0. \quad (9.4)$$

We have arrived at a structure which bears certain similarities and certain dissimilarities to the edge-of-the-wedge theorem.

Similarities. f_j is a solution of the overdetermined system (9.4) on the domain in $\mathbb{R}^{2n}$ which can be taken as the tube over the positive orthant in $\{y\}$. $\mathbb{R}^n$ is a Cauchy surface for this system and all the f_j have the same CD $= f^0$.

Dissimilarity. The systems $\vec{P}^j$ are different.

We begin by treating each system $\vec{P}^j$ individually. This means that we study the CP for the system $\vec{P}^j f = 0$ for solutions f_j in $L + \Gamma_j$ where Γ_j is a cone and L is a linear CS. (For the system (9.3) $L = \mathbb{R}^n$ and the cone can be taken as either $\{y_{j'} \geq 0\}_{\text{all } j'}$, or as $y_j \geq 0$ and $y_{j'}$ in a full neighborhood of 0, for $j' \neq j$.) Our goal is to find necessary conditions on the wave front set (see Section 5.3) of a (vector) function $\vec{f}^0$ on L to be the CD of a (local) solution f_j of $\vec{P} f_j = 0$ on $L + \Gamma_j$.

If the same $\vec{f}^0$ is the CD for several systems $\{\vec{P}^j\}$ then each $\vec{P}^j$ puts restrictions on the wave front set of $\vec{f}^0$. If all these conditions imply that the analytic wave front set of $\vec{f}^0$ is empty then $\vec{f}^0$ is locally holomorphic. If L is noncharacteristic for each $\vec{P}^j$ then by the Cauchy–Kowalewski theorem a locally holomorphic $\vec{f}^0$ extends to solutions f^j of $\vec{P}^j f^j = 0$ for each j on a full neighborhood of zero (Theorem 9.14). This can be regarded as the edge-of-the-wedge theorem for partial differential equations. (It is not as strong as

the usual edge-of-the-wedge theorem for $\bar{\partial}$ because, in that case, there is an extension theorem from (9.2) with each $y_j \geq 0$ to $\{(x, y)\}$ with $y \in$ positive orthant; while in the present description we start with y_j both positive and negative.)

Remark. *In our study of this problem we shall give a meaning to the concept "wave front set of a vector function." When this wave front set is empty all the components of the vector are locally holomorphic.*

The separate analyticity theorem fits into the Radon ansatz. The analytic extendability depends on properties of f^0 on the lower dimensional subsets $x_{j'} =$ const. for $j' \neq j$.

For another illustration of this form of the Radon ansatz let f^0 be a function defined in the exterior of the unit ball B in $\mathbb{R}^n$. Suppose the restriction of f^0 to any line L tangent to B is the CD of a function which is entire on the complexification of L. Then f^0 extends to an entire function on $\mathbb{C}^n$ [4, 5].

We shall treat another example in Section 9.5. Let $x = (x_1, \ldots, x_n)$ and let y be a single variable. Let $f(x, y)$ be a function which is defined and smooth in the closed cylinder $|x| \leq 1$. Let P be an elliptic operator of order $2m$ with smooth coefficients. Suppose the DD defined by f for P on each sphere S_y centered at $(0, y)$, radius 1, is the DD of a solution F_y of $PF_y = 0$ in the ball B_y bounded by S_y such that F_y is "tangent to f" on S_y, meaning that the first $m+1$ normal derivatives of f and F_y coincide on S_y. Then $Pf = 0$.

When $n = 1$ a stronger result is valid for holomorphic functions. The holomorphic case represents the origin of the general problem (see [1, 64]).[1]

Similar results apply to the noncharacteristic CP. In the characteristic case we must consider solutions which are global in the variables of L; we must also impose bounds at infinity on $\vec{f}^0$.

To formulate this precisely, let us divide the variables into x, y. x is a variable on $L = \mathbb{R}^n$ and y is a variable in an open cone Γ. We now study solutions of $\vec{P}f = 0$ on the tube $L + \Gamma$ (y can be local). We impose growth conditions on f of the form

$$f^{(j)}(x, y) = \mathcal{O}\left(e^{\Phi(cx)}\right) \tag{9.5}$$

where Φ is a smooth convex function and $f^{(j)}$ represents a derivative of f of any order; the constant in the $\mathcal{O}$ depends on j.

We are thus led to Φ wave front sets and microglobal analysis as discussed in Section 5.3. These ideas are developed in Section 9.3.

Up to now we have been concerned with actual solutions and actual assumption of the PD, meaning the functions f are smooth up to the PS. We can weaken the concepts of solution and of attaining the boundary data in several ways.

[1] The probelm was solved by Turmanov [148].

(1) *f is a solution of $\vec{P}f = 0$ on the interior of $L + \Gamma$ but f is not assumed to be smooth in $L + \bar{\Gamma}$.* Instead we assume that

$$\partial f(x, y) = \mathcal{O}\left(e^{\xi(y)}\right) \tag{9.6}$$

uniformly for x in compact sets. ∂ is a differentiation of arbitrary order and ξ is a function which does not approach ∞ "too rapidly" as $y \to 0$.

Under suitable circumstances we can make sense of

$$\vec{f}^0 = \lim_{y^0 \to 0} \{\text{CD } (f) \quad \text{on } y = y^0\}. \tag{9.7}$$

Then we can study the wave front set of $\vec{f}^0$ as before (see Section 9.4).

(2) *f is smooth in $L+\bar{\Gamma}$ but f is not assumed to be in the kernel of $\vec{P}$.* Instead

$$\vec{P}f \to 0 \quad \text{on } L. \tag{9.8}$$

When (9.8) holds we say *f is a solution of $\vec{P}f = 0$ on L to infinite order*, or f is an *asymptotic solution of $\vec{P}f = 0$.*

The expression "$\to 0$ on L" in (9.8) takes many different forms. One natural interpretation is

$$\partial_y \vec{P} f(x, y) \to 0 \quad \text{on } L$$

in the topology of $\mathcal{E}(L)$ for any

$$\partial_y = \frac{\partial^m}{\partial y_1^{m_1} \dots \partial y_l^{m_l}}. \tag{9.9}$$

We refer to this interpretation of (9.8) as *strong asymptotic solution.*

Strong asymptotic solutions satisfy

$$\partial_{x,y} \vec{P} f = \mathcal{O}(|y|^N) \quad y \to 0 \tag{9.10}$$

for any N and any $\partial_{x,y}$. The constants in $\mathcal{O}$ depend on $\partial_{x,y}$ and are uniform for compact x sets. Since differentiation is continuous on the space $\mathcal{E}(L)$ the equivalence of (9.10) with strong asymptoticity is readily established.

Inequality (9.10) is rooted in a C^∞ theory. There is an analogous question that can be studied for Denjoy–Carleman classes (Section 5.3). We replace (9.10) by

$$\partial_{x,y} \vec{P} f = \mathcal{O}(e^{-\xi(y)}) \tag{9.11}$$

with ξ as in (9.6). For example, $\xi(y) = y^{-1}$ is related to the Gevrey 2 class. More precisely, for suitable $\vec{P}$ inequality (9.11) implies a restriction on a Denjoy–Carleman wave front set of the CD of f. The particular Denjoy–Carleman class depends uniquely on ξ and $\vec{P}$.

One might think that (9.11) should imply that f can be split in the form

$$f = h + \phi \tag{9.12}$$

where $\vec{P}h = 0$ on $L + \Gamma$ and ϕ and *all* its derivatives are $\mathcal{O}[\exp(-\xi(y))]$ as $y \to 0$. However, such a splitting is impossible because (9.12) implies

$$\mathrm{CD}\,(f) = \mathrm{CD}\,(h).$$

But we have noted that (9.11) implies a restriction on a nonanalytic Denjoy–Carleman wave front set for $\mathrm{CD}\,(f)$ (unless $\xi \equiv \infty$ near $y = 0$) while for $\mathrm{CD}\,(h)$ restriction is put on the analytic wave front set. (All this is for $\vec{P}$ elliptic.)

In our study of asymptotic solutions, L is noncharacteristic for $\vec{P}$. In this situation, instead of requiring (9.10) or (9.11) for all $\partial_{x,y}$, we can posit the weaker criterion

$$\mathrm{CD}^*(\vec{P}f) = \mathcal{O}(e^{-\xi(y)}) \tag{9.11*}$$

uniformly on compact sets of x. (Sometimes we add the condition $\partial_x \mathrm{CD}^*(\vec{P}f) = \mathcal{O}(e^{(-\xi(y))})$ for all ∂_x of order 1 and 2 for convenience.) (9.11*) is referred to as *asymptotic solution of order* $\exp(-\xi)$.

$\mathrm{CD}^*(\vec{P}f)$ is defined by a complicated system of "normal derivatives." When $\vec{P}$ is first order then $\mathrm{CD}^*(\vec{P}f) = \mathrm{CD}\,(\vec{P}f)$ is the value of $\vec{P}f$. For $\vec{P} = P$ a single operator, which is the only higher degree case considered here, we can usually set $\mathrm{CD}^*\,(Pf) = \mathrm{CD}\,(Pf)$.

We can view the concept of "vanishing to infinite order" in another way.

Let Ω be a smooth n-dimensional surface in $\mathbb{R}^{n+l}$. Locally we can introduce coordinates $(\omega, \vec{N})$ in a tubular neighborhood of Ω with ω a coordinate on Ω and $\vec{N} = (N_1, \ldots, N_l)$ the normal coordinate. If $F(x)$ is a C^∞ function defined near Ω then we can define the *jet of infinite order* of F on Ω as

$$J(F) = \left\{ \frac{\partial^m}{\partial \vec{N}^m} F \Big|_\Omega \right\}. \tag{9.13}$$

Here

$$\frac{\partial^m}{\partial \vec{N}^m} = \frac{\partial^m}{\partial N_1^{m_1} \ldots \partial N_l^{m_l}}.$$

$J(F)$ is a set of functions $F^{(m)} \in C^\infty(\Omega)$.

If Ω is a hypersurface then this definition makes sense if F is only defined and C^∞ on one side of Ω. If codim $\Omega > 1$ then we can make the same definition if at each $\omega^0 \in \Omega$ there are cones Γ_ω contained in the normals to Ω at points $\omega \in \Omega$ near ω^0 such that F is defined and C^∞ on $\bigcup \Gamma_\omega$. We assume the Γ_ω depend smoothly on ω.

Given any jet $\{F^{(m)}\}$ the formal power series

$$F_*(x) = \sum \frac{F^{(m)}(\omega)\vec{N}^m}{m!} \tag{9.14}$$

satisfies

$$J(F_*) = \{F^{(m)}\}.$$

"Vanishing to infinite order" on Ω means $\mathrm{CD}\,(F_*) \equiv 0$.

In case $n = 0$, $l = 1$, Ω is a point, $\{F^{(m)}\}$ is a sequence of complex numbers. A theorem of É. Borel asserts that there exists an $F \in \mathcal{E}$ with $J(F) = \{F^{(m)}(\Omega)\}$. Otherwise stated the map

$$F \to \sum \frac{F^{(m)}(\Omega)N^m}{m!} \tag{9.15}$$

from $\mathcal{E}$ into the space $\mathcal{E}^\infty(\Omega)$ of formal power series at Ω is surjective.

This result is studied in Chapter 13 of FA; it is extended to $l > 1$ with Ω still a point. Essentially the same method applies in case Ω is a linear subvariety of $\mathbb{R}^{n+l}$. By taking a partition of unity we can extend the result to arbitrary smooth Ω.

Proposition 9.1 *For any smooth Ω the map (9.15) of $\mathcal{E}$ into $\mathcal{E}^\infty(\Omega)$ is surjective.*

We refer to Proposition 9.1 as the "general Borel theorem."

We shall discuss Proposition 9.1 as well as analogs for Denjoy–Carleman classes in Section 9.4 (see (9.96)ff.).

(3) *Instead of positing $P_j f = 0$ for all j we can impose inequalities of the form*

$$|(P_1^{m_1} P_2^{m_2} \dots P_r^{m_r} f)(x)| \le c^{|m|+1} A_m$$

where $|m| = m_1 + \cdots + m_r$ and $\{A_m\}$ is a given sequence. The study of the uniqueness of the CP for such classes of functions bears a close similarity to the study of quasianalytic and nonquasianalytic classes. It is carried out in Chapter 13 of FA.

9.2 Hartogs–Lewy extension

Unless explicitly stated to the contrary we assume in this section that $\vec{P}(D)$ has constant coefficients.

Let us begin by analyzing the proof of the general Hartogs' theorem for $\vec{P}(D)$ as given in [53]. Let Ω be a compact convex smooth hypersurface in $\mathbb{R}^n$ and let f^0 be defined and satisfy $\vec{P}(D)f^0 = 0$ on some interior neighborhood of Ω. We extend f^0 to a function g on all of $\Delta =$ interior Ω and call

$$h_j = P_j(D)g. \tag{9.16}$$

The h_j are of compact support $\subset$ interior Ω and satisfy

$$P_j h_k = P_k h_j. \tag{9.17}$$

Moreover

$$\text{support}\, h_k \subset \text{interior}\, \Omega.$$

By Fourier transform (9.14) becomes

$$\hat{P}_j \hat{h}_k = \hat{P}_k \hat{h}_j. \tag{9.18}$$

We set

$$\hat{u} = \hat{h}_k / \hat{P}_k. \tag{9.19}$$

By (9.18) this definition is independent of k.

Now suppose that $\vec{P}$ is truly overdetermined, meaning that the $\hat{P}_j$ have no common factor. Then the complex algebraic variety of common zeros of the $\hat{P}_j$ has codimension > 1. By (9.19) this means that $\hat{u}$ is holomorphic off a set of complex codimension > 1. It is standard that this implies that $\hat{u}$ is entire.

Equations (9.17) represent a part of the relations amongst the P_j. (They are only part, e.g. if $\vec{P} = (P_1, P_2, P_3)$ where P_1 and P_3 have no common factor and $P_2 = P_1 P_3$ then (9.18) contains the relation $P_1 h_2 = P_2 h_1$ but does not contain the relation $h_2 = P_3 h_1$.)

We denote by $\boxed{P^1}$ a matrix defined by a set of generators for the *module of relations* of the P_j. ($\vec{P}$ is considered as a column matrix so a relation is a row vector $\vec{P}^1_l$ satisfying $\vec{P}^1_l \cdot \vec{P} = 0$. The module of relations is finitely generated.) In general we call $\boxed{P^k}$ a matrix whose rows form a set of generators for the rows of $\boxed{P^{k-1}}$ (we have set $\boxed{P^0} = \vec{P}$).

The $\boxed{P^k}$ define a complex which is a resolution of $\vec{P}$. We call this the *syzygy or Koszul complex*; we denote it by $Z(\vec{P})$. In case $\vec{\hat{P}}$ is strongly free, meaning that the codimension of the variety V of common zeros of the P_j is r (see Section 3.1), we showed that the first r cohomology groups of the complex vanish.

We call $\vec{P}$ "k free" if

$$H^j(Z(\vec{P})) = 0 \quad j = 1, \dots, k. \tag{9.20}$$

The above argument shows that if codim $V > 1$ then the first cohomology group of the syzygy complex in the space of entire functions vanishes.

Finally the standard division inequalities of [34] and [119] show, since Δ is convex, that

$$\text{support}\, u \subset \text{interior}\, \Omega. \tag{9.21}$$

This means we have exactness of the Koszul (syzygy) complex at the first step when acting on $\hat{\mathcal{E}}'(\Delta)$.

We set

$$f = g - u. \tag{9.22}$$

(9.21) and the definition of g show that f is an extension of f^0 to the interior of Ω; (9.19) and (9.16) show that $\vec{P}(D)f = 0$ so f is the desired extension.

Now, we do not really need to assume that f^0 is defined on a whole interior neighborhood of Ω. Suppose only that f^0 is a jet (e.g. C^∞) on Ω of order ∞. As above, we identify C^∞ jets of infinite order on Ω with formal power series in the normal to Ω with coefficients which are C^∞ functions on Ω; upon thinking of f^0 as a formal power series it makes sense to form $\vec{P}f^0$. Suppose $\vec{P}f^0 = 0$. By the general Borel theorem there is a C^∞ function g on $\mathbb{R}^n$ whose jet $J(g) = f^0$. The vanishing of $\vec{P}f^0$ as a jet means that $\vec{P}g$ vanishes to infinite order on Ω.

We define h_j by (9.16) in the interior of Δ and extend h_j to be zero outside Δ. Equation (9.18) holds so again $\hat{u}$ as defined by (9.19) is entire. Of course the support of u can now be all of Δ. But the fact that $P_j f^0 = 0$ means that the functions h_j are of class C^∞ on all of $\mathbb{R}^n$. The division inequalities give $u \in C^\infty$ and vanishes outside Δ.

Hence

$$\begin{aligned} f = g - u \equiv f^0 \quad &\text{to infinite order on } \Omega \\ \vec{P}(D)f = 0 \quad &\text{in interior } \Delta \end{aligned} \tag{9.23}$$

so, as before, we have an extension.

We can go further. Our first illustration includes local extension.

Let Ω_0 be a "cap." This means that there is a bounded domain Δ with

$$\Omega = \text{bd } \Delta = \Omega_0 \cup \Omega_1. \tag{9.24}$$

Ω_0 is convex and

(α) Ω_0 contains a point p which does not lie in the closure of hull (Ω_1), *the convex hull of Ω_1 (Figure 9.1).*

Suppose f^0 is a smooth jet of infinite order on Ω_0 satisfying $\vec{P}(D)f^0 = 0$ on Ω_0. We extend f^0 to a C^∞ function g which is defined and of compact support

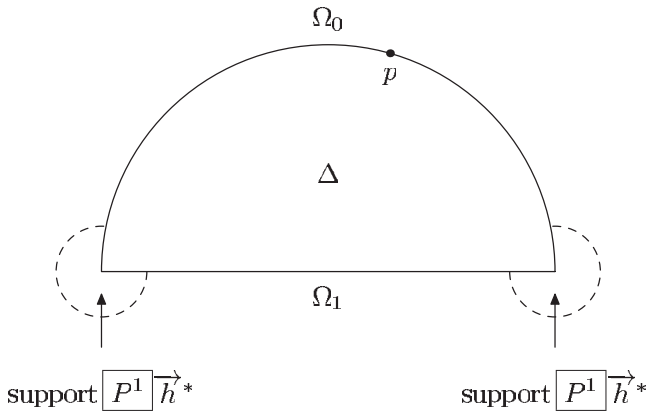

Figure 9.1

on all of $\mathbb{R}^n$. We could choose g to have its support in an ϵ neighborhood of Ω_0. Then

$$\vec{h} = \vec{P}(D)g \equiv 0 \qquad \text{to infinite order on } \Omega_0.$$

We know that $\boxed{P^1}\vec{h} = 0$. We modify $\vec{h}$ on $\mathbb{R}^n$ to be zero outside Δ near Ω_0 except near $\Omega_0 \cap \Omega_1$. Since $\vec{h}$ vanishes to infinite order on Ω_0 this modification $\vec{h}^*$ of $\vec{h}$ has the property

$$\text{support}\boxed{P^1}\vec{h}^* \subset \epsilon \text{ neighborhood of } \Omega_0 \cap \Omega_1 \text{ outside } \Delta. \tag{9.25}$$

We seek an $\vec{h}_1$ satisfying

$$\begin{aligned} \text{support}\,\vec{h}_1 &\subset \text{hull}\,[\text{support}\boxed{P^1}\vec{h}^*] \\ \boxed{P^1}\vec{h}_1 &= \boxed{P^1}\vec{h}^*. \end{aligned} \tag{9.26}$$

Once we have such an $\vec{h}_1$ we can set

$$\vec{h}_2 = \vec{h}^* - \vec{h}_1 \tag{9.27}$$

so

$$\begin{aligned} \boxed{P^1}\vec{h}_2 &= 0 \\ \vec{h}_2 &= \vec{h}^* \quad \text{near } p. \end{aligned} \tag{9.28}$$

We can now proceed as before using $\vec{h}_2$ in place of $\vec{h}$. We showed in (9.16)ff. that $\boxed{P^1}\vec{h}_2 = 0$ implies that we can write $\vec{h}_2 = \vec{P}u$. $f = g - u$ satisfies $\vec{P}f = 0$ inside Δ near p because

$$\begin{aligned} \vec{P}f &= \vec{P}g - \vec{h}_2 \\ &= \vec{P}g - \vec{h}^* + \vec{h}_1. \end{aligned} \tag{9.29}$$

$\vec{h}_1$ vanishes near p and $\vec{P}g = \vec{h}^*$ inside Δ near p.

We remain with two problems:

(1) Show that u vanishes to infinite order on Ω near p.
(2) Show that $\vec{h}_1$ exists.

(1) We must be careful because $\vec{h}_2$ does not vanish to infinite order on all of Ω; rather it vanishes to infinite order only on the part of Ω_0 which is away from hull (support $\vec{h}_1$). The function u defined as the function of compact support satisfying $\vec{P}u = \vec{h}_2$ has support in hull (support $\vec{h}_2$) by Paley–Wiener theory. The geometric condition (α) is not strong enough. We need

(β) The convex hull of $\Omega_0 \cup$ (ϵ neighborhood of $\Omega_0 \cap \Omega_1$) agrees with the convex hull of Ω_0 near p.

Thus support $u \subset \Delta$ near p. Since $\vec{h}_2 \in C^\infty$ so is u. Thus u vanishes to infinite order on Ω near p.

(2) It remains to prove the existence of $\vec{h}_1$ satisfying (9.26). Note that $\boxed{P^2}\,\boxed{P^1}\vec{h}^* = 0$ where $\boxed{P^2}$ represents generators of the module of relations of the rows of $\boxed{P^1}$. We need

(γ) The second cohomology group of the syzygy complex with coefficients in $\hat{\mathcal{E}}'(K)$ vanishes whenever K is a smooth bounded convex domain which is the closure of its interior.

In Section 3.1 we conjectured that this cohomology group vanishes if $\vec{P}$ is 2 free.

Let us investigate the case when Ω_0 is not convex; in this case we need another concept.

We assume for simplicity that $\Omega = \Omega_0 \cup \Omega_1$ as before and that Ω_0 is convex except possibly near some point p which does not lie in the convex hull of Ω_1 (Figure 9.2). We can proceed exactly as before (assuming (α) and (β)) to define u and f. Since $\vec{P}u = \vec{h}_2$ we know that $\vec{P}u$ vanishes outside Ω_0 near p. But, from Paley–Wiener theory, u may not vanish outside Δ near p; rather u vanishes outside $\operatorname{bd}\operatorname{hull}\Omega_0$ near p.

To go from $\operatorname{bd}\operatorname{hull}\Omega_0$ to $\operatorname{bd}\Omega_0$ we need "unique continuation":

(δ) $\vec{P}$ has $(\Omega, \operatorname{hull}(\Omega))$ unique continuation at p, meaning whenever $u \in C^\infty$ is a solution of $\vec{P}(D)u = 0$ outside Ω near p and u vanishes outside hull(Ω) near p then u vanishes outside Ω near p.

We have proven

Theorem 9.2 *Let Ω be a smooth bounded hypersurface. Suppose the first two cohomology groups of the syzygy complex vanish and the geometric conditions (α), (β) hold and that $\vec{P}$ has $(\Omega, \operatorname{hull}(\Omega))$ unique continuation at $p \in \Omega$. If f^0 is a smooth C^∞ jet of infinite order on Ω_0 and $\vec{P}(D)f^0$ vanishes on Ω_0 except on $\Omega_0 \cap \Omega_1$, then there is a solution f of $\vec{P}f = 0$ on Δ near p which agrees with f^0 near p.*

Remark. *If Ω_0 is convex near p then the unique continuation hypothesis is vacuous.*

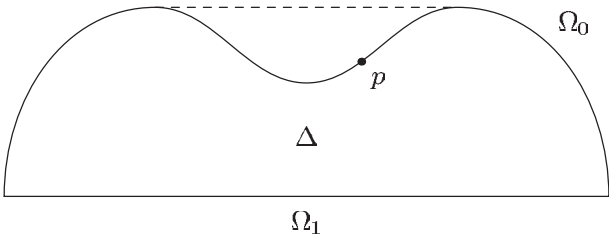

Figure 9.2

An interesting example related to Theorem 9.2 is Ω = cylinder. We use variables $x = (x_1, \dots, x_n), y = (y_1, \dots, y_l)$. Ω is defined by $y \in \Omega_0$ where Ω_0 is the boundary of a smooth, strictly convex domain Δ_0.

Theorem 9.2 does not apply to this situation because Δ is not bounded. If $n = 1$ we could cut Ω for $|x| = a$ and apply Theorem 9.2. Assuming we had uniqueness we could let $a \to \infty$ and obtain an extension to Δ.

Problem 9.1 *Derive an analog of Theorem 9.1 for cylinders.*

Let us examine Problem 9.1 for jets f^0 on Ω which are small at infinity. Suppose there exists an extension g as above with

$$g \in \mathcal{E}(y) \otimes \mathcal{S}(x). \tag{9.30}$$

This means that $g \in C^\infty(y, x)$ and g and its derivatives are polynomially small at $x = \infty$. If $\vec{P} f^0 = 0$ then

$$\vec{P} g = \vec{h},$$

when looked at from inside Ω, vanishes to infinite order on Ω and so can be continued as a C^∞ function on $\mathbb{R}^{n+l}$ vanishing outside Ω; that is, $\vec{h} \in \mathcal{E}_0(\Delta_0) \otimes \mathcal{S}$ where $\mathcal{E}_0(\Delta_0)$ is the space of C^∞ functions of y which vanish outside Ω_0.

We can take the Fourier transform of $\vec{h}$ in all variables. Now $\hat{\vec{h}} \in \hat{\mathcal{E}}_0 \otimes \mathcal{S}$ and satisfies (9.18).

Theorem 9.3 *Assume that for any fixed* $\hat{x}$

$$V_{\hat{x}} = \{P_j(i\hat{y}, i\hat{x}) = 0\} \quad j = 1, 2, \dots, r \tag{9.31}$$

has codimension > 1. *Then any infinite order* C^∞ *jet* f^0 *on* $\Omega_0 \times \{x\}$ *satisfying* $\vec{P} f^0 = 0$ *which has an extension* $g \in \mathcal{E}(y) \otimes \mathcal{S}(x)$ *has an extension* $f \in \mathcal{E}(\Delta_0) \otimes \mathcal{S}(x)$ *satisfying* $\vec{P}(D) f = 0$.

To prove Theorem 9.3 we define $\hat{u}$ as in (9.17). For each fixed $\hat{x}$ we have $\hat{u}(\hat{y}, \hat{x}) \in \hat{\mathcal{E}}_0(\Delta_0)$ as before. It is easy to obtain estimates on derivatives of $\hat{u}$ and on products of $\hat{u}$ with polynomials in $\hat{x}$. It follows that $\hat{u} \in \tilde{\mathcal{E}}(\Delta_0) \otimes \mathcal{S}$ so that $f = g - u$ is the desired extension of f^0.

Problem 9.1 can be reformulated as follows:

$$\textit{Is Theorem } 9.3 \textit{ valid if } g \in \mathcal{E}(y) \otimes \mathcal{E}(x)?$$

Finally we wish to discuss what happens when f^0 is defined on a lower dimensional part Ω_0 of boundary Δ. A jet on Ω_0 involves derivations in the multidimensional normal. If $\dim_{\mathbb{R}} \Omega_0 < \dim_{\mathbb{C}} V$ then we are in a WP-like situation. The given data on Ω_0 must satisfy growth conditions in the order of normal derivatives. (The Korevaar–Wiegerinck theorem is of this form.)

In this section we shall assume that

$$\dim_{\mathbb{R}} \Omega_0 \geq \dim_{\mathbb{C}} V.$$

To illustrate the problems that might arise, suppose $r = n$ and $\vec{P}$ is strongly free. Thus dim $V = 0$ and we might choose dim $\Omega_0 = 0$. Let Ω_0 consist of two points p, q and let Δ be a closed strictly convex set whose boundary contains p and q. f^0 consists of two formal power series, at p and at q. We could choose g to be a C^∞ function defined in a neighborhood of Δ whose formal power series at p and q agree with f^0 and $\vec{P}g$ vanishes to infinite order at p, q. Our methods give nothing new. (This is no surprise because it is obvious that there is, generally, no solution of $\vec{P}f = 0$ on all of Δ which extends f^0.) The reason is that the line joining p, q has dimension 1; we cannot have two constant coefficient "independent" operators (i.e. vanishing of first cohomology of the syzygy complex) on a line.

For the next example let Δ be the closed unit cube in $\mathbb{R}^n$ and let B_s be its skeleton of codimension s. f^s is a given formal C^∞ jet on B_s and g^s is a C^∞ extension of f^s to Δ. This means that g^s is a C^∞ function on a neighborhood of Δ whose formal power series jet on B_s agrees with f^s. Actually the definition of f^s at the corners of B_s is somewhat complicated so, in this heuristic argument, we use the existence of g^s to define the compatibility at the intersection of the faces of B_s.

Suppose f^s is in the kernel of $\vec{P}(D)$, meaning, as above, that g^s can be chosen so that $\vec{P}g^s$ vanishes to infinite order on B_s. We want to extend f^s to a (unique) solution on all of Δ. We begin with a formal argument.

Suppose that V has codimension $> s$. Let A_{s-1} be a face of Δ of codimension $s-1$. It seems reasonable that $\vec{P}(D)$ should induce an overdetermined system of constant coefficient partial differential equations on the restrictions of solutions of $\vec{P}(D)$ to A_{s-1}.

By our assumption f^s defines a solution of $\vec{P}(D)$ on the boundary of A_{s-1}. Hence we can apply Theorem 9.2 to extend f^s to a unique solution on A_{s-1}. Proceeding in this fashion we extend f^s to all of B_{s-1} and, eventually, to Δ.

Now for the details.

We begin with an explanation of the CP for the strongly free system $\vec{P}(D) = (P^1, \dots, P^r)$ using linear CSs. (A detailed study of this CP is found in Chapter IX of FA.)

We begin with a compact convex polyhedron Δ. By this we mean that all faces of Δ are convex polyhedra and the intersection of two faces of codimension s is a face of codimension $s+1$.

The CP for a single operator $P^1 = P^1(D)$ is defined by a hypersurface H and a nontangential direction at each point of the hypersurface. This nontangential direction t defines the "normal derivatives" for the CP for P^1, meaning that the CD of a function f is

$$\mathrm{CD}\,(f) = \left(f\big|_H, \left.\frac{\partial f}{\partial t}\right|_H, \dots, \left.\frac{\partial^{l-1} f}{\partial t^{l-1}}\right|_H \right)$$

where $l =$ order P^1. In dealing with Δ we have several CSs for P^1, namely the hypersurfaces H which are the hyperplanes $L^1_{j_1}$ ("hyperplane" entails passing

through the origin) whose translates contain the faces $A^1_{j_1}$ of the skeleton of codimension 1 of Δ. It is important that we choose a single direction for all the hyperplanes. We use coordinates x_{j_1} on $L^1_{j_1}$ and t_1 in the given nontangential direction. We call y_{j_1} the orthogonal to $L^1_{j_1}$.

Let $l_1 =$ degree P^1. We have seen on several occasions that the CP is dual to Lagrange interpolation. We wish to make that relation more precise.

Let $P^1 f = 0$. By the fundamental principle (Section 1.4) f has a formal Fourier representation

$$f(x_{j_1}, y_{j_1}) = \sum_k \int \exp i[x_{j_1} \cdot \hat{x}_{j_1} + y_{j_1} \cdot \hat{y}^k_{j_1}(\hat{x}_{j_1})]\, \hat{f}(\hat{x}_{j_1}, \hat{y}^k_{j_1})\, d\hat{x}_{j_1}. \tag{9.32}$$

$\hat{y}^k_{j_1}$ are the roots of $P^1(\hat{x}_{j_1}, \hat{y}_{j_1}) = 0$ so $V^1 = \{(\hat{x}_{j_1}, \hat{y}^k_{j_1})\}$ is the variety of zeros of P^1. We assume these roots, and similar sets of roots that we come in contact with later, are distinct. In FA it is shown that the change from distinct roots to nondistinct roots does not involve any essential difficulties.

We can write[2]

$$\frac{\partial^m f}{\partial t_1^m}(x_{j_1}, 0) = \sum_k \int \exp(i x_{j_1} \cdot \hat{x}_{j_1}) \hat{t}^m_1(\hat{x}_{j_1}, \hat{y}^k_{j_1})\, \hat{f}(\hat{x}_{j_1}, \hat{y}^k_{j_1})\, d\hat{x}_{j_1}. \tag{9.33}$$

$\hat{t}_1$ is a linear function of $\hat{x}_{j_1}, \hat{y}_{j_1}$ which involves $\hat{y}_{j_1}$ nontrivially because t_1 is not tangent to $L^1_{j_1}$. For each $\hat{x}_{j_1}$ the monomials $\hat{t}^m_1(\hat{x}_{j_1}, \hat{y}^k_{j_1})$ form a basis, the *Lagrange basis*, for functions on the zero-dimensional variety

$$V^1(\hat{x}_{j_1}) = \{(\hat{x}_{j_1}, \hat{y}^k_{j_1})\}.$$

This is the essence of Lagrange interpolation: Any function on a finite set of points in the complex plane, such as $\{\hat{y}^k_{j_1}(\hat{x}_{j_1})\}$, has a unique interpolation in terms of the monomials $\{\hat{y}^m_{j_1}\}_{m=0,\dots,l_1-1}$; $l_1 =$ degree P^1. Since $\hat{t}_1$ involves $\hat{y}_{j_1}$ nontrivially we can replace $\{\hat{y}^m_{j_1}\}$ by $\{\hat{t}^m_1\}$.

Put in other terms, given a function $\hat{h}$ on $V^1(\hat{x}_j)$ we can think of the values $\hat{h}(\hat{y}^k_{j_1})$ as the expression of $\hat{h}$ in terms of the δ *basis*, meaning $\{\delta_{\hat{y}^k_{j_1}}\}$ or, more precisely, $\{\delta_{(\hat{x}_{j_1}, \hat{y}^k_{j_1})}\}$. (Actually this basis is in the dual space but this is an unimportant point.) CD (f) is the Fourier transform in $\hat{x}_{j_1}$ of

$$\sum_k \hat{f}(\hat{x}_{j_1}, \hat{y}^k_{j_1}) \delta_{\hat{y}^k_{j_1}(\hat{x}_{j_1})} \hat{t}^m_1.$$

Since, on $V^1(\hat{x}_{j_1})$,

$$\hat{t}^m_1 = \sum \hat{t}^m_1(\hat{y}^k_{j_1}(\hat{x}_{j_1})) \delta_{\hat{y}^k_{j_1}(\hat{x}_{j_1})}$$

[2]We shall make calculations modulo powers of i. The reader can easily supply the correct powers.

the Vandermonde-like matrix $[t_1^m(\hat{y}_{j_1}^k)]_{mk}$ represents the (inverse) change from the δ basis to the Lagrange basis.

If we apply $P^2(D)$ to (9.32) we obtain

$$P^2(D)f(x_{j_1}, y_{j_1}) = \sum \int \exp i[x_{j_1} \cdot \hat{x}_{j_1} + y_{j_1} \cdot \hat{y}_{j_1}^k]\, P^2(\hat{x}_{j_1}, \hat{y}_{j_1}^k)\, \hat{f}(\hat{x}_{j_1}, \hat{y}_{j_1}^k)\, d\hat{x}_{j_1} \tag{9.34}$$

so that

$$\frac{\partial^m}{\partial t_1^m}[P^2(D)f](x_{j_1}, 0) = \sum \int \exp(i x_{j_1} \cdot \hat{x}_{j_1}) \hat{t}_1^m(\hat{x}_{j_1}, \hat{y}_{j_1}^k) P^2(\hat{x}_{j_1}, \hat{y}_{j_1}^k) \hat{f}(\hat{x}_{j_1}, \hat{y}_{j_1}^k)\, d\hat{x}_{j_1}. \tag{9.35}$$

Now, multiplication by a function is always a diagonal operator in the δ basis. Equation (9.35) shows that $\mathrm{CD}\,[P^2(D)f]$ can be interpreted as the Fourier transform in $\hat{x}_{j_1}$ of the operator of multiplication by $P^2(\hat{x}, \hat{y})$ on functions on V^1, acting on $\hat{f}$, when expressed in the Lagrange basis. Since the system $\vec{P}$ is strongly free P^s is generically nonzero on V^1 so the determinant of this transformation is $\not\equiv 0$. The same holds for $P^p, p \geq 2$.

We denote by $\boxed{P_{j_1}^p(\hat{x}_{j_1})}$ the matrix representing multiplication by P^p in the Lagrange basis on $V^1(\hat{x}_{j_1})$. Thus

$$\boxed{P_{j_1}^p(\hat{x}_{j_1})}\, \widehat{\mathrm{CD}}\,(f) = \widehat{\mathrm{CD}}\,[P^p(D)f]. \tag{9.36}$$

We now pass to the variety V^{12} of common zeros of P^1, P^2. Let $\{L_{j_1 j_2}^2\}$ be the set of linear planes of codimension 2 which are parallel to the faces of codimension 2 of Δ and are contained in $L_{j_1}^1$. We use, as before, x_{j_2} for coordinates on $L_{j_1 j_2}^2$ and y_{j_2} as orthogonal coordinates. If $l_2 =$ degree P^2 then by Bezout's theorem the number of points on $V^{12}(\hat{x}_{j_2})$ is $l_1 l_2$, but some of these points may lie at infinity. If the $L_{j_1}^1, L_{j_2}^2$ are noncharacteristic in the sense described in Section 4.1, then there are no points at infinity. Moreover we can choose a direction t_2 so that the monomials $\{\hat{t}_1^{m_1} \hat{t}_2^{m_2}\}_{m_1 < l_1, m_2 < l_2}$ form a Lagrange basis on $V^{12}(\hat{x}_{j_2})$. The Fourier transform of

$$\sum_k \hat{t}_1^{m_1}(\hat{x}_{j_2}, \hat{y}_{j_2}^k(\hat{x}_{j_2})) \hat{t}_2^{m_2}(\hat{x}_{j_2}, \hat{y}_{j_2}^k(\hat{x}_{j_2})) \hat{f}(\hat{x}_{j_2}, \hat{y}_{j_2}^k(\hat{x}_{j_2}))$$

represents $\mathrm{CD}\,(f)$ on $L_{j_2}^2$. $f \to P^p(D)f$ for $p > 2$ is represented by an $l_1 l_2 \times l_1 l_2$ matrix $\boxed{P_{j_1 j_2}^p(D)}$ with nonvanishing determinant as in (9.36).

This process continues in an obvious fashion.

Conversely, if we are given l_1 C^∞ functions $\vec{h}_{j_1}$ on $L_{j_1}^1$ we can use them to define a solution h of $P^1(D)h = 0$. h is a jet with C^∞ coefficients on $L_{j_1}^1$. We can construct h either by a formal power series argument or by writing $\vec{h}_{j_1}$ as a Fourier transform, then solving for $\hat{f}(\hat{x}_{j_1}, \hat{y}_{j_1}^k)$ by changing from the Lagrange

basis to the δ basis on each $V^1(\hat{x}_{j_1})$, and then applying (9.33). From the latter point of view we see clearly that when $P^1 h = 0$

$$P^p h = 0 \text{ if and only if } \boxed{P^p_{j_1}(D)}\,\vec{h}_{j_1} = 0.$$

The same result is true for CD on $L^s_{j_1 j_2 \ldots j_s}$. We start with $l_1 l_2 \cdots l_s$ functions of the form $\vec{h}_{j_1 \ldots j_s}$ and construct, by iteration of the case $s = 1$, the solution h of $P^1 h = \cdots = P^s h = 0$ with this CD. $h = h_{j_1 \cdots j_s}$ is a jet with C^∞ coefficients on $L^s_{j_1 \cdots j_s}$. We have

$$\vec{P} h = 0 \text{ if and only if } \boxed{P^p_{j_1 \ldots j_s}(D)}\,\vec{h}_{j_1 \ldots j_s} = 0 \text{ for } p > s. \tag{9.37}$$

All this deals with CD on $L^s_{j_1 \ldots j_s}$. The faces of Δ are translates of such linear spaces. The change from linear spaces to translates is reflected in formulas like (9.33) where $\exp(i x_{j_1} \cdot \hat{x}_{j_1})$ becomes $\exp i[x_{j_1} \cdot \hat{x}_{j_1} + y^0_{j_1} \cdot \hat{y}_{j_1}]$ where $y^0_{j_1}$ represents the amount of translation. $y^0_{j_1}$ has no effect on the algebra of transformation from Lagrange basis to δ basis so (9.37) is valid for the CP on the faces of Δ.

There is one significant aspect in which the faces of Δ differ from linear spaces, namely the faces are compact sets with boundary. This does not affect the formal power series solution so long as $\vec{h}_{j_1 \ldots j_s}$ is C^∞ on the closed face. But when two faces, say $A^1_{j_1 \ldots j_s}$ and $A^2_{j_1 \ldots j_s}$, intersect we need a

Compatibility condition. *At each point $u \in A^1_{j_1 \ldots j_s} \cap A^2_{j'_1 \ldots j'_s}$ the punctual formal power series of the solutions h^1, h^2 with respective CD $\vec{h}^1_{j_1 \ldots j_s}, \vec{h}^2_{j'_1 \ldots j'_s}$ are the same.*

The *punctual formal power series* at a point u is the usual formal power series at u.

We are now in a position to state our main result.

Theorem 9.4 *Let us be given (potential) CD $\vec{h}^q_{j_1 \ldots j_s} \in C^\infty$ on the skeleton $\{A^q_{j_1 \ldots j_s}\}$ of Δ of codimension s with $s < r$. Necessary and sufficient conditions for there to be a C^∞ solution h of $\vec{P} h = 0$ with $CD(h) = \vec{h}^q_{j_1 \ldots j_s}$ on each $A^q_{j_1 \ldots j_s}$ are*

(1) *The CD are compatible.*
(2) $\boxed{P^p_{j_1 \ldots j_s}(D)}\,\vec{h}^q_{j_1 \ldots j_s} = 0$ *for any $p > s$ and any q.*

Proof The necessity of (2) is clear; the necessity of (1) is contained in our above construction.

For the sufficiency, let $A_{j_1 \ldots j_{s-1}}$ be a face of codimension $s - 1$ and let $\{A^q_{j_1 \ldots j_{s-1} j'_s}\}$ be the set of boundary codimension s faces. Each $\vec{h}^q_{j_1 \ldots j_{s-1} j'_s}$ defines a formal power series solution of $\vec{P}(D)$ $h^q_{j_1 \ldots j_{s-1} j'_s} = 0$ on $A^q_{j_1 \ldots j_{s-1} j'_s}$. Compatibility says that the h^q fit together to form a C^∞ jet solution $h_{j_1 \ldots j_{s-1}}$ of $\vec{P}(D)$ on boundary $A_{j_1 \ldots j_{s-1}}$.

We call $\vec{h}_{j_1\dots j_{s-1}} = \mathrm{CD}\,(h_{j_1\dots j_{s-1}})$ where the CD of this solution means the CD for the system $(P^1,\dots,P^{s-1})$ on $A_{j_1\dots j_{s-1}}$. $\vec{h}_{j_1\dots j_{s-1}}$ is defined only on $\mathrm{bd}\,A_{j_1\dots j_{s-1}}$. It satisfies the equations

$$\boxed{P^p_{j_1\dots j_{s-1}}(D)}\,\vec{h}_{j_1\dots j_{s-1}} = 0$$

there for $p = s, s+1,\dots,r$, since it is the CD of a solution of $\vec{P}(D)$. The operators $P^p_{j_1\dots j_{s-1}}(D)$ depend only on the variables in $L^{s-1}_{j_1\dots j_{s-1}}$ which are the variables on $A_{j_1\dots j_{s-1}}$. Since $s < r$ there are at least two such operators $\boxed{P^p}$; their construction shows that $\{\det \boxed{P^p}\}$ have no common factor. The argument used to prove Theorem 9.2 applies. (There is no essential difference between matrix and scalar-valued differential operators.) It follows that the $\vec{h}_{j_1\dots j_{s-1}}$ extend to C^∞ solutions of $\{\boxed{P^p}\}$ on all of $A_{j_1\dots j_{s-1}}$.

We have thus reduced the s of the hypothesis of Theorem 9.4 to $s-1$. We have shown that hypothesis (1) is valid. Compatibility presents no problem because on a codimension s face $A^q_{j_1\dots j_s}$ of intersection of two codimension $s-1$ faces $A^{q'}_{j'_1\dots j'_{s-1}}$ and $A^{q''}_{j''_1\dots j''_{s-1}}$ we are given a solution $h^{q'q''}_{j_1\dots j_s}$ of $\vec{P}(D)$ which defines both CD $\vec{h}^{q'}_{j'_1\dots j'_{s-1}}$ and $\vec{h}^{q''}_{j''_1\dots j''_{s-1}}$ on $A^q_{j_1\dots j_s}$.

We can continue the process and obtain a unique solution h of $\vec{P}(D)h = 0$ with the prescribed CD. This completes the proof of Theorem 9.4.

One can regard our treatment of the extension problem as a "cohomology approach"; the proof of Theorem 9.2 works within the framework of extensions which are not solutions and then attempts to "correct" these extensions.

Another approach is to construct a solution H from the function h we wish to extend and then show that H actually extends h.

One example of this method has proved useful, e.g. for holomorphic functions of several complex variables, and goes as follows. (We do not need Δ to be a polyhedron.) Since $\mathrm{codim}\,V = r$, by projecting V on a codimension $r-1$ plane and using the fact that the projection of an algebraic variety is a Zariski open set of an algebraic variety, we can find an operator, say $P^0(D)$, in the ideal of $\{P^p\}$ which depends only on the variables on the codimension $r-1$ plane, say L^{r-1}. Some translates L^{r-1}_t of L^{r-1} intersect boundary Δ in a set of codimension 1 in L^{r-1}. Suppose the given data h to be extended is defined on all of boundary Δ. Then h defines data on boundary $(L^{r-1}_t \cap \Delta)$. Under suitable conditions we can apply the fundamental solution of $P^0(D)$ to this data to obtain a solution h_t of $P^0(D)$.

Of course, there is no reason to believe $h_t = h$ on $\mathrm{bd}\,(L^{r-1}_t \cap \Delta)$. We vary t to a value t_0 where $L^{r-1}_{t_0}$ has a high order of contact with Δ, so that $P^0(D)$ becomes a tangential operator to Δ. Then, in a suitable limit sense, $h_{t_0} = h$ there. We then need a unique continuation argument based on operators in the ideal generated by P^p to show h_t agrees with h for all t.

Actually this argument works better, technically, in terms of exterior Δ. For $h_t = h$ on $\operatorname{bd}(L_t^{r-1} \cap \Delta)$ is, under suitable conditions, equivalent to the vanishing of $P^0(D)h_t$ outside $L_t^{r-1} \cap \Delta$.

For the Cauchy–Riemann system $\bar\partial$ there are operators, e.g. $P^0 \equiv \partial/\partial\bar z_m$, whose support is on spaces of dimension 2, which is smaller than $n-r+1 = n/2+1$ (n is even) if $n \geq 4$ (two complex variables). This accounts for the extension results for holomorphic functions from suitable subsets of real codimension $> r-1$.

This type of argument was used by [118] in his study of boundary values of holomorphic functions in several variables.

It seems that all results obtained thus far by this method depend on the fact that P^0 is a first-order operator. In particular the pleuriharmonic system $\{\partial^2/\partial z_j \partial \bar z_k\}$ seems unexplored. One reason is the fact that we need a high order of contact of $L_{t_0}^{n-1}$ with $\operatorname{bd}\Delta$.

Problem 9.2 *Find extension results for $\{\partial^2/\partial z_j \partial \bar z_k\}$ for smooth Δ.*

We can study the $\bar\partial$ system in the framework of formal power series from another point of view. Let Ω be a smooth (piece of) hypersurface in $\mathbb{C}^n$. Suppose Ω is defined locally near x^0 by an equation $\psi = 0$ with $\operatorname{grad}\psi(x^0) \neq 0$. A basis for the tangential Cauchy–Riemann operator $\bar\partial_b$ near x^0 is

$$L_{jl} \equiv \psi_{\bar z_l} \frac{\partial}{\partial \bar z_j} - \psi_{\bar z_j} \frac{\partial}{\partial \bar z_l} \tag{9.38}$$

assuming that $\psi_{\bar z_l} \neq 0$ for any l. We can usually achieve this by a change of variables; otherwise there is a somewhat different form for the basis.

The tangential Cauchy–Riemann operators plus one operator, called the $\bar\partial$ *normal operator* $\partial/\partial\bar N$, span the same space as $\{\partial/\partial\bar z_j\}$. We can choose

$$\frac{\partial}{\partial \bar N} = \sum \psi_{z_l} \frac{\partial}{\partial \bar z_l}.$$

Since

$$\frac{\partial}{\partial \bar N}\psi = \sum \psi_{z_l}\psi_{\bar z_l} = \sum \psi_{x_j}^2 \neq 0$$

$\partial/\partial\bar N$ is not tangential.

The formal power series analog of Lewy's theorem is

Proposition 9.5[$\bar\partial$] *Suppose that Ω is a (local) C^∞ hypersurface defined by an equation $\psi = 0$ with $\operatorname{grad}\psi(x^0) \neq 0$. A function f_0 is the zeroth-order term in a C^∞ jet of order infinity (formal power series) solution f of $\bar\partial f = 0$ near x^0 if and only if $\bar\partial_b f_0 = 0$.*

Proof We shall construct f in the form

$$f = \sum f_k \psi^k. \tag{9.39}$$

We choose a coordinate system θ, ψ near p^0 where θ is the coordinate on $\psi = \text{const}$. The f_k depend only on θ. We must satisfy

$$0 = \frac{\partial f}{\partial \bar{z}_j} = \sum \frac{\partial f_k}{\partial \bar{z}_j} \psi^k + \sum (k+1) f_{k+1} \psi^k \frac{\partial}{\partial \bar{z}_j} \psi.$$

These equations will surely be satisfied if

$$\frac{\partial f_k}{\partial \bar{z}_j} + (k+1) f_{k+1} \frac{\partial \psi}{\partial \bar{z}_j} = 0$$

on Ω for all j, k.

For those j for which $\partial\psi/\partial\bar{z}_j \neq 0$ we set

$$f_{k+1} = -\frac{\partial f_k/\partial \bar{z}_j}{(k+1)\partial\psi/\partial\bar{z}_j}. \tag{9.40}$$

Since ψ is a real function and $\operatorname{grad}\psi(x^0) \neq 0$ some denominator in (9.40) is different from zero. We must show that f_{k+1} is defined independently of j. This is just the statement that $L_{jl} f_k = 0$ for all j, l. Our hypothesis is that $L_{jl} f_0 = 0$ so we must show that this property persists under iteration: Assume it holds for f_k. We apply $\partial/\partial\bar{z}_l$ to (9.40); there results (using the notation $\psi_{,l} = \partial\psi/\partial\bar{z}_l$, etc.)

$$-(k+1)\psi_{,j} f_{k+1,l} = \frac{\psi_{,j} f_{k,jl} - \psi_{,jl} f_{k,j}}{\psi_{,j}}. \tag{9.41}$$

We must show that this expression is symmetric in j, l. This is the identity

$$\psi_{,l}\psi_{,j} f_{k,jl} - \psi_{,jl}\psi_{,l} f_{k,j} = \psi_{,j}\psi_{,l} f_{k,jl} - \psi_{,jl}\psi_{,j} f_{k,l}$$

which is clearly a consequence of our induction assumption

$$L_{jl} f_k = 0.$$

This verifies $L_{jl} f_{k+1} = 0$.

This proves sufficiency of $\bar{\partial}_b f_0 = 0$ for the existence of f.

For the necessity, suppose there were a formal power series

$$f = \sum g_k(\theta)\psi^k \tag{9.42}$$

solution of $\bar{\partial} f = 0$ with $g_0 = f_0$. We apply the tangential operator L_{jl} to obtain

$$\sum \left[L_{jl} g_k(\theta)\right] \psi^k = 0$$

since $L_{jl}\psi = 0$. In particular $L_{jl} g_0 = L_{jl} f_0 = 0$ so $\bar{\partial}_b f_0 = 0$.

It remains to prove the uniqueness. The operators L_{jl} do not affect the order of terms in (9.42) so they cannot be used to form an iteration scheme. We apply $\partial/\partial\bar{N}$ to (9.42)

$$\sum \left[\frac{\partial g_k}{\partial \theta} \cdot \frac{\partial \theta}{\partial \bar{N}} + (k+1) g_{k+1} \frac{\partial \psi}{\partial \bar{N}} \right] \psi^k = 0.$$

The terms $\partial\theta/\partial\bar{N}$ and $\partial\psi/\partial\bar{N}$ are to be regarded as formal power series. The g_k are C^∞ functions of θ; they do not depend on ψ, nor do their θ derivatives. The term of order k in ψ in the above sum involves g_{k-l} times the term of order l in $\partial\theta/\partial\bar{N}$ and g_{k+1-j} times the term of order j in $\partial\psi/\partial\bar{N}$. In particular the term g_{k+1} appears with the coefficient which is the zeroth-order term in $(k+1)\partial\psi/\partial\bar{N}$ and all other terms involve $g_{k'}$ with $k' \leq k$.

It follows that we can make an iteration scheme as long as $\partial\psi/\partial\bar{N} \neq 0$, which we have shown.

This completes the proof of Proposition $9.5^{\bar{\partial}}$.

The proof of Proposition $9.5^{\bar{\partial}}$ depends on two properties of $\bar{\partial}$:

(a) $\partial/\partial\bar{z}_j$ are first order.
(b) The $\partial/\partial\bar{z}_j$ commute.

From (a) we determine that the linear combination of $\{\partial/\partial\bar{z}_j\}$ which are tangent to Ω are those that annihilate ψ.

(b) is used following (9.41) to establish $L_{jl} f_{k+1} = 0$.

Let us examine what happens when we drop condition (b). We replace $\bar{\partial}$ by a Lie algebra $\Lambda = \{\lambda\}$. Certainly

$$\lambda_{jl} = (\lambda_l \psi)\lambda_j - (\lambda_j \psi)\lambda_l$$

annihilates ψ and hence is tangent to Ω. We assume, ab initio, that $\lambda_{jl} f_0 = 0$ for all j, l.

Let us repeat the argument in (9.39)ff. Everything proceeds as before (assuming $\lambda_j \psi \not\equiv 0$ for some j) until the equation following (9.41).

We write $\psi_{,l} = \lambda_l \psi$ etc. but we now take care of the order in writing $\psi_{,jl} = \lambda_l \lambda_j \psi$ etc. The equation in question takes the form

$$\psi_{,l}\psi_{,j} f_{k,jl} - \psi_{,jl}\psi_{,l} f_{k,j} = \psi_{,j}\psi_{,l} f_{k,lj} - \psi_{,lj}\psi_{,j} f_{k,l}.$$

We must show that

$$M_{jl} \equiv \psi_{,j}\psi_{,l}(\lambda_l\lambda_j - \lambda_j\lambda_l) - (\psi_{,jl}\psi_{,l}\lambda_j - \psi_{,lj}\psi_{,j}\lambda_l)$$

annihilate f_k. Since Λ is a Lie algebra, $\lambda_l\lambda_j - \lambda_j\lambda_l \in \Lambda$. Hence M_{jl} is a first-order operator. It is clear that $M_{jl}\psi = 0$ so M_{jl} is tangent to Ω. However, it might not be the case that M_{jl} is a linear combination of the operators λ_{ab} which were the original operators in $\Lambda \cap$ (tangent space to Ω).

This means that we must impose the conditions

$$M_{jl} f_0 = 0.$$

We must show that this condition persists in our iteration scheme (9.40) defining f_{k+1}.

It may happen that, just as we needed to add $\{M_{jl} f_0 = 0\}$ to $\lambda_{ab} f_0 = 0$ to guarantee the persistence of the latter equations under (9.40), we may have to add more tangential equations (belonging to Λ) to insure the persistence of $\{M_{jl} f_k = 0\}$ under (9.40).

In this way we prove

Theorem 9.5 *Let Λ be a finite dimensional Lie algebra of first-order differential operators with C^∞ coefficients such that $\lambda\psi(x^0) \neq 0$ for some $\lambda \in \Lambda$. In order for a function f_0 on Ω to be the zeroth-order term of a C^∞ jet of order infinity f of $\{\lambda f = 0\}_{\lambda\in\Lambda}$ near x^0 it is necessary and sufficient that $Mf_0 = 0$ for all M which are linear combinations of $\lambda \in \Lambda$ with coefficients which are smooth functions, such that M lies in the tangent space of Ω.*

Remark. *Theorem 9.5, when combined with the argument centered around (9.23), shows that f_0 is the exotic CD for the system of equations $\Lambda f = 0$. The CP is exotic in the sense that Ω has too high a dimension to be a CS for $\Lambda f = 0$ in the usual sense, so we need the equations $\{Mf_0 = 0\}$ to replace the lowering of dimension.*

What happens for operators of order >1?

Suppose $P_j(x, D)$ have analytic coefficients and Ω is a (local) real analytic hypersurface. We call m_j the order of P_j and suppose $m_1 = N$ is the minimal m_j. Finally assume that Ω is noncharacteristic for P_1.

We may make a change of variables to transform Ω into $y = 0$. (We denote the new variables by $x = (x_1, \ldots, x_{n-1})$ and y.) This change does not affect the orders of P_j nor the noncharacteristic nature of P_1.

Since Ω is noncharacteristic for P_1 we can write

$$P_1 \equiv \frac{\partial^N}{\partial y^N} + P_{11}\frac{\partial^{N-1}}{\partial y^{N-1}} + \cdots + P_{1N} \tag{9.43}$$

where P_{1l} is a linear differential operator of order $\leq l$ in x. Let f be a jet of infinite order as above, so f is a formal power series in y with coefficients f_l which are functions of x. It is easily seen from this expression that the k-th-order term in $P_1 f$ is of the form

$$(P_1 f)_k = \sum_{l\leq k+N} c_k^l f_l.$$

The c_k^l are linear differential operators in x with coefficients which are analytic in x, y and polynomials in k. In particular

$$c_k^{k+N} = (k+N)(k+N-1)\cdots(k+1).$$

Since $c_k^{k+N} \neq 0$ if $P_1 f = 0$ we can solve successively for f_l, $l \geq N$ in terms of $f_0, \ldots, f_{N-1}$. This is the CP for a single equation.

Next we impose the equation $P_2 f = 0$. If P_2 is of degree N' then we have an expression for $f_{N'}$ in terms of $f_0, \ldots, f_{N'-1}$ depending on P_2. If $N' = N$ this gives two expressions for $f_{N'}$ in terms of $f_0, \ldots, f_{N-1}$. If $N' > N$ then by expressing $f_N, \ldots, f_{N'-1}$ in terms of $f_0, \ldots, f_{N-1}$ using P_1 we again obtain two expressions for $f_{N'}$ in terms of $f_0, \ldots, f_{N-1}$. These must be compatible for f to be a solution of $P_1 f = P_2 f = 0$.

It seems that there are infinitely many relations. Sometimes, using special techniques, we can reduce the number of compatibility conditions to a finite number; sometimes they can be explicitly calculated. In particular

(1) If the P_j have constant coefficients and Ω is linear then the structure of the CD can be determined algebraically (see Chapter IX of FA).
(2) For operators of the form $\bar{\partial}^m \partial^l$ (symmetric powers rather than the usual skew-symmetric powers) the explicit compatibility conditions were determined by [161].

Problem 9.3 *Find other examples in which the compatibility is finite and can be explicitly determined.*

If M is a nice real submanifold of $\mathbb{C}^n$ then, under suitable conditions, the equations induced by the Cauchy–Riemann system $\bar{\partial}$ make M into a *CR manifold*. CR manifolds have been the object of much study (see e.g. [22]). We want to produce an analogous structure for general overdetermined systems $\vec{P}(x, D)$.

One way of defining the restriction Q_M of an operator $Q(x, D)$, defined on $\mathbb{R}^N$, on a submanifold M is by setting

$$Q_M f = QF|_M$$

where F is a smooth extension of f from M to $\mathbb{R}^N$. This definition is meaningful only if Q annihilates the ideal of smooth functions which vanish on M.

Since $\bar{\partial}$ is a first-order system it is easy, via Leibnitz' rule, to determine the set of restrictions to M of the operators of the form $\sum h_j(x)\partial/\partial\bar{z}_j$. This defines the usual CR structure. But for general systems we need a different idea.

We noted in Section 1.1 that sets are defined by equations or by points. The formulation of CR manifolds given above depends on the equations of M. For the general theory we need a more intrinsic formulation.

We can formulate the CR structure on M in terms of Cauchy surfaces. We suppose $m = \dim M > n$. Let $p \in M$ and call T_p the tangent space to M at p. We assume that T_p contains S_p, which is a CS for $\bar{\partial}$. Since $\dim S_p = n < m$ there are $m - n$ linearly independent relations on T_p which define S_p.

Now comes the crucial point: there is an identity between T_p and the restriction to p of first-order differential operators on M. Thus the $m-n$ linear relations on T_p define $m-n$ relations amongst first-order differential operators on M at p.

Since p is arbitrary this leads to $m-n$ local relations amongst first-order operators on M. This is the CR structure.

We want to show that this concept of CR manifold agrees with the usual one, which is defined by the linear combinations of the $\hat{\bar{\partial}}_j$ which lie in T_p. (We identify elements of T_p with linear functions on $\mathbb{R}^{2n}$ which are constant in directions orthogonal to T_p.) Let Q^2 be such a tangent vector. Call Q^1 the restriction of Q^2 to S_p. Then $Q = Q^2 - Q^1$ vanishes on S_p. Conversely, if Q vanishes on S_p, then, in accordance with our treatment of the CP, we write (in $\mathbb{R}^{2n}$)

$$Q = Q^2 - Q^1$$

where Q^1 depends only on the variables of S_p and Q^2 is a linear combination of the $\hat{\bar{\partial}}_j$. Since Q and Q^1 are in T_p the same is true of Q^2. We have given a one-to-one correspondence between Q and Q^2 which verifies our claim.

All this works because T_p agrees with M at p to the first-order. Note that T_p is a linear subspace of $\mathbb{R}^{2n}$. To deal with higher order systems $\vec{P}$ (e.g. order 2) on $M \subset \mathbb{R}^N$ we have to replace T_p by a quadric T_p^2 which has order of contact 2. We thus have to construct $S_p^2 \subset T_p^2$ which is a CS for $\vec{P}$ and then extend relations from T_p to M.

The ideas are closely connected to those of the nonlinear Fourier transform (Chapter 5). The theory is very complicated and will be studied in later works.

9.3 Wave front sets and the Cauchy problem

In Section 9.1 we formulated the edge-of-the-wedge theorem and the Hartogs–Bernstein extension theorem in terms of wave front sets. We start with a plane L (variable x, dimension n) called the "edge" and a proper cone Γ (variable y, dimension m), so $L+\Gamma$ contains L in its closure. We are given an overdetermined system of linear partial differential equations with constant coefficients

$$\vec{P}f = 0 \tag{9.44}$$

in x, y space. f is a C^∞ solution on the "wedge" $L + \Gamma$ and f is smooth up to L.

Let V be the variety of common zeros of the $P_j(i\hat{x}, i\hat{y})$. We assume that V is an s sheeted covering of $\hat{L}$ and that L is noncharacteristic for $\vec{P}$; moreover there exists a CP for $\vec{P}$ with s data on L given by "normal derivatives" $\partial_k(D)$ which are polynomials in $\{\partial/\partial y_l\}$. s can be thought of as the order of $\vec{P}$ relative to L. (We recall that "noncharacteristic" means that $|\hat{y}| \leq c(1+|\hat{x}|)$ on V.)

These assumptions can be modified, but we make them to capture the essence of the problem without introducing unnecessary complications.

We want to study the wave front set of $\vec{f}^0$ which is the CD on L of a solution of (9.44) in the wedge. We studied the wave front sets of scalar functions in Section 5.3. In the present case we shall be able to compute the wave front sets of certain

linear combinations $\vec{\alpha}^j \cdot \vec{f}^0$ where the components α_k^j are suitable operators. With enough $\vec{\alpha}^j$ we can compute the wave front sets of the components f_k^0.

Let $\hat{\rho}^0$ be a ray in real $\{\hat{x}\} = \hat{L}$. This means that

$$\hat{\rho}^0 = \{\hat{x} = \hat{x}(t) = t\hat{x}^0, t > 0\}$$

where $\hat{x}^0 \neq 0$. (We usually take $|\hat{x}^0| = 1$.) We say that $\hat{\rho}^0$ has an *elliptic cover* in V if

$$\max_{\substack{(\hat{x}(t),\hat{y})\in V \\ \hat{x}(t)\in\hat{\rho}^0}} |\Im\hat{y}| \geq c|\hat{x}(t)|$$

for some $c > 0$.

An *elliptic ray* above $\hat{\rho}^0$ (which is not exactly a ray) is a real algebraic curve $\hat{\rho} \subset V$ which covers (projects on) $\hat{\rho}^0$ such that, on $\hat{\rho}$,

$$\hat{y}(t) = t\hat{y}^0 + \mathcal{O}(t^b) \qquad \hat{y}^0 \neq 0, \quad b < 1, \quad \hat{y}(t) \text{ above } \hat{x}(t) \tag{9.45a}$$

$$\Im\hat{y}^0 \neq 0. \tag{9.45b}$$

When $\hat{y}(t)$ satisfies (9.45) we call $\hat{y}(t)$ an *elliptic root* over $\hat{\rho}^0$.

Let f be the solution of $\vec{P}f = 0$ with $\mathrm{CD} = \vec{f}^0$. By the fundamental principle (see FA or Section 1.4) we can represent f as the Fourier transform of a measure $\hat{f}$ on V. The measure $\hat{f}(\hat{x}, \hat{y})$ is exponentially decreasing in $(\Im\hat{x}, \Im\hat{y})$. The amount of exponential decrease is given by

$$\left|\hat{f}(\hat{x},\hat{y})\right| = \mathcal{O}\left[(|\hat{x}| + |\hat{y}|)^{-N} e^{-c(|\Im\hat{x}|+|\Im\hat{y}|\psi(\Im\hat{y}))}\right] \tag{9.46}$$

for all N. ψ is the characteristic function of the intersection of the unit sphere with the complement $c(\Gamma')$ of the dual cone Γ' of Γ.[3] Although $\hat{f}$ is not unique the fundamental principle provides us with some $\hat{f}$ satisfying (9.46). Suppose that the elliptic ray lies in the tube

$$\tau(c(\Gamma')) = \{\Im\hat{y} \in c(\Gamma')\}.$$

Then $\hat{f}$ is exponentially decreasing on $\hat{\rho}$ so the contribution of the Fourier transform of f coming from $\hat{\rho}$ is real analytic.

In order to formulate this in terms of wave front sets we need to know that if $\hat{\rho}$ is an elliptic ray above $\hat{\rho}^0$ there is a whole angle of such rays. We formalize this as

Proposition 9.6 *Let $\hat{\rho}^0$ be a ray in $\mathbb{R}^n$ = real $\hat{L}$. Suppose $\hat{\rho}^0$ has an elliptic cover. Then there is an elliptic ray $\hat{\rho}$ above $\hat{\rho}^0$. We can vary $\hat{\rho}^0$ continuously to rays $\hat{\rho}_\lambda^0$ in a complex cone and there is a corresponding continuous variation*

[3]Some care must be taken here because of the boundaries of Γ and $c(\Gamma')$. We leave this to the reader.

$\{\hat{\rho}_\lambda\}$ of covering elliptic rays. The constants b in the corresponding (9.45) can be chosen independent of λ and $|\Im \hat{y}^0_\lambda|$ is bounded from below.

Proof We start with the case $r = 1$. Let $P(\hat{x}, \hat{y})$ be a polynomial and $\hat{\rho}^0$ be a ray in $\mathbb{R}^n$ = real $\hat{L}$. We can express $\hat{\rho}^0$ parametrically in the form

$$\hat{\rho}^0 = \{\hat{x} = t\hat{x}^0\}$$

where t is a variable on $\mathbb{R}^+$ and $|\hat{x}^0| = 1$.

If $\hat{y}$ is a single variable and $\{\hat{x}\}$ is noncharacteristic for P then we can write

$$P(\hat{x}(t), \hat{y}) = \hat{y}^s + Q_1(\hat{x}(t))\hat{y}^{s-1} + \cdots + Q_s(\hat{x}(t)).$$

Noncharacteristic means $\deg Q_j \leq j$.

We want to study the roots $\hat{y}^j(t)$ of $P(\hat{x}(t), \hat{y})$ at $t = \infty$. Since $\deg Q_l \leq l$,

$$|Q_l(\hat{x})| \leq c|\hat{x}|^l \qquad |\hat{x}| \to \infty.$$

Thus

$$\begin{aligned} |P(\hat{x}, \hat{y})| &\geq |\hat{y}|^s - c\sum |\hat{x}|^l |\hat{y}|^{s-l} \\ &= |\hat{y}|^s \left[1 - c\sum \left(\frac{|\hat{x}|}{|\hat{y}|}\right)^l\right]. \end{aligned}$$

Clearly this cannot vanish if $|\hat{y}| \gg |\hat{x}|$.

Call Q^0_l the principal part of Q_l which is the homogeneous part of degree l. Then

$$P(\hat{x}(t), \hat{y}) = \hat{y}^s + Q^0_1(\hat{x}^0)t\hat{y}^{s-1} + \cdots + Q^0_s(\hat{x}^0)t^s + Q^1(t, \hat{y})$$

where Q^1 contains terms in $t, \hat{y}$ of degree $< s$. We write

$$P(\hat{x}(t), \hat{y}) = \Pi(\hat{y} - \gamma_j t) + Q^1(t, \hat{y}).$$

It is a standard fact that the roots of a polynomial depend continuously on the coefficients. Varying the ray $\hat{\rho}^0$ corresponds to varying $\hat{x}^0$. If we change $\hat{x}^0$ by a small amount in the complex unit sphere then the $Q^0(\hat{x}^0)$ change accordingly, hence so do the γ_j.

Since $P(\hat{x}(t), \hat{y})$ is a polynomial in the two variables $t, \hat{y}$ we can write the Puisseaux expansion for the roots $\hat{y}^j(t)$. It is well known and readily verified that such expansions take the form

$$\hat{y}^j(t) = \gamma_j t + \mathcal{O}(t^b)$$

where $b < 1$. If all γ_j were real then

$$|\Im \hat{y}(t)| \leq c|t^b|$$

so $\hat{\rho}^0$ would not have an elliptic cover. We have assumed that $\hat{\rho}^0$ has an elliptic cover so, in particular, not all $\gamma_j = 0$ so not all $Q_l^0 \equiv 0$.

If γ_1 is not real then

$$\hat{\rho} = \{\hat{x}(t), \hat{y}^1(t)\}$$

is an elliptic ray above $\hat{\rho}^0$.

Instead of using the Puisseaux expansion we can argue as follows: we write

$$P(\hat{x}(t), \hat{y}) = t^s \left[\left(\frac{\hat{y}}{t}\right)^s + Q_1^0(\hat{x}^0) \left(\frac{\hat{y}}{t}\right)^{s-1} + \cdots + Q_s^0(\hat{x}^0) + \frac{1}{t} P^1\left(\hat{x}, \frac{\hat{y}}{t}\right) + \cdots + \frac{1}{t^s} P^s\left(\hat{x}, \frac{\hat{y}}{t}\right) \right]$$

where P^j is the part of Q^1 which is homogeneous of degree $s - j$ in $\hat{y}$. The term in [ ] is a polynomial in $\hat{y}/t$ of degree s. For t large the coefficients differ from those of P^0 by ϵ. By the continuity of the roots of a polynomial, the roots of $P(\hat{x}(t), \hat{y})$ are very close to $\{\gamma_j\}$. In particular at least one root is nonreal.

The Puisseaux expansion can be regarded as a more precise estimate on the variation of roots of a polynomial equation.

We have explained that changing $\hat{x}^0$ by a small amount in the unit sphere changes γ_1 by a small amount and hence keeps γ_1 from becoming real. On the other hand, an open set in $\{\hat{x}^0\}$ constitutes a cone.

This completes the proof of Proposition 9.6 for $r = 1$.

Let us pass to the general case $r \geq 1$ and $\dim Y \geq 1$. Let $\hat{\rho}^0$ be a ray $\hat{x} = t\hat{x}^0$ in real $\{\hat{x}\}$. For each t there are s points (with multiplicity) with $(\hat{x}(t), \hat{y}^j(t)) \in V$.

Let V_l be the closure of the projection of V on the complex $(\hat{x}, \hat{y}_l)$ plane. Recall [155] that the projection of an algebraic variety is a *Zariski open set* of an algebraic variety. (A Zariski open set of an algebraic variety is an algebraic variety minus a lower dimension subvariety.) Since $\hat{L} = \{\hat{x}\}$ is noncharacteristic for V it is also noncharacteristic for each V_l. V_l is a hypersurface and is thus defined by a single polynomial equation $S_l(\hat{x}, \hat{y}_l) = 0$. It is a standard (genericity) result, which is readily verified, that we can make a real linear change of $\hat{y}$ variables to make each S_l of degree s in $(\hat{x}, \hat{y}_l)$ with leading term $a_l(\hat{y}_l)^\Delta$, $a_l \neq 0$ (so we can assume $a_l = 1$ for every l).

We apply the Puisseaux expansion to S_l above, $\hat{\rho}^0$. As above, this yields

$$\hat{y}_l^j(t) = \gamma_l^j t + \mathcal{O}(t^b) \qquad t \to +\infty$$

with $\gamma_l^j \neq 0$ for some j, l and $b < 1$.[4]

We have assumed that $\hat{\rho}^0$ has an elliptic cover. This means that for some j, l there is a γ_l^j, say $\gamma_1^{j_0}$, which is not real.

[4]Our notation is such that $\hat{y}_l^j$ is the l-th coordinate of the j-th root. The reader should be careful to distinguish between superscripts and powers.

We are now in the same position for $S_1(\hat{x}, \hat{y}_1)$ as we were in above for $P(\hat{x}, \hat{y})$ when $\hat{y}$ was a single variable. We can vary $\hat{\rho}^0$ to an open set of rays $\hat{\rho}^0_\lambda$ forming a cone and still have

$$|\Im \hat{y}_1^{j_0}(t)| \geq c|t| \geq c'(|\hat{x}| + |\hat{y}|)$$

uniformly in λ. We denote these curves by $\tau_{\lambda 1}$.

Since the number s of points $(\hat{x}(t), \hat{y})$ above $\hat{x}(t)$ is finite we can lift the curves $\tau_{\lambda 1}$ from V_1 to curves $\tau_\lambda \subset V$ which cover (project on) $\tau_{\lambda 1}$. On τ_λ

$$\left| y_l^j(t) \right| = \gamma_l^j t + \mathcal{O}(t^b) \quad b < 1$$

for all j, l and

$$\left| \Im \hat{y}_1^{j_0}(t) \right| \geq c'' t \geq c(|\hat{x}(t)| + |\hat{y}(t)|).$$

This means that τ_λ are elliptic rays in V above $\hat{\rho}^0_\lambda$. The proof of Proposition 9.6 is complete.

Theorem 9.7 *Suppose there are s elliptic rays (belonging to the different sheets of the covering) above $\hat{\rho}^0$ for which $\Im \hat{y}(\hat{x}) \in c(\Gamma')$. Then $\hat{\rho}^0$ does not belong to the analytic wave front set of $\vec{f}^0$, meaning of any component of $\vec{f}^0$.*

Proof We have seen several times that although $\hat{f}_k^0(\hat{x})$ is not uniquely determined one choice is given by

$$\sum_l \hat{\partial}_k(\hat{y}_l(\hat{x}))\hat{f}(\hat{x}, \hat{y}_l(\hat{x})) = \hat{f}_k^0(\hat{x}). \tag{9.47}$$

The ∂_k are the "normal derivatives" which define the CP with data on L or, equivalently, $\hat{\partial}_k$ are the polynomials which give a harmonic basis for functions on the sets $V_{\hat{x}} = \{\hat{y} \big| (\hat{x}, \hat{y}) \in V\}$.

By the above there is a proper cone γ in complex $\hat{L}$ containing $\hat{\rho}^0$ in its interior such that $\hat{f}(\hat{x}, \hat{y}_l(\hat{x}))$ is exponentially decreasing for $\hat{x} \in \gamma$ for all l. By Theorem 5.14 $\hat{\rho}^0$ does not belong to the wave front set of f_k^0. Since this is valid for each k Theorem 9.7 is established.

Combining Theorem 9.7 with Theorem 5.15 yields

Corollary 9.8 *If the union of the rays $\hat{\rho}^0$ satisfying the hypothesis of Theorem 9.7 contains the complement of a proper real cone Γ_1 then $\vec{f}^0(x)$ is holomorphic for x small and $\Im x \in$ interior Γ_1'.*

There is a weakness in Theorem 9.7 because it deals with only one cone $\Gamma \subset Y$ so we need all the points in V above $\hat{\rho}^0$ to be elliptic and have the imaginary parts of their $\hat{y}$ coordinates in $c(\Gamma')$. We now investigate what happens if we have several systems $\vec{P}^p$. We assume that each V^p is an s sheeted covering of $\hat{L}$ and there are CPs for each $\vec{P}^p$ of a similar nature to the CP we described for $\vec{P}$. In particular L is noncharacteristic for each system $\vec{P}^p$. The "normal

derivatives" $\vec{\partial}^p = \{\partial_k^p\}$ can depend on p as can the cone Γ^p in Y. But L and the decomposition of the space into (x, y) do not depend on p. Moreover the CD $\vec{f}^0$ does not depend on p. In particular s is independent of p.

As before, let $\hat{\rho}^0$ be a ray in real $\hat{x}$ space. We now assume that amongst all the rays $\{\hat{\rho}_m^p\}$ lying above $\hat{\rho}^0$ with $\hat{\rho}_m^p \subset V^p$ and $\Im \hat{y}_m^p(\hat{x}) \in c[(\Gamma^p)']$ there are s *independent elliptic rays* $\hat{\rho}_{m^0}^p$.

In order to define "independent rays" we examine (9.47) for each p

$$\sum_l \hat{\partial}_k^p(\hat{y}_l^p(\hat{x}))\hat{f}(\hat{x}, \hat{y}_l^p(\hat{x})) = \hat{f}_k^{0p}(\hat{x}). \tag{9.48}$$

The problem is that only some of the $\hat{y}_l^p$, say $\hat{y}_{l_0}^p$, are "good," i.e. satisfy the above ellipticity condition. Thus it is not at all clear why the sum should be exponentially decreasing for $\hat{x} \in \hat{\rho}^0$.

Our hypothesis that there is a noncharacteristic CP for each $\vec{P}^p$ with data on L means that for each p we can invert (9.48)

$$\hat{f}(\hat{x}, \hat{y}_l^p(\hat{x})) = \sum_k \alpha_{lk}^p(\hat{x})\hat{f}_k^{0p}(\hat{x}). \tag{9.49}$$

The $\alpha_{lk}^p(\hat{x})$ are algebraic functions which are not $\equiv \infty$. We want to have at least s good roots $\hat{y}_{l_0}^p(\hat{x})$ amongst all the $\hat{y}_k^q$ for varying q, k such that we can invert (9.49) using only those roots.

Definition *The s roots $\{\hat{y}_{l_0}^p(\hat{x})\}$ are independent over $\hat{\rho}^0$ if*

$$\det\left[\alpha_{l_0 k}^p(\hat{x})\right] \not\equiv 0 \qquad \hat{x} \in \hat{\rho}^0.$$

The indices of the matrix $\alpha_{l_0 k}^p$ are k and the pair p, l_0.

We can proceed as follows. The independence of the roots $\hat{y}_{l_0}^p$ would allow us to solve (9.49) for the $\hat{f}_k^{0p}$ if $\hat{f}_k^{0p}$ were independent of p. This need not be the case because the Fourier transform is not unique. To emphasize the dependence on p we have used the notation $\hat{f}_k^{0p}$. Of course the $\hat{f}_k^{0p}$ for fixed k have the same Fourier transform, namely f_k^0, since we assume the CD $\vec{f}^0$ is independent of p. How can we account for the difference between the $\hat{f}_k^{0p}$ for various p?

We know that $f_k^0 \in \mathcal{E}$ so f_k^0 can be represented in many ways as the Fourier transform of a measure $\hat{f}_k^0$ of the form $d\hat{\mu}(\hat{x})/\kappa(\hat{x})$ where μ is a measure of total bounded variation on $\mathbb{C}^n$ and κ belongs to an AU structure for $\mathcal{E}$. We are faced with the problem of determining when $d\hat{\mu}/\kappa$ is zero as an element of $\hat{\mathcal{E}}$.

We want to change our viewpoint and think of $\hat{\mu}/\kappa$ as a measure on $\mathbb{R}^{2n}$ with real coordinates $\hat{\xi}, \hat{\eta}$ for $\hat{x} = \hat{\xi} + i\hat{\eta}$. We can take the $\mathbb{R}^{2n}$ Fourier transform

$${}^2f_k^0(\xi, \eta) = \int e^{i\xi\cdot\hat{\xi} + i\eta\cdot\hat{\eta}} \frac{d\hat{\mu}(\hat{x})}{\kappa(\hat{x})}.$$

Clearly ${}^2f_k^0 \in C^\infty$ and all its derivatives are uniformly bounded. Moreover, since κ increases in $\hat{\eta}$ faster than $\exp(c\hat{\eta})$ for any c, ${}^2f_k^0(\xi,\eta)$ can be extended to an entire function in $z = \eta + i\zeta$.

It is difficult to describe the space of ${}^2f_k^0$ which are of the above form because the Fourier transforms of spaces of measures are quite complicated. We can remove this difficulty by passing to distributions. If D is any differential operator with constant coefficients on $\mathbb{R}^{2n}$ then the Fourier transform of $D[\mu(\hat{\xi},\hat{\eta})/\kappa(\hat{\xi},\eta)]$, using the exponentials $\exp(ix\cdot\hat{x}) = \exp[ix\cdot(\hat{\xi}+i\hat{\eta})]$, is still in $\mathcal{E}(x)$ since $\mathcal{E}$ is closed under differentiation and multiplication by polynomials. A suitable closure of the linear span of $\{D(\hat{\mu}/\kappa)\}_{D,\hat{\mu},\kappa}$ can be identified with the tensor product space

$$ {}^2\hat{\mathcal{E}} = \mathcal{O}_c{}'(\hat{\xi}) \otimes \hat{\mathcal{H}}_c{}'(\hat{\eta}). $$

$\mathcal{O}_c'$ is the space introduced by Schwartz [135] of rapidly (i.e. faster than polynomially) decreasing distributions and $\hat{\mathcal{H}}_c{}'(\hat{\eta})$ is the analogous space in which "polynomially" is replaced by "exponentially."

We have remarked on several occasions that we obtain the same space if we use as representations for $\hat{\mathcal{E}}$ either measures, or distributions of finite order, or smooth functions.

As Schwartz has shown, the Fourier transform of $\mathcal{O}_c'$ is $\mathcal{O}_M$ which is the space of functions which are polynomially bounded as distributions, meaning some iterated integral is polynomially bounded. (The subscripts c and M refer to the fact that $\mathcal{O}_c'$ is the space of distributions which convolve $\mathcal{S}$ into itself and $\mathcal{O}_M$ is the space of multipliers on $\mathcal{S}$.) $\mathcal{H}_c'$ is the space of entire functions $F(z)$ which satisfy inequalities of the form

$$ |F(z)| \leq a(1+|z|)^a G(\Im z) $$

for some constant a and some function G. The topologies of these spaces are defined in a natural manner. (There might be some confusion because we originally stated that the ${}^2f_k^0$ were bounded on (ξ,η). That was because we originally represented ${}^2f_k^0$ as the Fourier transform of $d\mu/\kappa$ where μ is a measure. Our present viewpoint is that μ/κ is a distribution so "bounded" is replaced by "polynomially bounded.")

We can identify

$$ \mathcal{E} = {}^2\mathcal{E}\big|_{z=i\xi} \tag{9.50} $$

as can be seen from the definition of $\{{}^2f_k^0\}$. (ξ is real so $z = i\xi$ means $\zeta = \xi$, $\eta = 0$.)

This relation may seem puzzling at first because the functions in $\mathcal{E}$ are only C^∞ while those in ${}^2\mathcal{E}$ are holomorphic in z. However, if $g \in \mathcal{E}(\xi)$ is thought of as a function on $V = \{z = i\xi\}$ then we can extend g to a function ${}^2g(\xi,z)$ which is holomorphic in z (in fact, constant in z for each fixed ξ) by setting

$$ {}^2g(\xi,z) = {}^2g(\xi,\eta+i\zeta) = g(\xi). \tag{9.51} $$

Although ${}^2g(\xi, z)$ is holomorphic in z it does not satisfy the growth conditions prescribed by ${}^2\mathcal{E}$. These require, in particular, at most polynomial growth on $\zeta = 0$ in ξ, η as is clear from the definitions. However, $g(\xi)$ can be arbitrarily large so ${}^2g(\xi, \xi, 0) = g(\xi)$ need not be polynomially bounded.

Nevertheless (9.51) provides a semi-local extension of g from V (see FA). This means that ${}^2g(\xi, z)$ satisfies the desired inequalities as long as we are within a fixed finite distance from V. Then, as we shall explain below, the cohomology methods of FA or other cohomology methods can be used to extend ${}^2g(\xi, z)$ to all (ξ, z) satisfying the conditions prescribed by ${}^2\mathcal{E}$. Moreover the extension can be made continuous (though not linear).

This method of extension is the main ingredient in the proof of

Theorem 9.9 *If $\hat{u} \in {}^2\hat{\mathcal{E}}$ is $\equiv 0$ as an element of the dual of $\hat{\mathcal{E}}'$ then*

$$\hat{u} = \bar{\partial}\hat{\vec{w}} = \sum \left(\frac{\partial}{\partial \hat{\xi}_j} + i \frac{\partial}{\partial \hat{\eta}_j} \right) \hat{w}_j$$

for some $\hat{\vec{w}} = (\hat{w}_j) \in {}^2\hat{\mathcal{E}}^n$.

Remark 1 *It is Theorem 9.9 which allows us to alleviate the difficulties caused by the nonuniqueness of $\hat{f}_k^0$.*

Remark 2 *We can formulate Theorem 9.9 in terms of $\bar{\partial}$ cohomology with bounds. We regard $\hat{u}$ as an n form for the $\bar{\partial}$ complex and $\hat{\vec{w}}$ as an $n-1$ form. Any n form is closed so Theorem 9.9 asserts that*

$$\hat{\mathcal{E}} = H^n(\bar{\partial}, \hat{\mathcal{E}}),$$

meaning the n-th cohomology group for $\bar{\partial}$ with growth conditions prescribed by $\hat{\mathcal{E}}$. This cohomology approach is closely related to the Penrose transform (compare Section 1.5).

Proof of Theorem 9.9 *By the Fourier transform and (9.50) we must prove that if ${}^2g \in {}^2\mathcal{E}$ vanishes on $z = i\xi$ then ${}^2g$ is of the form*

$${}^2g = \sum (\xi_j + iz_j)\, {}^2h_j \tag{9.52}$$

for some $\vec{{}^2h} = ({}^2h_j) \in {}^2\mathcal{E}$. Actually we claim that more is true: we call V the "Cauchy–Riemann variety"

$$V = \{z_j = i\xi_j, \quad j = 1, \ldots, n\}.$$

The restriction map

$$R_V : {}^2g \to g(\xi, 0, \xi)$$

is a topological isomorphism of ${}^2\mathcal{E}/\hat{\bar{\partial}}\left({}^2\mathcal{E}\right)^n$ onto $\mathcal{E}(\xi)$.

This statement can be regarded as the fundamental principle (see Section 1.4) for ${}^2\hat{\mathcal{E}}$ for the operator $\bar{\partial}$, or, in the spirit of Remark 2, as a cohomological description of $\mathcal{E}$. Equation (9.52) is a special case of this fundamental principle; namely, (9.52) is the description of the kernel of R_V.

The proof of the fundamental principle as given in FA depends on two constructions:

A. Let $\vec{P} = (P_1, \dots, P_r)$ and call V the variety of common zeros of the P_j. Assume for simplicity (as in case of $\vec{P} = \hat{\bar{\partial}}$) that V is a nonsingular variety and the polynomial ideal generated by the P_j is the set of all polynomials that vanish on V. Then

(1) Holomorphic functions f on V can be extended to functions $\tilde{f}_\alpha$ which are holomorphic in cubes Q_α of fixed size (e.g. side $Q_\alpha = 10n^2$) centered at each lattice point α. (If $\alpha \notin V$ we can set $\tilde{f}_\alpha = 0$.)

We can bound $\tilde{f}_\alpha$ suitably in terms of bounds of f on $V \cap 10Q_\alpha$.

This is called *semi-local extension*.

(2) Any entire function h which vanishes in V can be written in the form

$$h = \sum u_{\alpha j} P_j \quad \text{on } Q_\alpha.$$

The u_α are holomorphic on Q_α.

If h belongs to the Fourier transform of some suitable AU space $\mathcal{W}$ then the $u_{\alpha j}$ satisfy the growth conditions of $\hat{\mathcal{W}}$, meaning

$$|u_{\alpha j}(z)| \le ck(z)$$

for all α, j and all k in an AU structure for $\mathcal{W}$. c depends on k but not on α, j.

B. A cohomology argument which allows passage from the semi-local theory to global (AU) theory.

Remark. *In our case "Holomorphic" in* A *and* B *means holomorphic in* z.

We have explained the semi-local extension in (9.51) above. As for the expression of a function ${}^2h(\xi, z)$ which vanishes on $\hat{\bar{\partial}} = 0$ in terms of $\hat{\bar{\partial}}_j$, let us start with the case of $n = 1$. We assume ${}^2h(\xi, z)$ vanishes on $\xi + iz = 0$. For fixed $\xi = \xi^0$ the only possible singularities of the quotient ${}^2h(\xi^0, z)/(\xi^0 + iz)$ as a function of z are at $z = i\xi^0$. The denominator has a simple zero there and the numerator, which is holomorphic in z, vanishes there. Thus the quotient is holomorphic in z for each ξ^0.

That ${}^2h(\xi, z)/(\xi + iz)$ is C^∞ in (ξ, z) can be proven in many ways, e.g. by using Taylor's formula with remainder in much the same manner as our proof of Theorem 2.2.

As we have explained in several places in this book, division by a polynomial has no serious effect on the bounds we use. In the present situation there is no difficulty in obtaining suitable bounds for ${}^2h(\xi, z)/(\xi + iz)$.

To pass to general n, for every m we introduce the variety

$$V_m = \{z_1 = i\xi_1, \ldots, z_m = i\xi_m\}.$$

We call ${}^2\mathcal{E}(V_m)$ the space of functions of $\xi, z_{m+1}, \ldots, z_n$ thought of as functions on V_m which are C^∞ in all variables and holomorphic in $z_{m+1}, \ldots, z_n$, and satisfy the growth conditions of ${}^2\mathcal{E}$ restricted to V_m.

We want to show that the restriction map

$$R_{V_m, V_{m+1}} : {}^2g \to {}^2g\big|_{V_{m+1}}$$

is a surjective topological isomorphism of

$${}^2\mathcal{E}(V_m)\big/(\xi_{m+1} + iz_{m+1}){}^2\mathcal{E}(V_m) \to {}^2\mathcal{E}(V_{m+1}).$$

To verify this, suppose first that ${}^2g(\xi, z) \in {}^2\mathcal{E}(V_m)$ vanishes on V_{m+1}. V_{m+1} is the codimension 1 subvariety of V_m defined by $z_{m+1} = i\xi_{m+1}$. Since V_m is a linear variety and $z_{m+1} = i\xi_{m+1}$ is a linear equation it can be proven exactly as in the case $n = 1$ that ${}^2h(\xi, z) = {}^2g(\xi, z)/(\xi_{m+1} + iz_{m+1})$ is C^∞ on V_m and satisfies the desired growth and analyticity conditions.

Using a simple variant of (9.51) we obtain semi-local extensions of ${}^2h$ and ${}^2g$ to all (ξ, z).

Now, suppose ${}^2g \in {}^2\mathcal{E}$ vanishes on V. For each m we call ${}^2g_m$ we the restriction of ${}^2g$ to V_m. We know that ${}^2g_{n-1}$ vanishes on $V_n = V$ so we can write

$${}^2g_{n-1} = (\xi_n + iz_n)^2 h_{n-1}.$$

We denote by ${}^2g^*_{n-1}$ and ${}^2h^*_{n-1}$ semi-local extensions of ${}^2g_{n-1}$ and ${}^2h_{n-1}$ respectively. By construction

$${}^2g_{n-2} - {}^2g^*_{n-1} = {}^2g_{n-2} - (\xi_n + iz_n)^2 h^*_{n-1}$$

vanishes on V_{n-1}. Hence, using a simple extension of our above discussion of $R_{V_m, V_{m+1}}$ to semi-local "functions," we can write

$${}^2g_{n-2} - (\xi_n + iz_n)^2 h^*_{n-1} = (\xi_{n-1} + iz_{n-1})^2 h^*_{n-2}$$

on V_{n-2}.

We can now iterate the construction and obtain a semi-local isomorphism of R_V.

There remains the cohomology argument passing from semi-local to global. In FA this argument depended on using a weighted form of Cauchy's integral and then using an iteration one real variable at a time. In the present situation we can carry out this argument in the z variables. The ξ variables give no trouble because the functions we deal with are C^∞ in ξ so we can use partitions of unity.

In this way the proof of Theorem 9.9 is completed.

Let us return to the ideas of (9.48)ff.

We revise (9.49) by arbitrarily choosing one p for each k and denoting that $\hat{f}_k^{0p}$ simply by $\hat{f}_k^0$. We can apply Theorem 9.9 because, as we have remarked, $f_k^0 = f_k^{0p}$ is independent of p; that is, $\hat{f}_k^{0p}$ are different representations for the Fourier transform of the CD f_k^0. Thus we can write (9.49) in the form

$$\hat{f}(\hat{x}, \hat{y}_{l_0}^p(\hat{x})) = \sum_k \alpha_{l_0 k}^p(\hat{x}) \hat{f}_k^0(\hat{x}) + \sum_k \alpha_{l_0 k}^p(\hat{x}) \bar{\partial} \hat{\vec{w}}_k^p(\hat{x}).$$

We apply the inverse matrix $\left[\beta_{l_0 p}^k\right] = \left[\alpha_{l_0 k}^p\right]^{-1}$ to obtain an expression of the form

$$\hat{f}_k^0(\hat{x}) = \sum_{l_0 p} \beta_{l_0 p}^k(\hat{x}) \hat{f}(\hat{x}, \hat{y}_{l_0}^p(\hat{x})) + \sum_{l_0 p} \beta_{l_0 p}^k(\hat{x}) \alpha_{l_0 k}^p(\hat{x}) \bar{\partial} \hat{\vec{w}}_k^p(\hat{x}). \tag{9.53}$$

The $\beta_{l_0 p}^k$ are algebraic functions.

We call V *simple above* $\hat{\rho}^0$ if the s roots in V above $\hat{\rho}^0$ are distinct outside a compact set.

Lemma 9.10 *Suppose each V^p is simple above $\hat{\rho}^0$. Then there is a complex cone Δ containing $\hat{\rho}^0$ in its interior on which the functions $\alpha_{lk}^p(\hat{x}), \beta_{lp}^k(\hat{x})$ are holomorphic outside of some compact set.*

Proof According to (9.48)ff. the $\alpha_{lk}^p(\hat{x})$ are defined in terms of $\hat{\partial}_k^p(\hat{y}_l^p(\hat{x}))$. $\hat{\partial}_k^p$ are polynomials and $(\hat{x}, \hat{y}_l^p(\hat{x}))$ are the points in V^p above $\hat{x}$. Our assertion is equivalent to the regularity of $\hat{y}_l^p(\hat{x})$ and the independence of the roots $\{\hat{y}_{l_0}^p(\hat{x})\}$ (which is assumed). The places where roots can become nonholomorphic are at points of coincidence. Thus, all comes down to showing that the roots $\hat{y}_l^p(\hat{x})$ remain distinct on a cone Δ.

We return to the notation of Proposition 9.6. For a single polynomial P with distinct roots it follows from the continuity of roots as functions of the coefficients that we can distort $\hat{x}^0$ without affecting the distinct nature of the roots. Actually a somewhat stronger statement is true: if γ_0 is a simple root of $P^0(\hat{x}^0, \hat{y}/t)$ then for large t there is a simple root of P above $\hat{x}(t)$ which is close to $\gamma_0 t$. By varying $\hat{x}^0$ in the unit sphere we obtain the desired cone Δ.

This proves Lemma 9.10 when $r = 1$. The reduction from $r > 1$ to $r = 1$ proceeds along the lines of the proof of Proposition 9.6.

From Lemma 9.10 we deduce that all the functions $\alpha_{l_0 k}^p$, $\beta_{l_0 p}^k(\hat{x})$, $\hat{y}_{l_0}^p(\hat{x})$ are holomorphic in a complex cone Δ containing $\hat{\rho}^0$ outside a compact set.

We can write

$$\hat{\vec{w}}_k^p(\hat{x}) = \hat{\vec{w}}_{k1}^p(\hat{x}) + \hat{\vec{w}}_{k2}^p(\hat{x})$$

where

$$\text{support } \hat{\vec{w}}_{k1}^p \subset \bar{\Delta}$$

$$\text{support } \hat{\vec{w}}_{k2}^p \subset \text{complement } \Delta.$$

Since support $\bar\partial\hat{\vec w}^p_{k2} \subset$ complement Δ it follows from Theorem 5.14 that

$$\text{wave front set } \sum \beta^k_{l_0p}(\hat x)\widehat{\alpha^p_{l_0k}}(\hat x)\bar\partial\hat{\vec w}^p_{k2}(\hat x) \text{ omits } \hat\rho^0.$$

Remark. *Strictly speaking, Theorem 5.14 requires that $\hat{\vec w}$ be a measure while Theorem 9.9 produces $\hat w$ as a distribution. A simple modification of the proof of Theorem 5.14 allows us to pass from measures to distributions.*

On the other hand, all $\gamma^p_k(\hat x)$ are regular on support $\hat{\vec w}^p_{k1}(\hat x)$. Thus

$$\beta^k_{l_0p}(\hat x)\alpha^p_{l_0k}(\hat x)\bar\partial\hat{\vec w}^p_{k1}$$

when considered as a linear function on $\hat{\mathcal E}'(\hat x)$ is zero since $\bar\partial$ kills holomorphic functions. Putting things together,

$$\text{wave front set } \vec f^0 \text{ omits } \hat\rho^0.$$

We have thus established our main result

Theorem 9.11 *Suppose $\vec f^0$ is the CD for functions f^p which satisfy $\vec P^p f^p = 0$ on the tube over Γ^p. Let $\hat\Gamma$ be a cone in $\hat x$ space and suppose that above every real ray $\hat\rho^0$ in $\hat\Gamma$ there are s independent elliptic rays $\hat\rho^p_l$ for which $\Im\hat y^p_l(\hat x) \in c\left[(\Gamma^p)'\right]$ and each V^p is simple over $\hat\rho^0$. Then the wave front set of $\vec f^0$ is contained in the complement of $\hat\Gamma$.*

Combining Theorems 9.11 and 5.15 we obtain

Theorem 9.12 *Under the hypotheses of Theorem 9.11, if $\hat\Gamma$ contains the complement of a proper convex cone ${\hat\Gamma_1}'$ then $\vec f^0$ is holomorphic in the tube over Γ_1.*

Remark. *Although the notion of wave front set is local, if we trace the proof of Theorem 9.11 we find that, in actuality, $\vec f^0$ decreases faster than every exponential when $\hat\Gamma_1'$ is empty. Thus $\vec f^0$ is entire so the global Cauchy–Kowalewskii theorem applies (Chapter IX of FA).*

We deduce, via the Cauchy–Kowalewski theorem,

Theorem 9.13 *If the cone $\hat\Gamma$ in Theorem 9.12 is all of $\mathbb{R}^n$ then each solution f^p extends to an entire solution $\tilde f^p$ on all (x, y).*

Theorem 9.12 contains the Hartogs–Bernstein separate analyticity theorem for $\bar\partial$ (see Section 5.2). We set $L =$ real space and $\hat P^p = \partial/\partial\bar z_p$. We call Γ_1 the positive orthant and $\hat\Gamma = c(\hat\Gamma_1')$; Γ^p is the cone $\{y_p \geq 0\}$ for $p = 1, \ldots, n$. For any real ray $\hat\rho^0 = \{t\hat x^0\}$ contained in $\hat\Gamma$ one of the components, say $\hat x^0_p$, is negative. In $\hat L + \hat\Gamma_p$ we have

$$-\Im\hat y_p = -\hat x_p \geq c|\hat x| \qquad \text{on } \hat\rho.$$

Theorem 9.12 shows that f^0 is holomorphic in the tube over Γ_1.

We can formulate a "separate analyticity" theorem for a general elliptic system $\vec{P}$. As we have seen in the proof of Theorems 9.10 and 9.11 the "generic" projection V_p of V on $\hat{L} + \{\hat{y}_p\}$ is a hypersurface which is an s sheeted elliptic cover of $\hat{L}$. ("Generic" means all linear transforms of the coordinates $\{\hat{y}_p\}$ except for a lower dimensional set satisfy the condition.) Let $P^p(\hat{x}, \hat{y}_p) = 0$ define the closure $\bar{V}_p$ of the projection.

$\bar{V}_p$ can be defined as the smallest algebraic variety in $\hat{L} + \{\hat{y}_p\}$ whose inverse projection contains V. It is a standard theorem in algebraic geometry that V_p is a Zariski open subset of $\bar{V}_p$, meaning $\bar{V}_p - V_p$ is a union of a finite number of algebraic varieties of dimension lower than $\dim V_p$.

In particular P^p (extended to be constant in $(\hat{y}_1, \dots, \hat{y}_{p-1}, \hat{y}_{p+1}, \dots, \hat{y}_m)$) belongs to the ideal $I(\vec{P})$ generated by the P_j. Moreover P^p generates the ideal $I_p(\vec{P})$ of all polynomials Q in $I(\vec{P})$ which do not depend on $(\hat{y}_1, \dots, \hat{y}_{p-1}, \hat{y}_{p+1}, \dots, \hat{y}_m)$ and vanish on V, since such a Q vanishes on $\bar{V}_p$.

Let $\vec{P}f = 0$. By the fundamental principle f has a Fourier representation by a measure μ supported on V. The projection–slice theorem says that the restriction f_p of f to $L + \{y_p\}$ is the Fourier transform of the projection μ_p of μ on $\hat{L} + \{\hat{y}_p\}$. μ_p is the Radon transform of μ on the spread in $\hat{Y}$ whose leaves are $\hat{L} + \hat{y}_p^0$.

Clearly $P^p(D)f_p = 0$. We call $P^p(D)h = 0$ the equations induced by $\vec{P}(D)$ on $L + \{y_p\}$.

We deal with the simplest general "separate analyticity" theorem. The reader should have no difficulty in finding more refined forms.

Since L is a hyperplane in $L + \{y_p\}$ the only "reasonable" CP for $P^p(D)$ is the usual CP. We can verify the hypothesis Theorem 9.11. Thus we have

Proposition 9.14 *Let $\vec{P}(D)$ be an elliptic system with CS L of dimension n. V is an s sheeted cover of $\hat{L}$, the s sheets being distinct at infinity over real $\hat{L}$. Let $\hat{\rho}$ be a real ray in $\hat{L}$ and suppose that there is a $p = p(\hat{\rho})$ such that*

$$\Im \hat{y}_l^p(\hat{x}) \in c[(\Gamma^p)']$$
$$|\Im \hat{y}_l^p(\hat{x})| \geq c|\hat{x}| \qquad \hat{x} \in \hat{\rho}, \text{ all } l.$$

If $\vec{f}_p^0(x)$ is CD for $P^p(D)$ then $\hat{\rho}$ is not in the analytic wave front set of any component of $\vec{f}^0$.

If this holds for all $\hat{\rho}$ then $\vec{f}^0$ is real analytic.

Our assumption that there is a single p which works for all l is used to establish the independence of the good roots.

Problem 9.4 *Find other hypotheses to guarantee the independence of the good roots.*

We can make things slightly more precise as follows: V is an s sheeted covering of $\hat{L}$. It is not difficult to see that if the coordinates $\hat{y}, \dots, \hat{y}_s$ are chosen

generically then, for each p, P^p is a polynomial of degree s for which L is noncharacteristic.

We can use L as a CS for each P^p. Since P^p is a single operator we use the usual CP. The differential operators are (identity, $\partial/\partial y_p, \ldots, \partial^{s-1}/\partial y_p^{s-1}$). $\vec{f}_p^0$, which is CD(f_p) for P^p, can be easily expressed in terms of $\vec{f}^0$ in conformity with the definition of CP (see (1.110)).

Lewy (see [116]) was the first to consider different operators $\vec{P}^p$. He showed that ($n = 2$) if f^1 is harmonic in $y > 0$ and $(\Delta + 1)f^2 = 0$ in $y < 0$ and $\vec{f}^0 = \text{CD}\ (f^1) = \text{CD}\ (f^2)$ then $\vec{f}^0$ is entire. Lewy's theorem was extended to $n > 2$ by Kinderlehrer and Nirenberg, and by the author (see [55, 56]).

Let us show how Lewy's theorem for arbitrary n follows from Theorem 9.12. The varieties are ($x = (x_1, \ldots, x_{n-1})$)

$$V^1 = \{x^2 + y^2 = 0\}$$
$$V^2 = \{x^2 + y^2 = 1\}.$$

Formally the Fourier transform of the CD $\vec{f}^0$ satisfies

$$\begin{aligned}
\hat{f}^1\left(\hat{x}, i\sqrt{\hat{x}^2}\right) + \hat{f}^1\left(\hat{x}, -i\sqrt{\hat{x}^2}\right) &= \hat{f}_0^0(\hat{x}) \\
i\sqrt{\hat{x}^2}\hat{f}^1\left(\hat{x}, i\sqrt{\hat{x}^2}\right) - i\sqrt{\hat{x}^2}\hat{f}^1\left(\hat{x}, -i\sqrt{\hat{x}^2}\right) &= \hat{f}_1^0(\hat{x}) \\
\hat{f}^2\left(\hat{x}, \sqrt{1-\hat{x}^2}\right) + \hat{f}^2\left(\hat{x}, -\sqrt{1-\hat{x}^2}\right) &= \hat{f}_0^0(\hat{x}) \\
\sqrt{1-\hat{x}^2}\hat{f}^2\left(\hat{x}, \sqrt{1-\hat{x}^2}\right) - \sqrt{1-\hat{x}^2}\hat{f}^2\left(\hat{x}, -\sqrt{1-\hat{x}^2}\right) &= \hat{f}_1^0(\hat{x}).
\end{aligned} \tag{9.54}$$

We want to show that the wave front set of $\vec{f}^0$ is empty. By symmetry it suffices to show that the positive $\hat{x}_1$ axis does not belong to the wave front set. We write $\hat{x}_1$ for $(\hat{x}_1, 0)$.

The roots are all elliptic. The "good" roots, that is the roots for which there is exponential decrease, are $-i\hat{x}_1$ for $\hat{f}^1$ and $\sqrt{1-\hat{x}_1^2}$ for $\hat{f}^2$. Thus we want to express $\vec{\hat{f}}^0$ in terms of $\hat{f}^1(\hat{x}_1, -i\hat{x}_1)$ and $\hat{f}^2(\hat{x}_1, \sqrt{1-\hat{x}_1^2})$.

We can invert (9.54) to obtain

$$\begin{aligned}
\hat{f}^1\left(\hat{x}_1, i\hat{x}_1\right) &= \frac{1}{2i\hat{x}_1}\left[i\hat{x}_1\hat{f}_0^0(\hat{x}_1) - \hat{f}_1^0(\hat{x}_1)\right] \\
\hat{f}^2\left(\hat{x}_1, -\sqrt{1-\hat{x}_1^2}\right) &= \frac{1}{2\sqrt{1-\hat{x}_1^2}}\left[\sqrt{1-\hat{x}_1^2}\hat{f}_0^0(\hat{x}_1) + \hat{f}_1^0(\hat{x}_1)\right].
\end{aligned} \tag{9.55}$$

The determinant of the coefficients is

$$\frac{1}{4}\left(\frac{1}{\sqrt{1-\hat{x}_1^2}} + \frac{1}{i\hat{x}_1}\right) \tag{9.56}$$

which is not $\equiv 0$. Thus the good roots are independent, which completes the proof of Lewy's theorem.

The above considerations involved analytic wave front sets which correspond to elliptic zeros and exponential decrease. There is no difficulty in extending Theorem 9.11 to Gevrey wave front sets. Equation (9.45) are replaced by (Gevrey α hypoelliptic zeros)

$$|\Im\hat{y}(\hat{x})| \geq c(|\hat{x}| + |\hat{y}|)^{1/\alpha}. \tag{9.45$^\alpha$}$$

It is not difficult to show that (9.45^1) is equivalent to (9.45).

However we cannot derive an analog of Theorem 9.13 unless we also have Gevrey α hyperbolicity (see Chapter VIII of FA). This means that there is a p^0 for which, on V^{p^0},

$$|\Im\hat{y}(\hat{x})| \leq c\left[|\Im\hat{x}| + (|\hat{x}| + |\hat{y}|)^{1/\alpha}\right]. \tag{9.45$^{\alpha *}$}$$

Inequality (9.45$^{\alpha *}$) replaces the Cauchy–Kowalewski theorem which was used in the elliptic case.

If (9.45$^\alpha$) and (9.45$^{\alpha *}$) hold then f^{p^0} can be extended to a solution of $\vec{P}^{p^0} f^{p^0} = 0$ in all (x, y) (locally or globally).

Conversely let α be the smallest exponent for which (9.45$^\alpha$) holds. If (9.45$^{\alpha *}$) does not hold then there exists potential CD $\vec{f}^{p^0} \in C^\infty$ satisfying all the above conditions but $\vec{f}^{p^0}$ does not extend to a Gevrey α solution of $\vec{P}^{p^0} f^{p^0} = 0$ on all (x, y).

Note that when $\alpha = 1$ condition (9.45$^{\alpha *}$) is vacuous while for $\alpha = \infty$ (which means the Gevrey space is C^∞) condition (9.45$^\alpha$) is vacuous and (9.45$^{\alpha *}$) becomes the usual hyperbolicity condition; that is, $(|\hat{x}| + |\hat{y}|)^{1/\alpha}$ is replaced by 1.

The above ideas pertain to the noncharacteristic CP. As explained in Section 5.4 it is natural to deal with Φ wave front sets as for suitable Φ they define uniqueness spaces for the CP. We assume $\Phi(x) = \Phi(|x|)$, i.e. Φ is rotationally invariant.

To clarify matters let us examine the heat equation

$$\left(\frac{\partial}{\partial y} - \frac{\partial^2}{\partial x^2}\right) f = 0. \tag{9.57}$$

The corresponding variety is

$$V = \{\hat{y} = i\hat{x}^2\}. \tag{9.58}$$

For $\hat{x}$ real

$$|\Im\hat{y}| \geq c\left(|\hat{y}| + |\hat{x}|^2\right). \tag{9.59}$$

Thus above the real $\hat{x}$ axis

$$\hat{f}(\hat{x}, i\hat{x}^2) = \mathcal{O}\left(e^{-c|\hat{x}|^2}\right). \tag{9.60}$$

Moreover $\Im\hat{y} > 0$ so we are in the dual cone to $y > 0$. This part of the integral for the Fourier inverse of $\hat{f}$ contributes a function which is an entire function of order 2 in x.

However, consider the portion of V which lies over the part of the complex $\hat{x}$ plane where $\hat{x} = \xi + i\eta$ and ξ is close to η. $\Im\hat{y} = \hat{x}^2 = \xi^2 - \eta^2$ is small there so the decrease of $\hat{f}$ is governed by η, which means

$$\Im(\hat{x}, \hat{y}) \sim |\hat{x}|. \tag{9.61}$$

The inverse Fourier transform of this part of $\hat{f}$ is an entire function in x. We cannot say that this entire function satisfies any growth condition. Indeed it is proven in [43] and FA that any entire function $h(x)$ is the CD for a (nonunique) solution of the heat equation for all y.

This result is proven (for $n = 1$) as follows. We use the fact that the union of the lines $\xi = \pm\eta$ constitute a sufficient set for the space $\mathcal{H}$ of entire functions. This means that h can be represented as

$$h(x) = \int e^{ix\hat{x}} d\mu(\hat{x})$$

where support $\mu \subset \{\xi = \pm\eta\}$ and μ is exponentially decreasing. The solution f of the heat equation with CD $(f) = h$ is

$$f(x, y) = \int e^{ix\hat{x} - y\hat{x}^2} d\mu(\hat{x}).$$

Since $\hat{x}^2$ is pure imaginary on $\xi = \pm\eta$ the integral converges and defines a solution of the heat equation.

Suppose that f is a solution of the heat equation satisfying

$$f^{(j)}(x, y) = \mathcal{O}\left(e^{\Phi(\alpha x)}\right) \tag{9.62}$$

for all j, α uniformly for y in compact sets. This means that $f \in \mathcal{E}_x(\Phi) \otimes \mathcal{E}_y$. By examining the AU structure for this space of f we find (see Chapter V of FA)

$$\hat{f}(\hat{x}, \hat{y}) = \mathcal{O}\left[(1 + |\hat{x}| + |\hat{y}|)^{-N} e^{-\Psi(N\Im\hat{x}) - N|\Im\hat{y}|}\right] \tag{9.63}$$

for all N. Ψ is the conjugate of Φ. That is (we assume Φ is even), $\Psi(t) = \max_u [|tu| - \Phi(u)]$. If f is a solution of the heat equation then support $\hat{f} \subset V$ so

$$\hat{f}(\hat{x}, i\hat{x}^2) = \mathcal{O}\left[(1 + |\hat{x}|)^{-N} e^{-\Psi(N\eta) - N|\xi^2 - \eta^2|}\right]. \tag{9.64}$$

For the space $\mathcal{H}_x(\Phi)$ of entire functions bounded by $\exp \Phi(\alpha|x|)$ for all α the role of the function $\Psi(\Im\hat{x})$ is replaced by $\Psi(|\hat{x}|)$. The right side of (9.64) is

certainly $\mathcal{O}\left[\exp -\Psi((N|\hat{x}|))\right]$ if either

(1) $|\eta| \geq c|\xi|$.
(2) $|\xi^2 - \eta^2| \geq \Psi(c|\hat{x}|)$.

If (1) does not hold then $|\xi^2 - \eta^2| \sim \xi^2 \sim |\hat{x}|^2$. This term in the exponential in (9.64) contributes $\mathcal{O}\left[\exp -\Psi((N|\hat{x}|))\right]$ if $\Psi(t) \leq ct^2$. We have shown (using the fact that the conjugate of t^2 is $\frac{1}{4}t^2$)

Proposition 9.15 *If $t^2 = \mathcal{O}(\Phi(ct))$ then any solution of the heat equation in $\mathcal{E}_x(\Phi) \otimes \mathcal{E}_y$ belongs to $\mathcal{H}_x(\Phi) \otimes \mathcal{E}_y$. In particular the (analytic) Φ wave front set of the CD is empty.*

We want to use this result on the heat equation as a paradigm for a characteristic analog of Theorems 9.11, 9.12, and 9.13. Notice that V is of degree 2 and $|\Im\hat{y}| \geq c|\hat{x}|^2$ for $\hat{x}$ real. This is a strong analog of an elliptic cover of the real axis. In the noncharacteristic case we used the weaker form $|\Im\hat{y}| > c|\hat{x}|$. We shall explore the general setting of both possibilities. We shall find that, in fact, there are three notions of ellipticity.

Suppose that for $(\hat{x}, \hat{y}) \in V$

$$|\hat{y}| \leq c(1 + |\hat{x}|)^a. \tag{9.65}$$

Let a be the minimal exponent with this property; it can be shown that a is rational. If $a > 1$ we say that L is a *characteristic of order a for V*. If $a \leq 1$ we call L a *noncharacteristic of order a*. $a = 1$ corresponds to an ordinary noncharacteristic.

For a polynomial $P(\hat{x}, \hat{y})$ of the form

$$P(\hat{x}, \hat{y}) = \hat{y}^m + Q_1(\hat{x})\hat{y}^{m-1} + \cdots + Q_m(\hat{x})$$

($\hat{y}$ is a single variable) it is easy to see that

$$a = \max \frac{\text{degree}\, Q_j}{j}.$$

We shall assume in what follows that the P_j are all of the form of the above P with $a = \max a_j$. In this case we say that $\hat{L}$ is a characteristic of order a for $\vec{P}$.

We say that V is a *strong elliptic cover of order a* of the ray $\hat{\rho}^0 \subset \hat{L}$ if

$$\max_{(\hat{x}(t), \hat{y}) \in V} |\Im\hat{y}| \geq c(1 + t^a). \tag{9.66}$$

As usual $\hat{\rho}^0$ is defined by $\hat{x}(t) = t\hat{x}^0$ with $|\hat{x}^0| = 1$.

We call $\hat{\rho} \subset V$ a *strong elliptic ray over $\hat{\rho}^0$ of order a* if $\hat{\rho}$ is an algebraic curve of the form $(\hat{x}(t), ct^a + \text{lower order})$ for t large, and c is not real.

We assume again that V is an s sheeted cover of $\hat{L}$. The analog of Proposition 9.6 is

Proposition 9.16 *Suppose $\hat{\rho}^0$ has a strong elliptic cover. Then there is a strongly elliptic ray $\hat{\rho}$ over $\hat{\rho}^0$. We can vary $\hat{\rho}^0$ in a complex cone and there is a corresponding variation of the covering $\hat{\rho}$.*

Proof As in Proposition 9.6 we can reduce the general case to the case $r = 1$.

We write

$$P(\hat{x}(t), \hat{y}) = \hat{y}^s + Q_1^0(\hat{x}^0)t^a\hat{y}^{s-1} + \cdots + Q_s^0(\hat{x}^0)t^{as} + Q^1(\hat{x}(t), \hat{y}).$$

$\hat{y}^s$ and the $Q_j^0\hat{y}^{s-j}$ constitute the *pseudo-principal part* of P, which means $Q_j^0(\hat{x})$ is the homogeneous part of Q_j when Q_j is of degree aj. (By definition aj is an integer for some j.) The degree of Q^1 in $\hat{y}$ is $< s$ and the coefficient of $\hat{y}^j$ is of degree $< aj$.

We can write the pseudo-principal part of P in the form

$$\begin{aligned} P^0(\hat{x}(t), \hat{y}) &= t^{as}\left[\left(\frac{\hat{y}}{t^a}\right)^s + Q_1^0(\hat{x}^0)\left(\frac{\hat{y}}{t^a}\right)^{s-1} + \cdots + Q_s^0(\hat{x}^0)\right] \\ &= t^{as}\Pi\left(\frac{\hat{y}}{t^a} - \gamma_j\right). \end{aligned}$$

Our assumption is that some γ_j (say γ_1) is not real. As in the proof of Proposition 9.6 the nonreality of γ_1 cannot be disturbed by Q^1.

We have shown that the root

$$\hat{y} \sim \gamma_1 t^a + \text{lower order}$$

defines a strongly elliptic ray over $\hat{x}^0$. Varying $\hat{x}^0$ allows us to vary $\hat{y}$ keeping γ_1 not real. This proves Proposition 9.16 when $r = 1$.

As mentioned, the passage to $r > 1$ is completed as in Proposition 9.6.

Recall that the hypothesis $t^2 = \mathcal{O}(\Phi(ct))$ in Proposition 9.15 was necessary to guarantee that for any N there is an N^1 such that

$$\Psi(N^1|\Im\hat{x}|) + N^1|\Im\hat{y}| \geq \Psi(N|\hat{x}|). \tag{9.67}$$

We do not know if the analog of Lemma 9.10 is valid in the characteristic case, so we shall assume it.

We can prove the following results exactly as in the noncharacteristic case.

Theorem 9.17 *Let $\hat{L} = \{\hat{x}\}$ be characteristic of order $\leq a$ for the algebraic varieties V^p which are s sheeted covers of $\hat{L}$. Suppose there exist s linearly independent, strongly elliptic roots $\hat{y}_{l_0}^{p\pm}(\hat{x})$ above $\pm\hat{\rho}^0$ amongst the V^p. Assume the V^p are simple over complex cones $\pm\Delta$ containing $\pm\hat{\rho}^0$ and also that*

$$\Psi(t) \leq c(1 + t^a).$$

Let $\{f_j(x)\}$ be CD for solutions $f^p \in \mathcal{E}_x(\Phi) \otimes \mathcal{E}(\Gamma^p_\pm)$ of $\vec{P}^p f^p = 0$. If $\hat{y}^{p\pm}_{l_0}(\hat{x})$ are contained in $c[(\Gamma^p_\pm)']$ then the line $\pm\hat{\rho}^0$ does not belong to the analytic Φ wave front set of any f_j.

Theorem 9.18 *If the union of the $\pm\hat{\rho}^0$ in the previous theorem is all of $\mathbb{R}^n$ then $\vec{f}^0 \in \mathcal{H}^*(\Phi)$.*

Remark. *Theorem 9.18 is the analog of Theorem 9.13; we do not have an analog of Theorem 9.12 because, as mentioned in Section 5.3, we do not have a good concept of Φ holomorphicity in a cone.*

We have dealt with strong ellipticity which refers to roots $\hat{y}_l$ with $|\Im\hat{y}_l(\hat{x})| \geq c|\hat{x}|^a$. We now wish to treat ordinary ellipticity, meaning $|\Im\hat{y}_l(\hat{x})| \geq c|\hat{x}|$.

To understand the pitfalls one can encounter let us examine the variety V defined by

$$\hat{y} = \hat{x}^2 + i\hat{x}.$$

We write $\hat{x} = \xi + i\eta$. We have

$$\Im\hat{y} = 2\xi\eta + i\xi.$$

This means that $|\Im\hat{y}| = |\xi|$ when $\eta = 0$ but $\Im\hat{y} = 0$ when $\eta = -1/2$; in particular there is no complex cone containing the positive $\hat{x}$ axis on which $\Im\hat{y}$ is large.

Now consider the variety in $\mathbb{C}^3$

$$\hat{y} = \hat{x}_2^2 + i\hat{x}_1.$$

We set $\hat{x}_j = \xi_j + i\eta_j$:

$$\Im\hat{y} = 2\xi_2\eta_2 + \xi_1.$$

On the ξ_1 axis

$$|\Im\hat{y}| = |\xi_1| = |\hat{x}|.$$

In fact, if $|\xi_2| \leq \frac{1}{4}|\xi_1|^{1/2}$, $|\eta_2| \leq \frac{1}{4}|\xi_1|^{1/2}$, $|\eta_1| \leq \frac{1}{4}|\xi_1|$ then

$$|\Im\hat{y}| \geq c|\xi_1| \geq c(|\hat{y}| + |\hat{x}|).$$

The point is that the ellipticity of the root $\hat{y} = i\hat{x}_1$ can be destroyed by adding polynomials in $\hat{x}_1$ but if we add a polynomial in $\hat{x}_2, \ldots, \hat{x}_n$ of degree a then a sort of ellipticity persists in a complex set of the form

$$|\hat{x}_2| + \cdots + |\hat{x}_n| \leq c|\xi|^{1/a}$$
$$|\hat{x}_1| \leq c|\xi|.$$

We are not able to obtain a full cone in $\Re\hat{x}$. According to the ideas of Section 5.3 it seems that this (weak) ellipticity can only treat C^∞ and Gevrey wave front sets.

There is a third type of ellipticity which deals with noncharacteristics of order $a < 1$.

Again we examine the heat equation but this time we interchange the roles of x and y so the corresponding polynomial is $i\hat{x} + \hat{y}^2$. We study solutions f which belong to $\mathcal{E}_x \otimes \mathcal{E}_y(\Phi)$. Such a function has a Fourier transform $\hat{f}$ supported on $V : \{\hat{y} = \pm\sqrt{-i\hat{x}}\}$ satisfying inequalities of the form

$$|\hat{f}(\hat{x}, \pm\sqrt{-i\hat{x}})| \leq c(1 + |\hat{x}|)^{-N} e^{-N|\Im\hat{x}| - \Psi(N\Im\pm\sqrt{-i\hat{x}})}$$

for all N.

For $\hat{x}$ real $|\Im\sqrt{-i\hat{x}}| \sim \sqrt{|\hat{x}|}$, and a similar estimate persists as long as $\hat{x}$ lies in an angle $< \pi/2$ around the positive and negative real rays. If $\Psi(t) \leq ct^2$ then when we are outside such an angle the term $|\Im\hat{x}|$ majorizes $c\Psi(\sqrt{|\hat{x}|})$ so the Ψ term plays no role there.

We conclude

$$|\hat{f}(\hat{x}, \pm\sqrt{-i\hat{x}})| \leq c(1 + |\hat{x}|)^{-N} e^{-N|\Im\hat{x}| - \Psi(N\sqrt{|\hat{x}|})}. \tag{9.68}$$

This means that $f(x)$ belongs to the Denjoy–Carleman space defined by $\exp(-\Psi(\sqrt{|\hat{x}|}))$ (see Section 5.3). This class is QA when

$$\int \frac{\Psi(\sqrt{t})}{1+t^2}\, dt \sim \int \frac{\Psi(t)}{1+t^3}\, dt = \infty.$$

This corresponds to Taeklind's uniqueness theorem for solutions of the heat equation: solutions in $\mathcal{E}_x \otimes \mathcal{E}_y(\Phi)$ are uniquely determined by their CD if and only if the above integral diverges. (For a detailed study of this question see Chapter IX of FA.)

Problem 9.5 *Study the Denjoy–Carleman wave front sets of simultaneous CD for solutions f^p of $\vec{P}^p f^p = 0$ on $\mathbb{R}^n \times \Gamma^p$ where $f^p \in \mathcal{E}_x \otimes \mathcal{E}_y(\Phi, \Gamma^p)$.*

9.4 Solutions of the Cauchy problem in a generalized sense

We mentioned three types of generalized solutions of the CP in Section 9.1. Let us examine (1).

The problem can be formulated as follows: let f be a C^∞ solution of $\vec{P}f = 0$ in the interior of $L + \Gamma$. Suppose

$$f(x, y) = \mathcal{O}\left(e^{\xi(y/\epsilon)}\right) \quad y \in \Gamma, y \to 0 \tag{9.69}$$

for all $\epsilon > 0$ with similar inequalities for all derivatives of f. ξ is a continuous positive convex function of $|y|$ which $\to \infty$ as $y \to 0$.

A. Under what conditions does the CD of f on $y =$ const. have a limit when $y \in \Gamma, y \to 0$?
B. Can one give a meaning to the wave front set of such limit CD?
C. In case the answer to B is positive, is there an analog of Theorem 9.11?

We shall assume that L is noncharacteristic and the CP for $\vec{P}$ has the same structure as in the previous section. There is no difficulty in extending our ideas to Φ wave front sets in the characteristic case.

As in the previous section we shall discuss analytic wave front sets and corresponding elliptic zeros and exponential decrease. We leave to the reader the extension of our ideas to Gevrey wave front sets.

To attack these problems we introduce the space $\mathcal{E}(\Gamma, \xi)$ of C^∞ functions $h(y)$ defined for $y \in$ interior Γ such that for any A, j, ϵ

$$\sup_{\substack{y\in\Gamma \\ |y|\le A}} \left|h^{(j)}(y)\right| e^{-\xi(y/\epsilon)} < \infty. \tag{9.70}$$

$h^{(j)}$ represents some derivative of h. The topology of $\mathcal{E}(\Gamma, \xi)$ is defined by using the sup in (9.70) as the semi-norms.

Remark. *The difference between requiring (9.69) for all $\epsilon > 0$ or for some $\epsilon > 0$ is insignificant because the properties of ξ which interest us are "stable."*

To understand the Fourier transform of $\mathcal{E}'(\Gamma, \xi)$ consider the case $n = 1$ and $\Gamma =$ positive y axis. The usual heuristic approach to the growth of functions in $\hat{\mathcal{E}}'(\Gamma, \xi)$ is to study bounds on the Fourier transform of $\left\{\delta_y^{(k)} \exp[-\xi(y)]\right\}$ which are independent of $y \in (0, y^0]$ since, roughly speaking, the Hahn–Banach theorem implies that every element of $\mathcal{E}'(\Gamma, \xi)$ is an integral of such punctual distributions. The change from $\hat{\delta}_y$ to $\hat{\delta}_y^{(k)}$ only involves multiplication by $\hat{y}^k$ so we can restrict our considerations to $k = 0$.

Writing $t = \Im \hat{y}$, we have

$$\begin{aligned} \left|\widehat{\delta_y e^{-\xi(y)}}(\hat{y})\right| &= \left|e^{iy\hat{y}-\xi(y)}\right| \\ &= e^{-yt-\xi(y)}. \end{aligned} \tag{9.71}$$

For t negative the right side is dominated by $c \exp(y^0|t|)$ for all y. For $t > 0$ we must compute

$$\min_{y>0}[yt + \xi(y)] = \eta(t). \tag{9.72}$$

$\eta(t)$ is a *concave* function. For example, if $\xi(y) = y^{-p}$, $p > 0$ then $\eta(t) = t^q$ where $1/q - 1/p = 1$ so $0 < q < 1$. The inverse of (9.72) is given by

$$\xi(y) = \max[\eta(t) - yt]. \tag{9.73}$$

We conclude that, for t positive,

$$\left|\widehat{\delta_y e^{-\xi(y)}}\right| \le e^{-\eta(t)}.$$

This heuristic argument points the way to the actual result. Using the methods of Chapter 5, Example 7 of FA we conclude that the space $\mathcal{E}(\Gamma, \xi)$ is LAU

(see also Section 1.4). More precisely (we now allow $n \geq 1$ and ξ a function of $|y|$)

$\hat{\mathcal{E}}'(\Gamma, \xi)$ consists of all entire functions $\hat{f}(\hat{y})$ which satisfy

$$|\hat{f}(\hat{y})| \leq c(1+|\hat{y}|)^c \begin{cases} \exp[-\eta(c\Im\hat{y})] & \Im\hat{y} \in \Gamma' \\ \exp(c|\Im\hat{y}|) & \text{all } \hat{y} \end{cases} \tag{9.74}$$

for some constant $c > 0$.

An AU structure (see Section 1.4) for $\mathcal{E}(\Gamma, \xi)$ consists of all continuous positive functions which dominate the right side of (9.74) for each c.

We are interested in the kernel of $\vec{P}$ in

$$\mathcal{W}' = \mathcal{E}(L) \otimes \mathcal{E}(\Gamma, \xi).$$

$\mathcal{W}$ is LAU with the natural AU structure.

By (9.74) the growth condition in $\hat{\mathcal{W}}$ is

$$|\hat{f}(\hat{x}, \hat{y})| \leq c(1+|\hat{x}|+|\hat{y}|)^c \begin{cases} \exp[c|\Im\hat{x}| - \eta(c\Im\hat{y})] & \Im\hat{y} \in \Gamma' \\ \exp[c(|\Im\hat{x}| + |\Im\hat{y}|)] & \Im\hat{y} \notin \Gamma'. \end{cases} \tag{9.75}$$

Since L is noncharacteristic the roots $\hat{y}_l(\hat{x})$ of $\vec{\hat{P}}(\hat{x}, \hat{y}) = 0$ satisfy

$$|\Im\hat{y}| \leq |\hat{y}| \leq c|\hat{x}|.$$

This suggests that on V the space $\hat{\mathcal{W}}$ is closely related to the space $\hat{\mathcal{D}}_x(\eta) \otimes \hat{\mathcal{E}}'_y(\Gamma)$. $\mathcal{D}(\eta)$ is the Denjoy–Carleman space related to η (see Section 5.3). Precisely, we define the space $\mathcal{D}(\eta)$ to consist of all C^∞ functions ϕ of compact support for which there is a c such that

$$\hat{\phi}(\hat{x}) = \mathcal{O}\left[e^{-\eta(c\Re\hat{x}) + c|\Im\hat{x}|}\right]. \tag{9.76}$$

The topology of $\mathcal{D}(\eta)$ is defined in the natural (Schwartz) manner. $\mathcal{D}(\eta)$ is reduced to $\{0\}$ unless

$$\int \frac{\eta(t)}{1+t^2} dt < \infty \tag{9.77}$$

in which case it is dense in most spaces.

Thus the dual $\mathcal{D}'(\eta)$ is defined only in the nonquasianalytic case (9.77).

The growth in $\hat{\mathcal{E}}'_y(\Gamma)$ is given by (9.74) with η replaced by 1. Thus the growth in $\hat{\mathcal{D}}_x(\eta) \otimes \hat{\mathcal{E}}'_y(\Gamma)$ is

$$|\hat{h}(\hat{x}, \hat{y})| \leq c(1+|\hat{y}|)^c e^{-\eta(c\Re\hat{x}) + c|\Im\hat{x}|} \begin{cases} 1 & \Im\hat{y} \in \Gamma' \\ e^{c|\Im\hat{y}|} & \Im\hat{y} \notin \Gamma'. \end{cases} \tag{9.78}$$

The important observation is that the fact that $y = 0$ is noncharacteristic implies that, on V,

$$\begin{aligned}\eta(\Im\hat{y}) &\le \eta(c|\Re\hat{x}| + c|\Im\hat{x}|)\\ &\le c\eta(c|\Re\hat{x}|) + c|\Im\hat{x}|\end{aligned}$$

if we assume that

$$\eta(t + t') \le \eta(t) + ct', \tag{9.79}$$

i.e.

$$\frac{d\eta}{dt}(t) \le c.$$

In this case the term $e^{-\eta(c\Re\hat{x})}$ of (9.78) is dominated by $\exp[c|\Im\hat{x}| - \eta(c\Im\hat{y})]$ on V so the right side of (9.78) is majorized by the right side of (9.75) on V; that is, the functions $k_{\mathcal{W}}$ in the AU structure for $\mathcal{W}$, which define the $\hat{\mathcal{W}}$ norms on V, are greater than those of $\hat{\mathcal{D}}_x(\eta) \otimes \hat{\mathcal{E}}'_y$. Thus $\hat{\mathcal{W}}(V) \supset \left[\hat{\mathcal{D}}_x(\eta) \otimes \hat{\mathcal{E}}'_y\right](V)$.

By duality, we have proven (see the discussion of comparison theorems in Chapter VIII of FA)

Theorem 9.19

$$\textit{kernel } \vec{P} \textit{ in } \mathcal{W}' \subset \textit{kernel } \vec{P} \textit{ in } \mathcal{D}'(\eta) \otimes \mathcal{E}(\Gamma) \tag{9.80}$$

when η satisfies (9.77) and (9.79).

It follows that when η satisfies (9.77) and (9.79) we are in a situation quite comparable to the ordinary theory exposed in the previous section. If $f(x, y) \in \mathcal{D}'(\eta) \otimes \mathcal{E}(\Gamma)$ then the distributions $f(x, y^0)$ have a limit in $\mathcal{D}'(\eta)$ as $y^0 \to 0$ in Γ. In particular the meaning of CD of solutions of $\vec{P}(D)f = 0$ is well defined. There is no difficulty in computing wave front sets and establishing the analog of Theorem 9.11.

Remark. *The convergence of the integral (9.77) can be shown (though not easily) to be equivalent to*

$$\int_0^1 \log \xi(y)\, d|y| < \infty. \tag{9.81}$$

For $\vec{P} = \bar{\partial}$, $n = 1$, Theorem 9.19 is due to Levinson [115] (see also [55, 56] and [19]).

In the QA case, i.e. when the integral (9.78) diverges, the space $\mathcal{D}(\eta)$ is reduced to $\{0\}$. It seems that there is no meaning to the CP or to CD. There is no analog of the edge-of-the-wedge theorem. In fact it is generally the case that if Γ_1 and Γ_2 are two proper cones which meet only at the origin then any pair of

functions f_j in the kernels of $\vec{P}$ on $\mathcal{W}'_j = \mathcal{E}(L) \otimes \mathcal{E}(\Gamma_j, \xi)$ can be approximated in the topology of $\mathcal{W}_1 + \mathcal{W}_2$ by global solutions f of $\vec{P}$ (see [55] for details).[5]

We can, however, restore meaning to the CP and the edge-of-the-wedge theorem if we replace $\mathcal{W}'$ by

$$\mathcal{W}'_\Phi = \mathcal{E}_L(\Phi) \otimes \mathcal{E}(\Gamma, \xi)$$

for suitable Φ. Then $\hat{\mathcal{D}}(\eta)$ is replaced by $\hat{\mathcal{D}}(\Psi; \eta)$ which is the space of entire functions $\hat{\phi}$ satisfying

$$\hat{\phi}(\hat{x}) = \mathcal{O}\left((1 + |\hat{x}|)^c e^{\Psi(c\Im\hat{x}) - \eta(c\Re\hat{x})}\right). \tag{9.82}$$

Problem 9.6 *What is the analog of Theorem 9.19 in the QA case?*

We now pass to (2) of Section 9.1. In order to make the question more precise, let ξ be a function defined for small $y \in \Gamma$ such that ξ is convex and $\xi \to \infty$ at the origin faster than $-N \log |y|$ for all N.[6] We define the space

$$\mathcal{E}^\infty_{\vec{P},\xi}(\Gamma) = \left\{ f \in \mathcal{E}(L + \Gamma) \,\middle|\, d[\vec{P}(D)f(x,y)] = \mathcal{O}(e^{-\xi(cy)}) \right\} \tag{9.83}$$

for all differentiations d of arbitrary order; c is uniform on compact sets of x and the constant in $\mathcal{O}$ depends on d. If d is restricted to be a "normal differential operator" defining CD^* as in (9.11*)ff., then we write $\mathcal{E}_{\vec{P},\xi}$ in place of $\mathcal{E}^\infty_{\vec{P},\xi}$. (We sometimes add x derivatives of order ≤ 2 for convenience.)

A function $f \in \mathcal{E}_{\vec{P},\xi}$ (resp. $f \in \mathcal{E}^\infty_{\vec{P},\xi}$) is called an *asymptotic solution* (resp. *strong asymptotic solution) of* $\vec{P}f = 0$ *of order* $e^{-\xi}$.

We shall investigate the following question:

What is the structure of the CD of $f \in \mathcal{E}_{\vec{P},\xi}(\Gamma)$*? In particular, is there an edge-of-the-wedge theorem?*

In order to understand what type of results are possible, let us examine a simplified version of our question. Let $n = 2$, and $\vec{P} = \bar{\partial}$ be the Cauchy–Riemann operator. To avoid complications we replace (x, y) by (θ, r) with $\{r \leq 1\}$ playing the role of Γ and $\{r = 1\}$ playing the role of the "edge" $\{y = 0\}$.

We write $f \in \mathcal{E}_{\vec{P},\xi}$ as a Fourier series

$$f(\theta, r) = \sum a_m(r) e^{im\theta}. \tag{9.84}$$

Since

$$\bar{\partial} = e^{i\theta} \frac{\partial}{\partial r} + \frac{ie^{i\theta}}{r} \frac{\partial}{\partial \theta} \tag{9.85}$$

[5]This result can be regarded as a sharpening of the classical Runge theorem which asserts that if A_1, A_2 are open sets in the complex plane whose closures are disjoint then functions f_2, f_2 which are holomorphic on A_1, A_2 respectively can be approximated simultaneously by polynominals.

[6]The reader should not confuse this ξ with the ξ used in (1).

we have

$$\begin{aligned}\bar{\partial} f &= \sum \left[a_m{}'(r) - \frac{m}{r} a_m(r) \right] e^{i(m+1)\theta} \\ &= \sum \left[\left(r^m \frac{d}{dr} r^{-m} \right) a_m(r) \right] e^{i(m+1)\theta}. \end{aligned} \tag{9.86}$$

From (9.83) we deduce that the Fourier series coefficients of $\bar{\partial} f(\theta, r)$ are bounded by $\exp\left[-c\xi(c(1-r))\right]$. Thus (ignoring the constant c)

$$\left(r^m \frac{d}{dr} r^{-m} \right) a_m(r) = \mathcal{O}(e^{-\xi(1-r)}). \tag{9.87}$$

Solving the ordinary differential inequality (9.87) yields

$$r^{-m} a_m(r) = a_m(1) + \mathcal{O}\left[\int_r^1 t^{-m} e^{-\xi(1-t)}\, dt \right]. \tag{9.88}$$

The left side of (9.88) represents the Fourier series coefficients of f. Since f is smooth (i.e. $f \in C^2$) up to $r = 1$, the Fourier coefficients $a_m(r)$ are bounded uniformly in $r \leq 1$. Hence the right side of (9.88) is also bounded when $m \leq 0$.

We want to show that $a_m(1)$ have a much greater vanishing as $m \to -\infty$. This entails finding good bounds on the integral.

For example, if $\xi(t) = -\infty$ for $t \in [0, 1-\alpha]$ then the integral vanishes for $r \geq \alpha$. Setting $r = \alpha$ we find

$$a_m(1) = \mathcal{O}(\alpha^{-m})$$

for $m < 0$. This means that $f(\theta, 1)$ extends to the function $\sum a_m(1) r^m \exp(im\theta)$ which is holomorphic for $r \in (\alpha, 1)$.

For another example, let

$$\xi(t) = \frac{1}{t}. \tag{9.89}$$

We want to use (9.88) to obtain an upper estimate on $a_m(1)$. We have $a_m(r) = \mathcal{O}(1)$ uniformly in r since this is the condition that $f \in C^2$. Thus we want to make r small since then (recall $m < 0$) r^{-m} is small. Making r smaller makes the integral in (9.88) larger (though not significantly).

Note that the maximum value of the integrand occurs (as seen by differentiation) at

$$t \sim 1 - \frac{1}{\sqrt{-m}}. \tag{9.90}$$

Thus there is no point in making r smaller than this value since smaller values would still encompass this value in the integral. For t defined by (9.90) the

integrand is

$$\sim \left(1 - \frac{1}{\sqrt{-m}}\right)^{-m} e^{-\sqrt{-m}}$$
$$\sim e^{-2\sqrt{-m}}. \tag{9.91}$$

The term $r^{-m}a_m(r)$ is also bounded essentially by $\exp(-\sqrt{-m})$. Thus

$$|a_m(1)| \le cr^{-m} + c|m|^{-1/2}e^{-2|m|^{1/2}}$$
$$\le ce^{-|m|^{1/2}} \quad (m \le 0). \tag{9.92}$$

Inequality (9.92) holds only for $m < 0$. In fact the condition (9.83) for $r < 1$ has no consequences for the growth of $a_m(1)$ as $m \to +\infty$.

We denote by $\mathbf{G}_2$ the Gevrey 2 space (see Section 5.3). Inequality (9.92) is the condition that the $\mathbf{G}_2$ wave front set of

$$f^0(\theta) = \mathrm{CD}\,(f)(\theta) = \sum a_m(1)e^{im\theta} \tag{9.93}$$

omits the negative m axis. (In Section 5.3 we dealt with Fourier integrals; this analog for Fourier series is, in fact, simpler.)

The usual Cauchy–Kowalewski theorem asserts that if we start with analytic CD for a differential operator on a noncharacteristic then there is an analytic solution of the equation near the noncharacteristic taking this CD. Using analytic wave front sets we can modify the statement to describe solutions in a one-sided neighborhood of the noncharacteristic.

We shall now derive

Cauchy–Kowalewski theorem for $\mathbf{G}_2$ (for the operator $\bar\partial$). *If $f^0(\theta) \in C^\infty$ and the Gevrey 2 wave front set of $f^0(\theta)$ omits the negative axis then there is a strong asymptotic solution $\tilde f^- \in \mathbf{G}_2$ of $\bar\partial$ of order $\exp(-c/(1-r))$ whose CD is $f^0(\theta)$.*

To prove this result observe that the functions

$$f^\pm(r,\theta) = \sum_{m=0}^{\pm\infty} a_m(1)r^m e^{im\theta} \tag{9.94}$$

are holomorphic in $r < 1$ (resp. $r > 1$) and have C^∞ boundary values since $a_m(1) = \mathcal{O}(|m|^{-k})$ for all k. Our object is to extend f^- to a Gevrey 2 function $\tilde f^-$ in the whole complex plane.

In order for $\tilde f^-$ to exist, its derivatives have to agree with those of f^- on $r = 1$. We have

$$\left|\frac{\partial^{j+k} f^-(1,\theta)}{\partial r^j \partial\theta^k}\right| \le c\sum_{-\infty}^{0} e^{-|m|^{1/2}}|m|^{j+k}$$
$$\le c^{j+k}[2(j+k)]!. \tag{9.95}$$

The last inequality comes from the fact that the sum is bounded by a fixed multiple of its maximum term, which can be computed by differentiation. From the binomial theorem we deduce

$$[2(j+k)]! \leq 2^{2(j+k)}(2j)!(2k)!. \tag{9.96}$$

We shall see the significance of (9.96) presently.

It is proven in Chapter XIII of FA and independently by B. S. Mityagin that any sequence $\{a_k\}_{k\geq 0}$ of complex numbers satisfying $|a_k| \leq c^k(2k)!$ can be represented in the form

$$a_k = h^{(k)}(0)$$

for some Gevrey 2 function h. Put in other terms, the map

$$h \to \{h^{(k)}(0)\} \tag{9.97}$$

is surjective from Gevrey 2 onto sequences satisfying the obvious necessary condition.

We want to extend $f^-(r,\theta)$ to a Gevrey 2 function. Since Gevrey spaces are non QA the problem is local so we can think of θ as a variable on a line segment. We are therefore led to the map

$$h(x,y) \to \left\{ \frac{\partial^j h(x,0)}{\partial y^j} \right\} \tag{9.98}$$

of the Gevrey 2 space $\mathbf{G}_2(x,y)$ into $\mathbf{G}_2(x) \otimes {}^0\mathbf{G}_2(y=0)$ where ${}^0\mathbf{G}_2$ is the space of sequences at the origin satisfying Gevrey 2 growth. Inequality (9.96) shows that

$$\left\{ \frac{\partial^j f^-(1,\theta)}{\partial r^j} \right\} \in \mathbf{G}_2(\theta) \otimes {}^0\mathbf{G}_2(r=1).$$

The harmonic majorant method used in Chapter 13 of FA shows that the real space is sufficient for $\mathbf{G}_2(x) \otimes {}^0\mathbf{G}_2(y=0)$. As explained in that work this sufficiency proves that the map (9.98) is surjective.

Let $\tilde{f}^- \in \mathbf{G}_2(r,\theta)$ be an extension of $\left\{ \frac{\partial^j f^-(1,\theta)}{\partial r^j} \right\}$. We can set $\tilde{f}^- = f^-$ in $r > 1$ since $f^- \in \mathbf{G}_2(r,\theta)$ there. All derivatives of $\bar{\partial}\tilde{f}^-$ at $r = 1$ are the same as those of $\bar{\partial}f^-$, and hence vanish. Moreover $\bar{\partial}f^+ = 0$ in the unit disk. Thus $\bar{\partial}(\tilde{f}^- + f^+)$ is a Gevrey 2 function in the unit disk all of whose derivatives vanish at $r = 1$. Moreover

$$\mathrm{CD}\,(\tilde{f}^- + f^+) = \mathrm{CD}\,(f).$$

It remains to show that a Gevrey 2 function (such as $\bar{\partial}(\tilde{f}^- + f^+)$) which vanishes together with all its derivatives at $r = 1$ is actually $\mathcal{O}[\exp(-c/(1-r))]$. By changing variables this is the same as saying (1 variable): $h \in \mathbf{G}_2$, $h^{(j)}(0) = 0$

implies $h(x) = \mathcal{O}[\exp(-c/x)]$. Using Taylor's formula with remainder this amounts to

$$\min_n \frac{x^n(2n)!}{n!} \sim \min_n (xn)^n \sim e^{-c/x}$$

which is the desired result.

We have shown that if f is an asymptotic solution in $r \leq 1$ of $\bar{\partial}$ of order $\exp(-c/(1-r))$ then there is a Gevrey 2 function $\tilde{f}$ whose formal power series at $r = 1$ agrees with that of f. In particular $\tilde{f}$ is an asymptotic solution of $\bar{\partial}$ with $\mathrm{CD}(\tilde{f}) = \mathrm{CD}(f)$.

We have established the equivalence amongst

(1) f is an asymptotic solution of order $\exp(-c/(1-r))$.
(2) The negative part of the Fourier series of $f(1,\theta) = \mathrm{CD}(f)$ belongs to Gevrey 2, i.e. the negative (m) axis does not belong to the Gevrey 2 wave front set of $\mathrm{CD}(f)$.
(3) There is a Gevrey 2 function $\tilde{f}$ on the disk which is a strong asymptotic solution of $\bar{\partial}$ of order $\exp(-c/(1-r))$ with $\mathrm{CD}(\tilde{f}) = \mathrm{CD}(f)$.

Our method has thus far been confined to the circle which is compact. What happens for the original (x, y)?

We start with a function $f \in \mathcal{E}_{\bar{\partial},\xi}(\Gamma)$ where Γ is the positive orthant in Y. We would like to represent f as a Fourier transform and apply the same idea as before. The problem is that $\hat{f}$ as given by the theory of AU spaces is not unique (Section 1.4) so we cannot use the same method as we used in the compact case to derive differential equations for $\hat{f}(\hat{x}^0, \hat{y})$ for each fixed $\hat{x}^0$.

It does not seem possible to overcome the nonuniqueness by establishing an analog of Theorem 9.9. Instead we restore uniqueness by using the nonlinear Fourier transform $F(\beta; \hat{x}; y)$ in x for fixed y. $F(\beta; \hat{x}; y)$ is the analog of the Fourier coefficients $a_m(r)$.

In defining the nonlinear Fourier transform we first cut off f smoothly, say near $|x| = 2$. Then $\bar{\partial}\phi f(x, y)$ is small for x small but may be large near $|x| = 2$. The term $\exp(-\beta x^2)$ (we are using the quadratic nonlinear Fourier transform) rectifies matters. Actually it is better to use $\exp(-\beta z^2)$ as the "rectifying factor" in the nonlinear Fourier transform as multiplication by $\exp(-\beta z^2)$ does not affect the smallness of $\bar{\partial}\phi f$.

Let us begin our quest for precise estimates with $n = 1$. We change our notation and write

$$G_f(\beta; \hat{x}; y) = \int \phi f(x, y) e^{-\beta z^2} e^{ix\hat{x}} dx. \tag{9.99}$$

Clearly

$$G_f(\beta; \hat{x}; y) = e^{\beta y^2} F_{\phi f}(\beta; \hat{x} - 2\beta y; y) \tag{9.100}$$

where $F_{\phi f}$ is the usual nonlinear Fourier transform of ϕf in x for y fixed. By integration by parts

$$\int \bar{\partial}\left[\phi f(x,y)e^{-\beta z^2}\right] e^{ix\hat{x}}dx = \left[i\frac{\partial}{\partial y} - i\hat{x}\right] G_f(\beta;\hat{x};y). \tag{9.101}$$

The left side of (9.101) can be written in the form

$$\int (\bar{\partial}\phi f)(x,y)e^{-\beta z^2+ix\hat{x}}dx = \mathcal{O}(e^{-\xi(y)+\beta y^2}) + \mathcal{O}(e^{-\beta}) \tag{9.102}$$

because

$$\bar{\partial}\phi f = \phi\bar{\partial}f + f\bar{\partial}\phi = \mathcal{E}_1 + \mathcal{E}_2 \tag{9.103}$$

where $\mathcal{E}_1 = \mathcal{O}[\exp(\xi(y))]$ and $\mathcal{E}_1$ has fixed (i.e. independent of y) support in x and $\mathcal{E}_2$ has fixed compact support in $\{|x| > 2-\epsilon\}$. Inequality (9.102) holds for y in a fixed (small) compact set.

We are now in a position to repeat our calculations of (9.85)ff. As in our study of the compact case we choose, for our first example, $\xi \equiv -\infty$ for $0 \le y \le \alpha$. Writing

$$\left(\frac{\partial}{\partial y} - \hat{x}\right) G_f = e^{+\hat{x}y}\frac{\partial}{\partial y}e^{-\hat{x}y}G_f \tag{9.104}$$

leads by (9.101), (9.102) to

$$ie^{-\hat{x}y}G_f(\beta;\hat{x};y) = iG_f(\beta;\hat{x};0) + \mathcal{O}(ye^{-\beta}) + \mathcal{O}\left[\int_0^y e^{-\hat{x}t}e^{-\xi(t)+\beta t^2}\,dt\right]. \tag{9.105}$$

The left side of (9.105) is $\mathcal{O}(\exp(\beta y^2 - \hat{x}y))$ by (9.99). Setting $y = \alpha$ makes the integral in (9.105) vanish; using (9.99) to estimate $G_f(\beta;\hat{x},0)$ we find

$$\begin{aligned} G_f(\beta;\hat{x};0) = F(\beta;\hat{x};0) &= \mathcal{O}\left(e^{-\hat{x}\alpha+\beta\alpha^2} + \alpha e^{-\beta}\right) \\ &= \mathcal{O}\left(e^{-c\beta}\right) \end{aligned} \tag{9.106}$$

for $\hat{x} > 0$ if α is small and $\hat{x} \ge c^0\beta$. We can now apply Theorem 5.6 to deduce that f is holomorphic in $\Im x > 0$, $|x|$ small. (Of course this is a trivial consequence of the fact that $\xi(t) \equiv \infty$ near $t = 0$ but we wanted to show how our ideas work.)

For a second example, let $\xi(y) = 1/y$. Now the $\mathcal{E}_1$ term of (9.102) and (9.103) is significant.

The term in (9.105) involving $\mathcal{E}_1$ is bounded by a constant times

$$\int_0^y e^{-\hat{x}t-1/t+\beta t^2}\,dt.$$

For y small and $c^0\beta \le \hat{x}$ the term βt^2 in the integrand is negligible. It is easily seen by differentiation that the integrand takes its maximum at $t = \hat{x}^{-1/2}$.

Arguing as in the compact case, an optimal choice for y is

$$y = \hat{x}^{-1/2}. \tag{9.107}$$

Thus the $\mathcal{E}_1$ term is bounded by

$$c\hat{x}^{-1/2}e^{-2\hat{x}^{1/2}}. \tag{9.108}$$

The left side of (9.105) is $\mathcal{O}(\exp(\beta y^2 - \hat{x}y)) = \mathcal{O}(\exp(\beta/\hat{x} - \hat{x}^{1/2}))$.

Putting together the estimates gives, by (9.105),

$$\begin{aligned} F_{\phi f}(\beta;\hat{x};0) &= G_f(\beta;\hat{x};0) \\ &= \mathcal{O}\left(e^{\beta/\hat{x}-\hat{x}^{1/2}} + e^{-2\hat{x}^{1/2}} + e^{-\beta}\right). \end{aligned} \tag{9.109}$$

This is exponentially decreasing in β if

$$|\hat{x}|^{1/2} \geq \beta. \tag{9.110}$$

According to the ideas of Chapter 5 we conclude that the positive $\hat{x}$ axis does not belong to the Gevrey 2 wave front set of $f(x,0)$.

We have explained how things work for $\xi(y) = 1/y$ and for $\xi(y) \equiv \infty$ near $y = 0$. What happens for other ξ?

We want to derive a relation between functions $f \in \mathcal{E}_{\bar{\partial},\xi_M}$ and Denjoy–Carleman M wave front sets (see (9.121) for the definition of $\mathcal{E}_{\bar{\partial},\xi_M}$). The ideas we developed in our study of Gevrey 2 suggest that we examine

$$\max[-\hat{x}t - \xi(t)] = -\zeta(\hat{x}). \tag{9.111}$$

$\zeta(\hat{x})$ is the conjugate of ξ. The maximum occurs at

$$-\hat{x} = \xi'(t).$$

We denote this value of t by y_m so

$$y_m = (\xi')^{-1}(-\hat{x}). \tag{9.112}$$

From (9.105) and the above we derive the bound

$$F(\beta;\hat{x};0) = \mathcal{O}\left[e^{-\hat{x}y_m(\hat{x})+\beta y_m^2(\hat{x})} + e^{-\beta} + y_m e^{-\zeta(\hat{x})}\right]. \tag{9.113}$$

It is often the case that the max in (9.111) occurs when the terms $\hat{x}t$ and $\xi(t)$ are essentially equal (recall that $\xi(t) \to \infty$) in which case $\hat{x}y_m(\hat{x}) \sim c\zeta(\hat{x})$. Since y_m is small,

$$F(\beta;\hat{x};0) = \mathcal{O}(e^{-c\beta}) \tag{9.114}$$

on

$$\beta \le c\zeta(\hat{x}).$$

According to Section 5.3 this means that the positive $\hat{x}$ axis does not belong to the Denjoy–Carleman M wave front set of f for any M such that

$$\log \lambda_M(\hat{x}) \le c\zeta(c\hat{x}). \tag{9.115}$$

For example, if $\xi(t) = -t^{-\alpha}$ then

$$\xi'(y_m) = -\hat{x}$$

means

$$y_m = c\hat{x}^{-1/(\alpha+1)}.$$

Hence

$$\hat{x}y_m = c\hat{x}^{\alpha/(\alpha+1)} \sim \xi(cy_m) \tag{9.116}$$

so the above considerations apply.

Let us reverse matters and start with a Denjoy–Carleman class $\mathcal{E}_M$ with $M = \{m_j\}$ as in Chapter 5 ($n = 1$). Let $h(y) \in \mathcal{E}_M$ vanish together with all its derivatives at $y = 0$. We want to estimate $h(y)$ for y small.

We know that $|h^{(j)}(y)| \le c^{j+1}m_j$. The constant c plays no role in our results so we set $c = 1$. Thus for any j

$$|h(y)| = \left| \int_0^y dy_{j-1} \int_0^{y_{j-1}} \dots \int_0^{y_1} h^{(j)}(y_0)\, dy_0 \right| \le m_j \frac{y^j}{j!}. \tag{9.117}$$

Under suitable regularity of $\{m_j\}$ the best bound on $h(y)$ for y fixed comes from the minimum of the right hand side of (9.117). Indeed for Gevrey α which is defined by $m_j = (\alpha j)!$ ($\alpha \ge 1$)

$$m_j y^j / j! \cong j^{(\alpha-1)j}(cy)^j. \tag{9.118}$$

By differentiation we find that the minimum occurs (when $c = 1$) when

$$j \sim y^{-(\alpha-1)^{-1}} e^{-1} \tag{9.119}$$

so the minimum is

$$\sim e^{-cy^{-1/(\alpha-1)}}. \tag{9.120}$$

For $\alpha = 2$ the minimum value of (9.118) is

$$e^{-c/y}.$$

This suggests that, in general, we define ξ_M by

$$e^{-\xi_M(y)} = \min m_j \frac{y^j}{j!} \tag{9.121}$$

so that

$$|h(y)| \le ce^{-c\xi_M(cy)}, \tag{9.122}$$

similarly for the derivatives of h; thus $h \in \mathcal{E}_{\mathrm{id},\xi_M}$, which is the space defined by (9.83) when $\vec{P} =$ identity.

Note that the functions $\hat{u}(\hat{x})$ in $\hat{\mathcal{E}}_M{}'$ are bounded on the real axis essentially by $\lambda_M(c\hat{x})$ where

$$\lambda_M(\hat{x}) = \max \frac{|\hat{x}|^j}{m_j}. \tag{9.123}$$

This follows from the Hahn–Banach representation of elements $u \in \mathcal{E}_M{}'$ in the form

$$u = \sum c^j m_j^{-1} \frac{d^j}{dx^j} u_j \tag{9.124}$$

where $\{u_j\}$ are measures of fixed compact support and of uniformly bounded mass (see Chapter 13 of FA for a detailed study of $\mathcal{E}_M$).

Remark. *There are two ways of defining $\mathcal{E}_M$, namely by inequalities $|f^{(j)}(x)| \le ac^j m_j$ where this inequality holds either for some c (uniformly on compact sets), or for all c, in which case a depends on c. The differences between these two definitions is minor so we shall ignore them here.*

In (9.121) y is small, while in (9.123) $\hat{x}$ is large. Setting $|\hat{x}| = y^{-1}$ leads to

Proposition 9.20 $\exp[-\xi_M(y)] \sim [\lambda_{\{m_j/j!\}}(y^{-1})]^{-1}$ *as* $y \to 0$.

Note that for Gevrey α we have

$$\begin{aligned} e^{-\xi_M(y)} &= \min(\alpha j)! \frac{y^j}{j!} \\ &\sim e^{-cy^{-1/(\alpha-1)}}. \end{aligned} \tag{9.125}$$

The conjugate ζ_M of ξ_M defined by (9.111) is

$$-\zeta_M(\hat{x}) = \max[-\hat{x}t - \xi_M(t)]. \tag{9.126}$$

We find that the max occurs at

$$y_m = c\hat{x}^{-(\alpha-1)/\alpha} \tag{9.127}$$

so that

$$\begin{aligned} \zeta_M(\hat{x}) &\sim \hat{x}y_m \sim \xi_M(y_m) \\ &\sim c\hat{x}^{1/\alpha} \\ &\sim \log \lambda_M(c\hat{x}) \end{aligned} \tag{9.128}$$

for $M = \{(\alpha j)!\}$ as can be seen from (9.123).

Proposition 9.21 *If $f \in \mathcal{E}_{\bar{\partial},\xi}(\mathbb{R}^+)$ then the Denjoy–Carleman wave front set of f^0 corresponding to*

$$\log \lambda_M(\hat{x}) \sim \zeta(c\hat{x})$$

omits $\mathbb{R}^+$.

We have shown that $\log \lambda_M$ is the conjugate of ξ_M for Gevrey classes. T. Ḃanh has extended this calculation to general ξ as follows.

Note that $y_m(\hat{x}) \to 0$ as $\hat{x} \to \infty$. For all the classes we consider $\zeta_M(\hat{x}) = |\zeta_M(\hat{x})| \leq c|\hat{x}|$. Hence, for suitable small fixed y in (9.105), for $\beta \leq \zeta_M(\hat{x})$

$$-\hat{x}y_m(\hat{x}) + \beta y_m^2(\hat{x}) < -c\beta$$

for $\hat{x}$ large, so the term $\exp[-\hat{x}y_m(\hat{x}) + \beta y_m^2(\hat{x})]$ in (9.113) is bounded by $\exp(-c\beta)$, i.e. the only significant term there is $\exp(\zeta_M(\hat{x}))$.

Banh has shown that, in general,

$$\zeta_M(\hat{x}) \geq \log \lambda_M(\hat{x}).$$

Banh's argument goes as follows:

$$\begin{aligned}
-\zeta_M(\hat{x}) &= \max_t\{-\hat{x}t + \xi_M(t)\} \\
&= \max_t\left\{-\hat{x}t + \min_j\{\log(m_j t^j/j!)\}\right\} \\
&= \max_t\left\{-\hat{x}t + \min_j\{\log m_j + j\log t - \log j!\}\right\} \\
&\leq \min_j\left\{\max_t\{-\hat{x}t + j\log t\} + \log m_j - \log j!\right\} \\
&= \min_j\{-j + j\log j - j\log\hat{x} + \log m_j - \log j!\} \\
&\leq \min_j\{-j\log\hat{x} + \log m_j\} \quad \text{(by Stirling's formula)} \\
&= -\log\lambda_M(\hat{x}).
\end{aligned}$$

This proves

Theorem 9.22 (Banh) *Proposition 9.21 holds for general nonquasianalytic M.*

To complete our "Cauchy–Kowalewski" theory of asymptotic solutions, we want to establish the converse of Theorem 9.22.

Let $(n = 1)$ $\mathcal{H}_M^0$ be the space of formal power series

$$h(y) = \sum a_j y^j$$

where

$$|h^{(j)}(0)| = j!|a_j| \leq c\epsilon^j m_j.$$

We set

$$\mathcal{W}_M = \mathcal{E}_M(x) \otimes \mathcal{H}^0_M.$$

The dual of $\mathcal{H}^0_M$ is the space of formal series $\sum b_j \delta^{(j)}$ with

$$b_j = \mathcal{O}(c^j/m_j).$$

The usual AU technique shows that $\mathcal{W}'_M$ is LAU. $\hat{\mathcal{W}}'_M$ consists of entire functions $G(\hat{x}, \hat{y})$ satisfying

$$|G(\hat{x}, \hat{y})| \leq c\lambda_M(c\Re\hat{x})\lambda_M(c\hat{y}) \exp(c|\Im\hat{x}|) \tag{9.129}$$

for some c. An AU structure for $\mathcal{W}_M$ consists of all continuous functions k on complex $(\hat{x}, \hat{y})$ space which dominate the right side of (9.129) for all c.

Now suppose that $f^0(x)$ is a C^∞ function whose M wave front set lies in the positive $\hat{x}$ axis $\hat{\Gamma}$. According to the ideas of Section 5.3 we can represent f^0 as a Fourier transform of a measure $\hat{\mu}$ satisfying

$$\hat{\mu}(\hat{x}) = \mathcal{O}\left[\lambda_M^{-1}(c\Re\hat{x}) \exp(-c|\Im\hat{x}|)\right] \tag{9.130}$$

for any c outside a small proper cone $\hat{\Gamma}^\epsilon$ in $\mathbb{C}$ around the positive $\hat{x}$ axis. In $\hat{\Gamma}^\epsilon$ we have

$$\hat{\mu}(\hat{x}) = \mathcal{O}\left[(1 + |\hat{x}|)^{-c} \exp(-c|\Im\hat{x}|)\right] \tag{9.131}$$

for all c.

We now define (formally)

$$f(x, y) = \int e^{i(x+iy)\hat{x}}\, d\hat{\mu}(\hat{x}). \tag{9.132}$$

The contribution f^+ to f from $\hat{\Gamma}^\epsilon$ converges in $y > 0$ to a holomorphic function which is C^∞ on $\overline{L + \Gamma}$.

The contribution f^- from the integral over the complement $c(\hat{\Gamma}^\epsilon)$ of $\hat{\Gamma}^\epsilon$ defines a function which is holomorphic in $y < 0$. We claim that f^- defines an element of $\mathcal{W}_M$; i.e. a continuous linear function on $\mathcal{W}'_M$.

For f^- defined by (9.132), with the integral over $c(\hat{\Gamma}^\epsilon)$, we have, for $g \in \mathcal{W}'_M$,

$$f^- \cdot g = \int_{c(\hat{\Gamma}^\epsilon)} \hat{g}(\hat{x}, i\hat{x})\, d\hat{\mu}(\hat{x}).$$

A comparison of (9.129) and (9.130) shows that this integral converges and defines f^- as an element of $\mathcal{W}_M$ as long as $\lambda^2_M(t) \leq c\lambda_M(ct)$ which we assume.

We say M has the *extension property* if for any sequence $\{b_j\}$ satisfying $|b_j| \leq c^{j+1}m_j$ there is a $g \in \mathcal{E}_M$ with $g^{(j)} = b_j$. We have explained above that this implies that any $w \in \mathcal{W}_M$ can be extended to $\tilde{w} \in \mathcal{E}_M(x, y)$.

The argument following (9.98) can be applied to demonstrate

Theorem 9.23 *Suppose the M wave front set of f^0 lies in $\mathbb{R}^-$. Suppose moreover that the class M has the extension property. Then there is a function $f \in \mathcal{E}_{\bar\partial,\xi_M}(\mathbb{R}^+)$ whose CD is f^0.*

Thus far all of our considerations regarding asymptotic solutions have been confined to the Cauchy–Riemann operator $\bar\partial$ in one complex variable. We now show how to extend these ideas to n complex variables. In order to avoid notational complications we restrict our considerations to $n = 2$.

In the simpler problem involving holomorphic functions (actual solutions of $\bar\partial$) $h(z) = \sum a_{m_1m_2} z_1^{m_1} z_2^{m_2}$ in $|z_1| < 1$, $|z_2| < 1$ which are "nice" up to the distinguished boundary, we estimate $a_{m_1m_2}$ on the ray $\rho = \{m_1 = \lambda m_2\}$ for $m_2 > 0$, $\lambda > 0$ by using the boundedness of $a_{m_1m_2} r_1^{m_1} r_2^{m_2}$ along $r_1 = r_2^{\lambda'}$ for a suitable $\lambda' < 0$.

We know that $a_{\lambda m_2, m_2} r_2^{\lambda\lambda' m_2 + m_2}$ is bounded as $r_2 \to 1$. If $\lambda\lambda' + 1 < 0$ then this gives the exponential decrease of λ along the ray $m_1 = \lambda m_2$ and a cone of neighboring rays. Hence this ray is not in the analytic wave front set of h. In this way we show that the analytic wave front set of h lies in the positive quadrant.

The method promulgated in (9.86)ff. for asymptotic solutions replaces the above idea by estimates obtained using the separation of variables of $\bar\partial$ ($n = 1$) in polar coordinates. For $n = 2$, $\bar\partial$ defines a differential operator on the complex curve $z_1 = z_2^{\lambda'}$ whose "radial" part is $r_1 = r_2^{\lambda'}$.[7] It is easy to modify the above so as to obtain estimates for $a_{\lambda m_2, m_2}$. This leads readily to

Theorem 9.21* *For*

$$\xi(t) = \frac{1}{t_1 + t_2}$$

the Gevrey 2 wave front set of $f \in \mathcal{E}_{\bar\partial,\xi}(\Gamma)$ (Γ = positive orthant) lies in the positive orthant. Moreover the Cauchy–Kowalewski theory described above applies.

Naturally the same method applies to more general ξ.

We now explain how to establish

Theorem 9.24 *Let $f^\pm \in \mathcal{E}_{\bar\partial,\xi}$ in the tube over the positive (resp. negative) orthant $\Gamma^\pm$ with ξ as in Theorem 9.21*. Suppose $f^+ = f^-$ on the edge ($y = 0$). Then there is a function f on all $\{(x, y)\}$ (locally) which is an asymptotic solution of $\bar\partial$ of order $e^{-\xi}$ such that $f = f^\pm$ in the tube over $\Gamma^\pm$.*

It is not difficult to show that if $f^\pm$ are asymptotic solutions of $\bar\partial$ in $\Gamma^\pm$ respectively then they are asymptotic solutions of order $\exp(-\xi)$ for some ξ as

[7] To avoid complications involving covering spaces we can assume that λ' is a negative integer or the reciprocal of a negative integer λ'' (in which case we express this curve in the form $z_2 = z_1^{\lambda''}$). Thus $\{z_1\}$ or $\{z_2\}$ parametrizes the curve.

in Theorem 9.23. Thus

Theorem 9.25$^{\bar\partial}$ *The edge-of-the-wedge theorem holds for asymptotic solutions of* $\bar\partial$.[8]

Remark. *If we assume "strong asymptotic" rather than "asymptotic," then the edge-of-the-wedge theorem is easily established. Then the functions* $f^{\pm}$ *fit together to form a* C^∞ *function* $\tilde f$ *on the tube over the closure of* $\Gamma^+ \cup \Gamma^-$. *Such a* C^∞ *function has a* C^∞ *extension* $\underset{\sim}{f}$ *to the whole space. The derivatives of* f *at the edge can be calculated using* $\tilde f$; *that is, from the values of* f *on the tube over* $\Gamma^+ \cup \Gamma^-$ *since* $\Gamma^+ \cup \Gamma^-$ *contains dim*$\{y\}$ *linearly independent vectors. It follows that* f *is an asymptotic solution.*

This argument is valid for all systems $\vec P(D)$ *for which the "edge" is a CS.*

Proof of Theorem 9.24 The Cauchy–Kowalewski theory presented above shows that we can assume that the CD of $f^{\pm}$ vanish. For we can replace $f^{\pm}$ by $f^{\pm} - g$ where g is an asymptotic solution in a whole neighborhood of the origin of $\bar\partial$ of order ξ having the same CD as $f^{\pm}$.

To understand our ideas let us return to (9.86)ff. Since $a_m(1) = 0$ we use (9.88) to estimate

$$a_m(r) = \mathcal{O}\left[r^m \int_r^1 t^{-m} e^{-\xi(1-t)}\, dt\right].$$

This shows how $a_m(r) \to 0$ in terms of m and r.

Let us pass to two dimensions. We can now obtain bounds for $a_{m_1 m_2}(r_1, r_2)$ in both r_1 and r_2. Let us concentrate on (9.87). In particular, for $m_1 = m_2 = 0$ these equations read

$$\vec b_{00}(r) = \left\{\frac{\partial}{\partial r_j} a_{00}(r_1, r_2)\right\} = \mathcal{O}(e^{-\xi(1-r_1, 1-r_2)}). \tag{9.133}$$

$\vec b_{00}$ is a closed (exact) 1 form defined in the closure of the interior of the square $\{0 \le r_j \le 1\}$ (which is the r part of the bidisk) which is small at the distinguished boundary $\{r_1 = 1, r_2 = 1\}$.

For the edge-of-the-wedge theorem we are given two closed forms $\vec b_{00}^{\pm}$ which are defined in $\{r_1 \le 1, r_2 \le 1\}$ and in $\{r_1 \ge 1, r_2 \ge 1\}$ respectively and which vanish on the distinguished boundary at the rate defined by ξ. We want to extend $\vec b_{00}^{\pm}$ to a closed form $\vec b$ on all of $\mathbb{R}^2$.

To understand how to produce the extension, let us examine the analogous problem in cartesian coordinates. $\vec b_{00}^{\pm}(y)$ are now closed forms in the positive and negative quadrants respectively which $\to 0$ as $y \to 0$ at a rate determined by ξ. We want to find a suitable extension $\vec b_{00}$ to the whole y plane.

[8]This answers a question posed by C. Fefferman, L. Nirenberg, and P. Yang.

Remark. *The Fourier series approach that we have presented transforms the problem for $\bar\partial$ into a similar problem for d (and analogous problems for the operators d_m arising from the other Fourier series coefficients).*

Since a_{00} vanishes at the origin we can obtain $a_{00}(y_1, y_2)$ in the first quadrant as

$$a_{00}(y_1, y_2) = \int_0^{y_2} b^2_{00}(0, u)\, du + \int_0^{y_1} b^1_{00}(t, y_2)\, dt. \tag{9.134}$$

Extension of a_{00} to the second quadrant entails an extension of b^1. The simplest continuous extension is the reflection in the y_2 axis:

$$b^1_{00}(y_1, y_2) = b^1_{00}(-y_1, y_2).$$

By (9.134) this gives an extension a^{12}_{00} of a_{00} from the first to the second quadrant whose first derivatives satisfy the desired growth condition at the origin. The so constructed a^{12}_{00} may not be smooth because b^1_{00} may not be smooth along the y_2 axis. But this can be remedied by a simple approximation argument.

However, a^{12}_{00} might not coincide with a_{00} on the negative y_1 axis. To remedy this difficulty, we make a similar construction extending a_{00} to a^{32}_{00} from the third to the second quadrant. Note that $a^{12}_{00} = a_{00}$ on the positive y_2 axis and $a^{32}_{00} = a_{00}$ on the negative y_1 axis.

Since a^{12}_{00} and a^{32}_{00} and their first derivatives satisfy the desired growth at the origin we can take a (circular) partition of unity to obtain an extension a^2_{00} of a_{00} to the second quadrant with the desired properties.

Remark. *It is at this point that we use the vanishing of f at the origin. As mentioned above, this possibility is afforded us by our Cauchy–Kowalewski theory.*

The same construction yields an extension a^4_{00} to the fourth quadrant and hence the desired extension of a_{00} whose differential extends $b^{\pm}_{00}$.

There is no problem in carrying out this process in r, θ coordinates. For the other Fourier series coefficients there are several possibilities; we present a simple method due to Banh. A combination of this idea with the nonlinear Fourier analysis techniques of (9.99)ff. completes the proof of Theorem 9.24.

Banh writes

$$a_{mn}(r) = \int_0^1 s^{-m-n}[r_1 b^1_{mn}(sr) + r_2 b^2_{mn}(sr)]\, ds \tag{9.134*}$$

where

$$r_1^{-m} r_2^{-n} b^j_{mn}(r) = \frac{\partial}{\partial r_j} r_1^{-m} r_2^{-n} a_{mn}(r_1, r_2).$$

The reflection proceeds as before.

The $\bar\partial$ system is first order. We want to illustrate the difficulties one meets for operators of higher order.

The simplest operator is the Laplacian Δ for $n = 2$. All the major problems promulgated by the increased order (for a single operator) already occur in the "compact case," i.e. when (x, y) is replaced by (θ, r).

Instead of Δ we use the equivalent operator $r^2\Delta = (r\partial/\partial r)^2 + \partial^2/\partial\theta^2$. We have

$$\begin{aligned} r^2\Delta[a_m(r)e^{im\theta}] &= (a_m''(r) - m^2)e^{im\theta} \\ &= \left(r\frac{\partial}{\partial r} + m\right)\left(r\frac{\partial}{\partial r} - m\right)a_m(r)e^{im\theta}. \end{aligned} \tag{9.135}$$

We have written a_m' for $r\partial a_m/\partial r$.

The functions

$$\begin{aligned} \alpha_m(r) &= \left(r\frac{\partial}{\partial r} + m\right)\left(r\frac{\partial}{\partial r} - m\right)a_m(r) \\ &= \left[r^{-m}r\frac{\partial}{\partial r}r^m\right]\left[r^m r\frac{\partial}{\partial r}r^{-m}\right]a_m(r) \end{aligned} \tag{9.136}$$

and $\beta_m(r)$ which is obtained by applying $r\partial/\partial r$ to (9.136) are both $\mathcal{O}(\exp(\xi(1-r)))$.

From (9.136) we derive

$$\left[r^m r\frac{\partial}{\partial r}r^{-m}\right]a_m(r) = r^{-m}\left(\left[r^m r\frac{\partial}{\partial r}r^{-m}\right]a_m(r)\right)_{r=1} - r^{-m}\int_r^1 t^m\alpha_m(t)\frac{dt}{t}.$$

Hence

$$\begin{aligned} r^{-m}a_m(r) = a_m(1) &+ \frac{1-r^{-2m}}{2m}\left(r\frac{\partial}{\partial r}r^{-m}a_m(r)\right)_{r=1} \\ &+ \int_r^1 u^{-2m}\frac{du}{u}\int_u^1 t^m\alpha_m(t)\frac{dt}{t}. \end{aligned} \tag{9.137}$$

As in our treatment of $\bar\partial$ we can estimate the double integral in (9.137) in terms of ξ. For m negative this gives a bound on

$$a_m(1) + \frac{1-r^{-2m}}{2m}[-ma_m(1) + a_m'(1)]. \tag{9.138}$$

Since $r\partial/\partial r$ commutes with (9.135) we obtain a similar estimate for a_m' in place of a_m. (This is the CD* condition in (9.11*).) Since f is asymptotically harmonic $a_m''(1) = m^2a_m(1)$. The estimate derived from $(\partial f/\partial r)$ involves

$$a_m'(1) + \frac{1-r^{-2m}}{2m}[-ma_m'(1) + a_m(1)]. \tag{9.139}$$

To obtain an edge-of-the-wedge theorem we coordinate the estimates obtained from (9.138) and (9.139) which we obtain from the interior and from the exterior of the unit disk. This yields the edge-of-the-wedge theorem for functions which are asymptotically pluriharmonic; that is, asymptotically harmonic on any complex line (real parts of holomorphic functions). To establish the edge-of-the-wedge theorem for asymptotic solutions of general systems we need a new approach. Our method uses an analog of the classical procedure of replacing an equation of high order by a system of first-order equations.

Call V the variety associated to $\vec{P}(D)$. We assume, as in Chapter 4, that the CP is defined by the harmonic functions $h_j(\hat{y})$ which give a basis for each of the "stalks"

$$V_{\hat{x}} = \{\hat{y} | (\hat{x}, \hat{y}) \in V\} \tag{9.140}$$

of V which lie above $\hat{x}$.

We now use the harmonic decomposition (see (1.104)ff.) to write

$$\hat{y}_j h_k(\hat{y}) = \sum u^j_{kl}(\hat{x}) hl(\hat{y}) + v^j_k(\hat{x}, \hat{y}) \tag{9.141}$$

on $V_{\hat{x}}$. The u^j_{kl} are functions (polynomials) of $\hat{x}$ and $v^j_k(\hat{x}, \hat{y})$ belong to the ideal generated by $\{P_\alpha(\hat{x}, \hat{y})\}$ in the ring of polynomials.

The Fourier transform of (9.141) becomes

$$\frac{\partial}{\partial y_j}[\mathrm{CD}f](y^0)(x) = \boxed{u}^j \left(\frac{\partial}{\partial x}\right) \mathrm{CD}f(y^0)(x) + \vec{v}^j \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) f(x, y^0). \tag{9.142}$$

If $\vec{P}\, f = 0$ then $\vec{v}^j f = 0$ so (9.142) can be regarded as saying that $\boxed{u}^j$ are the generators of the l parameter abelian group defined by translation of the CD in y.

As usual it is easier to clarify our method in the compact case, meaning f is a periodic function of x (so we write θ instead of x, and r in place of y, and $r\partial/\partial r$ in place of $\partial/\partial y$).

Expanding $f(r, \theta)$ in a Fourier series,

$$f(r, \theta) = \sum a_m(r) e^{im\cdot\theta}. \tag{9.143}$$

Upon applying (9.142) to (9.143) we derive

$$\vec{v}^j \left(r\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta}\right) f = \sum \left[r_j \frac{\partial}{\partial r_j} a_m(r) - \boxed{u}^j (im) \cdot \vec{h}\left(r\frac{\partial}{\partial r}\right) a_m(r)\right] e^{im\cdot\theta}. \tag{9.144}$$

In order to bring equation (9.144) in line with equation (9.86) we have to diagonalize $\boxed{u}^j(im)$. As we have seen in Chapter 4 (in x, y coordinates) $\boxed{u}^j$ represents the operator of multiplication by $\hat{y}_j$ on the functions on the stalk V_m using the basis $\{h_l\}$ for this space of functions. Multiplication is diagonal in the δ function basis. The interchange of these basis is given by a Vandermonde-like matrix M.

We have explained in several places (see Chapter 4) that M does not affect the bounds with which we deal.

It is awkward to give an analog of the Cauchy–Kowalewski theory for asymptotic solutions in the tube over a single cone because the diagonalization in $(\hat{x}, \hat{y})$ or $(\hat{r}, m)$ space involves algebraic processes which make the translation to (x, y) [or (r, θ)] space somewhat unnatural. (There is a pseudo-differential formulation which we shall not discuss.)

However, if we place ourselves in the structure expressed in Theorem 9.11 then the algebraic complications disappear because all the characteristic roots of the $\boxed{u}^j(\hat{x})$ are accounted for. We can now repeat the argument used in the proof of Theorem $9.25^{\bar{\partial}}$ to establish.

Theorem 9.25 *The asymptotic edge-of-the-wedge theorem is valid under the hypotheses of Theorem 9.11 provided that the subvarieties of V of maximal dimension have multiplicity 1.*

Remark. *The condition "multiplicity 1" is used in the factorization.*

9.5 Contact manifolds for partial differential equations

The simplest example of a contact manifold arises from an ordinary equation

$$\frac{dy}{dx} = H(x, y). \tag{9.145}$$

At any point x, y we are given a direction $H(x, y)$. This direction can be used to define a line element $l(x, y)$ through (x, y) whose slope is $H(x, y)$. A curve γ defines a solution of (9.145) if γ is tangent to $l(x, y)$ at each point $(x, y) \in \gamma$.

Let us set a framework for this example. The line $l(x^0, y^0)$ is the (graph of the) solution of the equation

$$\frac{dy}{dx} = H(x^0, y^0) \tag{9.146}$$

which passes through (x^0, y^0). The underlying modus operandi of contact manifolds is: if a curve γ has a "high degree of contact" with a family of solutions of equations

$$P_{x^0y^0}(x, D)l_{x^0y^0} = 0$$

at each $x^0, y^0 \in \gamma$ then γ defines a solution of an associated equation

$$P(x, D)f = 0.$$

$P_{x^0y^0}$ is an approximation of P at x^0, y^0.

For ordinary differential equations, parametrization data is given at a point. It is for this reason that the approximate solutions $l(x, y)$ are uniquely determined by the point (x, y). For partial differential equations (or systems) $P(x, D)f = 0$, PD is generally given on submanifolds S of positive dimension. This suggests that

the proper geometric setting for a contact manifold theory involves replacing the original manifold $M = \{x\}$ by a family of PSs $\{S_t\}$. We term the set of S_t a *parametric loop space (PLS).*

$\{S_t\}$ is the "$\mathbf{x}$ variable." The $\mathbf{y}$ variable becomes the set of PD. It is convenient to make the assumption that all S_t are isomorphic and that the set $\{\mathbf{y}\}$ of PD is independent of t. We call $\mathbf{M}$ this space $\{(\mathbf{x}, \mathbf{y})\}$. $\mathbf{M}$ is a *parametric contact manifold (PCM).*

PD for P on S_t is a set of functions $\{f_t^j\}$ on the S_t; this is the PD_t of a function f if $\mathrm{PD}_t(f) = \{\partial_t^j f\big|_{S_t}\}$. Here ∂_t^j are differential operators determined by P and t. For each t the point $(S_t, \mathrm{PD}_t(f))$ belongs to $\mathbf{M}$. The set $\{(S_t, \mathrm{PD}_t(f))\}$ is the "curve" γ (graph of f) $\subset \mathbf{M}$ defined by f.

In analogy to our introduction to contact manifolds we use $(S_t, \mathrm{PD}_t(f)) = (\mathbf{x}, \mathbf{y})$ to define the solution of some equation $P_t(x, D)F_t = 0$ which is "tangent to P along S_t," meaning P_t is an approximation of P near S_t. The sets of PD for P and P_t are assumed to be the same. Thus we use f to define F_t which is the solution of the PP for P_t with $\mathrm{PD}_t(F_t) = \mathrm{PD}_t(f)$ on S_t. We can identify $(\mathbf{x}, \mathbf{y})$ with F_t as in the classical case. (In the classical case P_t is the constant coefficient operator (9.146).)[9]

Note that in our original example of contact manifold the approximate solution $l(x^0, y^0)$ existed globally. We thus postulate that F_t exists on a "large" set.

The second aspect of contact manifolds is that γ should be tangent to (the graph of) each $l(x^0, y^0)$ at x^0, y^0. This means that the order of contact of γ and $l(x^0, y^0)$ is more than mere passing through x^0, y^0. The analog for partial differential equations is that some t derivatives of order "greater than" the order of the "normal derivatives" defining the PP of f and F_t should agree on S_t.

Ansatz. *If γ is tangent to $\{F_t\}$ on S_t then $P(x, D)f = 0$.*

Of course, in order for the ansatz to be applicable we need enough S_t. We also need to know that the PP is well posed. In the present work "well posed" means in C^∞.

Rather than making precise conditions for the validity of the ansatz let us give some examples.

Example 1 *Cauchy problem. Let $P(x, D)$ be a hyperbolic operator. $\{S_t\}$ is a family of space-like hypersurfaces which depend smoothly on t. $P_t(x, D)$ is tangent to $P(x, D)$ along S_t, meaning the coefficients of P_t agree with those of P on S_t. (Coefficients are written on the left.)*

Suppose first that P is a single operator of order m, i.e. not a system. Then we can use t as a normal coordinate to S_t since dt is not tangent to S_t and, modulo tangential derivatives, all nontangential derivatives are equivalent. This means that, for any l, the information encoded in the first l derivatives on S_t in

[9] At first reading the reader can assume that $P_t = P$ for all t.

a nontangential direction ν of a function h_t which is C^∞ near S_t is independent of ν.

We have assumed that the coefficients of P_t and P are the same on S_t. Moreover, by the definition of F_t, $\mathrm{PD}_t(f) = \mathrm{PD}_t(F_t)$. By our above remark we can assume that the PD is defined by the first $m-1$ t derivatives. Since (see (9.148)) the equation $Pf = 0$ gives an expression for the m-th t derivative in terms of lower order t derivatives and derivatives tangent to S_t, it follows from the fact that P_t is the same as P on S_t that $P_t F_t = 0$ on S_t implies $Pf = 0$ on S_t.

We have shown that $Pf \equiv 0$ which verifies the ansatz.

Suppose now that $\vec{P}$ is a Cauchy hyperbolic system with CSs S_t. This means that S_t is of codimension > 1 and $\dim\{t\} = \operatorname{codim} S_t$. The PD (which is CD) consists of functions $\{g_t^\alpha\}_\alpha$ on S_t which represent the "normal derivatives" $\{\partial_\alpha f|_{S_t}\}$. Here ∂_α are partial differential operators which cannot be expressed in terms of tangential derivatives on S_t.

From the fact that S_t is a CS it follows (see Chapter IX of FA) that any t derivative ∂ of any order can be expressed in the form

$$\partial = \sum U_{\alpha t}^{\partial} \partial_\alpha + \sum V_{jt}^{\partial} P_j \quad \text{on } S_t. \tag{9.147}$$

The $U_{\alpha t}^{\partial}$ are differential operators which are tangent to S_t and V_{jt}^{∂} are general differential operators. The $U_{\alpha t}^{\partial}$ and V_{jt}^{∂} depend smoothly on t.

To understand this point let P be a single operator. We write it in the form

$$P(t,s) = \sum_{k=1}^{m} a_k(s) \frac{\partial^k}{\partial t^k} \tag{9.148}$$

where a_m is a function which does not vanish on S_{t^0} and, for $k < m$, a_k is a differential operator in s of order $\leq m-k$. (We are using coordinates t, s where s is a coordinate on S_t. (Recall that all S_t are diffeomorphic.)) This allows us to write, *à la* (9.147),

$$\frac{\partial^m}{\partial t^m} = -\sum_{k=1}^{m-1} \frac{a_k(s)}{a_m(s)} \frac{\partial^k}{\partial t^k} + \frac{1}{a_m(s)} P(t,s). \tag{9.149}$$

Differentiating (9.149) allows us to express all t derivatives of order $\geq m$ in terms of s derivatives and of t derivatives of order $< m$. This is (9.147) in the case of a single operator.

By a straightforward modification of our argument for single equations we can see that the ansatz is valid for overdetermined hyperbolic systems.

Example 2 *Dirichlet problem. The CP is local. We now switch to a global problem; we assume that the S_t are compact as in the case of the usual DP. Again $\{S_t\}$ is an l parameter family of PSs where $l = \operatorname{codim} S_t$.*

The original impetus for this work was the "strip problem": M is the strip in the complex plane $|\Im z| \leq 1$. f is a smooth function whose restriction to each circle $S_x = \{|z - x| = 1\}$ (x real) extends to a holomorphic function F_x in the disk $|z - x| < 1$. The strip problem is the problem of proving (or disproving) that f is holomorphic in the interior of the strip.

A positive answer has been given only under assumptions of strong regularity on f: namely, that f is real analytic and extends to a holomorphic function of two complex variables in a "large neighborhood" of the strip ([1, 64]).[10]

A similar problem was posed by the author for elliptic equations. The strip becomes the cylinder

$$\{(t, u) \big| \, |u|^2 \leq 1\}.$$

Here t is a single variable and $u \in \mathbb{R}^n$ and $x = (t, u)$. We present a positive answer to this problem below. A beautiful, simple proof for the Laplacian was given by M. Agranovsky and Narayamin E.K. (oral communication).

Let $P = P(x, D)$ be an elliptic operator of order $2m$. We illustrate our methods by solving this strip problem. Overdetermined systems can be treated in the same manner but require more complicated technique.

We assume that the DP is well posed for each sphere S_t centered on the t axis (in (t, u) space) of radius 1. Thus the contact manifold $= \{\mathbf{x}, \mathbf{y}\}$ where $\mathbf{x}$ represents S_t and $\mathbf{y}$ is an m-tuple of functions on S_t which represents the DD of a solution F_t of $PF_t = 0$ in the ball B_t bounded by S_t.

Our contact hypothesis for a function $f(t, u)$ is that

$$f \text{ has order of contact } m + 1 \text{ with } F_t \text{ on } S_t.$$

This means that

$$\frac{\partial^j (f - F_t)}{\partial r^j}(x) = \mathcal{O}(\epsilon^{m+1-j}) \quad \text{for dist}[x, S_t] \leq \epsilon, \ j \leq m. \tag{9.150}$$

Moreover (9.150) holds for all "tangential" derivatives of $f - F_t$, meaning we can replace $f - F_t$ by a tangential derivative of any order. We also assume that F_t extends to a solution of P on a neighborhood of the ball bounded by S_t.

The last statement can be made precise as follows. Let us think in terms of r, θ coordinates in the ball $B_{t^0} = \{(t - t^0)^2 + u^2 \leq 1\}$ or, in fact, in a neighborhood of B_{t^0}. r is the distance from the origin in B_{t^0} and θ is the angle measured, via the rotation group, from the point on the t axis. (Actually each sphere of radius 1 centered at the t axis meets the t axis in two points; we choose the larger one.)

[10] A complete solution to this problem has been given by Tumanov [148]. The author has developed a new method to deal with such problems. It applies to the analog of the strip problem in case the circle is replaced by the polydisc. It also applies to some other overdetermined systems of constant coefficient partial differential equations. Unfortunately, thus far we must assume that f is Gevrey 2.

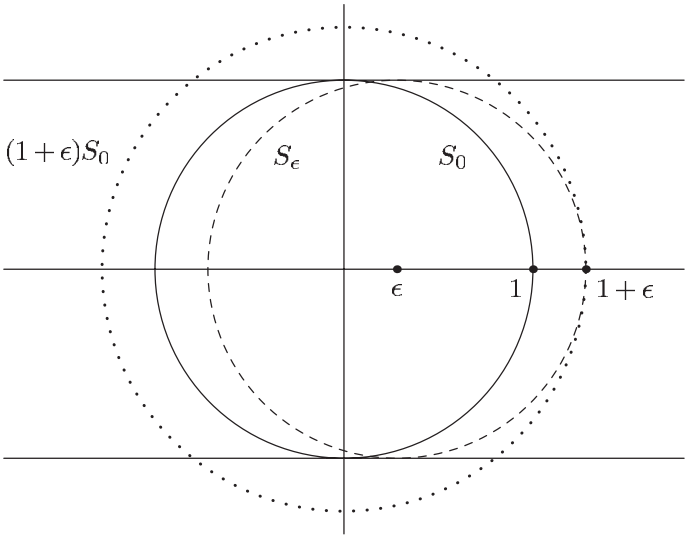

Figure 9.3

Since scalar multiplication and rotation commute the meaning of

$$D_\theta(f - F_t)(r, \theta) = \mathcal{O}(\epsilon^{m+1}) \qquad \text{for } |r - 1| \le \epsilon \tag{9.151}$$

is clear. Inequality (9.150) for tangential derivatives D_θ of any order is the meaning of order of contact m. To avoid confusion we shall usually write r_t, θ_t for the coordinate system related to S_t.

To go further we must take cognizance of the fact that

$$(1 + \epsilon)B_0 + (t, 0) \supset B_{t+\epsilon}. \tag{9.152}$$

This is best seen from Figure 9.3 (for $t = 0$).

The contact hypothesis is that f and F_0 have order of contact $m + 1$ on S_0 so that (9.150) holds. From (9.150) and Figure 9.3 we deduce

$$\frac{\partial^j (f - F_0)}{\partial r_0^j} = \mathcal{O}(\epsilon^{m+1-j}) \quad \text{on } S_\epsilon \tag{9.153}$$

for $j \le m + 1$ with similar inequalities where $f - F_0$ is replaced by $D_\theta(f - F_0)$.

We want to replace r_0 derivatives by r_ϵ derivatives on S_ϵ since the contact hypothesis deals with r_ϵ and θ_ϵ derivatives. Now, the change of coordinates from (r_0, θ_0) to (r_t, θ_t) is smooth in t. Moreover the transformation is the identity at $t = 0$. This means that, for small t,

$$\begin{pmatrix} \partial/\partial r_\epsilon \\ \partial/\partial \theta_\epsilon \end{pmatrix} = \begin{pmatrix} \partial/\partial r_0 \\ \partial/\partial \theta_0 \end{pmatrix} + J^{0\epsilon} \begin{pmatrix} \partial/\partial r_0 \\ \partial/\partial \theta_0 \end{pmatrix}. \tag{9.154}$$

Here $J^{0\epsilon}$ is a matrix with smooth entries which are $\mathcal{O}(\epsilon)$.

Let us transform (9.153) to $r_\epsilon, \theta_\epsilon$ derivatives. For first-order derivatives

$$\frac{\partial}{\partial r_0}(f - F_0) = \mathcal{O}(\epsilon^m)$$
$$D_{\theta_0}(f - F_0) = \mathcal{O}(\epsilon^{m+1}) \qquad \text{on } S_\epsilon.$$

By (9.154) we deduce

$$\frac{\partial}{\partial r_\epsilon}(f - F_0) = \mathcal{O}(\epsilon^m)$$
$$D_{\theta_\epsilon}(f - F_0) = \mathcal{O}(\epsilon^{m+1}) \qquad \text{on } S_\epsilon.$$

Next we examine the transformation of second derivatives. We write (9.154) in the form

$$\begin{pmatrix} \partial/\partial r_\epsilon \\ \partial/\partial\theta_\epsilon \end{pmatrix} = \begin{pmatrix} 1 + a_{11} & a_{12} \\ a_{21} & 1 + a_{22} \end{pmatrix} \begin{pmatrix} \partial/\partial r_0 \\ \partial/\partial\theta_0 \end{pmatrix} \tag{9.155}$$

where $a_{ij} = \mathcal{O}(\epsilon)$. On differentiating we find

$$\begin{pmatrix} \partial^2/\partial r_\epsilon^2 \\ \partial^2/\partial r_\epsilon \partial\theta_\epsilon \\ \partial^2/\partial\theta_\epsilon^2 \end{pmatrix}$$
$$= \begin{pmatrix} (1+a_{11})^2 & 2a_{12}(1+a_{11}) & a_{12}^2 \\ a_{21}(1+a_{11}) & (1+a_{11})(1+a_{22}) + a_{12}a_{21} & a_{12}(1+a_{22}) \\ a_{21}^2 & 2a_{21}(1+a_{22}) & (1+a_{22})^2 \end{pmatrix} \begin{pmatrix} \partial^2/\partial r_0^2 \\ \partial^2/\partial r_0\partial\theta_0 \\ \partial^2/\partial\theta_0^2 \end{pmatrix}$$
$$+ \begin{pmatrix} (1+a_{11})\frac{\partial a_{11}}{\partial r_0} + a_{12}\frac{\partial a_{11}}{\partial\theta_0} & (1+a_{11})\frac{\partial a_{12}}{\partial r_0} + a_{12}\frac{\partial a_{12}}{\partial\theta_0} \\ a_{21}\frac{\partial a_{11}}{\partial r_0} + (1+a_{22})\frac{\partial a_{11}}{\partial\theta_0} & a_{21}\frac{\partial a_{12}}{\partial r_0} + (1+a_{22})\frac{\partial a_{12}}{\partial\theta_0} \\ a_{21}\frac{\partial a_{21}}{\partial r_0} + (1+a_{22})\frac{\partial a_{21}}{\partial\theta_0} & a_{21}\frac{\partial a_{22}}{\partial r_0} + (1+a_{22})\frac{\partial a_{22}}{\partial\theta_0} \end{pmatrix} \begin{pmatrix} \partial/\partial r_0 \\ \partial/\partial\theta_0 \end{pmatrix}. \tag{9.156}$$

The a_{ij} are $\mathcal{O}(\epsilon)$ but their derivatives are only $\mathcal{O}(1)$.

Let us examine the $\partial^2/\partial\theta_\epsilon^2$ term. Note that $\partial/\partial r_0$ diminishes the order of vanishing of $f - F_0$ on S_0 by 1 and $\partial/\partial\theta_0$ preserves the order of vanishing. In the first matrix in (9.156) the bottom row is $\mathcal{O}(\epsilon^2, \epsilon, 1)$ so the order of $\partial^2/\partial\theta_\epsilon^2$ is preserved. Similarly the second row is $\mathcal{O}(\epsilon, 1, \epsilon)$ and the first row is $\mathcal{O}(1, \epsilon, \epsilon^2)$ so the orders of $\partial^2/\partial r_0^2, \partial^2/\partial r_0\partial\theta_0, \partial^2/\partial\theta_0^2$ are preserved when S_0 is transformed to S_ϵ.

But in the 3×2 matrix the orders are all $\mathcal{O}(1, 1)$. However, we are saved because these involve only first derivatives. We conclude from (9.153) that the second derivatives except $\partial^2/\partial r_\epsilon^2$ are $\mathcal{O}(\epsilon^m)$.

Now suppose, for example, that P is of order 4. The order of contact of F_0 and f is assumed to be 3. The above calculation shows that $\partial^2(f - F_0)/\partial\theta_\epsilon^2$ and

$\partial^2(f - F_0)/\partial\theta_\epsilon \partial r_\epsilon$ are $\mathcal{O}(\epsilon^2)$. But our hypothesis is that $f = F_\epsilon$ on S_ϵ as well as the first and second r_ϵ derivative of any θ_ϵ derivative. This leads to the crucial inequality

$$F_0 - F_\epsilon \text{ and its derivatives } \frac{\partial}{\partial\theta_\epsilon}, \frac{\partial^2}{\partial\theta_\epsilon^2}, \frac{\partial}{\partial r_\epsilon}, \frac{\partial^2}{\partial r_\epsilon \partial\theta_\epsilon}, \frac{\partial^2}{\partial\theta_\epsilon^2} \text{ are } \mathcal{O}(\epsilon^2). \tag{9.157}$$

By replacing f by F_ϵ we have gained because $F_0 - F_\epsilon$ is in the kernel of P. The well-posedness of the DP means that there is a

Dirichlet–Neumann (D–N) condition. *If G is a smooth solution of $PG = 0$ in the closed ball B_ϵ then for $k \geq m$ we can bound $\partial^k G/\partial r_\epsilon^k$ in terms of all derivatives of G of order $\leq k$ in $\theta_\epsilon, r_\epsilon$ which are of order $< m$ in r_ϵ.*

Remark. *We have left the question of "which norms" intentionally vague at this point. Many norms are possible.*

For example, let $P = \Delta$ be the Laplacian (and $\epsilon = 0$), so $m = 1$. The simplest form of the D–N condition asserts that we can bound $\partial G/\partial r_0$ in terms of G and $\partial G/\partial\theta_0$. This is easily seen by using Fourier series on the unit circle if $n = 2$ and by means of spherical harmonics if $n > 2$.

We have arrived at the inequality

$$\partial_\epsilon^2(F_\epsilon - F_0) = \mathcal{O}(\epsilon^2) \quad \text{on } S_\epsilon \tag{9.158}$$

where ∂_ϵ^2 is any second-order operator with smooth coefficients. We can replace F_ϵ by f in (9.158) by our hypothesis. Since ∂_ϵ^2 is an arbitrary second-order operator we can also replace ∂_ϵ^2 by ∂_0^2.

On S_0 we are given

$$\partial_0^2(f - F_0) = 0. \tag{9.159}$$

We apply the difference quotient in t to $\partial_0^2(f - F_0)\big|_{S_\epsilon} - \partial_0^2(f - F_0)\big|_{S_0}$. This yields

$$\frac{\partial}{\partial t}\partial_0^2(f - F_0) = 0 \quad \text{on } S_0. \tag{9.160}$$

We can write

$$\frac{\partial}{\partial t} = \alpha\frac{\partial}{\partial r_0} + \beta\frac{\partial}{\partial\theta_0} \quad \text{near } S_0. \tag{9.161}$$

Since $\partial_0^2(f - F_0) = 0$ on S_0 it follows that

$$\frac{\partial}{\partial\theta_0}\partial_0^2(f - F_0) = 0 \quad \text{on } S_0. \tag{9.162}$$

Hence

$$\alpha\frac{\partial}{\partial r_0}\partial_0^2(f - F_0) = 0 \quad \text{on } S_0. \tag{9.163}$$

It is clear from the geometry that the function α vanishes only at the "poles" of S_0, meaning the points (t, u) where $t = 0$. Hence by continuity we derive our desired result

$$\partial_0^3(f - F_0) = 0 \quad \text{on } S_0. \tag{9.164}$$

The same is clearly true on any S_t (with ∂_0 replaced by ∂_t).

We shall show how to deal with arbitrary m and how to iterate the argument to conclude that

$$\partial_t^p(f - F_t) = 0 \quad \text{on } S_t$$

for any p, t. Knowing this for all $p \leq 2m$ means that f satisfies the same differential equation of order $2m$ that is satisfied by F_t on S_t.

This proves

Theorem 9.26 *Let f be a smooth function in the strip $t \in (-\infty, \infty), \|u\| \leq 1$ $(u = (u_1, \ldots, u_n))$. Suppose on each sphere $S_{t_0} = \{(t - t_0)^2 + u^2 = 1\}$ the solution F_{t_0} of the DP with DD defined by f on S_{t_0} agrees with f to order $m + 1$ and that F_{t_0} extends to a solution of $PF_{t_0} = 0$ in a neighborhood of $B_{t_0} = \{(t - t_0)^2 + u^2 \leq 1\}$. Suppose, moreover, that the D–N condition holds. Then $Pf = 0$.*

We are left with the task of showing how our ideas extend to arbitrary p, m.

We used the condition $m = 2$ mainly in our calculation of how the change of coordinates from r_0, θ_0 to $r_\epsilon, \theta_\epsilon$ affects higher derivatives (see (9.156)ff.). The crucial properties we used were:

(1) In the first matrix in (9.156) the only time that $\partial^2/\partial r_0^2$ appears with a coefficient which is not $\mathcal{O}(\epsilon)$ is in the expression for $\partial^2/\partial r_\epsilon^2$.
(2) In the second matrix in (9.156) there is no term involving $\partial^2/\partial r_0^2$.

When we deal with the change of k-th order derivatives these properties are:

(1^k) The only time that $\partial^k/\partial r_0^k$ appears with a coefficient that is not $\mathcal{O}(\epsilon)$ is in the expression for $\partial^k/\partial r_\epsilon^k$ in the part of the calculation involving undifferentiated coefficients.
(2^k) In the part of the transformation which arises from differentiating coefficients only derivatives of order $< k$ appear.

There is no difficulty in continuing the argument following (9.156) to establish Theorem 9.25.

Problem 9.8 *Develop a complete theory of PDE contact manifolds.*

10

PERIODS OF EISENSTEIN AND POINCARÉ SERIES

Chapter 10 introduces the arithmetic theory related to the Radon transform. This centers around finding periods of functions like $E(s)$ (Eisenstein series) which are obtained from sums of the δ functions at lattice points on the three-dimensional light cone by taking Fourier transforms and performing various decompositions. Such periods belong to the realm of integral geometry.

In Section 10.1 we study Minkowski geometry in terms of the Lorentz group. Various types of coordinates which are useful for the computations of the periods are introduced in Section 10.2. In Section 10.3 we give explicit formulas for the periods of $E(s)$ over elliptic, hyperbolic, and parabolic cycles. In classical theory [142] the hyperbolic cycles are geodesics whereas we deal with arbitrary hyperbolic cycles. The periods are expressed in terms of ζ functions, β functions, and hypergeometric series.

In Section 10.4 we make similar computations for Poincaré series. The genesis of Poincaré series is the sum of δ functions on the orbit of a suitable one-parameter group which passes through "many" lattice points. The ζ functions that appeared in the periods of Eisenstein series become Kloosterman series; they are the usual Kloosterman sums for parabolic periods and "hyperbolic Kloosterman sums" for hyperbolic periods.

In Section 10.5 the light cone used to define usual Eisenstein series is replaced by a hyperboloid of one sheet. The periods are expressed in terms of hypergeometric functions.

In Section 10.6 similar calculations are made in four dimensions.

In Section 10.7 we show, to some extent, how to extend our calculations to groups other than $SL(2, \mathbb{R})$: for example, to symplectic groups.

10.1 The Lorentz group, Minkowski geometry, and a nonlinear projection–slice theorem

In this chapter we shall be concerned with certain explicit actions of groups on spaces. To clarify matters we present some background material.

Let G be a group of transformations of the manifold M. For $g \in G, x \in M$ we denote by gx the result of the action of g on x. $Gx = \{gx\}$ is called the *orbit* of x. Any two G orbits on M are either disjoint or identical. Thus the orbits form the leaves of a spread-like structure on M. (We say "spread-like" because not all the orbits may be isomorphic. Recall that a spread is a collection of nonintersecting isomorphic subvarieties L_α whose union is M.)

The subgroup G_x of G fixing x is called the *isotropy group* of x. If $y = gx$ then clearly

$$G_y = gG_xg^{-1}.$$

Thus the isotropy groups of points in the orbit Gx of x are conjugates of G_x. Conversely if $H = \gamma G_x\gamma^{-1}$ is conjugate to G_x then H is the isotropy group of γx.

Since G acts transitively on Gx the orbit can be identified with G/G_x. Precisely, the set of $g \in G$ taking x to a given point $y = g_0x$ in the orbit is g_0G_x.

A favorable situation occurs when G_x is its own normalizer. In this case the correspondence between cosets g_0G_x and conjugates $g_0G_xg_0^{-1}$ is one–one because

$$g_0G_xg_0^{-1} = g_1G_xg_1^{-1}$$

is equivalent to

$$g_0^{-1}g_1 \in (\text{normalizer } G_x) = G_x,$$

i.e. g_0 and g_1 belong to the same coset. Thus the points in the orbit can be identified with the conjugates of G_x.

Of course if the normalizer of G_x is much larger than G_x, e.g. if G is abelian, then the conjugates of G_x give little information on the orbit. But we shall meet situations in which normalizer G_x is only a "little larger" then G_x; such situations will still be favorable for our constructions.

Suppose there are two orbits of G on M, say Gx, Gy $(x \neq y)$, with $G_x = G_y$. Then we can identify Gx with Gy since both are identified with $G/G_x = G/G_y$. Explicitly $u \in Gx$ is identified with $v \in Gy$ if there is a g which maps $x \to u$ and $y \to v$.

The same identification can be made if G acts on 2 manifolds M_1, M_2 and $x \in M_1, y \in M_2$ with $G_x = G_y$. We call this *isotropic identification*.

In general there is a lack of uniqueness in isotropic identification because, as we mentioned above, there may be many points in Gx with the same isotropy group. Thus the isotropic identification of Gx with Gy depends on choosing *base points* x_0 in Gx and y_0 in Gy with $G_{x_0} = G_{y_0}$. We referred above to the "favorable situation" when G_x is its own normalizer. In that case isotropy groups determine the points in Gx and Gy so the isotropic identification is unique.

If G is a continuous group and G_x is an infinite subgroup then we can sometimes modify the condition $G_x =$ normalizer G_x to

$$(\text{normalizer } G_x)/G_x \quad \text{finite.}$$

Let us illustrate with $G =$ rotation group on $\mathbb{R}^n$, $n \geq 3$. If $x = (0, 0, \ldots, 0, 1)$ is the "north pole" then Gx is the unit sphere and G_x is the group of rotations leaving the x_n axis fixed. Thus G_x leaves two points on Gx fixed, namely the north and south poles.

If S^α is the sphere centered at the origin of radius α then when α is close to 1 the only continuous isotropic identification of S^1 with S^α entails the identification of the respective north poles since the north pole of S^1 is far from the south pole of S^α which is the only other point with which it can be identified via isotropic identification.

When $n = 2$, $G_x =$ identity, so the isotropy groups give no information. In this case we have to add reflections to G to form a group $\tilde{G}$ for which $\tilde{G}_x$ has only two fixed points on each orbit.

We mentioned that the set of orbits forms a spread-like structure on M. Suppose we have a subregion M^0 of M which is a union of orbits $\mathcal{O}^\alpha$ each of which has a point x^α with the same isotropy group H. Suppose moreover that $\{x^\alpha\}$ is a submanifold of M. Then there is a natural associated *cylindrical coordinate system* on M^0. This involves choosing coordinate systems on $\{x^\alpha\}$ and on G/H. For example, for the rotation group acting on $\mathbb{R}^n$, $n \geq 3$, this becomes polar coordinates. Many examples appear below.

We now give a detailed analysis of the three-dimensional Lorentz group, which is the orthogonal group of $y^2 - t_1^2 - t_2^2$. (More details on this group are given in Section 7.2.)

For many computations it is easier to regard three-dimensional Minkowski (Lorentz) space M^3 as the space of 2×2 symmetric real matrices $X = \left(\begin{smallmatrix} x & w \\ w & z \end{smallmatrix}\right)$. We call $\tilde{G} = SL(2, \mathbb{R})$ the group of 2×2 real matrices of determinant 1 and $G = \tilde{G}/\mu$ where $\mu = \{\pm I\}$. (The distinction between $\tilde{G}$ and G is minor for our purposes and we shall sometimes write G for $\tilde{G}$.) $\tilde{G} = \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\right\}$ acts on M^3 by

$$X \to gXg^t, \tag{10.1}$$

i.e.

$$\begin{pmatrix} x \\ z \\ w \end{pmatrix} \to \begin{pmatrix} a^2 & b^2 & 2ab \\ c^2 & d^2 & 2cd \\ ac & bd & (ad+bc) \end{pmatrix} \begin{pmatrix} x \\ z \\ w \end{pmatrix}. \tag{10.2}$$

Note that M acts trivially so this is a representation of G.

G acts on the ordinary plane $\mathbb{R}^2$ with coordinates (u, v) by linear action[1]

$$g(u, v) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

The isotropy group of $(1, 0)$ is $N = \{\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)\}$ (see (10.8)ff.). G clearly acts transitively on the plane minus the origin, so

$$\mathbb{R}^2 - \{(0, 0)\} = G/N.$$

Whenever a group acts on a space it acts on the functions on that space by

$$gf(x) = f(g(x)).$$

[1]We generally regard vectors as column vectors although for typographical reasons, when not in displayed formulas, they are written as row vectors.

(Actually this is an antirepresentation since $(g_1 g)f = g(g_1 f)$. To obtain a representation we should have defined $gf(x) = f(g^{-1}x)$ or $gf(x) = f(g^t x)$. We shall not concern ourselves with the difference. Note, however, that (10.1) is an actual representation.)

Since the action of G on $\mathbb{R}^2$ commutes with scalar multiplication the corresponding action on functions preserves homogeneity. The action on homogeneous polynomials of degree l is called the symmetric l-th power of the original representation.

We examine the symmetric square of the representation on $\mathbb{R}^2$. In terms of the basis vectors u^2, v^2, uv we find:

$$\begin{aligned} g(u^2) &= (au + bv)^2 = a^2u^2 + b^2v^2 + 2abuv \\ g(v^2) &= (cu + dv)^2 = c^2u^2 + d^2v^2 + 2cduv \\ g(uv) &= (au + bv)(cu + dv) = acu^2 + bdv^2 + (ad + bc)uv. \end{aligned} \tag{10.2*}$$

Comparison with (10.2) shows that this is the same as the representation (10.1). The basis (u^2, v^2, uv) corresponds to (x, z, w).

For later purposes let us note that the *quadratic transform* of the (u, v) plane, meaning

$$(u, v) \to (u^2, v^2, uv)$$

maps the (u, v) plane into the light cone $\{\det X = 0\}$. We showed (see Section 7.2) that, in fact, the positive light cone Γ^+ (the part of the light cone where $x + z \geq 0$) is exactly the quadratic transform of $\mathbb{R}^2 - \{0\}$.

Note that G acting on M^3 preserves the quadratic form (norm)

$$\|X\| = \det X = xz - w^2 = -\tfrac{1}{2}\mathrm{Tr}[JXJX] \tag{10.3}$$

where J is the Weyl (symplectic) element

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{10.4}$$

Expression (10.3) is a nondegenerate G invariant form so we use it to define Fourier transform.

Remark. *Although* $\det X$ *is a more natural way of writing the quadratic form, it is* $-\frac{1}{2}\mathrm{Tr}[JXJX]$ *which seems to have a more general meaning. For matrix groups in higher dimension* $\det X$ *is a polynomial of degree* >2 *in the matrix*

entries. On the other hand let J^* be a fixed matrix. Let $G^* = O(J^*)$ be the (transpose of the) orthogonal group of J^*, meaning

$$g^{*t} J^* g^* = J^*$$

for $g^* \in G^*$. G^* acts on the space S of symmetric matrices by

$$X \to g^* X g^{*t}.$$

The quadratic form

$$\mathrm{Tr}[J^* X J^* X]$$

is G^* invariant. For

$$\begin{aligned}\mathrm{Tr}[J^* g^* X g^{*t} J^* g^* X g^{*t}] &= \mathrm{Tr}[J^* g^* X J^* X g^{*t}] \\ &= \mathrm{Tr}[g^{*t} J^* g^* X J^* X] \\ &= \mathrm{Tr}[J^* X J^* X].\end{aligned}$$

Note that for the group $G = GL(2, \mathbb{R})$ and J as in (10.4)

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} J \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & -ad + bc \\ ad - bc & 0 \end{pmatrix}$$

so $\det g = 1$ is equivalent to $g^t J g = J$.

To make contact with the usual Minkowski form we set[2]

$$\begin{gathered} x = y + t_1, \quad z = y - t_1, \quad w = t_2 \\ \|X\|^2 = \det X = y^2 - t_1^2 - t_2^2 \\ \mathrm{tr}\, X = 2y. \end{gathered} \tag{10.5}$$

In this way we can realize G as a subgroup of the Lorentz group $SO(1,2)$. In fact we shall see (see (10.13)ff.) that G preserves the interior of the positive light cone $\det X \geq 0, y \geq 0$. Thus G is actually a subgroup of $SO^+(1,2)$, the connected component of the identity in $SO(1,2)$. (This is clear since G is connected.) In fact, from the structure of $SO^+(1,2)$ we can see that $G = SO^+(1,2)$.

The Minkowski quadratic form is associated to the inner product

$$\begin{aligned} X \cdot X' &= yy' - t_1 t_1' - t_2 t_2' \\ &= \tfrac{1}{2}(xz' + x'z) - ww'. \end{aligned} \tag{10.6}$$

[2]In order to be consistent with the notation in the rest of this book we use y for the time coordinate. The reader should distinguish between our notation $X \leftrightarrow (y, t_1, t_2)$ here and our previous notation in (10.2) (which will rarely occur).

In terms of the basis (y, t_1, t_2) the three-dimensional matrices representing $g \in G$ become

$$\begin{pmatrix} y \\ t_1 \\ t_2 \end{pmatrix} \to \begin{pmatrix} \frac{1}{2}(a^2+b^2+c^2+d^2) & \frac{1}{2}(a^2+c^2-b^2-d^2) & ab+cd \\ \frac{1}{2}(a^2+b^2-c^2-d^2) & \frac{1}{2}(a^2-b^2-c^2+d^2) & ab-cd \\ ac+bd & ac-bd & ad+bc \end{pmatrix} \begin{pmatrix} y \\ t_1 \\ t_2 \end{pmatrix}. \tag{10.7}$$

In $\tilde{G}$ there are three standard conjugacy classes of elements. Every element is conjugate to an element of one of these three subgroups:

$$\begin{aligned} K = \{\text{elliptic}\} &= \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \right\} \\ &\leftrightarrow \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta \\ 0 & -\sin 2\theta & \cos 2\theta \end{pmatrix} \right\} \\ A = \{\text{hyperbolic}\} &= \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \right\} \quad (a > 0) \\ &\leftrightarrow \left\{ \begin{pmatrix} \frac{1}{2}(a^2+a^{-2}) & \frac{1}{2}(a^2-a^{-2}) & 0 \\ \frac{1}{2}(a^2-a^{-2}) & \frac{1}{2}(a^2+a^{-2}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\} \\ N = \{\text{parabolic}\} &= \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \right\} \leftrightarrow \left\{ \begin{pmatrix} 1+n^2/2 & -n^2/2 & n \\ n^2/2 & 1-n^2/2 & n \\ n & -n & 1 \end{pmatrix} \right\}. \end{aligned} \tag{10.8}$$

It is sometimes convenient to replace A by the conjugate subgroup

$$\tilde{A} = \left\{ \begin{pmatrix} \cosh\zeta & \sinh\zeta \\ \sinh\zeta & \cosh\zeta \end{pmatrix} \right\} \leftrightarrow \left\{ \begin{pmatrix} \cosh 2\zeta & 0 & \sinh 2\zeta \\ 0 & 1 & 0 \\ \sinh 2\zeta & 0 & \cosh 2\zeta \end{pmatrix} \right\}.$$

The symbol $\leftrightarrow$ refers to the 3×3 matrix associated to g by (10.7).

To each of the subgroups there is an associated class of orbits in M^3 whose isotropy groups are conjugate to the subgroup. (There is also the trivial orbit $\{0\}$.) We proceed to describe the orbits.

(1) *Half-hyperboloid of two sheets.* This is the orbit of a positive or negative definite X. For such an X $\det X > 0$ which is the same as $y^2 - t_1^2 - t_2^2 > 0$. The orbit consists of all Y with $\det Y = \det X$ and $\operatorname{sgn} \operatorname{tr} Y = \operatorname{sgn} \operatorname{tr} X$.

It is easier to understand this from the three-dimensional viewpoint. We claim that the half-hyperboloid

$$y^2 - t_1^2 - t_2^2 = \alpha^2, \quad y > 0$$

is the orbit of the point $(\alpha, 0, 0)$ with $\alpha > 0$.

This sheet of the hyperboloid can be obtained by rotating the half-hyperbola in the y, t_2 plane

$$y^2 - t_2^2 = \alpha^2, \quad t_1 = 0, \quad y > 0$$

about the y axis. This half-hyperbola is clearly the orbit of the point $(\alpha, 0, 0)$ under the subgroup $\tilde{A}$. Rotation about the y axis is the group K.

Thus the orbit of $(\alpha, 0, 0)$ contains the positive sheet of the hyperboloid $\{\det Y = \alpha^2\}, \operatorname{tr} Y > 0$. Since the orbit is contained in $\{\det Y = \alpha^2\}$ we need only verify that G does not map $(\alpha, 0, 0)$ into the negative sheet. But this is a clear consequence of (10.7) since the y coordinate of $g(\alpha, 0, 0)$ is $\frac{1}{2}\alpha(a^2 + b^2 + c^2 + d^2)$ which manifestly has the same sign as α.

We have identified G with a subgroup of $SO(1, 2)$. Since G preserves the positive sheet of $\{\det Y = \alpha^2\}$ we have also verified our contention that G corresponds by (10.7) to a subgroup of $SO^+(1, 2)$ and since $SO^+(1, 2)$ has no three-dimensional proper subgroup, $G \approx SO^+(1, 2)$.

Note that $(2\alpha, 0, 0)$ corresponds under (10.5) to the symmetric matrix αI so the isotropy subgroup is clearly K. Thus the isotropy group of each point in the interior of the light cone is a conjugate of K.

There is another way to realize the orbit of I. For $\tilde{G}$ acts on the Poincaré upper half-plane, meaning the upper half of the complex plane $z = \xi + i\eta, \eta > 0$.[3] The action is

$$z \to \frac{az + b}{cz + d} = gz = \begin{pmatrix} a & b \\ c & d \end{pmatrix} z. \tag{10.9}$$

The isotropy group of $z = i$ is exactly $\pm K$. Thus we can identify $\tilde{G}/K$ with the upper half-plane.

As mentioned above, when a group H acts transitively on two manifolds M_1, M_2 and there are points $\gamma_1 \in M_1, \gamma_2 \in M_2$ which have the same isotropy group W then we can identify M_1 and M_2 as H spaces by corresponding $m_1 \in M_1$ to $m_2 \in M_2$ if there is as $h \in H$ with $h\gamma_j = m_j$ for $j = 1, 2$ (i.e. using γ_1, γ_2 as base points). In this way m_1 and m_2 have the same isotropy groups. In this way we can identify $\tilde{G}/K$ with any half-hyperboloid of two sheets. We termed this isotropic identification.

For the precise identification of the upper half-plane H with the positive half-hyperbola of matrices of determinant+1, we start with $i \leftrightarrow I$ since they have the common isotropy group K. We use the fact that $\tilde{G} = SK$ where $S = NA^+$ is the group of upper triangular matrices with positive diagonal entries. A point z in H can be identified with an $s \in S$, namely with that s which moves i to z. In particular

$$z = \xi + i\eta = \begin{pmatrix} \eta^{1/2} & \xi\eta^{-1/2} \\ 0 & \eta^{-1/2} \end{pmatrix} i. \tag{10.10}$$

[3]The reader should distinguish between the coordinate z in the upper half-plane and the coordinate z of the 2×2 matrix X (see (10.1)) in Minkowski space.

The corresponding point in Minkowski space is the image of I under the same transformation, i.e.

$$\begin{pmatrix} \eta^{1/2} & \xi\eta^{-1/2} \\ 0 & \eta^{-1/2} \end{pmatrix} I \begin{pmatrix} \eta^{1/2} & 0 \\ \xi\eta^{-1/2} & \eta^{-1/2} \end{pmatrix} = \begin{pmatrix} \eta + \xi^2\eta^{-1} & \xi\eta^{-1} \\ \xi\eta^{-1} & \eta^{-1} \end{pmatrix}$$
$$= \frac{2i}{z - \bar{z}} \begin{pmatrix} z\bar{z} & \frac{1}{2}(z + \bar{z}) \\ \frac{1}{2}(z + \bar{z}) & 1 \end{pmatrix}. \quad (10.11)$$

Thus the point $(t_1, t_2, y) \in$ orbit I is given by

$$y = \tfrac{1}{2}(\eta + \eta^{-1} + \xi^2\eta^{-1}), \quad t_1 = \tfrac{1}{2}(\eta - \eta^{-1} + \xi^2\eta^{-1}), \quad t_2 = \xi\eta^{-1}. \quad (10.12)$$

From either point of view the sheets of the hyperboloid of two sheets can be identified with NA. This is in conformity with the Iwasawa decomposition $\tilde{G} = NAK$.

(2) *(Half-)open light cone.* The light cone is defined by $\det X = 0$. The halves are sgn tr $X > 0$ and sgn tr $X < 0$. (There is also the origin which is an orbit.) The isotropy group of $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ is given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a^2 & ac \\ ac & c^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \quad (10.13)$$

Thus $a = \pm 1$ and $c = 0$. This means that g is an upper triangular matrix with diagonal elements ± 1, so the isotropy group is MN where $M = \{\pm I\}$.

The positive light cone, which is the orbit of $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$, can be identified exactly with the quadratic transform of the (a, c) plane under the quadratic transform mentioned following (10.2): that is, $x = a^2, z = c^2, w = ac$. For, $x + z > 0$ means one of x, z is > 0, say $x > 0$. Since $xz = y^2$, it follows that $z \geq 0$. We set $x = a^2, z = c^2$ so $w = \pm ac$. We have freedom to choose d arbitrarily, so we can make $\det \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = 1$. (If $x = 0$ the argument is similar.) Note that G acts transitively on this plane minus the origin and the isotropy group of $(1, 0)$ is N. The extra factor M comes from the quadratic transform.

As in the case of the half-hyperboloid of two sheets, we can show that the orbits are of the form

$$NA \begin{pmatrix} \pm 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

(3) *Hyperboloid of one sheet.* This is the orbit of X with $\det X < 0$. Let us compute the isotropy group $\tilde{K}$ of $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. The equation is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a^2 - b^2 & ac - bd \\ ac - bd & c^2 - d^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (10.14)$$

From this we conclude that (see (10.8)ff.)

$$\tilde{K} = \mu\tilde{A} = \left\{ \pm \begin{pmatrix} \cosh\zeta & \sinh\zeta \\ \sinh\zeta & \cosh\zeta \end{pmatrix} \right\}. \tag{10.15}$$

Thus $\tilde{A}$ should be regarded here as another real form of K (meaning, the natural complexifications of K and $\tilde{A}$ are the same) and should be properly denoted by $\tilde{K}$. In the present case $\tilde{K}$ is conjugate to the diagonal group A, by $g = \frac{1}{\sqrt{2}} \left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$, but for higher rank groups there is a large difference between diagonal (Cartan) subgroups and real forms of the maximal compact subgroup.

The hyperboloid $\det X = -1$ can be realized as the rotation of the sheet of the half-hyperbola $y^2 - t_1^2 = -1, t_1 > 0, t_2 = 0$ around the y axis. By (10.8) this half-hyperbola is the orbit of $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \leftrightarrow (0,1,0)$ under the group A. Thus the general hyperboloid of one sheet of the form $y^2 - t_1^2 - t_2^2 = -\alpha^2$ is given by

$$KA \cdot \begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix}$$

with $\alpha > 0$ and can be identified with KA.

There is also a two-dimensional model for this action of G on hyperboloids of one sheet. G acts by fractional linear transformations on the upper half-plane, and hence on the real axis. The isotropy group of 0 is clearly

$$S = \left\{ \begin{pmatrix} a & 0 \\ c & d \end{pmatrix} \right\}.$$

Actually G acts doubly transitively on $\bar{\mathbb{R}}$ = real axis with ∞ adjoined, i.e. transitively on $\bar{\mathbb{R}} \times \bar{\mathbb{R}} \setminus$ diagonal. For, given real numbers $\xi_1 \neq \xi_2$ we have

$$g_{\xi_1,\xi_2} = \frac{1}{(\xi_2 - \xi_1)^{1/2}} \begin{pmatrix} \xi_2 & \xi_1 \\ 1 & 1 \end{pmatrix} \text{ maps } 0 \to \xi_1, \infty \to \xi_2 \tag{10.16}$$

if $\xi_2 > \xi_1$. If $\xi_2 < \xi_1$ we can reverse the roles of $0, \infty$ either by using g_{ξ_2,ξ_1} or by multiplying g_{ξ_1,ξ_2} on the right by

$$z \to -\frac{1}{z} \qquad \text{that is } \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \text{ which maps } 0 \to \infty, \infty \to 0. \tag{10.17}$$

The isotropy group of $0, \infty$ consists of elements of S which leave ∞ fixed. This is exactly A. (Actually A leaves the two ordered pairs $0, \infty$ and $\infty, 0$ fixed.) Thus

G/A can be identified with the product of two projective lines from which the diagonal has been removed.

Topologically this is a cylinder with the circle as a base, which is the same as a hyperboloid of one sheet.

The G invariant metric in the Poincaré upper half-plane is given by

$$ds^2 = \frac{d\xi^2 + d\eta^2}{\eta^2}. \tag{10.18}$$

The geodesics are circles orthogonal to the real axis. Such a circle is determined by the two (distinct) points of intersection with the real axis. G preserves the metric and hence it acts as a group of transformations of the space of geodesics. The double transitive action of G on the real axis means that G acts transitively on the space of geodesics. One geodesic is the imaginary axis whose isotropy group is A. Thus G/A can be identified with the space of oriented geodesics.

The matrix $X = \left(\begin{smallmatrix} x & w \\ w & z \end{smallmatrix}\right)$ is naturally associated to the quadratic form

$$Q(u, v) = xu^2 + 2wuv + zv^2 \tag{10.19}$$

with

$$\text{discriminant } Q = -4 \det X.$$

Suppose Q is indefinite. Since Q is homogeneous it annihilates two lines $u = \omega v$. The two values of ω are given by

$$\omega^{\pm} = \frac{-w \pm \sqrt{w^2 - xz}}{x}.$$

G/A corresponds to fixing $w^2 - xz = \text{discriminant } Q = -4 \det X > 0$. If we set discriminant $Q = 1$ then

$$\omega^{\pm} = \frac{-w \pm 1}{x}$$

with $w = \pm 1$ if $x = 0$. For $x \neq 0$ we have

$$\omega^+ + \omega^- = -\frac{2w}{x}$$

$$\omega^+ - \omega^- = \frac{2}{x}.$$

ω^+, ω^- form an arbitrary pair of finite numbers, for we choose x so that $2/x = \omega^+ - \omega^-$ and then choose w to satisfy $\omega^+ + \omega^- = -2w/x$. However, if $\omega^+ = \infty$ then $x = 0$ so $w = \pm 1$. We have to replace $u = \omega v$ in (10.19) by $v = \lambda u$. The solutions are

$$\lambda = 0, \frac{-2w}{z}$$

and $-2w/z$ is arbitrary.

This yields another description of G/A, namely

$$G/A \sim \{\textit{ordered pairs of distinct lines through the origin on the } uv \textit{ plane}\}.$$

Since a pair of lines through the origin is determined by its pair of slopes, this description agrees with the description using the boundary of the upper half-plane.

We can clarify the geometry behind this model of G/A. From the three-dimensional point of view the positive light cone, the positive half-hyperboloid G/K, and the positive part (i.e. $y = \operatorname{tr} X > 0$) of the hyperboloid of one sheet all have the same limit at infinity (asymptotics), which is a circle. When we identify G/K with the upper half-plane this circle becomes the projectivized real axis.

The geometry can be described precisely as follows: $L = \{\left(\begin{smallmatrix} 0 & w \\ w & 0 \end{smallmatrix}\right)\}$ is the fixed line of A; it lies in the exterior of the light cone. The Minkowski orthogonal to this line is the plane $L^{\perp} = \{\left(\begin{smallmatrix} x & 0 \\ 0 & z \end{smallmatrix}\right)\}$. $L^{\perp}$ meets the light cone in the two lines spanned by $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and by $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. These lines are fixed by A as are their points as infinity. This pair of points at positive infinity, being A fixed, is the base point for G/A in the manifold of pairs of points at positive infinity. Applying G we obtain all pairs of such points since $gL^{\perp}$ is an arbitrary plane through the origin whose Minkowski normal has negative determinant.

We can also describe this pair of points at infinity in terms of geodesics on G/A or G/K. Geodesics for the Minkowski metric on G/K are the intersections of G/K with planes through the origin. These planes are $L^{\perp}$ and its transforms under G.

Such a plane meets $\{\det X = -1\}$ in a hyperbola or an ellipse. Those planes which are fixed by a conjugate of A intersect $\{\det X = -1\}$ in a hyperbola since this is true for the A fixed plane $\{\left(\begin{smallmatrix} x & 0 \\ 0 & z \end{smallmatrix}\right)\}$. In fact these planes form the same family as $\{L^{\perp}\}$ discussed above except that now we start with $L^{\perp}$ and before we started with L.

It is clear that the points at infinity on

$$L^{\perp} \cap \{\det X = -1\} \cap \{\operatorname{tr} X > 0\}$$

are the same as the points at infinity on $L^{\perp} \cap$ positive light cone.

Such an $L^{\perp}$ also intersects the half-hyperboloid $G/K = \{\det X = 1\} \cap \{\operatorname{tr} X > 0\}$ in a hyperbola having the same points at infinity as $L^{\perp} \cap$ positive light cone. These points at infinity correspond to the real endpoints of the geodesic when we use the upper half-plane model for G/K.

We illustrate this in detail for the plane $L^{\perp} = \{\left(\begin{smallmatrix} x & 0 \\ 0 & z \end{smallmatrix}\right)\}$ which meets $\{\det X = -1\} \cap \{\operatorname{tr} X > 0\}$ in the half-hyperbola

$$\{w = 0, xz = -1, x + z > 0\}.$$

The plane $L^{\perp}$ meets the (half-)hyperboloid of two sheets $G/K = \{\det X = 1\} \cap \{\operatorname{tr} X > 0\}$ in

$$\left\{ \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix}, \alpha > 0 \right\},$$

i.e.

$$\{w = 0, xz = +1, x > 0\}$$

which has the same points at infinity.

Under our identification (10.11) of G/K with the half-hyperboloid the intersection corresponds to the imaginary axis in the Poincaré upper half-plane. Moreover the fixed points of the hyperbolic transformation $z \to \alpha^2 z$ are $(0, \infty)$ which are the endpoints of the geodesic which is the imaginary axis.

We have shown the correspondence amongst

$$\begin{aligned}
\text{geodesic } \gamma \text{ on } G/K \quad &\longleftrightarrow \quad \text{time-like plane } P \text{ (i.e. } P^\perp \text{ lies outside the light cone)} \\
&\qquad\quad \text{through the origin in } \mathbb{R}^3 \\
&\longleftrightarrow \quad \text{time-like line } P^\perp \text{ through origin} \\
&\longleftrightarrow \quad \text{the endpoints of the geodesic } \gamma \\
&\longleftrightarrow \quad P \cap G/A = \text{points in } G/A \text{ corresponding to } \gamma \\
&\longleftrightarrow \quad \text{two points at } \infty \text{ on the common infinite part of} \\
&\qquad\quad G/K, (G/N)^+, (G/A)^+.
\end{aligned}$$

Remark. *We can now understand why we work with the three-dimensional representation of G on Minkowski space rather than the various two-dimensional models.*

(a) *The three-dimensional model unifies all the two-dimensional models of quotient spaces of G.*

(b) *The geodesics and horocycles (see Chapter 7) which appear as quadrics in two-dimensional models have a linear structure in the three-dimensional model, e.g. geodesics are intersections of $G/K = \{\det X = 1\}$ with planes through the origin. This enables us to make our calculations using Fourier analysis.*

10.2 Spreads and cylindrical coordinates in Minkowski geometry

Unlike the case of euclidean geometry there are different types of spreads in Minkowski geometry. (In the present context a spread is a family of parallel hypersurfaces called leaves.) As usual a (hyperplane) spread is defined by its orthogonal line or by a vector X of norm ± 1 on this line. (There is also the case of the orthogonal to a line on the light cone; we shall treat this later.)

We call X, or the associated spread, *elliptic* if X lies in the interior of the light cone; in the contrary case X is *hyperbolic*. Ellipticity is characterized by $\det X = +1$ or, what is the same thing, the quadratic form $Q(u, v)$ of (10.19) is definite. The nomenclature "elliptic, hyperbolic, parabolic" refers to the conjugacy class of elements in the isotropy group of X.

Let L_X^0 be the (Minkowski) plane orthogonal to X passing through the origin. The leaves of the spread are

$$L_X^y = L_X^0 + yX. \tag{10.20}$$

L_X^y is the plane through yX which is orthogonal to X.

In order to make explicit calculations we shall introduce *cylindrical coordinates* in each spread.

Since $\{yX\}$ and L_X^0 are orthogonal the Lorentz metric $ds^2 = dt^2 - dx^2 - dy^2$ splits into a component in $\{yX\}$ and an L_X^0 component. Thus we can deal with the metric and the resulting Fourier transform on the two components independently. These components are of lower dimension than the dimension of the ambient space (3) and hence are somewhat simpler to deal with. This passage from $\mathbb{R}^3$ to $\mathbb{R}^2 \oplus \mathbb{R}^1$ can be regarded as a form of *separation of variables*.

Our form of cylindrical coordinates is defined by using polar coordinates in both components. Actually we use y rather than $|y|$ in $\{yX\}$. Distance $\rho = \rho_X$ in L_X^0 is measured from the origin so the *distance polar coordinate* in $L_X^y = yX + L_X^0$ is the Minkowski distance measured from the *distance base point* yX.

To complete our definition of cylindrical coordinates we need an *angle base point* from which to determine angles.

(a) The elliptic case

In case X is *elliptic* $L_X^y \cap \Gamma$ is an ellipse.[4] If X lies in the interior of the forward light cone (so tr $X > 0$) then $L_X^y \cap \Gamma$ contains a unique point of the ray $\sigma_0^+ = \{\left(\begin{smallmatrix} a & 0 \\ 0 & 0 \end{smallmatrix}\right)\}_{a \geq 0}$ when $y > 0$. We use this point as the angle base point on $L_X^y \cap \Gamma$.

If tr $X < 0$ we use the ray $\sigma_0^- = \{\left(\begin{smallmatrix} a & 0 \\ 0 & 0 \end{smallmatrix}\right)\}_{a \leq 0}$.

We then define the *angle base ray* σ_X^y to be the ray (half-line) in L_X^y through yX and the angle base point.

When $y < 0$ we reverse the roles of $\sigma_0^{\pm}$.

Put in other terms, the angle base ray is the projection on L_X^y of σ_0^+ or of σ_0^-.

L_X^y is the union of the ellipses $\{\rho_X = \text{const.}\}$ on L_X^y. Each such ellipse intersects σ_X^y in a unique point, the angle base point on the ellipse, to which we assign the *angle polar coordinate* $\theta = \theta_X = 0$. For other points U on the ellipse the angle polar coordinate $\theta(U) = \theta_X(U)$ is the length of arc (suitably oriented) from the angle base point to U divided by $\rho_X(U)$. The perimeter of the ellipse on L_X^y defined by $\rho_X(U) = 1$ is 2π.

It is clear that the angle coordinate is preserved under translation by yX and is constant on rays. Constancy on rays is equivalent to invariance under multiplication by positive scalars.

We call $H = H_X$ the orthogonal group of X. H acts simply transitively on each orbit $\rho_X = \text{const.}$ on L_X^y. We can parametrize H by defining $h(\theta)$ to be

[4] Γ = light cone; we also use the letter Γ for the modular group and for the Γ function—hopefully there is no confusion.

the element of H that maps the angle base point p on L_X^y into the point on the ellipse whose angle polar coordinate is θ. It is clear that θ does not depend on y or on $\rho_X(p)$. If $U \in L_X^y$ then $h(\theta)U$ has angle polar coordinate $\theta + \theta_0$ where θ_0 is the angle polar coordinate of U.

Remark. *The explicit choice of the angle base ray is not of any particular significance. We chose it as above to simplify some calculations. The important property is that the angle should be invariant under translation by $y'X$ and by positive scalar multiplications on L_X^0.*

(b) The hyperbolic case

We now pass to the *hyperbolic case*, meaning $\det X = -1$. We need a whole line σ_0 in Γ rather than a ray to define the angle base point because we want the angle base line to meet both halves of the hyperbolas $\Gamma \cap L_X^y$. σ_0 is chosen so that L_X^y meets σ_0 for all y. For example, if $X = (0, 1, 0)$ then we could set $\sigma_0 = \{(t, t, 0)\}$. We then define the angle base line σ_X^y to be the line containing yX and $\sigma_0 \cap L_X^y$. It is clear that σ_X^y meets each branch of the hyperbola $\{\rho_X = \text{const.}\}$ in a unique point of σ_X^y which is the angle base point on this branch.

The (hyperbolic) angle polar coordinate of a point U_0 on L_X^y is the length of the arc of hyperbola $\rho_X(U) = \rho_X(U_0)$ on L_X^y (properly oriented) from the angle base line to U_0 deviated by $\rho_X(U_0)$.

We parametrize the orthogonal group $H = H_X$ of X as in the elliptic case.

We shall discuss some other aspects of the parametrization of H_X below.

Proposition 10.1 *$L_X^y \cap \Gamma$ is the set of points U in L_X^y the square of whose distance to the base point is*

$$\rho_X^2(U) = \begin{cases} -y^2, & X \text{ elliptic} \\ y^2, & X \text{ hyperbolic.} \end{cases} \tag{10.21}$$

Proof Let us begin with the elliptic case. For $X = I$, $L_X^y \cap \Gamma$ is defined by

$$y^2 = -\rho^2(t_I) \tag{10.22}$$

where $(y, t) = (y_I, t_I)$ are the usual cartesian coordinates. (Recall that in our notation $t = t_I$ is the two-dimensional coordinate on L_I^0 and ρ_I^2 is minus the square of the euclidean distance.) Let $g \in G$ map $I \to X$. Since g preserves distances and orthogonality

$$y_X^2(U) = -\rho_X^2(U) \tag{10.23}$$

if U is the image under g of $(y, t) \in L_I^y$. Since every $U \in \Gamma^+$ (positive light cone) is of the form $g(y, t_I)$ for such y, t_I we have established (10.21) for $U \in \Gamma^+$.

Since all concepts are invariant under $U \to -U$, (10.21) is valid for X elliptic.

For the hyperbolic case we begin with $X = \tilde{I} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ so that

$$L^0_{\tilde{I}} = \left\{ \begin{pmatrix} t_1 & t_2 \\ t_2 & t_1 \end{pmatrix} \right\}.$$

We write the analog of (10.22) in the form

$$\begin{aligned} 0 &= \det \begin{pmatrix} t_1 + y & t_2 \\ t_2 & t_1 - y \end{pmatrix} \\ &= t_1^2 - t_2^2 - y^2 \end{aligned} \tag{10.24}$$

where now $(y, t) = (y_{\tilde{I}}, t_{\tilde{I}})$. Equation (10.24) becomes

$$\begin{aligned} y_{\tilde{I}}^2 &= t_{\tilde{I}}^2 \\ &= \det \begin{pmatrix} t_1 & t_2 \\ t_2 & t_1 \end{pmatrix} \\ &= \rho_{\tilde{I}}^2(t_{\tilde{I}}) \quad \text{by (10.24).} \end{aligned}$$

This establishes (10.21) for $X = \tilde{I}$; the passage to general hyperbolic X goes as in the elliptic case.

For $U \in \Gamma \cap L_X^y$ we can express $\rho_X(U)$ in a more interesting way. We have explained that Γ^+ (forward light cone) is the quadratic transform of the (u, v) plane, meaning that elements of Γ^+ are of the form $\left(\begin{smallmatrix} u^2 & uv \\ uv & v^2 \end{smallmatrix}\right)$. We denote by $Q(u, v)$ the quadratic form defined by $X = \left(\begin{smallmatrix} x & w \\ w & z \end{smallmatrix}\right)$

$$Q(u, v) = xu^2 + 2wuv + zv^2. \tag{10.25}$$

For U in the forward light cone we have by (10.6)

$$\begin{aligned} X \cdot U = \begin{pmatrix} x & w \\ w & z \end{pmatrix} \cdot \begin{pmatrix} u^2 & uv \\ uv & v^2 \end{pmatrix} &= \frac{1}{2} xv^2 - wuv + \frac{1}{2} zu^2 \\ &= \frac{1}{2} \tilde{Q}(u, v) \end{aligned} \tag{10.26}$$

say.

If X is elliptic the square of the distance from yX to $U \in \Gamma$ is given by

$$(yX - U) \cdot (yX - U) = y^2 - 2yX \cdot U \tag{10.27}$$

because $X \cdot X = 1$ and $U \cdot U = 0$. In particular if U lies on L_X^y so that $U - yX$ is orthogonal to X we find

$$\begin{aligned} (yX - U) \cdot (yX - U) &= -U \cdot (yX - U) \\ &= -yX \cdot U. \end{aligned} \tag{10.28}$$

Combining this with (10.27) yields $y = 0$ or

$$y = X \cdot U \tag{10.29}$$

so that by (10.27)

$$\|yX - U\|^2 = -y^2 \tag{10.30}$$

which is clear when $y = 0$. This is another proof of Proposition 10.1 for X elliptic. The hyperbolic case is treated in the same manner.

Remark. *By (10.28) and (10.26) we also have*

$$\begin{aligned} \|yX - U\|^2 &= -yX \cdot U \\ &= -(X \cdot U)^2 \\ &= -\frac{y}{2}\tilde{Q}(u, v). \end{aligned} \tag{10.31}$$

In the hyperbolic case the minus sign changes to plus.

We now pass to the dual space $\hat{M}$ of Minkowski space M. The nondegenerate Lorentz metric leads to an identification of $U \in M$ when $\|U\| \neq 0$ with the element $U' \in \hat{M}$ defined by

$$\|U'\|^2 = \|U\|^2 = U' \cdot U.$$

(The case $\|U\| = 0$ will be presented below.) We introduce cylindrical polar coordinates in $\hat{M}$ exactly as in M. We often write U for U' when no confusion is possible.

Conversely for $\hat{U} \in \hat{M}$ we write $\hat{U}'$ or $\hat{U}$ for the element of M with which $\hat{U}$ is identified. In this way $\hat{U} \cdot U$ is defined.

Since the Lorentz metric separates in the orthogonal components $\{yX\}$, L_X^0 we can write

$$\hat{U} \cdot U = y\hat{y}\hat{X} \cdot X + \hat{U}^0 \cdot U^0 \tag{10.32}$$

where $\hat{U} \in L_{\hat{X}}^{\hat{y}}, U \in L_X^y$, and $\hat{U}^0$ and U^0 are the respective $\hat{L}_X^0$, L_X^0 components of $\hat{U}, U$.

In the simplest case $X = I$ the inner product

$$\hat{U}^0 \cdot U^0 = \rho_{\hat{X}}(\hat{U}^0)\rho_X(U^0)\cos[\theta_{\hat{X}}(\hat{U}^0) - \theta_X(U^0)] \tag{10.33E}$$

because L_I^0 is the plane $\{y = 0\}$ in (y, t_1, t_2) coordinates and ρ_I, θ_I are the usual polar coordinates. (The letter E in (10.33E) denotes "elliptic".)

When $X = \tilde{I} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ we have

$$\hat{U}^0 \cdot U^0 = \pm\rho_{\hat{X}}(\hat{U}^0)\rho_X(U^0)\cosh[\zeta_{\hat{X}}(\hat{U}^0) - \zeta_X(U^0)] \tag{10.33H}$$

because by (10.6) $L_{\tilde{I}}^0$ is the plane $t_1 = 0$ and $\rho_{\tilde{I}}, \zeta_{\tilde{I}}$ $(= \rho_X, \zeta_X)$ are the usual hyperbolic coordinates for the metric $y^2 - t_2^2$. The minus sign is taken when $\hat{U}^0, U^0$ belong to different branches of the hyperbola. (We assume that $\hat{U}, U$ lie in the interior of the light cone.)

Theorem 10.2 *Equations (10.33E) and (10.33H) are valid for arbitrary X with $\|X\|^2 = \pm 1$.*

Proof We start with X elliptic, $X \neq I$. There is a $g \in G$ which maps $I \to X$. g is not unique; we can multiply g on the right by $h_I \in H_I$ and on the left by $h_X \in H_X$. g maps the angle base ray σ_I in L_I^0 into a ray $\tilde{\sigma}_X \subset L_X^0$. We choose h_X so that $h_X \tilde{\sigma}_X = \sigma_X$ is the angle base ray in L_X^0. By relabeling we can assume $\tilde{\sigma}_X = \sigma_X$.

For precision we denote $U, \hat{U}$ by $U_X, \hat{U}_{\hat{X}}$ with a similar notation for $U^0, \hat{U}^0$. Let $U_I^0 \in L_I^0$ be chosen so that $gU_I^0 = U_X^0$. We define $\hat{U}_I^0$ similarly.

Since g preserves distance it preserves the duality inner product so that

$$\hat{U}_X^0 \cdot U_X^0 = \hat{U}_I^0 \cdot U_I^0. \tag{10.34}$$

Since we have defined angles in terms of arc length the angular distance between U_I^0 and σ_I is the same as the angular distance between U_X^0 and σ_X. Similarly the angle between $\hat{U}_I^0$ and $\hat{\sigma}_{\hat{I}}$ is the same as the angle between $\hat{U}_X^0$ and $\hat{\sigma}_{\hat{X}}$. This completes our proof of Theorem 10.2 in the elliptic case.

When $\det X = -1$ we proceed in essentially the same way except that we replace I by $\tilde{I} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. There is only one new twist, which results from the fact that a hyperbola consists of two connected curves. If $\hat{U}^0$ and U belong to different branches then there is no direct meaning to the angle between $\hat{U}^0$ and U^0. For $X = \tilde{I}$ we define this to be the angle between $w(\hat{U}_0)$ and U^0 where w is the reflection which interchanges the branches. $\hat{\zeta}_X(\hat{U}^0) - \zeta_X(U^0)$ is the hyperbolic angle between $w(\hat{U}^0)$ and U^0. w introduces a minus sign in the inner product.

In this way we complete the proof of Theorem 10.2.

(c) The parabolic case

Although Theorem 10.2 is sufficient for most of our applications which deal with hyperbolic and elliptic X, it is important to penetrate to the group theoretic structure of (10.33E) and (10.33H). This will enable us to deal with parabolic X and some more subtle problems.

In the t plane the relation

$$t \cdot \hat{t} = r\hat{r} \cos(\theta - \hat{\theta}) \tag{10.35}$$

can be understood as follows. The distance base point in the t plane is the origin and the angle base ray is the positive t_1 axis. The same is true for $\hat{t}$. We write

$$\begin{aligned} t &= h(\theta)t_B \\ \hat{t} &= h(\hat{\theta})\hat{t}_B \end{aligned} \tag{10.36}$$

where $t_B, \hat{t}_B$ are the angle base points corresponding to $t, \hat{t}$ respectively (points on the t_1 axis). Thus

$$\begin{aligned} t \cdot \hat{t} &= h(\theta)t_B \cdot h(\hat{\theta})\hat{t}_B \\ &= [h(\theta - \hat{\theta})t_B] \cdot \hat{t}_B. \end{aligned} \tag{10.37}$$

Equations (10.33E,H) follow from the properties:

(1) The angle base rays for t and $\hat{t}$ are the same under the identification of M with $\hat{M}$.

(2)
$$|h(\theta)t_B \cdot \hat{t}_B| = \begin{cases} |\hat{t}|\,|t\cos\theta| & \text{elliptic} \\ |\hat{t}|\,|t\cosh\theta| & \text{hyperbolic.} \end{cases}$$

These remarks lead to the principle:
To compute $\hat{W} \cdot W$ we have to determine $y_X(W), r_X(W)$ and the angle of W (similarly for $\hat{W}$).
We now apply these ideas to the parabolic case.

For simplicity of notation we set $X = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ so $H_X = N = \{\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)\}$. For any fixed y the plane

$$L_X^y = \left\{ \begin{pmatrix} x & w \\ w & y \end{pmatrix} \right\} \tag{10.38}$$

is N fixed. These planes form a spread corresponding to X.

We define the angle base ray on L_X^y to be $\left\{\left(\begin{smallmatrix} x & 0 \\ 0 & y \end{smallmatrix}\right)\right\}_{x \geq 0}$. This angle base ray has the "natural" property that it meets $\Omega = \{W \mid \det W = 1\}$ in the point $\left(\begin{smallmatrix} y^{-1} & 0 \\ 0 & y \end{smallmatrix}\right)$ which, by (10.11), corresponds to the point iy^{-1} on the imaginary axis when Ω is identified with the Poincaré half-plane. Moreover when Γ is represented as the quadratic transform of the (u, v) plane minus the origin, the angle base ray meets Γ^+ in the point $(0, \sqrt{y})$ on the v axis.

An arbitrary point W on L_X^y is of the form

$$W = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix} = \begin{pmatrix} x + n^2 y & ny \\ ny & y \end{pmatrix}. \tag{10.39}$$

Using our group theoretical definition of angle it makes sense to define n as the angle of the point W. Thus

$$\begin{aligned} x &= \frac{\det W}{y} \\ n &= \frac{w}{y}. \end{aligned} \tag{10.40}$$

It is convenient to call $x = \rho_X(W)$ the distance coordinate in L_X^y. According to our above ideas y, x, n are "cylindrical coordinates" of W. All this assumes $y \neq 0$; we shall deal with $y = 0$ below.

The formula for inner product in terms of these coordinates follows from (10.6) and (10.39):

$$\hat{W} \cdot W = \tfrac{1}{2}[y(\hat{x} + \hat{n}^2\hat{y}) + \hat{y}(x + n^2 y)] - n\hat{n}y\hat{y}. \tag{10.41}$$

It is this formula which replaces (10.33E,H).

For $y = 0$ an element $W \in L_X^0$ is of the form $\left(\begin{smallmatrix} x & w \\ w & 0 \end{smallmatrix}\right)$. Thus $\det W = -w^2$. For $w \neq 0$ we can use $\left(\begin{smallmatrix} 0 & w \\ w & 0 \end{smallmatrix}\right)$ as the angle base point. We have

$$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & w \\ w & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix} = \begin{pmatrix} 2nw & w \\ w & 0 \end{pmatrix} \tag{10.42}$$

so w is fixed (since y and $\det W$ are fixed this must be the case) and $n = x/2w$.

Combining (10.39) and (10.42) for $\hat{W} \in \hat{L}_{\hat{X}}^{\hat{y}}$ leads via (10.6) to

$$\hat{W} \cdot W = \hat{y}nw - \hat{n}\hat{y}w \tag{10.43}$$

when $y = 0, w \neq 0$. If $w = 0$ then $\left(\begin{smallmatrix} x & 0 \\ 0 & 0 \end{smallmatrix}\right)$ is N fixed so there is no angle coordinate; in this case

$$\hat{W} \cdot \begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix} = \frac{1}{2}\hat{y}x. \tag{10.44}$$

When we introduced the notions of angle polar coordinate we introduced a relation between this geometrically defined angle and the orthogonal group H_X of the quadratic form

$$Q_X(a, b) = (a, b)X \begin{pmatrix} a \\ b \end{pmatrix}. \tag{10.45}$$

We now present a more detailed treatment. Let X be elliptic and write $g(I) = gIg^t = X$. Then

$$gH_I g^{-1} = H_X \tag{10.46}$$

because

$$\begin{aligned} gh_I g^{-1} X (g^{-1})^t h_I^t g^t &= gh_I I h_I^t g^t \\ &= gIg^t \\ &= X. \end{aligned} \tag{10.47}$$

This means that the eigenvalues of h_X are the same as those of the h_I for which $gh_I g^{-1} = h_X$; of course the eigenvalues are independent of g. There are

two such eigenvalues generically so with a proper orientation (method of choosing an eigenvalue) they associate an "angle" θ to h_X, which is $-i$ times the log of this eigenvalue.

To clarify this point recall that we have defined the angle θ_X of a point U_X in L_X^y as the Lorentz distance measured along the ellipse $r_X = 1$ in L_X^y from the intersection U_X^1 of the ray through U_X with this ellipse to the angle base point on this ellipse. This Lorentz distance is the same as the usual polar angle measured in the plane L_I^y of the point U_I for which there is a $g \in G$ taking $I \to X$ and taking the angle base ray of L_I^y into the angle base ray of L_X^y and taking $U_I \to U_X$. Now $H_I = K$ is the usual rotation group

$$\left\{ \begin{pmatrix} \cos\theta_I & \sin\theta_I \\ -\sin\theta_I & \cos\theta_I \end{pmatrix} \right\}.$$

The polar angle θ_I of the point U_I in the plane L_I^y is the amount of rotation from the angle base ray to U_I. We call this rotation $h_I(\theta_I)$. $i\theta_I$ represents the logarithm of an eigenvalue (suitably oriented) of $h_I(\theta_I)$. (Actually there is a factor of 2 owing to our definition of representation of G as $X \to gXg^t$. This factor plays no essential role and can be "normalized" to 1.)

Combining these remarks with (10.47) yields

Proposition 10.3 *For elliptic X the geometrically defined angle θ_X of a point $U_X \in L_X^y$ is equal to $-i$ times log of the suitably oriented eigenvalue of the $h_X \in H_X$ for which $U_X \in h_X$ (angle base ray). The same is true when X is hyperbolic except that we do not need the factor $-i$ because the eigenvalues are real.*

Proposition 10.3 represents an interplay of geometry and group theory. We now introduce a further group theoretical definition of the angle.

The homogeneous quadratic polynomial $Q(a,b) = Q_X(a,b)$ factors into linear factors

$$Q(a,b) = (\alpha_1 a + \alpha_2 b)(\beta_1 a + \beta_2 b). \tag{10.48}$$

Let $h_X \in H_X$ which is the orthogonal group of X or Q. Then

$$\begin{aligned} Q[(a,b)h_X] &= Q(a,b) \\ &= (\alpha_1 a + \alpha_2 b)(\beta_1 a + \beta_2 b). \end{aligned} \tag{10.49}$$

We now introduce a principle which we call the *spinor principle* as it is the underlying principle in the construction of spinors. In general it asserts that whenever algebraic structures are uniquely determined by the factorization of a polynomial $P(x)$ then whenever L is a linear transformation preserving P then L "essentially" preserves the associated algebraic structures. (For spinors $P(x) = \sum x_j^2$ is factored into

$$P(x) = \left(\sum \sigma_j x_j \right)^2.$$

The σ_j generate a Clifford algebra which defines the spinor representation of the orthogonal group. The nontriviality of this representation is embedded in the word "essentially".)

In our case we observe that the factorization (10.49) is "essentially" unique. The only possibilities are multiplication of $(\alpha_1 a + \alpha_2 b)$ by a scalar λ and $(\beta_1 a + \beta_2 b)$ by λ^{-1} combined with a possible interchange of the factors. Now h_X acts on the vector (a, b) and hence there is an adjoint action on (α_1, α_2) which we continue to denote by h_X. Since h_X preserves Q we must have

$$\begin{aligned} h_X(\alpha_1, \alpha_2) &= \lambda(\alpha_1, \alpha_2) \text{ or } \tilde{\lambda}(\beta_1, \beta_2) \\ h_X(\beta_1, \beta_2) &= \lambda^{-1}(\beta_1, \beta_2) \text{ or } \tilde{\lambda}^{-1}(\alpha_1, \alpha_2). \end{aligned} \tag{10.50}$$

This means that (α_1, α_2) and (β_1, β_2) are eigenvectors of h_X with eigenvalues λ, λ^{-1} *except* for a possible interchange. The elliptic and hyperbolic groups are conjugate to K and to A. Neither K nor A interchanges the eigenvalues so this possibility does not occur because we are in $SL(2, \mathbb{R})/\pm I$ which corresponds to the proper Lorentz group.

We combine these ideas with Proposition 10.3 to deduce

Proposition 10.4 *The eigenvectors of $h_X \in H_X$ are the coefficients (α_1, α_2), (β_1, β_2) of the linear factors of Q_X. The factors are independent of h_X up to order and multiplication by constants μ, μ^{-1}. The eigenvalues are ($(-i$ times) in the elliptic case) the exponential of the angle of h_X.*

There is a geometry behind the factorization. The fact that (α_1, α_2) and (β_1, β_2) are eigenvectors of H_X means that the lines $\alpha_1 t_1 + \alpha_2 t_2 = 0$ and $\beta_1 t_1 + \beta_2 t_2 = 0$ are the fixed lines of h_X. In the elliptic case these lines are not real. But in the hyperbolic case they are real.

The hyperbolic rotation h_X acts on the fixed lines by multiplication by its eigenvalue. This is in conformity with our ideas of Section 3.4 which show how the action of H_X degenerates on the fixed lines.

Remark. *Since these lines are fixed the angle base point cannot be on them.*

We shall apply the spinor principle to give a new expression for hyperbolic angles.

Our introduction of coordinates in Minkowski space via the spreads $\{L_X^y\}$ leads to interesting coordinate systems on the Lorentz group G.

Suppose X_1, X_2 are elliptic. Let $gX_1 = X_2$. g is unique modulo right multiplication by $h_1 \in H_{X_1}$ and left multiplication by $h_2 \in H_{X_2}$. h_2 can be uniquely chosen by the requirement that g maps the angle base ray on $L^1_{X_1}$ onto the angle base ray $L^1_{X_2}$. We denote such g by $g^0(X_1, X_2)$.

When X_1 is fixed $\{g^0(X_1, X_2)\}$, modulo right multiplication by $h_{X_1} \in H_{X_1}$, gives a parametrization of the positive half of the hyperboloid $\{X \mid \|X\| = 1\}$. Since $g^0(X_1, X_2)$ takes an angle base ray into an angle base ray it maps the point

$\left(\begin{smallmatrix}\alpha_1 & 0\\ 0 & 0\end{smallmatrix}\right)$ on the angle base ray of X_1 which lies on Γ^+ on the corresponding point $\left(\begin{smallmatrix}\alpha_2 & 0\\ 0 & 0\end{smallmatrix}\right)$ for X_2. Thus, modulo right multiplication by $h_{X_1} \in H_{X_1}$, g^0 is of the form

$$g^0 = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \in NA.$$

Conversely if $g^0 \in NA$ then it maps the line $\{\left(\begin{smallmatrix}\alpha & 0\\ 0 & 0\end{smallmatrix}\right)\}$ onto itself. This means that

$$G = NAH_{X_1} \quad \text{(Iwasawa decomposition)}. \tag{10.51}$$

We can also parametrize the points of the half-hyperboloid $\{\det X = 1\}^+ = G/H_{X_1}$ using the spread $\{L^y_{X_2}\}_y$ for X_2 fixed. The intersections of the leaves $L^y_{X_2}$ with $\{\det X = 1\}^+$ cover the hyperboloid uniquely; they are orbits of H_{X_2}. This gives the polar decomposition

$$G = H_{X_2} A H_{X_1}$$

since A acts simply transitively on the ray $\{\left(\begin{smallmatrix}\alpha & 0\\ 0 & 0\end{smallmatrix}\right)\}_{\alpha>0}$ and hence on the leaves of the spread.

10.3 Eisenstein series and their periods

At several places in this book we have met the distribution $S = \sum \delta_n$ on $\mathbb{R}^1$. S is obtained from δ_0 by summing over the transforms of δ_0 under the action of $\mathbb{Z}$. More generally we have studied sums of the form $\sum f(x+n)$ where f is suitably small at infinity. Such a sum is manifestly $\mathbb{Z}$ invariant.

Let $\Gamma = SL(2,\mathbb{Z})$ be the modular subgroup of $G = SL(2,\mathbb{R})$; it is the group of 2×2 matrices of determinant 1 with integral entries. We want to use a similar summation process to construct interesting Γ invariant objects.

For the analog of $\sum \delta_n$ we pick some point p in $\mathbb{R}^3$ and form $\sum_\Gamma \delta_{\gamma p}$. In the case of $\mathbb{Z}$ action on $\mathbb{R}^1$ we started with δ_0; it would have made little difference had we started with δ_α for any $\alpha \in \mathbb{R}^1$ because $\mathbb{Z}$ is a subgroup of $\mathbb{R}^1$. But Γ is not a subgroup of $\mathbb{R}^3$ and the choice of p is significant.

We have seen in the previous section (see (10.8)ff.) that the isotropy group G_p of p in G depends in an important way on p. In particular, different p may lead to different conjugacy classes of G_p. In addition, there is a certain advantage if $\Gamma_p = \Gamma \cap G_p$ is large. For then we sum $\delta_{\gamma p}$ only over Γ/Γ_p and this helps for some convergence problems.

Once we have constructed the Γ invariant object $\hat{\Theta} = \sum \delta_{\gamma p}$ we want to form something "interesting" from it. Note that all the points γp lie on the variety

$$V_p = \{\hat{x} \mid \|\hat{x}\| = \|p\|\}.$$

This suggests that we form the Fourier transform (using the Lorentz inner product) Θ of $\sum \delta_{\gamma p}$ as the condition $\|\gamma p\| = \|p\|$ for all γ means that Θ satisfies

the differential equation

$$(\Box - \|p\|^2)\Theta = 0.$$

$\Box$ is the usual wave operator. Since G commutes with the Fourier transform, Θ is Γ invariant.

Being a solution of this equation Θ is determined to a large extent by its Cauchy data (CD) on G/K = positive half-hyperbola of two sheets. Since G/K is the Poincaré upper half-plane which is one of the primary objects of analysis we have "moved" $\hat{\Theta}$ to an interesting place.

We can go further. Whenever we have a function $F(x)$ which is invariant under the action of a group U on a manifold $\{x\}$ there is a way to decompose F into parts, each of which is U invariant. Let U^* be a group acting on $\{x\}$ which commutes with U. Then we can decompose F via the representations of U^*. For example, if U^* is commutative then we could decompose F into $\{F_\chi\}$ where F_χ transforms under U^* like the character χ of U^* (Fourier series or integrals on the orbits of U^*). F_χ is U invariant because U^* commutes with U.

Remark. *This method of decomposition is an illustration of a general theory of decomposition promulgated by von Neumann [125]. Let $\mathcal{A}$ be an algebra of bounded operators on a Hilbert space $\mathcal{H}$. $\mathcal{A}'$ is the commutator of $\mathcal{A}$, which is the set of bounded operators which commute with each $a \in \mathcal{A}$. When $\mathcal{A}'$ is simple then the decomposition of $\mathcal{H}$ into subspaces which are $\mathcal{A}$ invariant is accomplished by decomposition under $\mathcal{A}'$.*

For example, the set of $x \in \mathcal{H}$ which are eigenfunctions of all $a' \in \mathcal{A}$ with fixed eigenvalue $\lambda(a')$ is $\mathcal{A}$ invariant because

$$a'ax = aa'x = \lambda(a')ax.$$

In the present situation Γ (in fact all of G) commutes with scalar multiplication. This suggests that we form

$$\Theta_s(x) = \int_0^\infty \Theta(rx) r^s \frac{dr}{r}. \tag{10.52}$$

Θ_s is again Γ invariant. In case $\|p\| = 0$ separation of variables (to be clarified below) shows that the restriction of Θ_s to G/K is an eigenfunction of the noneuclidean Laplacian Δ. As such it is determined by its CD on suitable one-dimensional sets (details below). In particular we can choose these one-dimensional sets to be orbits $\mathcal{O}$ of elliptic, hyperbolic, or parabolic groups H.

Remark. *The data of Θ on $\mathcal{O}$ is given by the nonlocal operator (10.52) of harmonic decomposition under U^* (in our case this is the Mellin transform) and the CD on $\mathcal{O}$ for the noneuclidean Laplacian. Thus we can think of $\mathcal{O}$ as a WS for $\Box$ with these operators defining WD. According to our discussion at the beginning of this chapter we cannot give a completely local WP.*

Since Θ_s is Γ invariant its restriction to $\mathcal{O}$ is invariant under $\Gamma \cap H$. If $\Gamma \cap H$ is an infinite subgroup of Γ then, since H is an abelian group of rank 1, $\Gamma \cap H$

is an infinite cyclic group (perhaps times a finite group). Thus the restriction of Θ_s to $\mathcal{O}$ can be described by Fourier series coefficients, which constitute the relevant multiplicity Radon transform. These coefficients are called the *periods* of Θ_s on $\mathcal{O}$. They represent the primary object of study in this section.

Remark 1 *The periods represent the CD on $\mathcal{O}$ for the convolution equation defined by a generator of $\Gamma \cap H$ (see Section 1.4) in terms of the Fourier series basis.*

Remark 2 *We have noted that when p lies on the light cone then Θ_s is an eigenfunction of Δ. If p is not on the light cone then Θ_s satisfies a differential difference equation [112]. However, it is not necessary to define Θ_s as a weighted integral of $\Theta(r,x)$ using the weights r^s as in (10.52) and in [112]. Any weight will lead to a Γ invariant. In particular we could search for weights which produce eigenfunctions of Δ. Separation of variables tells us that such weights are Bessel functions [58].*

$\hat{\Theta}$ is a geometric object; it should be easier to make calculations using $\hat{\Theta}$ than Θ_s. In fact we shall apply the Parseval formula to perform the calculations of the periods in Fourier transform space where the results are much more transparent.

Our construction started with the point p which is invariant under Γ_p. In this case Θ_s is called an *Eisenstein series*—parabolic if $\|p\| = 0$, hyperbolic if $\|p\|^2 < 0$. We could have started with other interesting objects in place of δ_p, e.g. a character χ of a parabolic or a hyperbolic subgroup $\hat{J}$ of G on an orbit $\hat{O}$ of $\hat{J}$. If χ is $\Gamma \cap \hat{J}$ invariant so is $\chi\delta_{\hat{O}}$.[5] We can repeat our process of summation on $\Gamma/\Gamma \cap \hat{J}$, the Fourier transform, and decomposition under the Mellin transform in scalar multiplication. (See the discussion leading up to (10.52).) The resulting function on G/K is called a *Poincaré series* (parabolic or hyperbolic). We shall explain how to compute its periods in Section 10.4.

Our first task is to determine when Γ_p in infinite. This leads us into the algebraic number theory of quadratic fields $K(\sqrt{d})$ for which we now give some background. This background can be found, for example, in [92].

In most of our applications $d > 0$, which we shall assume, although there is no essential difficulty in dealing with imaginary quadratic fields. We shall sometimes make simplifying assumptions to avoid computational difficulties; these assumptions can be removed.

As usual we denote by $\Gamma = SL(2,\mathbb{Z})$ the modular group which is the group of 2×2 matrices of determinant 1 with integral entries. Γ_∞ is the subgroup

[5] For any set S with some fixed measure μ we denote by δ_S the linear function whose value on a function ϕ is

$$\delta_S \cdot \phi = \int_S \phi \, d\mu.$$

that leaves the point ∞ in the (closure of the) Poincaré half-plane fixed so

$$\Gamma_\infty = \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \right\}. \tag{10.53}$$

From our point of view the relation of Γ to real quadratic number fields arises in two ways:

(1) When G acts on the upper half-plane the fixed points of hyperbolic elements of Γ belong to real quadratic number fields. We call $\alpha = \left(\begin{smallmatrix} a & b \\ c & \delta \end{smallmatrix}\right)$. $\alpha z = z$ means

$$\begin{aligned} &\frac{az+b}{cz+\delta} = z \\ &cz^2 + (\delta - a)z - b = 0 \\ &z = \frac{a - \delta \pm \sqrt{(\delta - a)^2 + 4bc}}{2c}. \end{aligned}$$

Since $a\delta - bc = 1$

$$z = \frac{a - \delta \pm \sqrt{(a+\delta)^2 - 4}}{2c} = (\omega, \omega'). \tag{10.54}$$

An element $\alpha \in \Gamma$ is called hyperbolic if $|\operatorname{tr} \alpha| > 2$. This is the same as saying that $\alpha \neq I$ is conjugate in G to a real diagonal matrix, i.e. an element of the subgroup A (see (10.8)). Thus a hyperbolic element $\alpha \in \Gamma$ is associated to the real quadratic number field $K = Q(\sqrt{d})$ where

$$d = (\operatorname{tr} \alpha)^2 - 4.$$

The fixed points ω, ω' are conjugate in $Q(\sqrt{d})$.

(2) Since α leaves ω, ω' fixed it leaves the geodesic through ω, ω' fixed. The set of all $g \in G$ leaving ω, ω' fixed is a subgroup $G_{\omega\omega'}$ which is a conjugate of A. We have remarked that $A = G_{0\infty}$. Thus to write $G_{\omega\omega'}$ explicitly we conjugate by the element (10.16) sending $(0, \infty) \to (\omega', \omega)$. (Since $\omega' > \omega$ we set $\xi_2 = \omega', \xi_1 = \omega$ in (10.16).) We find, for $g \in G_{\omega\omega'}$ which is the conjugate of $\left(\begin{smallmatrix} a & 0 \\ 0 & a^{-1} \end{smallmatrix}\right)$

$$\begin{aligned} g_{\omega\omega'} &= \frac{1}{\omega' - \omega} \begin{pmatrix} \omega' & \omega \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 1 & -\omega \\ -1 & \omega' \end{pmatrix} \\ &= \frac{1}{\omega' - \omega} \begin{pmatrix} \omega' a - \omega a^{-1} & -\omega\omega'(a - a^{-1}) \\ a - a^{-1} & \omega' a^{-1} - \omega a \end{pmatrix}. \end{aligned} \tag{10.55}$$

In particular,

$$\begin{aligned} \operatorname{tr} g_{\omega\omega'} &= \frac{1}{\omega' - \omega}(\omega' - \omega)(a + a^{-1}) \\ &= a + a^{-1}. \end{aligned}$$

We have noted that, in terms of three-dimensional action, A leaves $\left(\begin{smallmatrix} 0 & y \\ y & 0 \end{smallmatrix}\right)$ fixed. Thus $G_{\omega\omega'}$ leaves

$$\begin{pmatrix} \omega' & \omega \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & y \\ y & 0 \end{pmatrix} \begin{pmatrix} \omega' & 1 \\ \omega & 1 \end{pmatrix} = \begin{pmatrix} 2\omega\omega' y & (\omega+\omega')y \\ (\omega+\omega')y & 2y \end{pmatrix} \tag{10.56}$$

invariant. For $y = 1$ this symmetric matrix corresponds (except for a factor $2y$) to the quadratic form

$$\omega\omega' u^2 + (\omega+\omega')uv + v^2 = (\omega u + v)(\omega' u + v).$$

When ω, 1 is an integral basis for an ideal $\mathfrak{A}$ in $Q(\sqrt{d})$ with $\omega' - \omega = N(\mathfrak{A})\sqrt{d}$ then this quadratic form (i.e. its coefficients) is divisible by $N(\mathfrak{A})$ [92, p. 213].

There is a general procedure due to Gauss which relates ideal classes in quadratic fields to binary quadratic forms with integral coefficients [92, p. 210ff.]. If $\mathfrak{A}$ is an integral ideal then $\mathfrak{A}$ has a basis α_1, α_2 where

$$\alpha_1\alpha_2' - \alpha_1'\alpha_2 = N(\mathfrak{A})\sqrt{d} > 0.$$

Using this basis we can relate to $\mathfrak{A}$ the integral quadratic form

$$Q(u,v) = \frac{(\alpha_1 u + \alpha_2 v)(\alpha_1' u + \alpha_2' v)}{|N(\mathfrak{A})|}. \tag{10.57}$$

We have

$$\det Q = -\tfrac{1}{4} \text{ discriminant } Q = -\tfrac{1}{4}d$$

(by (10.11)).

$\mathfrak{A} \leftrightarrow Q$ sets up a one–one correspondence between narrow ideal classes in the field K and equivalence classes of semi-integral quadratic forms of determinant $-\frac{1}{4}d$. "Semi-integral" means that the diagonal entries of the matrix are integral and the off-diagonal terms are integers or half-integers. We say that $Q \sim Q_1$ if Q is conjugate to Q_1 by an element in Γ. (Q is identified with a 2×2 real symmetric matrix of determinant $-d$.)

From our geometric point of view this means that narrow ideal classes in K correspond to orbits of Γ on the semi-integral points on the hyperboloid $\det X = -d$.

We have denoted the fixed points by ω, ω' and in (10.56) we have associated to the pair the symmetric matrix

$$\begin{pmatrix} 2\omega\omega' & \omega+\omega' \\ \omega+\omega' & 2 \end{pmatrix}.$$

To be able to handle the general quadratic form (10.57) it is preferable to denote the fixed points by

$$\omega = \alpha_1/\alpha_2, \quad \omega' = \alpha_1'/\alpha_2'. \tag{10.58}$$

Instead of the matrix

$$(\omega' - \omega)^{-1/2} \begin{pmatrix} \omega & \omega' \\ 1 & 1 \end{pmatrix}$$

we use

$$\frac{1}{N(\mathfrak{A})^{1/2}} \begin{pmatrix} \alpha_1 & \alpha_1' \\ \alpha_2 & \alpha_2' \end{pmatrix} \tag{10.59}$$

which has determinant $\sqrt{d}$. The fractional linear transform defined by (10.59) takes $0 \to \omega' N(\mathfrak{A})^{-\frac{1}{2}}, \infty \to \omega N(\mathfrak{A})^{-\frac{1}{2}}$.

The 2×2 symmetric matrix associated to (10.59) is the image of the "base point" $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ by (10.59), namely

$$\begin{aligned} &\frac{1}{N(\mathfrak{A})} \begin{pmatrix} \alpha_1 & \alpha_1' \\ \alpha_2 & \alpha_2' \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_1' & \alpha_2' \end{pmatrix} \\ &\qquad = \frac{1}{N(\mathfrak{A})} \begin{pmatrix} 2\alpha_1\alpha_1' & \alpha_1\alpha_2' + \alpha_2\alpha_1' \\ \alpha_1\alpha_2' + \alpha_2\alpha_1' & 2\alpha_2\alpha_2' \end{pmatrix}. \end{aligned} \tag{10.60}$$

Except for the harmless factor 2 this agrees with Hecke's quadratic form (10.57) [92, p. 213].

Remark. *It can be shown that for any $d > 0$ and any $\omega \in Q(\sqrt{d})$ there is a $\gamma \in \Gamma$ whose fixed points are ω, ω'. This means that every integral quadratic form F of discriminant d corresponds to a conjugate pair of fixed points of some $\gamma \in \Gamma$. We can write ω in the form α_1/α_2 where α_1, α_2 are integers in $Q(\sqrt{d})$ and*

$$F(x, y) = \frac{\alpha_1\alpha_1' x^2 + (\alpha_1\alpha_2' + \alpha_2\alpha_1')xy + \alpha_2\alpha_2' y^2}{\|\mathfrak{A}\|}.$$

This correspondence between fixed points of elements of Γ and quadratic forms is not valid for definite quadratic forms (d negative). For, if $\gamma \in \Gamma$ is elliptic then $\operatorname{tr} \gamma = 0, \pm 1$. By (10.54)ff. $d = \operatorname{tr}^2\gamma - 4$. Thus $d = -4$ or $d = -3$ are the only possible negative discriminants. To obtain other imaginary quadratic fixed points (which are in the upper half-plane) we can use integral matrices of determinant $-d$ for $d < 0$. For example, the fixed points of $\left(\begin{smallmatrix} 0 & d \\ 1 & 0 \end{smallmatrix}\right)$ are $\pm\sqrt{d}$. In this way we can obtain the ideal classes in $Q(\sqrt{d})$ for $d < 0$ in terms of such fixed points. This construction is closely related to the theory of Hecke operators [91]. This aspect of the Hecke operators is discussed in Chapter 8.

Before going further we give the classical description of the set of primitive lattice points on the light cone; they correspond to primitive Pythagorean triples (meaning k, l, m have no common factor). If $k^2 = l^2 + m^2$ then one of l, m (say m) is $\equiv 0, 2 \bmod 4$ and the others are $\equiv \pm 1 \bmod 4$. Write $m = 2m'$ so

$$\left(\frac{k-l}{2}\right)\left(\frac{k+l}{2}\right) = m'^2. \tag{10.61}$$

Let p be an odd prime which divides m'. Then p^2 divides the left side of (10.61) so either $p^2|k-l$, or $p^2|k+l$, or $p|k-l$ and $p|k+l$. In the latter case $p|k$ and $p|l$. Going back to the Pythagorean relation we could divide by p and obtain another Pythagorean triple. It follows that for primitive Pythagorean triples either $p^2|k-l$ or $p^2|k+l$.

By the same reasoning if $p^\alpha|m'$ then $p^{2\alpha}|k-l$ or $p^{2\alpha}|k+l$. A similar result is true for $p=2$. This means that if

$$m' = p_1^{\alpha_1}\dots p_r^{\alpha_r}$$

then

$$\begin{aligned} k+l &= 2p_{q(1)}^{2\alpha_{q(1)}}\dots p_{q(s)}^{2\alpha_{q(s)}} = 2u^2 \\ k-l &= 2p_{q'(1)}^{2\alpha_{q'(1)}}\dots p_{q'(r-s)}^{2\alpha_{q'(r-s)}} = 2v^2 \end{aligned}$$

where $[q(1),\dots,q(s),q'(1),\dots,q'(r-s)]$ is a partition of $1,\dots,r$. Hence

$$k = u^2+v^2, \quad l = u^2-v^2, \quad m = 2uv. \tag{10.62}$$

This is the solution of the primitive Pythagorean identity when m is even. When l is even the result is similar.

The important point for us is that if k^2 is a norm in the Gaussian field in a nontrivial way, i.e. $lm \neq 0$, then so is k. This fact is also a consequence of factorization in $Q(i)$.

We have observed in (10.13) that points on the positive light cone can be written in the form $\left(\begin{smallmatrix} u^2 & uv \\ uv & v^2 \end{smallmatrix}\right)$. In view of the identification (10.5) this coincides with the parametrization of the lattice points just given. Thus the quadratic parametrization can be thought of as an integral parametrization of (half) the points on G/MN.

From the group theoretical point of view it is natural to study the Γ orbit of a point with integral coordinates on the light cone, e.g. $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. We have

$$\begin{pmatrix} p & a \\ q & b \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} p & q \\ a & b \end{pmatrix} = \begin{pmatrix} p^2 & pq \\ pq & q^2 \end{pmatrix}. \tag{10.63}$$

By our identification of this with (klm)

$$k = \tfrac{1}{2}(p^2+q^2), \quad l = \tfrac{1}{2}(p^2-q^2), \quad m = pq. \tag{10.64}$$

Except for the factor $\frac{1}{2}$ which can be removed by starting from $\beta = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ we obtain exactly the primitive Pythagorean triples with m even, since any relatively prime (p,q) forms the first column of an element of Γ.

We have discussed the distribution

$$\hat{\Theta} = \sum \delta_{klm}$$

with the sum taken over all primitive lattice points on the light cone, i.e. all primitive Pythagorean triples. It is clear that $\hat{\Theta}$ is Γ invariant. We use the Minkowski inner product to define the Fourier transform so Θ is also Γ invariant. Since Γ commutes with scalar multiplication so does the Mellin transform $\tilde{E}_s$ of Θ in scalar multiplication (see (10.52)).

The Fourier transform of δ_x is $\exp(ix \cdot \hat{x})$. Hence

$$\Theta = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} e^{ix \cdot \gamma \beta \gamma^t}. \tag{10.65}$$

(Actually this represents only half of the primitive lattice points; the other half is dealt with in the same manner.)

We write $x = r\tilde{x}$ where $\tilde{x} \in G/K$, meaning $\|\tilde{x}\| = 1$. Since G/K is the G orbit of I, $\tilde{x}$ is of the form $\tilde{x} = gg^t$. Hence

$$\tilde{E}_s(\tilde{x}) = A(s) \sum (gg^t \cdot \gamma \beta \gamma^t)^{-s} \tag{10.66}$$

where $A(s)$ involves the Γ function and elementary functions.[6]

Now, gg^t represents a real positive definite symmetric matrix of determinant 1. Thus g represents a point in G/K for which the upper half-plane coordinates of gg^t are given in (10.11). The inner product of a symmetric matrix with a point on the light cone is

$$\begin{aligned} gg^t \cdot \gamma \beta \gamma^t &= \begin{pmatrix} \eta + \xi^2 \eta^{-1} & \xi \eta^{-1} \\ \xi \eta^{-1} & \eta^{-1} \end{pmatrix} \cdot \begin{pmatrix} p^2 & pq \\ pq & q^2 \end{pmatrix} \\ &= (\eta + \xi^2 \eta^{-1}) q^2 - 2\xi \eta^{-1} pq + \eta^{-1} p^2 \\ &= \eta^{-1} [(\xi^2 + \eta^2) q^2 - 2\xi pq + p^2] \\ &= \eta^{-1} |qz - p|^2. \end{aligned} \tag{10.67}$$

As usual $z = \xi + i\eta$. We can rewrite this calculation in the Gauss–Hecke spirit discussed above (see (10.56)):

$$\begin{aligned} gg^t &= \frac{2i}{(\xi + i\eta) - (\xi - i\eta)} \begin{pmatrix} (\xi + i\eta)(\xi - i\eta) & \frac{1}{2}[(\xi + i\eta) + (\xi - i\eta)] \\ \frac{1}{2}[(\xi + i\eta) + (\xi - i\eta)] & 1 \end{pmatrix} \\ &= \frac{2i}{z - \bar{z}} \begin{pmatrix} z\bar{z} & \frac{1}{2}(z + \bar{z}) \\ \frac{1}{2}(z + \bar{z}) & 1 \end{pmatrix} \end{aligned} \tag{10.68}$$

[6]The reader should distinguish between Γ = modular group, the Γ function, and $\Gamma =$ light cone.

which leads to an alternative derivation of (10.67).

We have established

$$\begin{aligned}\tilde{E}_{is}(z) &= A(s) \sum_{(p,q)=1} \frac{\eta^{is}}{|qz-p|^{2is}} \\ &= A(s)E(z,is). \end{aligned} \tag{10.69}$$

$E(z,is)$ is called the *Eisenstein series*. Sometimes we modify the Eisenstein series to $\mathbf{E}(z,is)$ which is defined as the sum over all pairs $(p,q) \neq (0,0)$, which means that we sum over all orbits of $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 2\alpha \end{smallmatrix}\right)$ with $\alpha \in \mathbb{Z}$, $\alpha \neq 0$. As mentioned above this is half the lattice points on the light cone. Clearly

$$\mathbf{E}(z,s) = \zeta(2s)E(z,s). \tag{10.70}$$

We have shown

Proposition 10.5 *The Mellin transform of* Θ *evaluated on the upper half-plane is an elementary factor times* $\mathbf{E}(z,is)$.

Our construction of the Eisenstein series comprised three procedures:

(1) Form $\hat{\Theta} = \delta_{(\Gamma/\Gamma_{\hat{p}})\hat{p}}$ with $\hat{p} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$.

(2) Fourier transform to obtain $\Theta(x) = \widehat{\hat{\Theta}(x)} = \sum_{\gamma \in \Gamma/\Gamma_{\hat{p}}} \exp(ix \cdot \gamma\hat{p})$

(3) Take the Mellin transform of Θ in scalar multiplication.

It is more convenient for us to interchange (2) and (3). The commutation of Fourier and Mellin transforms is studied in detail in [63]. In the present case they essentially commute (clarified below).

To understand this point, we start with a point $\hat{p} \neq 0$. The Fourier transform of $\delta_{\hat{p}}$ is $\exp(ix \cdot \hat{p})$. The Mellin transform (under the action of scalar multiplication by $\mathbb{R}^+$) of $\hat{\delta}_{\hat{p}}$ is given formally by

$$\begin{aligned}\hat{\underset{\vee}{\delta}}_{\hat{p}}(x,s) &= \int e^{irx\cdot\hat{p}} r^{is} \frac{dr}{r} \\ &= (-ix \cdot \hat{p})^{-is}\Gamma(is). \end{aligned} \tag{10.71}$$

The analysis necessary to validate this formalism is detailed in Chapter 7. (We put the symbol $\underset{\vee}{}$ for Mellin transform to the left of the symbol $\hat{}$ for Fourier transform to indicate that the Fourier transform is performed first.)

On the other hand the Mellin transform of $\delta_{\hat{p}}$ (under scalar multiplication) is

$$\underset{\vee}{\delta}_{\hat{p}} = \hat{r}^{is}\delta_{\sigma(\hat{p})} \tag{10.72}$$

where $\sigma(\hat{p})$ is the ray $\{\hat{r}\hat{p}\}$ through $\hat{p}$ with measure $\hat{r}\,d\hat{r}$. The change in measure from $d\hat{r}/\hat{r}$ to $\hat{r}\,d\hat{r}$ is a consequence of the fact that $\delta_{\hat{p}}$ is defined using volume

measure whereas the measure on $\sigma(\hat{p})$ is linear. (Equation (10.72) means that $\delta_{\sigma(\hat{p})}$ is multiplied by $\hat{r}^{is}$ at the point $\hat{r}\hat{p}$.) The Fourier transform of $\underset{\vee}{\delta}_{\hat{p}}$ is

$$\begin{aligned}\hat{\underset{\vee}{\delta}}_{\hat{p}}(x,s) &= e^{ix\cdot\hat{x}}\cdot\hat{r}^{is}\delta_{\sigma(\hat{p})}\\ &= \int e^{ix\cdot\hat{r}\hat{p}}\hat{r}^{i(s+1)}\,d\hat{r}\\ &= \hat{\underset{\vee}{\delta}}_{\hat{p}}(x,s+2)\end{aligned}\tag{10.73}$$

by (10.71). We conclude

Lemma 10.6 *The Fourier transform and the Mellin transform in scalar multiplication commute except for translations in s.*

We change our notation to ignore this change in s: namely, we define $E(z,s)$ by interchanging (2) and (3). We shall use $d\hat{r}/\hat{r}$ as the measure on $\sigma(\hat{p})$.

Remark. *We shall generally work with $E(z,s)$ rather than $\mathbf{E}(z,s)$ because the set of primitive lattice points is easier to study from a group theoretical, i.e. Γ, point of view. If c is a scalar then*

$$\underset{\vee}{\delta}_{c\hat{p}}(s) = c^{-is}\underset{\vee}{\delta}_{\hat{p}}(s)$$

so the change from E to $\mathbf{E}$ in the Fourier transform space involves multiplication by $\zeta(is)$.

Let $\Omega = G/K = \{t^2 - x^2 - y^2 = 1\}$. As we have seen Ω^+ can be identified as a $G(= SL(2,\mathbb{R}))$ space with the Poincaré upper half-plane and is an important space for analysis.

Let us return to the notation of (10.20). $L_X^y \cap \Omega$ is either empty or a point or a conic (or half a conic). We call $H = H_X$ the orthogonal group of X. H acts on L_X^y and on Ω, and hence on $L_X^y \cap \Omega$.

Consider the simplest case $X = I$. Clearly, for any y^0,

$$L_I^{y^0} \cap \Omega = \{y = y^0, t_1^2 + t_2^2 = (y^0)^2 - 1\}.$$

In this case $H_I = K$ is the group of rotations around the y axis and so acts simply transitively on $L_I^{y^0} \cap \Omega$. The distance base point is $y^0 I$. Moreover

$$L_I^{y^0} \cap \Omega = \{\text{points on } L_I^{y^0} \text{ the square of whose distance to } y^0 I \text{ is } -[(y^0)^2 - 1]\}.$$

By making a Lorentz transformation we deduce, if X lies in the interior of the light cone,

$$\begin{aligned}L_X^y \cap \Omega = \{&\text{points on } L_X^y \text{ the square of whose distance to the distance base}\\ &\text{point } yX \text{ is } -[y^2 - 1]\}.\end{aligned}\tag{10.74}$$

This is the analog of Proposition 10.1 for Ω in place of Γ. Moreover H_X acts simply transitively on $L_X^y \cap \Omega$.

If $\tilde{I} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ then by (10.5)

$$L_{\tilde{I}}^{y^0} \cap \Omega = \{t_2 = y^0, y^2 - t_1^2 = 1 + (y^0)^2\}.$$

Now

$$H_{\tilde{I}} = A = \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \right\}.$$

$H_{\tilde{I}}$ acts simply transitively on each of the half-hyperbolas of $L_{\tilde{I}}^{y^0} \cap \Omega$.

If we make a Lorentz transformation we find, if $\det X < 0$,

$$L_X^y \cap \Omega = \{\text{points on } L_X^y : \quad \text{distance to } yX = [1 + y^2]^{1/2}\}. \tag{10.75}$$

Moreover $L_X^y \cap \Omega$ consists of two components on each of which H_X acts simply transitively.

Remark. *$L_X^y \cap \Omega$ is a geodesic in the Poincaré upper half-plane when $y = 0$.*

Finally if $X = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ then we examine $L_X^y \cap \Omega$ where L_X^y is defined in (10.38)ff. Thus for $y \neq 0$

$$L_X^y \cap \Omega = \{xy - w^2 = 1\} \tag{10.76}$$

is a parabola on which $H_X = \{\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)\}$ acts simply transitively since for $y \neq 0$

$$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{-1} & 0 \\ 0 & y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix} = \begin{pmatrix} y^{-1} + n^2 y & ny \\ ny & y \end{pmatrix}. \tag{10.77}$$

The right side of (10.77) is an arbitrary symmetric matrix satisfying (10.76) since we can set $n = w/y$ and then

$$x = (1 + w^2)/y = y^{-1} + n^2 y$$

in conformity with (10.77).

For $y = 0, \det \left(\begin{smallmatrix} x & w \\ w & 0 \end{smallmatrix}\right) = -w^2 \neq 1$ so $L_X^0 \cap \Omega$ is empty.

Remark. *We now compute the periods of the Eisenstein series over elliptic, hyperbolic, and parabolic cycles. At the end of this section we explain the abstract idea behind the computations.*

We want to compute the periods of the Eisenstein series $E(z, s)$ on $L_X^y \cap \Omega$. The periods are defined, formally, by

$$\chi_H \delta_{L_X^y \cap \Omega} \cdot E(z, s) = \int_{z \in L_X^y \cap \Omega} E(z, s) \chi(h)\, dh \tag{10.78}$$

where, for X elliptic or hyperbolic, χ_H is a character of H which is $\Gamma_H = \Gamma \cap H$ invariant. (There is an analogous expression in the parabolic case.) Of course when Γ_H is infinite (which can only happen in the hyperbolic case) (10.78) has no meaning as it stands because the integral over $L_X^y \cap \Omega$ is an infinite number of copies of the integral over $\mathcal{D}(\Gamma_H; L_X^y \cap \Omega)$ which is the fundamental domain for Γ_H acting on $L_X^y \cap \Omega$. We thus interpret (10.78) as an integral over $\mathcal{D}(\Gamma_H; L_X^y \cap \Omega)$.

We have given a geometric interpretation of the Fourier transform of $E(z, s)$ by means of (10.72) (summed over Γ). On the other hand, the factor $\chi_H \delta_{L_X^y \cap \Omega}$ is geometric. Thus one factor of (10.78) is simpler in x space than in $\hat{x}$ space while the reverse is true for the other factor. For making calculations we have found it more illuminating to apply Parseval's formula and make calculations in $\hat{x}$ space.

Remark 1 *According to our formalism we can regard the periods of the Eisenstein series as a "double Radon transform." Thus (10.72) shows how the Fourier transform of the Eisenstein series is formed as the Mellin transform (multiplicity Radon transform) of* $\hat{\Theta}$ *while the periods over a cycle form another multiplicity Radon transform.*

Remark 2 Θ *is a more complicated object then* $\chi_H \delta_{L_X^y \cap \Omega}$ *so our application of Parseval's formula is in keeping with the idea of simplifying the most complicated object.*

Actually the Fourier transform of $\chi_H \delta_{\mathcal{D}(\Gamma_H; L_X^y \cap \Omega)}$ is complicated; we can simplify it using an *unfolding lemma* which allows us to replace $\delta_{\mathcal{D}(\Gamma_H; L_X^y \cap \Omega)}$ by $\delta_{L_X^y \cap \Omega}$ whose Fourier transform can be analyzed by classical methods.

We present here a simplified unfolding lemma; a more comprehensive result is given in Lemma 10.7 below. In the unfolding lemma we start with a function f^0 and define

$$f = \sum_{\gamma \in \Gamma^0 \backslash \Gamma} \gamma f^0$$

where Γ^0 is a subgroup of Γ leaving f^0 invariant. We want to integrate f over $\mathcal{D} = \mathcal{D}(\Gamma_H; L_X^y \cap \Omega)$. Roughly speaking we expect that

$$\int_{\mathcal{D}} f = \int_{L_X^y \cap \Omega} \sum_{\tilde{\gamma} \in \Gamma^0 \backslash \Gamma / \Gamma_H} \tilde{\gamma} f^0 \tag{10.79}$$

because

$$L_X^y \cap \Omega = \cup_{\gamma_H \in \Gamma_H} \gamma_H \mathcal{D}.$$

For the simplest precise result, suppose we can find a set $\{\tilde{\gamma}\} = \tilde{\Gamma}$ which are distinct mod (left) Γ^0 and such that $\{\tilde{\gamma}\gamma_H\}_{\tilde{\gamma} \in \tilde{\Gamma}, \gamma_H \in \Gamma_H}$ is uniquely a set of representatives for $\Gamma^0 \setminus \Gamma$. By "uniquely" we mean that all products $\tilde{\gamma}\gamma_H$ are

distinct modulo Γ^0. In this case

$$\begin{aligned}\int_{\mathcal{D}} f &= \int_{\mathcal{D}} \sum_{\gamma\in\Gamma^0\backslash\Gamma} \gamma f^0 \\ &= \int_{\mathcal{D}} \sum_{\tilde\gamma,\gamma_H} \tilde\gamma\gamma_H f^0 \\ &= \sum_{\gamma_H} \int_{\gamma_H\mathcal{D}} \sum_{\tilde\gamma} \tilde\gamma f^0 \\ &= \int_{L^y_X\cap\Omega} \sum \tilde\gamma f^0. \end{aligned} \tag{10.80}$$

$\tilde\Gamma$ is clearly a set of representatives for the double cosets, confirming our naive idea (10.79), provided the unique representation is valid.

The unique representation has a natural interpretation: think of Γ_H as acting on the homogeneous space $\Gamma^0\backslash\Gamma$ (on the right).

The orbits of Γ_H are *generic* (compare Section 3.1), meaning if $\gamma_H \neq \gamma_H^1$ then for no $\gamma\in\Gamma, \gamma_0\in\Gamma^0$ do we have

$$\gamma\gamma_H = \gamma^0\gamma\gamma_H^1. \tag{10.81}$$

We can rewrite (10.81) as[7]

$$\gamma_1^0\gamma = \gamma\gamma_H^2. \tag{10.82}$$

This is the same as saying that no element of Γ_H except the identity fixes a point of $\Gamma^0\backslash\Gamma$. In this case the set $\tilde\Gamma$ consists of one point from each orbit. Since clearly $\gamma^0\neq 1$, (10.82) also says that the left action of Γ^0 on Γ/Γ_H is generic.

We have proven

Lemma 10.7 (Unfolding lemma.) *Let f^0 be a function or distribution on $L^y_X\cap\Omega$; we call Γ^0 the subgroup of elements of Γ leaving f^0 fixed. Suppose Γ_H acts generically on $\Gamma^0\backslash\Gamma$. Assume the series $f=\sum_{\gamma\in\Gamma^0\backslash\Gamma}\gamma f^0$ converges suitably. Then for any Γ_H invariant function χ we have*

$$\int_{\mathcal{D}} \chi f = \int_{L^y_X\cap\Omega} \chi \sum_{\gamma\in\Gamma^0\backslash\Gamma/\Gamma_H} \gamma f^0. \tag{10.83}$$

Remark. *We shall meet instances in which the action of Γ_H is not generic.*

Suppose $\tilde\gamma$ is a representative for a double coset $\Gamma^0\tilde\gamma\Gamma_H$ for which there is a subgroup $\Gamma_H(\tilde\gamma)$ of Γ_H which preserves the left coset $\Gamma^0\tilde\gamma$. Then the set of cosets

[7] In γ_H^2 the 2 is a superscript, not a power.

$\{\Gamma^0\tilde{\gamma}\gamma_H\}_{\gamma_H\in\Gamma_H}$ consists of the cosets $\{\Gamma^0\tilde{\gamma}\gamma_H^*\}_{\gamma_H^*\in\Gamma_H(\tilde{\gamma})\backslash\Gamma_H}$ repeated *order* $\Gamma_H(\tilde{\gamma})$ times. Hence

$$\int_{L_X^y\cap\Omega}\tilde{\gamma}f^0=\int_{\mathcal{D}}\sum_{\gamma_H\in\Gamma_H}\tilde{\gamma}\gamma_H f^0$$

$$=|\Gamma_H(\tilde{\gamma})|\int_{\mathcal{D}}\sum_{\gamma_H\in\Gamma_H(\tilde{\gamma})\backslash\Gamma_H}\tilde{\gamma}\gamma_H f^0 \tag{10.84}$$

where $|\Gamma_H(\tilde{\gamma})|=$ order $\Gamma_H(\tilde{\gamma})$.

Equation (10.84) shows how to modify the heuristic (10.79) for nongeneric double cosets $\tilde{\gamma}$; that is, we have to multiply $\tilde{\gamma}f^0$ in the right side of (10.79) by $|\Gamma_H(\tilde{\gamma})|$.

Some special methods for dealing with such orbits appear below.

We are now in a position to calculate the periods of the Eisenstein series. We distinguish the following:

(a) $X\in$ interior light cone (elliptic cycles).
(b) $X\in$ exterior light cone (hyperbolic cycles).
(c) $X\in$ light cone (parabolic cycles).

Remark. *Our nomenclature "elliptic, hyperbolic, parabolic" is geared to the subgroups of the continuous group $G=SL(2,\mathbb{R})$. There is an analogous nomenclature related to Γ: for example, a geodesic joining two parabolic fixed points of Γ is called "parabolic."*

(a) Elliptic cycles

To understand how things work let us start with the simplest case when $X=I$. Thus by (10.74)

$$L_X^y\cap\Omega=\{\rho^2=-(y^2-1)\} \tag{10.85}$$

where $\rho=$ Lorentz distance on L_I^y to the distance base point yI. In particular $L_I^y\cap\Omega$ is empty if $y<1$. When $y>1$ the intersection is a circle; in the Poincaré half-plane it is a hyperbolic circle centered at i.

We begin by computing the simplest period, which corresponds to the trivial character of H_I.

Parseval's formula calls for us to take the Fourier transform of $\delta_{L_I^y\cap\Omega}$ which is, by classical results and Theorem 10.2,

$$\hat{\delta}_{L_I^y\cap\Omega}(x)=cJ_0[(y^2-1)^{1/2}|\hat{\rho}_I(\hat{x})|]e^{iy\hat{y}}. \tag{10.86}$$

Here J_0 is the standard Bessel function. We have applied Theorem 10.2 (which is obvious in this case) which allows us to compute $\hat{\delta}_{L_I^y\cap\Omega}$ by taking its Fourier transform in the plane L_I^y and then multiplying by $\exp(iy\hat{y})$.[8]

[8]We shall usually write $\hat{\rho}$ for $|\hat{\rho}|$ since there is little chance of confusion.

Since $L_I^y \cap \Omega$ is compact there is no need to apply the unfolding lemma; we can integrate over the whole of $L_I^y \cap \Omega$. The same is true when I is replaced by an arbitrary elliptic X.

The period of $E(z, is)$ over $L_I^y \cap \Omega$ is evaluated by Parseval's formula

$$\begin{aligned} \delta_{L_I^y \cap \Omega} \cdot E &= \textit{is the standard Bessel } \hat{\delta}_{L_I^y \cap \Omega} \cdot \hat{E} \\ &= A\delta_{L_I^y \cap \Omega} \cdot \sum \hat{r}^{is} \delta_{\sigma(\hat{p})} \end{aligned} \tag{10.87}$$

where Δ, as usual, is a simple factor. By use of a suitable analytic continuation (see Chapter 7 for details in an analogous situation) we can interchange the order of summation and integration.

Since $\hat{\delta}_{L_I^y \cap \Omega}$ depends only on $\hat{r}, y$, to evaluate (10.87) we need only evaluate the distance polar coordinate on $\delta_{\sigma(\hat{p})}$ in every $\hat{L}_I^{\hat{y}}$. This is $\hat{\rho}_I(\hat{r}\hat{p})$ when $\hat{y} = \hat{y}(\hat{r}\hat{p})$ by Proposition 10.1. Equation (10.86) leads to

$$\hat{\delta}_{L_I^y \cap \Omega} \cdot \hat{r}^{is} \delta_{\sigma(\hat{p})} = \int e^{iy\hat{y}(\hat{r}\hat{p})} J_0[(y^2-1)^{1/2} \hat{\rho}_I(\hat{r}\hat{p})] \hat{r}^{is} \frac{d\hat{r}}{\hat{r}}. \tag{10.88}$$

It is convenient to interpret the integral in (10.88) geometrically. $\hat{r}^{is}\delta_{\sigma(\hat{p})}$ is supported on $\sigma(\hat{p}) = \{\hat{r}\hat{p}\}$. The ray $\sigma(\hat{p})$ meets $\hat{L}_I^{\hat{y}}$ in the point with cylindrical coordinates $\hat{y} = \hat{y}_I(\hat{r}\hat{p})$ and $\hat{L}_I^0$ coordinate $\hat{r}_I(\hat{r}\hat{p})$. The value of $\hat{\delta}_{L_I^y \cap \Omega}$ at this point is $\exp[iy\hat{y}_I(\hat{r}\hat{p})] J_0[(y^2-1)^{1/2}\hat{\rho}_I(\hat{r}\hat{p})]$. Equation (10.88) now follows from Theorem 10.2 which "separates" the y_I and L_I^0 coordinates.

Since $\hat{r}\hat{p}$ lies on the light cone we can apply Proposition 10.1 which shows that

$$\hat{\rho}_I^2(\hat{r}\hat{p}) = -\hat{y}_I^2(\hat{r}\hat{p}).$$

Notation. Henceforth we shall generally denote lattice points on the light cone as $\mathbf{p}, \mathbf{q}$, etc., to avoid confusion.

We now fix $\hat{r}$. We write $\hat{\mathbf{p}}$ in the form

$$\hat{\mathbf{p}} = \begin{pmatrix} p^2 & pq \\ pq & q^2 \end{pmatrix}$$

so $\hat{y}_I(\hat{r}\hat{p}) = \hat{r}(p^2+q^2)$. In summing over $\hat{\mathbf{p}}$ it is convenient to sum over all $\hat{\mathbf{p}}$ for which $\hat{y}(\hat{\mathbf{p}}) = p^2+q^2$ is fixed, i.e. all $\hat{\mathbf{p}}$ lying in a fixed $L_I^{\hat{y}}$, since the contributions from each of these is the same. Such $\hat{\mathbf{p}}$ can be identified with the integers in the quadratic field $Q(i)$ of fixed norm $\|\hat{\mathbf{p}}\|$. The contribution of all $\hat{\mathbf{p}}$ with $\|\hat{\mathbf{p}}\|$ fixed at m corresponding to a fixed $\hat{r}$ is

$$N(m) e^{iy\hat{r}m} J_0[\hat{r}m(y^2-1)^{1/2}] \hat{r}^{is} \frac{d\hat{r}}{\hat{r}}. \tag{10.89}$$

where $N(m)$ is the number of $\hat{\mathbf{p}}$ of norm m.

It is more transparent to interpret the geometry in a somewhat different manner.

With $\hat{y}$ fixed $L^{\hat{y}}_{\hat{X}} \cap$ light cone is a circle. These circles contain lattice points when $\hat{y} = m$ is of the form $p^2 + q^2$. We group these points together because $\hat{\delta}_{L^y_I \cap \Omega}$ is constant on $L^{\hat{y}}_X$. The number of such lattice points is $N(m)$. We now integrate over $\{\sigma(\hat{\mathbf{p}})\}_{p^2+q^2=m}$. Geometrically this can be regarded as "sliding" $L^{\hat{y}}_{\hat{X}}$ perpendicular to the $\hat{X}$ axis and integrating the sums of the contributions from $\{\sigma(\hat{\mathbf{p}})\}$ as we slide.

We now carry out the integration over $\hat{r}$ which takes the form of the Mellin transform. We change variables to replace $\hat{r}m \to \hat{r}$. This introduces a factor m^{-is}. We arrive at the formal expression

$$\begin{aligned}\delta_{L^y_I \cap \Omega} \cdot E(z, is) &= A\zeta_i(is) \int e^{iy\hat{r}} J_0[\hat{r}(y^2-1)^{1/2}]\hat{r}^{is}\,\frac{d\hat{r}}{\hat{r}} \\ &= A\zeta_i(is)\,{}_2F_1\left(\frac{is}{2}, \frac{is+1}{2}; 1; -(y^2-1)(-iy)^{-2}\right). \qquad (10.90)\end{aligned}$$

Here ${}_2F_1$ is the usual hypergeometric function. The evaluation of the integral is in Bateman [11, vol. I, p. 49 (16)]. The values of our parameters do not satisfy Bateman's condition so the evaluation of the Bessel function integral must proceed using an elaborate analytic continuation.

$$\zeta_i(s) = \sum N(m) m^{-s} \qquad (10.91)$$

is the ζ *function* of the Gaussian field.

Remark. *Our change of variables $\hat{r}m \to \hat{r}$ affected the terms $\hat{\rho}_I(\hat{r}\hat{\mathbf{p}})$ and $\hat{y}_I(\hat{r}\hat{\mathbf{p}})$. Proposition 10.1 now plays its role. For points on the light cone $\rho_X(\hat{\mathbf{p}}) = \hat{y}_X(\hat{p})$ for all X. After the change of variables, the integral becomes a product of terms involving $\hat{r}$ (which are products of Bessel functions, exponential functions, and $\hat{r}^{is}$) with terms depending purely on $\hat{\mathbf{p}}$. It is this "separation" into arithmetic and transcendental factors which provides a nice form to our results.*

For the next level of sophistication we compute the periods of E over the same cycle for nontrivial characters $\chi_j(\theta) = \exp(ij\theta)$ of $H_I = K$. Theorem 10.2 tells us that (10.86) must be modified to

$$\widehat{e^{ij\theta}\delta}_{L^y_I \cap \Omega} = ce^{ij\hat{\theta}} J_j[(y^2-1)^{1/2}\hat{\rho}(\hat{x})]e^{iy\hat{y}}. \qquad (10.92)$$

θ and $\hat{\theta}$ are angles measured to the respective angle base points.

We can now repeat the geometric method we used for the computation of (10.88). The contribution from $\hat{\mathbf{p}}$ is modified by the introduction of a factor $\exp(ij\hat{\theta}(\hat{\mathbf{p}}))$. Again we group together those $\hat{\mathbf{p}}$ for which $\|\hat{p}\|$ is constant; that is, those $\hat{\mathbf{p}}$ which lie on $L^{\hat{y}} \cap$ light cone for fixed $\hat{y}$ so that $\hat{\delta}_{L^y_I \cap \Omega}$ is constant. Now

the angle polar coordinates $\hat{\theta}(\hat{\mathbf{p}})$ are significant; the crucial property of $\hat{\theta}(\hat{\mathbf{p}})$ is that it is constant on $\sigma(\hat{\mathbf{p}})$. This allows us to make the computation in the same manner as in the case $j = 0$. The ζ function is to be replaced by the ζ *function with Grossencharakter* [93]

$$\zeta(I, s, j) = \sum \frac{e^{ij\hat{\theta}(\hat{\mathbf{p}})}}{(p^2 + q^2)^s}. \tag{10.93}$$

The hypergeometric function becomes

$${}_2F_1\left[\tfrac{1}{2}(is + j), \tfrac{1}{2}(is + j + 1); j + 1; -(y^2 - 1)(-iy)^{-2}\right]. \tag{10.94}$$

We now replace I by an arbitrary elliptic X; we denote by $Q = Q_X$ the quadratic form defined by X. From the point of view of Minkowski geometry there is no difference between I and X.

By making the Lorentz transformation g which takes $I \to X$ and the line $\{\left(\begin{smallmatrix} x & 0 \\ 0 & 0 \end{smallmatrix}\right)\}$, which is the angle base point line, into itself, we find that the equation for Ω becomes

$$\Omega : y_X^2 - Q(t_X) = 1. \tag{10.95}$$

Here y_X is the usual parameter on the line $\{y_X X\}$ and

$$t_X = g t_I$$

where t_I is the usual parameter on L_I^y.[9] As before this means that $L_X^y \cap \Omega$ is the ellipse

$$L_X^y \cap \Omega = \{y^2 = Q(t_X) - 1\}.$$

In particular $L_X^y \cap \Omega$ is empty if $Q(t_X) < 1$.

The changes made in passing from I to X are simple. We replace $y_I(\hat{r})$ by $y_X(\hat{r})$ and we replace $u^2 + v^2$ by $Q(u, v)$ where $\left(\begin{smallmatrix} u^2 & uv \\ uv & v^2 \end{smallmatrix}\right)$ is a point on the light cone.

The classical definition of the Bessel function gives, as in (10.86),

$$\widehat{\chi(\theta)\delta_{L_{X\cap\Omega}^y}}(\hat{x}) = c\chi(\hat{\theta}) J_j[(y_X^2 - 1)^{1/2} \rho_X(\hat{x})] \exp(iy\hat{y}_X). \tag{10.96}$$

$\chi(\hat{\theta}) = \chi_j(\hat{\theta})$ is of the form $\exp(ij\hat{\theta}_{\hat{X}})$.

By following our computation for the case $X = I$ we deduce

[9] We remind the reader that, in conformity with the notation in the book, the parameter on the slices is t and y is the parameter in the orthogonal line.

Theorem 10.8 *The periods of E over the elliptic cycle $L_X^y \cap \Omega$ are given by*

$$\chi_j \delta_{L_X^y \cap \Omega} \cdot E(z, is) = A\zeta(X, j, is) \times {}_2F_1\left[\frac{1}{2}(is+j), \frac{1}{2}(is+j+1); j+1; -(y_X^2-1)(-iy_X)^{-1}\right]. \tag{10.97}$$

$\zeta(X, j, s)$ is the ζ function with Grossencharakter of the quadratic form Q, namely

$$\zeta(X, j, s) = \sum \frac{\chi_j(\hat{\mathbf{p}})}{Q(\hat{\mathbf{p}})^s}. \tag{10.98}$$

Remark. *Our computation assumes* $\det X = 1$. *Multiplication by* $\|X\|^s$ *allows us to pass to arbitrary elliptic* X.

(b) Hyperbolic cycles

The main difference between the computations for hyperbolic and elliptic cycles is that $L_X^y \cap \Omega$ is noncompact when X is hyperbolic. Let us compute the periods corresponding to those X for which $\Gamma_X = \Gamma_H = \Gamma \cap H$ is infinite. Γ_X is an infinite cyclic group times a finite group W corresponding to roots of unity. We want to use the unfolding lemma 10.7 which entails the replacement of E by E_X in which the sum in (10.65) becomes a sum over the double coset space $\Gamma_\infty \setminus \Gamma / \Gamma_X$.

In order to apply the unfolding lemma we have to show that Γ_X acts with no fixed points on $\Gamma_\infty \setminus \Gamma$. Suppose that there are $\gamma \in \Gamma$ and $\eta \in \Gamma_X$ such that the left Γ_∞ cosets of γ and $\gamma\eta$ are the same, i.e.

$$\lambda\gamma = \gamma\eta$$

for some $\lambda \in \Gamma_\infty$. In particular $\lambda \in N$ and thus the right action of Γ_X on $N \setminus G$ has fixed points.

$N \setminus G$ is the u, v plane minus the origin; G acts by linear transformations. The only elements of G that have fixed points must have one, and hence both, eigenvalues 1, and hence are conjugate to elements of N. In particular no element of Γ_X can have a fixed point.

This proves that we can apply the unfolding lemma.

We could start our discussion of the geometry behind the computation of the periods by examining the hyperbolic matrices $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ or $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Unfortunately these matrices do not fit into our criterion that H_X/Γ_{H_X} be compact since $\Gamma_X = \Gamma \cap H_X = \{I\}$ for each of these matrices X. For this reason such periods have to be treated in a different way which we postpone to later.

We fix X with $\det X = -1$ and Γ_X infinite. $L_X^{\hat{y}} \cap$ light cone is now a hyperbola when $\hat{y} \neq 0$. As in the geometric interpretation of the elliptic case we choose $\hat{y} = m$ such that this intersection contains lattice points $\{\hat{\mathbf{p}}\}$. For any such $\hat{\mathbf{p}}$ the hyperbola contains $\{\gamma\hat{\mathbf{p}}\}_{\gamma \in \Gamma_X}$ which is an infinite set. Our definition of the

Eisenstein series is a sum over $\Gamma_\infty \backslash \Gamma$, but when we unfold the integral defining the period from the (compact) fundamental domain $\mathcal{D}_X$ for the action of Γ_X on $L_X^y \cap \Omega$ we replace the sum over $\Gamma_\infty \backslash \Gamma$ by the sum over $\Gamma_\infty \backslash \Gamma / \Gamma_X$. This is equivalent to choosing only one $\hat{\mathbf{p}}$ from each orbit of Γ_X because $\hat{\mathbf{p}}$ can be identified with $\gamma \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ for some $\gamma \in \Gamma_\infty \backslash \Gamma$. (Actually it is $\gamma \in \Gamma / \Gamma_\infty$ and the double coset space is $\Gamma_X \backslash \Gamma / \Gamma_\infty$ because we are in Fourier transform space. This is only a formal difference and is insignificant.)

Choosing one $\hat{\mathbf{p}}$ from each Γ_X orbit means that we choose $\hat{\mathbf{p}}$ in some fixed fundamental domain $\hat{\mathcal{D}}_X^y$ for Γ_X on each $L_X^{\hat{y}} \cap$ light cone.

We form the period corresponding to $\chi =$ trivial character. We call $N(m) = N_X(m) =$ number of orbits of Γ_X on $L_{\hat{X}}^m \cap$ (lattice points on light cone). Since $\hat{D}_X^y$ is compact $N_X(m)$ is finite.

If χ is a general character of H_X which is trivial on Γ_X then $\chi = \chi_j$ is of the form

$$\chi_j(h) = e^{ij\theta(h)/\theta_X^0}. \tag{10.99}$$

$\theta(h) = \theta_X(h)$ is the parameter of h which is the log of one (suitably chosen) of its eigenvalues. (In (10.33H) we called this parameter $\zeta = \zeta_X$ but we use θ here to avoid confusion with the ζ function.) We denote by $\theta^0 = \theta_X^0$ the parameter value of the chosen generator of the unit group of X (modulo roots of unity).

It remains to form the integral corresponding to (10.88). The J Bessel function is the Fourier transform of the characteristic function of a circle; the Fourier transform of the characteristic function of a (half-)hyperbola is a K Bessel function [11, vol. II]. The analog of (10.92) is

$$\widehat{e^{ij\theta_X/\theta_X^0}\delta_{L_X^y \cap \Omega}}(\hat{x}) = ce^{ij\hat{\theta}/\theta_X^0} K_{ij/\theta_X^0}[(y^2+1)^{1/2}\rho_X(\hat{x})]e^{iy\hat{y}_X}. \tag{10.100}$$

(The change from $y^2 - 1$ of (10.96) to $y^2 + 1$ in (10.100) results from the change of X from elliptic to hyperbolic.)

There is no difficulty in following the ideas of the elliptic case to deduce (see [11, vol. II, p. 50 (26)] for the computation of the relevant integral)

Theorem 10.9 *The periods of E over the hyperbolic $L_X^y \cap \Omega$ are given by*

$$\chi_j \delta_{L_{D_X}^y \cap \Omega} \cdot E(z, is) = A\zeta(X, j, is) \times {}_2F_1\left[is + ij/\theta_X^0, \frac{1}{2} + ij/\theta_X^0; \frac{1}{2} + is; \frac{-iy - (y^2+1)^{1/2}}{-iy + (y^2+1)^{1/2}}\right]. \tag{10.101}$$

$\zeta(X, j, s)$ is the ζ function with Grossencharakter

$$\zeta(X, j, s) = \sum \frac{e^{ij\hat{\theta}_X(\hat{p})/\theta_X^0}}{Q(\hat{p})^s} \tag{10.102}$$

the sum being taken over lattice points $\hat{p} \in$ light cone modulo Γ_X.

Remark 1 *For $y = 0$ this result is due to Hecke and Siegel* [142, p. 107ff.].

Remark 2 *There is an interesting structure to the hyperbolic periods (10.101) and the elliptic periods (10.97) and parabolic periods (10.113). They are products of arithmetically defined functions such as ζ functions or L^N which are sums determined by the lattice points and transcendental functions (Bessel or hypergeometric) which depend only on $j, y,$ and s. Our geometric method clarifies this structure. (See the end of this section for further explanations.)*

Remark 3 *The appearance of the ζ function has been explained above; it arises from the sum of the lattice points on the sets where $L_X^y \cap$ light cone is constant. Other arithmetic functions that appear in the (more complicated) examples studied below arise from similar considerations. On the other hand the meaning of the transcendental function ${}_2F_1$ and its analogs which appear below is best understood by means of differential equations.*

(c) Parabolic cycles

We return to the notation of (10.38)ff. By (10.11) $L_X^y \cap \Omega$ can be interpreted as the horocycle (orbit of N) $\{\eta = y^{-1}\}$ in the Poincaré half-plane $\{z = \xi + i\eta, \eta > 0\}$. Since the Eisenstein series is Γ invariant it is $\Gamma_N = \{\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)\}_{n\in\mathbb{Z}}$ invariant so we want to form its periods over $(L_X^y \cap \Omega)/\Gamma_N$.

As before we want to verify the applicability of the unfolding lemma. This means checking whether Γ_∞ acts (on the right) without fixed points on $\Gamma_\infty \setminus \Gamma$.

In fact there is a fixed point on $\Gamma_\infty\backslash\Gamma$, namely the left Γ_∞ coset corresponding to the identity. This is fixed by all of Γ_∞. (Actually we shall examine the left action of Γ_∞ on Γ/Γ_∞.)

Γ_∞ is the subgroup of Γ that leaves $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ fixed. Thus Γ/Γ_∞ can be interpreted as the image of $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ under Γ. For

$$\gamma = \begin{pmatrix} p & a \\ q & b \end{pmatrix}$$

we have

$$\gamma \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \gamma^t = \begin{pmatrix} p^2 & pq \\ pq & q^2 \end{pmatrix} \tag{10.103}$$

so this set of matrices, which is (essentially) the set of primitive lattice points on the light cone, represents the right Γ_∞ cosets of Γ. It is the quadratic transform of the set of primitive lattice points in the u, v plane.

We now apply $\left(\begin{smallmatrix}1 & m\\ 0 & 1\end{smallmatrix}\right) \in \Gamma_\infty$ to obtain

$$\begin{pmatrix}1 & m\\ 0 & 1\end{pmatrix}\gamma\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\gamma^t\begin{pmatrix}1 & 0\\ m & 1\end{pmatrix} = \begin{pmatrix}(p+mq)^2 & q(p+mq)\\ q(p+mq) & q^2\end{pmatrix}$$
$$= \begin{pmatrix}p+mq & a\\ q & b\end{pmatrix}\cdot\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}. \tag{10.104}$$

This can equal $\gamma\left(\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right)\gamma^t$ only if either

(1) $m = 0$,
(2) $q = 0$.

In case (1) $\left(\begin{smallmatrix}1 & m\\ 0 & 1\end{smallmatrix}\right) = I$ while in case (2) $\gamma = \left(\begin{smallmatrix}p & r\\ 0 & s\end{smallmatrix}\right)$ so $\gamma \in \pm\Gamma_\infty$, which means we are in the right Γ_∞ coset of I. (As we have mentioned the $\pm$ is unimportant.)

By (10.104) a double coset is determined by fixing $p+mq$. There is a unique m such that $0 \le p+mq < |q|$. We use this m in (10.104) as the unique representative of the double Γ_∞ coset when $q \neq 0$. These representations take the form

$$\hat{\mathbf{p}} = \begin{pmatrix}p^2 & pq\\ pq & q^2\end{pmatrix}_{\substack{0\le p<|q|\\ q\neq 0\\ (p,q)=1}}. \tag{10.105}$$

We denote by $E_N(z,s)$ the sum of those terms defining the Eisenstein series for which $q \neq 0$. In accordance with Lemma 4.25 we can unfold the corresponding double cosets; this allows us to integrate the "reduced sum" $E_N^0(z,s)$, meaning the sum over such double cosets, over all of $L_X^y \cap \Omega$. The terms with $q = 0$ form a single sum over Γ_∞ so they can be unfolded to a single integral over $L_X^y \cap \Omega$.

For points $\hat{\mathbf{p}}$ as in (10.105)

$$\sigma(\hat{\mathbf{p}}) = \left\{\begin{pmatrix}\hat{r}p^2 & \hat{r}pq\\ \hat{r}pq & \hat{r}q^2\end{pmatrix}\right\}. \tag{10.106}$$

In accordance with Parseval's theorem

$$\int_{\Gamma_\infty\backslash L_X^y\cap\Omega} E_N(z,s) = \delta_{L_X^y\cap\Omega}\cdot E_N^0(z,s)$$
$$= \hat{\delta}_{L_X^y\cap\Omega}\cdot\sum \delta_{\sigma(p)}\hat{r}^{is}. \tag{10.107}$$

We call $\hat{\mathbf{x}} = \left(\begin{smallmatrix}\hat{y} & \hat{w}\\ \hat{w} & \hat{x}\end{smallmatrix}\right)$. For the angle base point on $L_X^y \cap \Omega$ we have $x = y^{-1}$ in (10.39). This gives

$$\hat{\delta}_{L_X^y\cap\Omega}(\hat{\mathbf{x}}) = \int e^{i/2(y\hat{y}+\hat{x}/y+n^2\hat{x}y-2ny\hat{w})}\,dn. \tag{10.108}$$

We can evaluate the integral formally by completing the square; this can be justified in various ways by analytic continuation.

We replace n by $n + n_0$ and choose n_0 to remove the linear term in n. We find $n_0 = \hat{w}\hat{x}^{-1}$. The n integral becomes a Gaussian integral (formally) which equals $c(\hat{x}y)^{-1/2}$. This leads to (modulo constants)

$$\begin{aligned}\hat{\delta}_{L_X^y \cap \Omega}(\hat{\mathbf{x}}) &= \frac{1}{\sqrt{y\hat{x}/2}} e^{i/2(y\hat{y} + y^{-1}\hat{x} - y\hat{w}^2\hat{x}^{-1})} \\ &= \frac{e^{i/2y^{-1}\hat{x}}}{\sqrt{y\hat{x}}} e^{i/2y\hat{x}^{-1}(\hat{x}\hat{y} - \hat{w}^2)} \\ &= \frac{e^{i/2y^{-1}\hat{x}}}{\sqrt{y\hat{x}}} e^{i/2y\hat{x}^{-1}\|\hat{\mathbf{x}}\|}. \end{aligned} \tag{10.109}$$

Note that N acting on $\hat{\mathbf{x}}$ preserves $\|\hat{\mathbf{x}}\|$ and $\hat{x}$. Thus this formula is of the same nature as the corresponding formulas in the elliptic and hyperbolic cases, namely it is a function of the invariants $\|\hat{\mathbf{x}}\|$ and $\hat{x}$.

We can combine (10.109) with (10.106). We use the fact that $\|\hat{\mathbf{x}}\|$ vanishes on the light cone and our usual method of grouping terms in $L_{\hat{X}}^{\hat{x}}$, namely groups are defined by $\{\hat{x} = \text{const.}\}$. We set $\hat{x} = q^2\hat{r}$ in accordance with (10.106) to obtain

$$\begin{aligned}\delta_{L_X^y \cap \Omega} \cdot E_N^0(s) &= \sum_{p,q} \int \frac{e^{1/2q^2\hat{r}/y}}{q\sqrt{y\hat{r}}} \hat{r}^s \frac{d\hat{r}}{\hat{r}} \\ &= y^{s-1} \sum \varphi(q) q^{-2s} \int e^{i/2\hat{r}} \hat{r}^{s-1/2} \frac{d\hat{r}}{\hat{r}} \\ &= B y^{s-1} \Gamma(s - \frac{1}{2}) \frac{\zeta(2s-1)}{\zeta(2s)} \end{aligned} \tag{10.110}$$

[150, p. 6]. The factor $\varphi(q)$ (Euler's φ function) arises from (10.105). (B is an elementary factor; the meaning of B can change in various formulas.)

We are not finished because we have ignored the term coming from $q = 0$. From the series representation (10.69) or else by the same computation as above this is $A\eta^s$.

Recall that $y = \eta^{-1}$. A careful check of the elementary factors A in (10.110) and (10.69) shows

$$\delta_{\Gamma_\infty \backslash \Omega \cap L^y} \cdot E = \eta^s + \frac{\Gamma(s - \frac{1}{2})}{\Gamma(s)} \frac{\zeta(2s-1)}{\zeta(2s)} \eta^{1-s} \tag{10.111}$$

which is the classical formula (see e.g. [70]).

For the other Fourier series coefficients of E we attenuate the measure dn by the character $\exp(2\pi ijn)$ where j is an integer. This has the effect of changing

the coefficient of n in (10.108) from $-2y\hat{w}$ to $-2y\hat{w}+4\pi j$, which is the same as changing $\hat{w}$ to $\hat{w}-2\pi jy^{-1}$.

We now replace $\hat{w}$ in (10.109) by $\hat{w}-2\pi jy^{-1}$; this produces an additional term

$$-\frac{i}{2}y\hat{x}^{-1}(-4\pi j\hat{w}y^{-1}+4\pi^2j^2y^{-2})=2\pi ij\hat{w}/\hat{x}-2\pi^2j^2i\eta\hat{x}^{-1}$$

in the exponential.

We have to integrate this against $\delta_{\sigma(\hat{\mathbf{p}})}$. As in the case $j=0$ the term $\|\hat{\mathbf{x}}\|$ disappears. The angle (10.32)

$$\hat{n}=\frac{\hat{w}}{\hat{x}}=\frac{\hat{r}pq}{\hat{r}q^2}=\frac{p}{q}$$

is constant along $\sigma(\hat{\mathbf{p}})$. As in (10.110) the integral is

$$e^{2\pi ijn}\delta_{L_X^y\cap\Omega}\cdot E_N^0(s)=\sqrt{\eta}\sum\frac{1}{q}e^{2\pi ijp/q}\int e^{i/2[q^2\hat{r}\eta-4\pi^2j^2\eta/\hat{r}q^2]}\hat{r}^{s-1/2}\frac{d\hat{r}}{\hat{r}}. \tag{10.112}$$

To evaluate the integral we write

$$q^2\hat{r}\eta-\frac{4\pi^2j^2\eta}{\hat{r}q^2}=2\pi j\eta\left(\frac{q^2\hat{r}}{2\pi j}-\frac{2\pi j}{\hat{r}q^2}\right).$$

We change variables in (10.112)

$$\frac{q^2\hat{r}}{2\pi j}\to\hat{r}$$

to obtain

$$\begin{aligned}e^{2\pi ijn}\delta_{L_X^y\cap\Omega}\cdot E_N^0&=B(2\pi j)^{s-1/2}\sum q^{-2s}e^{2\pi ijp/q}\int e^{2\pi j\eta(\hat{r}-\hat{r}^{-1})}\hat{r}^{s-1/2}\frac{d\hat{r}}{\hat{r}}.\\&=B(2\pi j)^{s-1/2}K_{s-\frac{1}{2}}(2\pi j\eta)L^N(s,j).\end{aligned} \tag{10.113}$$

K is the Bessel function and L^N is the L function:

$$L^N(s,j)=\sum_{q>0}q^{-2s}\sum_{\substack{0\le p<q\\(p,q)=1}}e^{2\pi ijp/q}. \tag{10.114}$$

We have ignored the term corresponding to $q = 0$. But this is a multiple of η^s which is N invariant and cannot contribute to the Fourier series coefficients of E for $j \neq 0$.

The inner sum was studied by Ramanujan [150, p. 10].

Analogous formulas hold for those conjugates of N whose intersection with Γ is infinite.

We have examined the periods of the Eisenstein series E over elliptic, hyperbolic, and parabolic cycles which are compact. We now examine noncompact cycles. These are orbits of isotropy groups H_X for which $\Gamma_X = \Gamma \cap H_X$ is finite.

Let $X = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ so H_X is the hyperbolic group $A = \left\{\left(\begin{smallmatrix} a & 0 \\ 0 & a^{-1} \end{smallmatrix}\right)\right\}$. The orthogonal complement of X is $L_X^0 = \left\{\left(\begin{smallmatrix} x & 0 \\ 0 & z \end{smallmatrix}\right)\right\}$ so

$$L_X^y = \left\{ W = \begin{pmatrix} x & y \\ y & z \end{pmatrix} \right\}. \tag{10.115}$$

The distance base point on L_X^y is the A fixed point yX. For the angle base ray we can choose $\left\{\left(\begin{smallmatrix} t & 0 \\ 0 & t \end{smallmatrix}\right)\right\}_{t \geq 0}$ on L_X^0 which becomes $\left\{\left(\begin{smallmatrix} t & y \\ y & t \end{smallmatrix}\right)\right\}$ on L_X^y. Since

$$\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} t & y \\ y & t \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} = \begin{pmatrix} a^2 t & y \\ y & a^{-2} t \end{pmatrix} \tag{10.116}$$

the A orbit of $\left(\begin{smallmatrix} t & y \\ y & t \end{smallmatrix}\right)$ consists of all $W = \left(\begin{smallmatrix} x & y \\ y & z \end{smallmatrix}\right) \in L_X^y$ with $xz > 0$. (We need the angle base ray $\left\{\left(\begin{smallmatrix} t & y \\ y & -t \end{smallmatrix}\right)\right\}_{t \geq 0}$ for the points with $xz < 0$; as this is treated in a similar manner we shall ignore it.)

Comparing (10.115) and (10.116) shows that

$$\begin{aligned} t^2 &= xz \\ a^2 &= (x/z)^{1/2} \end{aligned} \tag{10.117}$$

are the (squares of the) cylindrical coordinates of W. Since $\det W = 1$ we have $t^2 = y^2 + 1$.

Note that the angle base point on $L_X^0 \cap \Omega$ is I which corresponds to the point i on the Poincaré half-plane so that $L_X^0 \cap \Omega$ corresponds to the i orbit of A which is the imaginary axis.

To compute the periods of E over the cycle $L_X^y \cap \Omega$ we have to compute the Fourier transform of $\delta_{L_X^y \cap \Omega}$ on Γ. For points $\hat{W} = \left(\begin{smallmatrix} u^2 & uv \\ uv & v^2 \end{smallmatrix}\right)$ on Γ

we have

$$\begin{aligned}\hat{\delta}_{L_X^y\cap\Omega}(W) &= \int e^{\frac{i}{2}[t(a^2v^2+a^{-2}u^2)-2uvy]}\,\frac{da}{a}\\ &= e^{-iuvy}\int e^{\frac{it}{2}(a^2v^2+a^{-2}u^2)}\,\frac{da}{a}\\ &= e^{-iy\hat{y}}\int e^{\frac{it\hat{y}}{2}(a^2vu^{-1}+a^{-2}uv^{-1})}\,\frac{da}{a}\\ &= e^{-iy\hat{y}}\int e^{\frac{it\hat{y}}{2}(a^2\hat{a}^{-2}+a^{-2}\hat{a}^2)}\,\frac{da}{a}\\ &= \frac{1}{2}e^{-iy\hat{y}}\int e^{\frac{it\hat{y}}{2}(\alpha+\alpha^{-1})}\,\frac{d\alpha}{\alpha}\\ &= Be^{-iy\hat{y}}K_0(it\hat{y}/2)\end{aligned}\tag{10.118}$$

[11, Vol. II, p. 82]. (B is a simple factor.)

For the other periods we use the measure $a^{i\lambda}da/a$ on the cycle. This changes the result to

$$\widehat{a^{i\lambda}\delta_{L_X^y\cap\Omega}}(W) = Be^{-iy\hat{y}}\hat{a}^{i\lambda}K_{-i\lambda}(it\hat{y}/2).\tag{10.119}$$

To compute the periods of E we have to integrate this over $\hat{W}\in\sigma(\hat{\mathbf{p}})$ for $\hat{\mathbf{p}} = \begin{pmatrix} p^2 & pq \\ pq & q^2\end{pmatrix}$ a primitive lattice point on $\hat{\Gamma}$ and then sum over $\{\hat{\mathbf{p}}\}$.

From (10.116) we find that the cylindrical coordinates of $\hat{r}\hat{\mathbf{p}}$ are

$$\begin{aligned}\hat{t} &= \hat{r}|pq|\\ \hat{y} &= \hat{r}pq\\ \hat{a}^2 &= |p/q|.\end{aligned}\tag{10.120}$$

The contribution to the periods from each fixed $\hat{\mathbf{p}}$ (for $pq>0$ with an analogous formula for $pq<0$) is

$$\begin{aligned}&B\int K_{-i\lambda}(itpq\hat{r}/2)e^{-iy\hat{r}pq}|p/q|^{i\lambda/2}\hat{r}^{is}\,\frac{d\hat{r}}{\hat{r}}\\ &= B(p/q)^{i\lambda/4}(tqp)^{-is}\int K_{-i\lambda}(i\hat{r})e^{-iy\hat{r}/t}\hat{r}^{is}\,\frac{d\hat{r}}{\hat{r}}\\ &= Bp^{i\lambda/2-is}q^{-i\lambda/2-is}t^{-is}{}_2F_1\left[is-i\lambda,-i\lambda+1/2;is+1/2;\frac{y-t}{y+t}\right]\end{aligned}\tag{10.121}$$

[11, Vol. II, p. 50].

It is difficult to sum this over all p,q with $(p,q)=1$. However, we have noted that the difference between using all $(p,q)\neq(0,0)$ and relatively prime p,q only changes $E(z,s)$ by a factor $\zeta(s)$. We drop the requirement $(p,q)=1$ and absorb this $\zeta(s)$ factor in B. We write $\mathbf{E}$ for the modified E.

We have shown that, properly interpreted, the period of **E** is

$$B\zeta\left(\frac{-i\lambda}{4}+is\right)\zeta\left(\frac{i\lambda}{2}+is\right)t^{-is}{}_2F_1\left[is-i\lambda,-i\lambda+1/2;is+1/2;\frac{y-t}{y+t}\right]. \tag{10.122}$$

Problem 10.1 *What are the periods over other cycles for which Γ_X is finite?*

We have found the Fourier series of $E(z,s)$ on various hyperbolic, elliptic, and parabolic cycles. For any such cycle S the Fourier series determines $E(z,s)$ on S and hence on S^H which is the orbit containing S of the one-parameter group H.

We want to go from S or S^H to all of $\Omega = G/K$ and finally to all of $\mathbb{R}^3$. By our construction $E(z,s)$ is the restriction of a homogeneous solution of the wave equation to Ω. Since $\Box$ separates in hyperbolic and spherical coordinates $E(z,s)$ is an eigenfunction of the noneuclidean Laplacian Δ on Ω (see Chapter 7 for details).

In the usual $z=\xi+i\eta$ coordinates on Ω

$$\Delta = \eta^2\left(\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\eta^2}\right). \tag{10.123}$$

Let f be a solution of $\Delta f = s(s-1)f$ of the form $f(z)=g(\xi)h(\eta)$ then

$$\begin{aligned}\frac{\partial^2}{\partial\xi^2}g &= -k^2 g\\ \left(\frac{\partial^2}{\partial\eta^2}+s(1-s)\eta^{-2}\right)h &= k^2 h.\end{aligned} \tag{10.124}$$

These equations show that for $s\neq 0,1$

$$\begin{aligned}g(\xi)=\phi(k,\xi) &= c^+e^{ik\xi}+c^-e^{-ik\xi}\\ h(\eta)=\psi(k,\eta) &= c\eta B_{s-1/2}(k\eta)\end{aligned} \tag{10.125}$$

where B denotes a generic Bessel function [11, vol. II, p. 2(15)].

For $s=0,1$, $h(\eta)=a^+e^{k\eta}+a^-e^{-k\eta}$.

If we are given any function $g(\xi)$ thought of, for example, as a function on $\eta=1$, we can expand it in a Fourier integral and then multiply the factor $\exp(ik\xi)$ in the integrand by $\psi(k,\eta)$ which is normalized to be 1 on $\eta=1$ to obtain an eigenfunction of Δ on Ω with eigenvalue $s(1-s)$ which equals g on $\eta=1$. Then we multiply by r^s where r is the Minkowski distance in $\mathbb{R}^3$ to obtain a solution of the wave equation in the interior of the light cone.

Remark. *There is a certain nonuniqueness in the procedure we have given. For Δ and $\Box$, hence the operator in (10.124), are second-order operators. Thus we have to provide 2 CD on a set $\eta=\eta^0$ in order to obtain a unique eigenfunction*

of Δ with fixed eigenvalue on Ω and we also need 2 CD on Ω to obtain a unique solution of $\Box$ in the forward light cone. This introduces a small complication in our discussion which we shall ignore as it is easily treated.

The Fourier series coefficients of E on suitable geodesics were shown to be simply expressible in terms of ζ functions with Grossencharakter on real quadratic fields. If we have two such geodesics $\mathbf{g}_1$ and $\mathbf{g}_2$ we can start with CD on $\mathbf{g}_1$, express E on Ω in terms of this CD, i.e. in terms of the ζ functions corresponding to the associated quadratic field, and then use this expression to evaluate the Fourier series coefficients on $\mathbf{g}_2$. In this way we have related the ζ functions of the quadratic fields corresponding to $\mathbf{g}_1$ and $\mathbf{g}_2$.

Problem 10.2 *Find precise relations amongst the ζ functions of various quadratic fields. In particular can one use this procedure to find finite relations?*

Remark. *We have dealt with ζ functions with Grossencharakter. The Grossencharakter defining ζ functions can be interpreted as characters involving the infinite part of the adelized quadratic number field. There are also characters involving finite primes—these lead to L functions. We hope to "adelize" our theory and hence give a new setting for L functions.*

Abstract formulation of computation of periods

We wish to put this method of computation in an abstract form.

In our above examples we started with $\hat{\Theta}$ which is (essentially) the sum of the δ functions of the lattice points on the light cone. We then decomposed $\hat{\Theta}$ under the Mellin transform and formed the resulting Fourier transform $\Theta_s = E(z,s)r^s$. Finally we evaluated the "inner product" of $E(z,s)$ with $\delta_{L_X^y \cap \Omega}$ via Parseval's theorem.

Our first observation is that the Γ invariance of $\hat{\Theta}$ and of the Lorentz inner product is used only to establish the Γ invariance of E. However, the H_X invariance of the inner product is more basic to our computation. For it implies that $\hat{\delta}_{L_X^y \cap \Omega}$ is H_X invariant for each y since $\delta_{L_X^y \cap \Omega}$ is H_X invariant. We also use the orthogonal splitting of the inner product into y and L_X^0 components.

The splitting means that $\hat{\delta}_{L_X^y \cap \Omega}$ is of the form $U_y(\mathbf{x}) \exp(iy \cdot \hat{y})$ where $\mathbf{x}$ is the parameter on $\hat{L}_X^0$. The H_X invariance of $\delta_{L_X^y \cap \Omega}$ means that U_y is H_X invariant.

For a general set-up we replace the light cone by an algebraic variety V which we assume contains "many" lattice points $\mathbf{p}$.[10] The ambient space is of the form $\mathbb{R}^N = L_X^0 \oplus Y$; we write coordinates in the form $(\mathbf{x}, \mathbf{y})$. There is a linear group H_X acting on L_X^0, hence acting on $\mathbb{R}^N$, preserving the $\mathbf{y}$ coordinate. Moreover there is a nondegenerate quadratic form on $\mathbb{R}^N$ which is H_X invariant; we use this quadratic form to define Fourier transform.

[10] Although in the case of the Eisenstein series "lattice point" is to be taken literally, in the case of Poincaré series (see below) the "lattice points" form lines so are not literal "lattice points" (which is insignificant for this heuristic treatment).

We regard V as being in the Fourier transform space. There is a group $\hat{\mathbf{R}}$ of linear transformations of $\hat{\mathbb{R}}^N$ which acts on V and also preserves the $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ splitting of $\hat{\mathbb{R}}^N$, meaning $\hat{\mathbf{r}} \in \hat{\mathbf{R}}$ defines maps $\hat{\mathbf{r}}_{\hat{\mathbf{x}}} : \hat{L}_X^0 \to \hat{L}_X^0$ and $\hat{\mathbf{r}}_{\hat{\mathbf{y}}} : \hat{Y} \to \hat{Y}$. We write

$$\hat{\mathbf{r}}_{\hat{\mathbf{x}}}(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = \hat{\mathbf{r}}_{\hat{\mathbf{x}}}(\hat{\mathbf{x}})$$

and

$$\hat{\mathbf{r}}_{\hat{\mathbf{y}}}(\hat{\mathbf{x}}, \hat{\mathbf{y}}) = \hat{\mathbf{r}}_{\hat{\mathbf{y}}}(\hat{\mathbf{y}}).$$

We define the general (Fourier transform of) Eisenstein series

$$\hat{E}_s = \sum s\delta_{\sigma(\hat{\mathbf{p}})}.$$

As before $\sigma(\hat{\mathbf{p}})$ is the $\hat{\mathbf{R}}$ orbit of $\hat{\mathbf{p}}$ and $\delta_{\sigma(\hat{\mathbf{p}})}$ is the measure $d\hat{\mathbf{r}}$ (Haar measure) on this orbit. s is a one-dimensional representation of $\hat{\mathbf{R}}$. The sum is over the lattice points on V. (We leave the study of higher dimensional representations to the reader.)

Let Ω be an algebraic variety in $\mathbb{R}^N$ which is H_X invariant. For any $\mathbf{y} \in Y$ the measure $\delta_{L_X^{\mathbf{y}} \cap \Omega}$ is H_X invariant. Hence its Fourier transform takes the form

$$\hat{\delta}_{L_X^{\mathbf{y}} \cap \Omega} = e^{i\mathbf{y}\cdot\hat{\mathbf{y}}} U_{\mathbf{y}}(\hat{\mathbf{x}})$$

where U is H_X invariant.

The simplest period of E_s is, by Parseval's formula,

$$\delta_{L_X^{\mathbf{y}} \cap \Omega} \cdot E_s = \sum \int e^{i\mathbf{y}\cdot\hat{\mathbf{y}}(\hat{\mathbf{r}}\hat{\mathbf{p}})} U_{\mathbf{y}}(\hat{\mathbf{r}}\hat{\mathbf{p}}) s(\hat{\mathbf{r}})\, d\hat{\mathbf{r}}.$$

In order to make sense of this formula we need

Assumption. *There are a finite number of H_X orbits $\mathcal{O}_1, \ldots, \mathcal{O}_\beta$ on $\hat{L}_X^0$ such that for any $\hat{\mathbf{x}}$ lying in the projection of V on $\hat{L}_X^0$ there is an $\hat{\mathbf{r}}_{\hat{\mathbf{x}}}$ and a point $\hat{\mathbf{1}}_l \in \mathcal{O}_l$ with*

$$\hat{\mathbf{x}} = \hat{\mathbf{r}}_{\hat{\mathbf{x}}} \hat{\mathbf{1}}_l.$$

l is uniquely determined by $\hat{\mathbf{x}}$.

Otherwise put:

$$\hat{\mathbf{R}}_{\hat{\mathbf{x}}} \left(\bigcup \mathcal{O}_l \right) \supset \text{projection } V \text{ on } \hat{L}_X^0.$$

We now make the further assumption that $\mathbf{y}$ satisfies

$$\mathbf{y} \cdot \hat{\mathbf{y}} \text{ is a function of } \hat{\mathbf{x}} \text{ on } V,$$

e.g. $\mathbf{y} = 0$.

With these assumptions we can evaluate the period. To accomplish this we can write, since $\hat{\mathbf{y}}$ is determined by $\hat{\mathbf{x}}$,

$$\hat{\mathbf{p}} = \hat{\mathbf{r}}_{\hat{\mathbf{p}}} \hat{\mathbf{1}}_{l(\hat{\mathbf{p}})}$$

for some $\hat{\mathbf{r}}_{\hat{\mathbf{p}}} \in \hat{\mathbf{R}}$. The above formula for the period becomes

$$\delta_{L_X^{\mathbf{y}} \cap \Omega} \cdot E_s = \sum \int e^{i(\mathbf{y}\cdot\hat{\mathbf{y}})[\hat{\mathbf{r}}\hat{\mathbf{r}}_{\hat{\mathbf{p}}}\hat{\mathbf{1}}_{l(\hat{\mathbf{p}})}]} U_{\mathbf{y}}(\hat{\mathbf{r}}\hat{\mathbf{r}}_{\hat{\mathbf{p}}}\hat{\mathbf{1}}_{l(\hat{\mathbf{p}})}) s(\hat{\mathbf{r}})\, d\hat{\mathbf{r}}.$$

We make the change of variables

$$\hat{\mathbf{r}}\hat{\mathbf{r}}_{\hat{\mathbf{p}}} \to \hat{\mathbf{r}}.$$

Since s is a one-dimensional character of $\hat{\mathbf{R}}$

$$s(\hat{\mathbf{r}}) \to s(\hat{\mathbf{r}}) s(\hat{\mathbf{r}}_{\hat{\mathbf{p}}}^{-1}).$$

We arrive at

$$\delta_{L_X^{\mathbf{y}} \cap \Omega} \cdot E_s = \sum_{\hat{\mathbf{p}}} s(\hat{\mathbf{r}}_{\hat{\mathbf{p}}}^{-1}) \int e^{i(\mathbf{y}\cdot\hat{\mathbf{y}})(\hat{\mathbf{r}}\hat{\mathbf{1}}_{l(\hat{\mathbf{p}})})} U_{\mathbf{y}}(\hat{\mathbf{r}}\hat{\mathbf{1}}_{l(\hat{\mathbf{p}})}) s(\hat{\mathbf{r}})\, d\hat{\mathbf{r}}.$$

Of course we have to sum over $\hat{\mathbf{p}}$ in a suitable "double coset space." It seems that, in order to make sense of this, we need to know that there is a discrete group Γ of linear transformations of $\mathbb{R}^N$ acting "close to" transitively on the lattice points of V and commuting with the Fourier transform.

We have given a general computation of the period corresponding to the trivial character of H_X. For the other periods we need to know that the actions of $\hat{\mathbf{R}}$ and H_X on V commute. This allows us to define an angle which is $\hat{\mathbf{R}}$ invariant.

Since H_X commutes with the Fourier transform our ideas on angles allow us to write

$$\widehat{\chi\delta_{L_X^{\mathbf{y}} \cap \Omega}} = e^{i\mathbf{y}\cdot\hat{\mathbf{y}}} \bar{\chi}(\hat{\theta}(\hat{\mathbf{x}}, \hat{\mathbf{y}})) U_{\mathbf{y},\chi}(\hat{\mathbf{x}})$$

where $U_{\mathbf{y},\chi}$ is H_X invariant. Here χ is the character of a one-dimensional representation of H_X.

A straightforward extension of the computation for $\chi \equiv 1$ yields

$$\chi\delta_{L_X^{\mathbf{y}} \cap \Omega} \cdot E_s = \sum \bar{\chi}(\hat{\theta}(\hat{\mathbf{p}})) s(\hat{\mathbf{r}}_{\hat{\mathbf{p}}}^{-1}) \int e^{i(\mathbf{y}\cdot\hat{\mathbf{y}})(\hat{\mathbf{r}}\hat{\mathbf{1}}_{l(\hat{\mathbf{p}})})} U_{\mathbf{y},\chi}(\hat{\mathbf{r}}\hat{\mathbf{1}}_{l(\hat{\mathbf{p}})}) s(\hat{\mathbf{r}})\, d\hat{\mathbf{r}}. \quad (10.126)$$

Our derivation shows the types of conditions which are needed to compute the periods.

10.4 Poincaré series and their periods

Our construction of E began with the point $\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$ which is N invariant. We then formed the sum of the δ functions of $\{\gamma\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)\}$ for $\gamma \in \Gamma_\infty \backslash \Gamma$, then took Fourier and Mellin transforms. The δ function of $\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$ becomes, essentially, the function η^s on the $z = \xi + i\eta$ upper half-plane. η^s is N invariant.

The convergence of the series for E is somewhat dependent on the fact that the function $\Im z$ is N invariant so we sum only over $\Gamma_\infty \backslash \Gamma$ rather than all of Γ. It suffices to know that $\Im z$ is Γ_∞ invariant. This suggests that we look for other interesting Γ_∞ invariant functions and sum them over $\Gamma_\infty \backslash \Gamma$.

A natural example is $\exp(2\pi i z)$. The sum of $\gamma \exp(2\pi i z)$ over $\gamma \in \Gamma_\infty \backslash \Gamma$ is the seminal example of a *Poincaré series*. Instead of $\exp(2\pi i z)$ we could start with $\exp(2\pi i \nu z)$ where ν is rational. This function is fixed by the infinite subgroup

$$\Gamma_\infty^l = \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \mid n \equiv 0 \bmod l \right\}.$$

l is the denominator of ν.

We want to express the Poincaré series using our three-dimensional formalism. Note that $\exp(i\nu z)$ is transformed under $N = \Gamma_\infty^R$ (the real one-parameter group containing Γ_∞) by multiplication by the character $\exp(i\nu n)$. A reasonable guess for the replacement of $\delta_{\left(\begin{smallmatrix}1&n\\0&0\end{smallmatrix}\right)}$ is the function $\exp(i\nu n)$ on an orbit $\{\left(\begin{smallmatrix}1&n\\0&1\end{smallmatrix}\right)\hat{\mathbf{x}}\}$ of a suitable point $\hat{\mathbf{x}}$ in the light cone.

From the point of view of linear action of G on the (u, v) plane we could choose $\hat{\mathbf{x}}$ corresponding, for example, to $x^0 = \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$. Then $N \cdot x^0$ is the line $v = 1$ and our starting function is $\exp(i\nu u)\delta_{v=1}$. The quadratic transform (see (10.13)ff.) maps the point $\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$ into the matrix $\left(\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right)$. We have

$$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix} = \left\{ \begin{pmatrix} n^2 & n \\ n & 1 \end{pmatrix} \right\}$$

so a reasonable guess for our initial function is

$$\hat{\phi} = \delta \left\{ \begin{pmatrix} n^2 & n \\ n & 1 \end{pmatrix} \right\} e^{2\pi i \nu n}. \tag{10.127}$$

Remark. *An examination of the action of G on $\mathbb{R}^2 \backslash \{0\}$ shows certain difficulties in obtaining a meaningful result from summation over $\Gamma_\infty \backslash \Gamma$. For the image under Γ of the line $v = 1$ is easily seen to be dense in the plane, so it is difficult to construct a nontrivial distribution from the sum of δ functions of these lines. (In fact the image of any point $(u^0, 1)$ with u^0 irrational is dense.)*

It appears that it is possible to use a subtle "ergodic theory" argument to define the sum as a distribution. But such an approach would be very complicated. Rather, we define the periods by means of an iterated integration procedure which is well defined.

It may be that other iteration schemes lead to different answers.

The support of $\hat{\phi}$ is contained in the light cone so its Fourier transform ϕ is a solution of the wave equation. To obtain from ϕ a significant function on $\Omega = G/K$ as in the case of the Eisenstein series we first homogenize ϕ (take its Mellin transform under scalar multiplication).

We have

$$\phi\begin{pmatrix} x & w \\ w & z \end{pmatrix} = \int e^{\frac{1}{2}[x+n^2z-2nw]+2\pi i\nu n}\, dn. \tag{10.128}$$

To compute ϕ on Ω in upper half-plane coordinates we set, according to (10.11),

$$\begin{pmatrix} x & w \\ w & z \end{pmatrix} = \begin{pmatrix} \eta+\eta^{-1}\xi^2 & \xi\eta^{-1} \\ \xi\eta^{-1} & \eta^{-1} \end{pmatrix}.$$

Thus the Mellin transform ϕ_s of ϕ in scalar multiplication is, by our usual treatment of the Γ function,

$$\begin{aligned} \phi_s(\xi,\eta) &= \int \phi\left[r\begin{pmatrix} x & w \\ w & z \end{pmatrix}\right] r^s\, \frac{dr}{r} \\ &= A\int e^{2\pi i\nu n}\left[\frac{1}{2\eta}(n^2-2n\xi+\eta^2+\xi^2)\right]^{-s} dn \\ &= A\int e^{2\pi i\nu n}\left\{\frac{1}{2\eta}[n-(\xi+i\eta)][n-(\xi-i\eta)]\right\}^{-s} dn. \end{aligned} \tag{10.129}$$

(As usual A is an elementary expression which may vary.) When $s = 0$ the integral does not seem to make any sense. Perhaps we could obtain an interesting answer using analytic continuation.

When s is a positive integer we can evaluate the integral using the calculus of residues. In particular for $s = 1$, which corresponds to harmonicity,

$$\phi_1(\xi,\eta) = Be^{2\pi i\nu z} - B'e^{2\pi i\nu\bar{z}} \tag{10.130}$$

which is the expected result. (Recall $z = \xi + i\eta$.)

The homogeneous part of ϕ_s is an eigenfunction of the noneuclidean Laplacian Δ with eigenvalue $s(1-s)$. The method of separation of variables given in (10.123)ff. shows that for $s \neq 0, 1$ ϕ_s is of the form

$$\phi_s(\xi,\eta) = ce^{i\nu\xi}B_{s-\frac{1}{2}}(2\pi\nu\eta) \tag{10.131}$$

where B is a Bessel function.

The function $\exp(2\pi i\nu z)$ on the upper half-plane is holomorphic, hence harmonic, and thus corresponds to $s = 0$ or $s = 1$. In fact the harmonics on the upper half-plane which transform according to $\exp(2\pi i\nu n)$ under N form a two-dimensional space spanned by $\exp(2\pi i\nu z)$ and $\exp(2\pi i\nu\bar{z})$. For, the transformation property under N implies that such functions are of the form $\alpha(y)\exp(2\pi i\nu x)$. Harmonicity makes $\alpha(y) = c\exp(2\pi\nu y)$ or $c\exp(-2\pi\nu y)$.

Our next task is to sum ϕ_s over $\Gamma_\infty \backslash \Gamma$. This represents a generalization of Poincaré series. However, for $s = 1$ we do not obtain exactly the usual Poincaré series because we have not separated the holomorphic and the antiholomorphic parts of ϕ_1. The difference between $\exp(i\nu z)$ and ϕ_1 is not very significant for most purposes. (In Section X.6 we shall show how to obtain $\exp(i\nu z)$ alone.)

To define Poincaré series precisely, we must introduce *multiplier systems*. Let $v(\gamma)$ be a complex-valued function (independent of z) on Γ of absolute value 1. We call v a multiplier system for Γ of weight l (see Knopp [104]) if for any $\gamma_1, \gamma_2, \gamma_3 \in \Gamma$ with $\gamma_3 = \gamma_1\gamma_2$ we have

$$v(\gamma_1\gamma_2)(c_3z + d_3)^l = v(\gamma_1)v(\gamma_2)(c_1\gamma_2 z + d_1)^l(c_2z + d_2)^l. \tag{10.132}$$

We have written

$$\gamma_j = \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix}.$$

The function $f(z)$ on the upper half-plane is called a *modular form* of weight l and multiplier system v if, in addition to regularity and growth conditions,

$$f(\gamma z) = v(\gamma)(cz + d)^l f(z) \tag{10.133}$$

for all $\gamma \in \Gamma$.[11]

The present section treats only the case $l = 0$. Thus v is a character of the group Γ. In Section 10.6 we shall show how to pass to general integral l. It can be shown that v is a character when l is an integer.

Let $\nu = \nu_0 + \kappa$ where ν_0 is an integer and $0 \le \kappa < 1$. We define the Poincaré series, formally, by

$$\Phi_v(z, s) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \frac{\phi_s(\gamma z)}{v(\gamma)}. \tag{10.134}$$

Our Fourier method of definition leads to a procedure to establish convergence for $\Re s$ sufficiently large. The underlying idea in our convergence procedure involves "splittings" as in (10.130). We can verify convergence in certain regions for each of the summands. We then (hopefully) can show that the sums have meromorphic continuations to overlapping regions.

We claim that Φ_v is defined: that is, it does not depend on the representatives in $\Gamma_\infty \backslash \Gamma$.

[11] In the literature the weight l is variously called $-l, l/2, -l/2$.

$S = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma$ is the generator of Γ_∞. We have, by (10.131),

$$\begin{aligned} \phi_s(S\gamma z) &= \phi_s(\gamma z + 1) \\ &= e^{i\kappa}\phi_s(\gamma z). \end{aligned} \tag{10.135}$$

Moreover

$$v(S\gamma) = v(S)v(\gamma).$$

Thus $\Phi(v)$ is well defined as long as

$$v(S) = e^{2\pi i\kappa}. \tag{10.136}$$

Φ_v is not Γ invariant but rather for $\gamma' \in \Gamma$

$$\begin{aligned} \Phi_v(\gamma' z, s) &= \sum \frac{\phi_s(\gamma\gamma' z)}{v(\gamma)} \\ &= \sum \frac{\phi_s(\gamma\gamma' z)}{v(\gamma\gamma')} v(\gamma') \\ &= v(\gamma')\Phi_v(z). \end{aligned} \tag{10.137}$$

We shall be interested in the transformation of Φ_v under $\gamma' \in \Gamma_H$ where H is a one-parameter subgroup of G and $\Gamma_H = \Gamma \cap H$. As in the previous section the interesting cases occur when H is hyperbolic or parabolic and Γ_H is an infinite cyclic group (times a finite group of roots of unity which creates no difficulties and will generally be ignored). If γ_0 is a suitably chosen generator of (the infinite cyclic part of) Γ_H (e.g. $\gamma_0 = S$ for $H = \mathbb{N}$) then (10.137) shows that $\Phi_v(z)$ is multiplied by $v(\gamma_0)$ under the action of γ_0. Thus Fourier series for Φ_v on the orbits of H must be defined in terms of characters χ of H which satisfy

$$\gamma_0\chi = v(\gamma_0)\chi. \tag{10.138}$$

In particular for $H = \mathbb{N}$

$$S\chi = e^{2\pi i\kappa}\chi. \tag{10.139}$$

The computation of the Fourier coefficients of Eisenstein series depended on the unfolding lemma 10.7. In the present situation we need a more subtle form:

Lemma 10.10 *Let f^0 be a suitable function or distribution which transforms under the subgroup Γ^0 of Γ by multiplication by v. Suppose the series*

$$f = \sum_{\gamma \in \Gamma^0 \backslash \Gamma} \frac{\gamma f^0}{v(\gamma)}$$

converges suitably. Assume that Γ_H acts generically on $\Gamma^0 \setminus \Gamma$. Then for any function χ on $L_X^y \cap \Omega$ which transforms under $\Gamma_H = \Gamma \cap H$ like v^{-1} we have

$$\int_{\mathcal{D}} \chi f = \int_{L_X^y \cap \Omega} \chi \sum_{\gamma \in \Gamma^0 \backslash \Gamma / \Gamma_H} \gamma f^0. \tag{10.140}$$

H is the orthogonal group of X and $\mathcal{D}$ is a fundamental domain for the action of Γ_H on $L_X^y \cap \Omega$.

Note that, as in (10.134)ff., f is well defined.

Proof The proof of Lemma 10.10 follows exactly along the lines of the proof of Lemma 4.25. The passage from $\mathcal{D}$ to $\gamma_H \mathcal{D}$ where $\gamma_H \in \Gamma_H$ multiplies f by $v(\gamma_H)$ as in (10.137)ff. The factor $v(\gamma_H)$ is canceled by the factor $v^{-1}(\gamma_H)$ coming from χ. This completes the proof.

We are now in a position to evaluate the parabolic periods of Φ_v. Let us examine how (10.112) is to be modified to fit the period situation.

(1) The integer j in (10.112) is now an integer $+\kappa$.
(2) The point $\hat{\mathbf{p}}$ which represents $\gamma\left(\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right)$ is to be replaced by

$$\gamma N \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \gamma \left\{ \begin{pmatrix} n^2 & n \\ n & 1 \end{pmatrix} \right\}.$$

There is a resultant integration over n; this integration must take into account the change of angle $\hat{n}$ as n varies.
(3) The multiplier system is to be introduced in the summation.

If $\gamma = \left(\begin{smallmatrix}p & a\\ q & b\end{smallmatrix}\right) \in \Gamma$ then

$$\gamma \begin{pmatrix} n^2 & n \\ n & 1 \end{pmatrix} = \begin{pmatrix} (pn+a)^2 & (pn+a)(qn+b) \\ (pn+a)(qn+b) & (qn+b)^2 \end{pmatrix} \tag{10.141}$$

which is the quadratic transform of

$$\gamma \begin{pmatrix} n \\ 1 \end{pmatrix} = \begin{pmatrix} pn+a \\ qn+b \end{pmatrix} = \begin{pmatrix} p' \\ q' \end{pmatrix} \tag{10.142}$$

Notation. $\gamma\left(\begin{smallmatrix}n^2 & n\\ n & 1\end{smallmatrix}\right)$ will be denoted by

$$\hat{\mathbf{p}}' = \begin{pmatrix} p'^2 & p'q' \\ p'q' & q'^2 \end{pmatrix}$$

while

$$\hat{\mathbf{p}} = \begin{pmatrix} p^2 & pq \\ pq & q^2 \end{pmatrix}.$$

In the calculation leading to (10.112) we have to replace the angle $\hat{w}/\hat{x} = p/q$ by the angle of (10.141) which is, by (10.40),

$$\frac{\hat{w}'}{\hat{x}'} = \frac{pn+a}{qn+b} = \frac{p'}{q'}. \tag{10.143}$$

This angle is constant on the line $\sigma(\hat{\mathbf{p}}(n))$ joining $\hat{\mathbf{p}}(n)$ to the origin but depends on n.

Equation (10.113) takes essentially the same form except that p, q become p', q' and we have the additional factor $\exp(2\pi i n\nu)$. We have to integrate with respect to n. These changes effect only the factor $L^N(s,j)$ so we compute the new factor $L'(s,j)$. The n integral is

$$\int (qn+b)^{-2s} e^{2\pi i j(pn+a)/(qn+b)+2\pi i n\nu}\, dn. \tag{10.144}$$

To evaluate (10.144) we form the partial fraction decomposition

$$\frac{pn+a}{qn+b} = \frac{p}{q} - \frac{1}{q(qn+b)}. \tag{10.145}$$

(This identity uses $1 = pb - aq$.) Now we make the change of variables $qn+b \to u$ to obtain

$$q^{-1} e^{2\pi i j p/q - 2\pi i \nu b/q} \int u^{-2s} e^{2\pi i j q^{-1} u^{-1} + 2\pi i \nu q^{-1} u}\, du. \tag{10.146}$$

To put the integral in standard form we write

$$k^2 = -4\pi^2 j q^{-2} \nu$$
$$\lambda^2 = \frac{\nu}{j}.$$

Thus the exponential becomes

$$e^{k[\lambda u + (\lambda u)^{-1}]}.$$

When we change variables $\lambda u \to u$ we introduce a factor λ^{2s-1} so the integral is

$$\lambda^{2s-1} \int u^{-2s} e^{k(u+u^{-1})}\, du = \left(\frac{\nu}{j}\right)^{s-1/2} K_{-2s+1}(-2k) \tag{10.147}$$

[11, Vol. II, p. 82 (23)].

It remains to sum this times the appropriate multipliers $v(\gamma)$ over the double cosets of Γ_∞ for $q \neq 0$ and over the coset Γ_∞ for $q = 0$ as in the case of the Eisenstein series and then multiply by the Bessel function of (10.13). Representatives for the double cosets are given in (10.105).

When q, ν, j are fixed we can sum over the double cosets. For $q \neq 0, 1$ the coefficient of the K function is, by (10.146),

$$K^{\mathcal{P}}(j,\nu,v) = \frac{1}{q} \sum_{\substack{0<p<q \\ (p,q)=1}} \frac{e^{2\pi i/q(pj-\nu b)}}{v\begin{pmatrix} p & a \\ q & b \end{pmatrix}}. \tag{10.148}$$

b is the unique inverse of p mod q which lies in $(0, q)$ and $a = (pb - 1)/q$. To compare (10.148) with p. 89 of [104]: our notation j, p, q, b corresponds to $n - \kappa, h, k, h'$. For the η function $\kappa = 1/24$.

The expression (10.148) is an example of a Kloosterman sum (see [104]). It is somewhat reminiscent of a Bessel function since it is a sum over p and $b \equiv p^{-1}$ (mod q). Such sums appear in formulas for parabolic Fourier series coefficients of modular functions in the work of [104, 127, 130, 132].

Problem 10.3 *Give a conceptual reason for such coefficients to be Bessel-like.*

Remark. *We remarked following (10.102) that the periods of the Eisenstein series are products of arithmetically defined functions with transcendental functions depending on j, y, s. The splitting in (10.146) as completed in (10.148) is somewhat "weaker" in that the transcendental factor K_{-2s+1} depends on q; that is, on the double coset.*

Hyperbolic cycles

To compute the periods of the Poincaré series over hyperbolic cycles we have to examine the integral (10.90) as modified by (10.100). As in our remark following (10.91) we make the change of variables $\hat{r}\hat{y}_X(\hat{\mathbf{p}}') \to \hat{r}$. Since $\rho_X(\hat{\mathbf{p}}') = \hat{y}_X(\hat{\mathbf{p}}')$ this separates the integral into a part depending only on $\hat{r}$ and a part depending on $\hat{\mathbf{p}}'$. As in the case of parabolic cycles the $\hat{r}$ terms do not depend on $\hat{\mathbf{p}}$ or n so they are the same for the Eisenstein and Poincaré series, namely the hypergeometric function ${}_2F_1$ of (10.101). We shall therefore restrict our attention to the $\hat{\mathbf{p}}'$ part.

The $\hat{\mathbf{p}}'$ terms differ from those of the Eisenstein series in that p, q are replaced by p', q', and there is an additional factor $\exp(2\pi i\nu n)$, and we have to integrate over n. Also $\hat{\theta}(\hat{\mathbf{p}}')$ has to be computed in terms of n; it does not depend on $\hat{r}$.

We shall explain below how j and θ^0 of (10.100) are to be modified. For the present we shall simply write j for the analog of $j/\hat{\theta}^0$ in (10.100).

As the computation is somewhat complicated let us begin with the simple case $X = \tilde{I} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. We deal with a fixed γ as summing over γ for $X = \tilde{I}$ leads to many convergence problems.

We have to compute the $\hat{y}_{\tilde{I}}$ and $\hat{\theta}_{\tilde{I}}$ coordinates of

$$\hat{\mathbf{p}}' = \begin{pmatrix} p'^2 & p'q' \\ p'q' & q'^2 \end{pmatrix}.$$

$L^0_{\tilde{I}}$ is the set of $U = \left(\begin{smallmatrix} \hat{x} & \hat{w} \\ \hat{w} & \hat{z} \end{smallmatrix}\right)$ such that $\tilde{I} \cdot U = 0$, i.e. the *antitrace* $\hat{z} - \hat{x} = 0$, so that $\hat{z} = \hat{x}$. If we, write

$$\hat{\mathbf{p}}' = \hat{y}\tilde{I} + U$$

with $U \in L^0_{\tilde{I}}$ then by (10.6)

$$\begin{aligned}\hat{y} &= -\tilde{I} \cdot \hat{\mathbf{p}}' \\ &= \tfrac{1}{2}(p'^2 - q'^2).\end{aligned} \tag{10.149}$$

To compute the angle $\hat{\theta}_{\tilde{I}}(\hat{\mathbf{p}}')$ we use $\{\left(\begin{smallmatrix}\alpha & 0\\ 0 & 0\end{smallmatrix}\right)\}_{\alpha \geq 0}$ as the angle base ray. We shall explain the significance of this choice below. The angle base point on $L^{\hat{y}}_{\tilde{I}}$ is

$$U_{\tilde{I}} = \begin{pmatrix} 2\hat{y} & 0 \\ 0 & 0 \end{pmatrix} = \hat{y}\tilde{I} + \hat{y}I.$$

The isotropy group of $\tilde{I}$ is

$$H_{\tilde{I}} = \left\{ \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix} \right\}. \tag{10.150}$$

For $\hat{\mathbf{p}}' \in L^{\hat{y}}_{\tilde{I}}$, by (10.149), the angle $\hat{\theta}(\hat{\mathbf{p}}')$ is determined by

$$\begin{pmatrix} \cosh\hat{\theta} & \sinh\hat{\theta} \\ \sinh\hat{\theta} & \cosh\hat{\theta} \end{pmatrix} (p'^2 - q'^2) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \hat{\mathbf{p}}'; \tag{10.151}$$

that is,

$$(p'^2 - q'^2) \begin{pmatrix} \cosh^2\hat{\theta} & \cosh\hat{\theta}\sinh\hat{\theta} \\ \cosh\hat{\theta}\sinh\hat{\theta} & \sinh^2\hat{\theta} \end{pmatrix} = \hat{\mathbf{p}}'. \tag{10.152}$$

Using $\cosh^2\hat{\theta} + \sinh^2\hat{\theta} = \cosh 2\hat{\theta}$ and $2\sinh\hat{\theta}\cosh\hat{\theta} = \sinh 2\hat{\theta}$ we find

$$\begin{aligned}(p'^2 - q'^2)\cosh 2\hat{\theta} &= p'^2 + q'^2 \\ (p'^2 - q'^2)\sinh 2\hat{\theta} &= 2p'q'\end{aligned}$$

from which follows

$$e^{2\hat{\theta}} = \frac{p' + q'}{p' - q'}. \tag{10.153}$$

Formula (10.153) makes sense only if the right side is >0. When it is <0 we use the angle base point $(p'^2 - q'^2)\left(\begin{smallmatrix}-1 & 0\\ 0 & 0\end{smallmatrix}\right)$. Two angle base points are necessary because of the two branches of the hyperbola $L^{\hat{y}}_X \cap \Omega$.

We shall deal only with the $+$ branch as the $-$ branch is handled in the same manner.

When we make the change of variable $\hat{r}\hat{y}(\hat{\mathbf{p}}) \to \hat{r}$ as described preceding (10.90) with the above modifications, the $\hat{\mathbf{p}}'$ part of the integral becomes[12]

$$\begin{aligned}\int e^{2\pi i\nu n}\hat{y}^{-is}e^{2\pi ij\hat{\theta}}dn &= \int e^{2\pi i\nu n}(p'^2-q'^2)^{-is}\left(\frac{p'+q'}{p'-q'}\right)^{\pi ij}\,dn\\ &= \int e^{2\pi i\nu n}[n(p+q)+(a+b)]^{-is+\pi ij}\\ &\quad\times[n(p-q)+(a-b)]^{-is-\pi ij}\,dn\\ &= (p^2-q^2)^{-is}\left(\frac{p+q}{p-q}\right)^{\pi ij}\\ &\quad\times\int e^{2\pi i\nu n}\left[n+\frac{a+b}{p+q}\right]^{-is+\pi ij}\left[n+\frac{a-b}{p-q}\right]^{-is-\pi ij}dn.\end{aligned} \tag{10.154}$$

An integral of the form

$$\int e^{-kn}(n+u)^l(n+v)^m\,dn$$

can be reduced to standard Whittaker form, i.e.

$$\begin{aligned}&= e^{ku}\int e^{-kn}n^l(n+v-u)^m\,dn\\ &= (v-u)^m e^{ku}\int e^{-kn}n^{l+1}\left(\frac{n}{v-u}+1\right)^m\frac{dn}{n}\\ &= (v-u)^{m+l+1}e^{ku}\int e^{-k(v-u)n}n^{l+1}(n+1)^m\frac{dn}{n}\\ &= \Gamma(l+1)(v-u)^m e^{ku}W(l+1,l+m+2;k(v-u))\end{aligned} \tag{10.155}$$

[11, Vol. I, p. 255 (2)].
Combining (10.154) with (10.155) gives

$$\begin{aligned}&\int e^{2\pi i\nu n}\hat{y}^{-is}e^{2\pi ij\hat{\theta}}\,dn\\ &\quad= \Gamma(-is-\pi ij+1)2^{-is+\pi ij}(p-q)^{-2\pi ij}e^{-2\pi i\nu\frac{a-b}{p-q}}\\ &\qquad\times W\left(-is-\pi ij+1,-2is+2;\frac{-4\pi i\nu}{p^2-q^2}\right).\end{aligned} \tag{10.156}$$

[12]Integration in n is over the full group N, and hence over $(-\infty,\infty)$. The integrals in (10.154)ff. are not absolutely convergent and must be understood in the sense of analytic continuation described in Section 7.2. The interval $(-\infty,\infty)$ has to be divided into $(-\infty,0)\cup(0,\infty)$ to apply this analytic continuation.

We wish to pass to a general hyperbolic X with $\det X = -1$. We shall make the necessary modifications from $X = \tilde{I}$.

$$\hat{\mathbf{p}}' = \hat{y}X + U$$

with $U \in L_X^0$ gives

$$\hat{y} = -X \cdot \hat{\mathbf{p}}'. \tag{10.157}$$

To compute the angle $\hat{\theta}(\hat{\mathbf{p}}')$ we have to understand the significance of our choice of $\{\left(\begin{smallmatrix}\alpha & 0\\ 0 & 0\end{smallmatrix}\right)\}$ as the angle base ray when $X = \tilde{I} = \left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$.

The eigenvectors of $h_{\tilde{I}} = \left(\begin{smallmatrix}\cosh\theta & \sinh\theta\\ \sinh\theta & \cosh\theta\end{smallmatrix}\right)$ (see (10.150)) are

$$\begin{aligned} \tilde{I} &\quad \text{eigenvalue } 1\\ \tilde{\xi}_1 = \begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} &\quad \text{eigenvalue } e^{2\theta}\\ \tilde{\xi}_2 = \begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix} &\quad \text{eigenvalue } e^{-2\theta}. \end{aligned} \tag{10.158}$$

The angle base point on $L_{\tilde{I}}^{\hat{y}}$ can be written as

$$U_{\tilde{I}} = \begin{pmatrix}2\hat{y} & 0\\ 0 & 0\end{pmatrix} = \frac{1}{2}\hat{y}(\tilde{\xi}_1 + \tilde{\xi}_2) + \hat{y}\tilde{I}. \tag{10.159}$$

Note that $\tilde{\xi}_1, \tilde{\xi}_2 \in L_{\tilde{I}}^0 \cap \Gamma$. We added $\hat{y}\tilde{I}$ to $\frac{1}{2}\hat{y}(\tilde{\xi}_1 + \tilde{\xi}_2)$ to make the sum in $L_{\tilde{I}}^{\hat{y}} \cap \Gamma$.

Note that $\tilde{\xi}_1, \tilde{\xi}_2 \in \Gamma$ so $\tilde{\xi}_1 \cdot \tilde{\xi}_1 = \tilde{\xi}_2 \cdot \tilde{\xi}_2 = 0$. On the other hand $\tilde{\xi}_1 \cdot \tilde{\xi}_2 = 2$.

Let $g \in G = SL(2, \mathbb{R})$ map $\tilde{I} \to X$. Thus $gH_{\tilde{I}}g^{-1} = H_X$. The eigenvectors of h_X are

$$\begin{aligned} X &\quad \text{eigenvalue } 1\\ \xi_1 = g\tilde{\xi}_1 &\quad \text{eigenvalue } e^{2\theta}\\ \xi_2 = g\tilde{\xi}_2 &\quad \text{eigenvalue } e^{-2\theta}. \end{aligned}$$

(Recall that gU means gUg^t for U a symmetric 2×2 matrix but gh is the product of g and h when h is a group element.)

We call $U_X = \frac{1}{2}(\xi_1 + \xi_2) + X = gU_{\tilde{I}}$. We can multiply g on the left by an element of H_X so we can assume that g fixes the angle base point $\hat{y}U_{\tilde{I}}$ (assuming that it is not an eigenvector of H_X).

We now return to the spinor principle in the form of Proposition 10.3. We use the notation of (10.44)ff. where Q is the quadratic form defined by X. Since

$\xi_1, \xi_2 \in$ light cone we can write

$$\begin{aligned}
\xi_1 &= \begin{pmatrix} \alpha_1^2 & \alpha_1\alpha_2 \\ \alpha_1\alpha_2 & \alpha_2^2 \end{pmatrix} \\
&= \text{symmetric square } \xi_1^{1/2} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \\
\xi_2 &= \begin{pmatrix} \beta_1^2 & \beta_1\beta_2 \\ \beta_1\beta_2 & \beta_2^2 \end{pmatrix} \\
&= \text{symmetric square } \xi_2^{1/2} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.
\end{aligned} \tag{10.160}$$

To compute the angle of $\hat{\mathbf{p}}'$ we write

$$\hat{y} h_{\hat{\theta}} U_X = \hat{\mathbf{p}}'. \tag{10.161}$$

The left hand side of (10.161) is, by (10.159),

$$\hat{y}\left(\tfrac{1}{2}e^{2\hat{\theta}}\xi_1 + \tfrac{1}{2}e^{-2\hat{\theta}}\xi_2 + X\right).$$

Now, $\xi_1, \xi_2 \in \Gamma$ and ξ_j are orthogonal to X (since $\tilde{\xi}_j$ are orthogonal to $\tilde{I}$). Moreover since $\tilde{\xi}_1 \cdot \tilde{\xi}_2 = 2$

$$\begin{aligned}
\hat{y}e^{2\hat{\theta}} &= \xi_2 \cdot \hat{\mathbf{p}}' \\
&= \tfrac{1}{2}\beta_1^2 q'^2 + \frac{1}{2}\beta_2^2 p'^2 - \beta_1\beta_2 p' q' \\
&= \tfrac{1}{2}(\beta_1 q' - \beta_2 p')^2.
\end{aligned} \tag{10.162}$$

$\xi_1^{1/2}$ and $\xi_2^{1/2}$ are the eigenvectors of h_X with respective eigenvalues $\exp(\theta), \exp(-\theta)$ since the analogous property holds for $\tilde{I}$. Thus by Proposition 10.4, up to a constant,

$$\begin{aligned}
Q(q', p') &= (\alpha_1 q' + \alpha_2 p')(\beta_1 q' + \beta_2 p') \\
X &= \begin{pmatrix} \alpha_1\beta_1 & \frac{1}{2}(\alpha_1\beta_2 + \alpha_2\beta_1) \\ \frac{1}{2}(\alpha_1\beta_2 + \alpha_2\beta_1) & \alpha_2\beta_2 \end{pmatrix}
\end{aligned} \tag{10.163}$$

so that, by (10.157),

$$\hat{y} = -X \cdot \hat{\mathbf{p}}' = -\tfrac{1}{2}(\alpha_1 q' - a_2 p')(\beta_1 q' - \beta_2 p'). \tag{10.164}$$

Combining (10.164) with (10.162) gives

$$e^{2\hat{\theta}} = -\frac{\beta_1 q' - \beta_2 p'}{\alpha_1 q' - a_2 p'}. \tag{10.165}$$

This is the spinor principle which computes the eigenvalue as the quotient of linear forms.

We are now in a position to repeat the computation in (10.154)ff. The integral is now (up to powers of -1), using (10.155),

$$\begin{aligned}
&\int \hat{y}^{-is} e^{2\pi i j\hat{\theta}} e^{2\pi i\nu n}\, dn \\
&\quad = \int \left[(\alpha_1 q' - \alpha_2 p')(\beta_1 q' - \beta_2 p')\right]^{-is} \left[\frac{\beta_1 q' - \beta_2 p'}{\alpha_1 q' - a_2 p'}\right]^{i\pi j} e^{2\pi i\nu n}\, dn \\
&\quad = \int e^{2\pi i\nu n} [n(\alpha_1 q - \alpha_2 p) + \alpha_1 b - \alpha_2 a]^{-is - i\pi j} \\
&\qquad \times [n(\beta_1 q - \beta_2 p) + \beta_1 b - \beta_2 a]^{-is + i\pi j}\, dn \\
&\quad = [(\alpha_1 q - \alpha_2 p)(\beta_1 q - \beta_2 p)]^{-is} \left[\frac{\beta_1 q - \beta_2 p}{\alpha_1 q - \alpha_2 p}\right]^{i\pi j} \\
&\qquad \times \int e^{2\pi i\nu n} [n + u]^{-is - i\pi j} [n + v]^{-is + i\pi j}\, dn \qquad (10.166) \\
&\quad = \Gamma(-is - \pi i j + 1)(\alpha_1\beta_2 - \alpha_2\beta_1)^{-is + \pi i j} (\alpha_1 q - \alpha_2 p)^{-2\pi i j} \\
&\qquad \times e^{-2\pi i\nu \frac{\alpha_1 b - \alpha_2 a}{\alpha_1 q - \alpha_2 p}} W\left(-is - \pi i j + 1; -2is + 2; \frac{2\pi i\nu(\alpha_1\beta_2 - \alpha_2\beta_1)}{(\alpha_1 q - \alpha_2 p)(\beta_1 q - \beta_2 p)}\right).
\end{aligned}$$

We have used (10.155) to evaluate the integral. (The discriminant of the quadratic form is 1 by our normalization of X.)

As in the case of Eisenstein series we have to sum this over $\Gamma_\infty \backslash \Gamma / \Gamma_X$. This makes most sense when Γ_X is infinite. We ignore the unimportant finite group of roots of unity. Let γ^0 be a generator of Γ_X with angle $\gamma^0 = \theta^0$. Then

$$\gamma^0 \colon\ e^{2\pi i j\hat{\theta}} \to e^{2\pi i j\hat{\theta}} e^{2\pi i j\theta^0}.$$

We have seen in (10.137) that for the Poincaré series $\mathbf{P}$ we have $\gamma^0\mathbf{P} = v(\gamma^0)\mathbf{P}$. Thus we require

$$e^{2\pi i j\theta^0} = v(\gamma^0)$$

as in the case of parabolic cycles.

As usual we fix $X \cdot \hat{\mathbf{p}}$ and sum over this finite set of $\hat{\mathbf{p}}$. We call

$$K_X^{\mathcal{H}}(j, \nu, v) = \sum \frac{1}{v(\gamma)} (\alpha_1 q - \alpha_2 p)^{-2\pi i j} e^{-2\pi i\nu \frac{\alpha_1 b - \alpha_2 a}{\alpha_1 q - \alpha_2 p}} \qquad (10.167)$$

the *hyperbolic Kloosterman sum*. As in (10.100) the sum is over all $\mathbf{p} = \left(\begin{smallmatrix} p & a \\ q & b \end{smallmatrix}\right)$ lying in a fixed fundamental domain for Γ_X acting on $L_X^{\hat{y}}$.

Problem 10.4 *Study these hyperbolic Kloosterman sums.*

We call $\mathcal{W}$ the product of all terms in (10.166) which do not depend on $\mathbf{p}$, and call $\mathcal{F}$ the hypergeometric function of (10.101). We have shown

Theorem 10.11 *The hyperbolic periods of the Poincaré series are*

$$\mathcal{W}\mathcal{F}K_X^{\mathcal{H}}(j,\nu,v).$$

Remark 1 *Our formulas show that the type of splitting of the hyperbolic periods of Poincaré series is similar to the type of splitting of parabolic periods (see the remark following (10.148)).*

Remark 2 *The computations of the periods of the Poincaré series follow the general lines set forth at the end of Section 3.6. An additional complication arises because of the integration over N which makes these periods "triple Radon transforms."*

10.5 Hyperbolic Eisenstein and Poincaré series

The Eisenstein and Poincaré series that we have studied might properly be called *parabolic Eisenstein series* or *parabolic Poincaré series*, the reason being that we started with measures on the light cone with a prescribed transformation under (left) N. We now pass to *hyperbolic Eisenstein and hyperbolic Poincaré series*.

The origin of these series is a measure on G/A which, in the three-dimensional representation, is a hyperboloid of one sheet. We have seen how integral points $\hat{\mathbf{p}}$ with $\det\hat{\mathbf{p}} < 0$ correspond to ideals in real quadratic number fields. In particular for such a $\hat{\mathbf{p}}$ the orbit $\Gamma\hat{\mathbf{p}}$ is discrete and the isotropy group $\Gamma_{\hat{\mathbf{p}}} = \Gamma \cap H_{\hat{\mathbf{p}}}$ is infinite. $\Gamma_{\hat{\mathbf{p}}}$ is an infinite cyclic group times a finite group of roots of unity; the finite group plays no essential role and will be ignored. The orbit $\Gamma\hat{\mathbf{p}}$ can be identified with the narrow ideal class of the ideal corresponding to $\hat{\mathbf{p}}$.

To simplify our calculations we divide $\hat{\mathbf{p}}$ by $\sqrt{|\det\hat{\mathbf{p}}|}$ so we assume $\det\hat{\mathbf{p}} = -1$.

We start with one such $\hat{\mathbf{p}}$ which we denote by $\hat{\mathbf{p}}^0$. The points $\gamma\hat{\mathbf{p}}^0$ are called $\hat{\mathbf{p}}^0$ *lattice points*. To define the hyperbolic Eisenstein series we begin with

$$\hat{\Theta} = \sum_{\Gamma/\Gamma_{\hat{\mathbf{p}}^0}} \delta_{\gamma\hat{\mathbf{p}}^0}.$$

We then form its Fourier transform

$$\begin{aligned}\Theta(X) &= \sum \hat{\delta}_{\gamma\hat{\mathbf{p}}^0}(X) \\ &= \sum e^{iX\cdot\gamma\hat{\mathbf{p}}^0}.\end{aligned} \tag{10.168}$$

Θ and $\hat{\Theta}$ are clearly Γ invariant. Our next step is the decomposition of Θ so as to extract interesting functions on $\Omega = G/K = \{X \mid \det X = 1\}$. In the

parabolic case this was accomplished by means of the Mellin transform in scalar multiplication. In the hyperbolic case the Mellin transform should be replaced by the Bessel transform. We now clarify this remark.

Since $\det \gamma\hat{\mathbf{p}}^0 = -1$, Θ satisfies the inhomogeneous wave equation

$$(\Box + 1)\Theta = 0 \tag{10.169}$$

where $\Box$ is the wave operator. If f is any solution of (10.169) then we can write

$$f(rz) = \int \phi_s(z) r^{-1/2} K_{s-1/2}(r)\, ds. \tag{10.170}$$

We have denoted by z a parameter on Ω and by r the Minkowski distance parameter. K_s is the usual Bessel function and

$$\Delta\phi_s = s(1-s)\phi_s \tag{10.171}$$

where Δ is the Laplacian on G/K. Thus (10.170) allows us to extract eigenfunctions of Δ from solutions of (10.169) in much the same way as the Mellin transform produced eigenfunctions of Δ from solutions of the homogeneous wave equation.

To understand the meaning of ϕ_s, let us start with the point $\mathbf{p}^0 = (0, 0, 1)$. (We are using coordinates (y, t_1, t_2).) On G/K we write $t_2 = \sinh\zeta \sin\theta$. According to equation (10.168) ϕ_s is obtained from the function $F(z) = \exp(i \sin\zeta \sin\theta)$ by decomposition under Bessel transforms.

ϕ_s is an eigenfunction of the Laplacian Δ on G/K with eigenvalue $s(1-s)$ as in the case of the parabolic Eisenstein series treated above. Moreover, ϕ_s is a function of the A-invariant t_2 which, by (10.12), is $\xi\eta^{-1}$ in coordinates $z = \xi + i\eta$.

In these coordinates

$$\Delta = \eta^2 \left(\frac{\partial^2}{\partial\xi^2} + \frac{\partial^2}{\partial\eta^2} \right).$$

To simplify the notation we write t for t_2. A simple computation shows that, if $g = g(t)$,

$$\Delta g(t) = (1 + t^2)g''(t) - 2tg'(t).$$

According to [11, vol. I, p. 121] the equation $\Delta g = s(1-s)g$, which is

$$(1 + t^2)g''(t) \; - 2tg'(t) - s(1-s)g(t) = 0$$

defines the Legendre function $P_s^0(it)$. This coincides with the expression of ϕ_s as a Bessel transform via [11, vol. II, p. 56].

We are faced with the problem of summing ϕ_s over Γ. Of course $\Gamma_{\hat{\mathbf{p}}^0} = \Gamma \cap A = \{\pm I\}$ so we cannot expect convergence. Suppose that we make a fractional linear transformation taking Γ to $\tilde{\Gamma}$ so that $\tilde{\Gamma}_{\mathbf{p}^0} = \tilde{\Gamma} \cap A = \{H^n\}$ where H is a

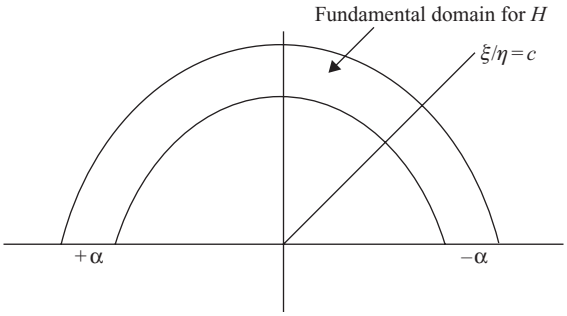

Figure 10.1

hyperbolic transformation. By the ideas of Section 10.3, H is the transform of a generator of the unit group of a real quadratic form.

The geometry is described in Figure 10.1

In the case of the parabolic Eisenstein series the usual fundamental domains for Γ which meet a fundamental domain for $\Gamma^0 = \{z \to z + n\}$ converge to the ξ axis. Only one such fundamental domain reaches to $\eta = \infty$. It is their geometry which accounts for the convergence of the parabolic Eisenstein series since we sum $\eta^s = (Imz)^s$ over the points $\{\gamma z\}_{\gamma \in \Gamma^\circ / \Gamma}$ and, for Re s large, $|\eta^s|$ is small.

In the present situation $\phi_s = P_s$ cannot be made small near $-\alpha$ or $+\alpha$ (see Figure 10.1). Moreover, the fundamental domains for $\tilde{\Gamma}$ crowd around both $\pm\alpha$, which correspond to $t = \pm\infty$.

In order to obtain convergence we split P_{-s} into Q_{-s} and Q_{s-1} where Q_s is the associated Legendre function [11, vol. I, p. 140]. This is the analog of Euler's decomposition

$$\cos sx = \tfrac{1}{2}(e^{is|x|} + e^{-is|x|}).$$

$\exp(is|x|)$ is not an eigenfunction of d^2/dx^2. Rather

$$\left(\frac{d^2}{dx^2} + s^2\right) e^{\pm isx} = \pm 2\delta,$$

meaning that $\exp(is|x|)$ is a fundamental solution for $(d^2/dx^2 + s^2)$. In a similar vein $Q_{-s}(it)$ is an eigenfunction of the Legendre operator except at $it = \pm 1$. Of course, this is a disadvantage, but it is compensated by the fact that

$$Q_{-s}(t) \sim c(s)t^{s-1} \quad t \to \infty.$$

The asymptotic formula can be deduced from the expression of Q as a hypergeometric series [11, vol. I, p. 122 (5)]

$$Q_s(t) = c(s)t^{-s-1}F\left(\frac{s}{2} + 1, \frac{s}{2} + \frac{1}{2}; s + \frac{3}{2}; t^{-2}\right).$$

It follows that we can sum $Q_{-s}(\gamma t)$ over $\gamma \in \tilde{\Gamma} \cap A/\tilde{\Gamma}$ when $\operatorname{Re} s < 0$. Similarly we can sum $Q_{s-1}(\gamma t)$ for $\operatorname{Re} s > 2$.

Although the series $\sum Q_{-s}(\gamma t)$ is $\tilde{\Gamma}$ invariant, it does not define an eigenfunction of Δ. Rather

$$[\Delta - s(1-s)] \sum Q_{-s}(\gamma t)$$

has support on the unions of the transforms by $\{\gamma\}$ of the rays $\{t = \pm 1\}$. Moreover, just as in the case of $\exp(\pm is|x|)$, the values of $[\Delta - s(1-s)] \sum Q_{-s}(\gamma t)$ and $[\Delta - s(1-s)] \sum Q_{s-1}(\gamma t)$ on these sets are negatives of each other.

To obtain actual eigenfunctions we continue $\sum Q_{-s}(\gamma t)$ and $\sum Q_{s-1}(\gamma t)$ meromorphically and then add the meromorphic continuations.

To carry out the continuation we recall that if we use $\hat{r}^s$ in place of th Bessel function K_s to extract an interesting function on G/K from γ_{001} (see (10.173) ff.) then we are in the situation treated by [112]. The Legendre function $P_{-s}(t)$ is replaced by t^{-s}. This means that the terms in the asymptotic expansion of $\sum Q_{-s}(\gamma t)$ become hyperbolic Eisenstein series in the sense of Kudla and Milson [112]. Using the recursion equation (10.174), Kudla and Milson demonstrated the meromorphic continuation of their Eisenstein series.

We conclude that $\sum Q_{-s}(\gamma t)$ has a meromorphic continuation to the whole complex plane so the eigenfunction

$$\sum \phi_s(\gamma t) = \sum Q_{-s}(\gamma t) + \sum Q_{s-1}(\gamma t)$$

is well defined

Problem 10.5 *Study this hyperbolic Eisenstein series. Does it differ from the parabolic Eisenstein series?*

For the origin of (10.170) we write

$$\Box = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} - \frac{1}{r^2}\Delta. \tag{10.172}$$

A product solution $A(z)B(r)$ (z is the parameter in G/K) of (10.169) satisfies

$$\left[\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + 1 + \frac{s(1-s)}{r^2}\right] B = 0$$

$$\Delta A = -s(1-s)A. \tag{10.173}$$

Thus A is an eigenfunction of Δ as in Equation (10.171) and B is of the form $r^{-1/2}K_{1/2-s}$ [11, vol. II, p. 2 (13)]. (10.170) is the decomposition of f into such product solutions.

We can understand the equations (10.173) as follows. The operator $\Box + 1$ involves two degrees of homogeneity (-2 and 0) under scalar multiplication (i.e. in r). By (10.172) the Bessel equation (10.173) involves two degrees of

homogeneity while the eigenvalue equation (10.171) is homogeneous of degree 0. (Δ is independent of r hence is homogeneous of degree 0.) Thus the radial equation and radial function are complicated but the z equation is simple.

We could have reversed our position and sought an expansion of the form $\int r^s \psi_s(z)\, ds$. In this case r^s is simple and satisfies a simple equation but the two degrees of homogeneity impinge on ψ_s which formally satisfies

$$\Delta\psi_s + s(s+1)\psi_s + \psi_{s-2} = 0. \tag{10.174}$$

This difference equation approach was studied in detail in [112].

Remark 3 *Equations (10.173) arise from the inhomogeneous wave equation (10.169) by putting the inhomogeneous term 1 in the radial equation for $B(r)$. The Kudla–Milson [112] approach puts 1 in the z equation.*

Let us return to the separation (10.170). There is an inversion formula which computes ϕ_s in terms of f [11, vol. II, p. 75 (75)]. Up to a normalizing constant

$$\phi_s(z) = t \sinh \pi t \int_0^\infty K_{it}(\hat{r})\hat{r}^{1/2} f(\hat{r}z)\, \frac{d\hat{r}}{\hat{r}} \tag{10.175}$$

where $s = \frac{1}{2} + it$.

Equations (10.170) and (10.175) involve the values of $\phi_s(z)$ only on $\Omega = G/K$. Of course when using separation of variables it is tacitly assumed that $\phi_s(z)$ is extended to $\{(z,s)\}$ by making it constant in r.

We shall, however, follow a somewhat different method. For $f = \delta_{\hat{\mathbf{p}}}$ we extend ϕ_s to $\{rz\}$ by

$$\phi_s(rz) = \left(s - \tfrac{1}{2}\right) \cos \pi s \delta_{\sigma(\hat{p})}(\hat{K}_{s-1/2}). \tag{10.176}$$

We have written $\delta_{\sigma(\hat{\mathbf{p}})}(K_{s-1/2})$ for the δ function of the ray $\sigma(\hat{\mathbf{p}}) = \{\hat{r}\hat{\mathbf{p}}\}$ with the measure $K_{s-1/2}(\hat{r})\hat{r}^{1/2} d\hat{r}/\hat{r}$.

The factor $(s - 1/2)\cos \pi s$ is unimportant; it could be absorbed in ϕ_s so we can assume that is done. We are in essentially the same position we were in for $\hat{\mathbf{p}} \in \Gamma$ except that $\hat{r}^s$ is replaced by $K_{s-1/2}(\hat{r})\hat{r}^{1/2}$.

Let $\hat{\mathbf{p}}^0$ in (10.168) be a hyperbolic unit vector such that $\Gamma_{\hat{\mathbf{p}}^0}$ is infinite. We assume that Γ_X is infinite and acts generically on $\Gamma_{\hat{\mathbf{p}}^0} \setminus \Gamma$ so we can apply the unfolding lemma 10.7. (We could weaken this hypothesis.)

We are unable to compute the hyperbolic periods of the hyperbolic Eisenstein series on $L_X^y \cap \Omega$ unless $y = 0$ in which case $L_X^0 \cap \Omega$ is a geodesic because, for $\hat{\mathbf{p}} \in G/A$, $\hat{y}_X(\hat{\mathbf{p}}) \neq \rho_X(\hat{\mathbf{p}})$ (see the remark following (10.91)).

The distance base point on L_X^0 is the origin so all points U in $L_X^0 \cap \Omega$ have $\rho_X(U) = 1$. For $U \in L_X^y$ we have, by (10.75),

$$\rho_X^2(U) = \hat{y}^2 + 1.$$

As in the case of the parabolic Eisenstein series, the sum over $\hat{\mathbf{p}}^0$ lattice points involves summing over the $\gamma\hat{\mathbf{p}}^0$ which lie in fundamental domains $\hat{D}$ for $\Gamma_X = \Gamma \cap H_X$ on the H_X orbits on G/A. Since H_X preserves X it preserves $\hat{y}$ and hence all points on each $\hat{D}$ have fixed $\hat{\rho}_X$.

The Fourier transform of $\exp(2\pi ij\theta/\theta^0)\delta_{L^0_X \cap \Omega}$ is given by (10.100). As in the case of parabolic Eisenstein series we perform the sum over Γ by summing first for fixed $\hat{\rho}_X(\gamma\hat{\mathbf{p}}^0)$ and then summing over $\{\hat{\rho}_X\}$. The angular structure is exactly the same as in the parabolic case.

We fix $\hat{\rho}_X(\gamma\hat{\mathbf{p}}^0) = m$. After we apply Parseval's formula as we did for the parabolic periods of the ordinary Eisenstein series the $\hat{r}$ integral becomes

$$\int K_{ij/\theta^0}(\hat{r}m) K_{s-1/2}(\hat{r})\hat{r}^{1/2}\,\frac{d\hat{r}}{\hat{r}}.$$

The evaluation of this integral in [11, vol. II, p. 95(36)] leads to

Theorem 10.12 *The hyperbolic X periods of the hyperbolic $\hat{\mathbf{p}}^0$ Eisenstein series are*

$$A\sum_m \sum_{\hat{\theta}} e^{2\pi ij\hat{\theta}/\theta^0}\, {}_2F_1\left[\frac{1}{2}(1+s+ij/\theta^0), \frac{1}{2}(1+s-ij/\theta^0), \frac{3}{2}; 1-\frac{1}{m^2}\right]. \tag{10.177}$$

A is expressed in terms of Γ functions and elementary functions. The $\hat{\theta}$ sum is over $\hat{X}$ angles of all $\hat{\mathbf{p}}^0$ lattice points in a fundamental domain on $\hat{\rho}_X = m$.

Remark 1 *Unlike the case of parabolic Eisenstein series the values of $\|\gamma\hat{\mathbf{p}}^0\|$ do not appear in (10.187) because $\|\gamma\hat{\mathbf{p}}^0\| = -1$ for all γ.*

Remark 2 *Formula (10.177) shows that we have a splitting of the period into an arithmetic factor which is the sum over $\hat{\theta}$ and a transcendental factor. This splitting is more analogous to the splitting of the periods of the Poincaré series (as in the remark following (10.148)) than to the Eisenstein series (see (10.100) and the remark following Theorem 10.9) because the transcendental factor ${}_2F_1$ involves m.*

Remark 3 *The computation of the periods does not follow the lines set forth at the end of Section 10.6 because the hyperbolic Eisenstein series is defined using the Bessel function on rays; the Bessel function is not a character. We are saved because of the property described in Remark 1.*

To compute the periods of the Eisenstein series over parabolic cycles we note that the N orbit of a point on G/A is a parabola. Using the ideas of Section 10.3(c) above, the computation of these periods entails the replacement of the Bessel factor $K_{ij/\theta^0}(\hat{r}m)$ in the above integral by a quadratic exponential. We can then apply [11, vol. II, Section 7.73] to compute the periods.

We can define hyperbolic Poincaré series by replacing $\delta_{\hat{\mathbf{p}}^0}$ which is $H_{\hat{\mathbf{p}}^0}$ invariant by a character on the $H_{\hat{\mathbf{p}}^0}$ orbit of $\gamma^1\hat{\mathbf{p}}^0 = \hat{\mathbf{p}}^1$ for a suitable $\gamma^1 \in \Gamma$. This character should be $\Gamma_{\hat{\mathbf{p}}^0}$ invariant.

Problem 10.6 *Compute periods of the hyperbolic Poincaré series.*

10.6 The four dimensional representation

Up to this point we have dealt with functions on G/K; these are functions on G which are right K invariant. In this section we drop this K invariance. Any function on G can be decomposed into a Fourier series under right K action since K is compact. In this way we can assume that there is a fixed l for which our functions satisfy

$$f(gk_\theta) = e^{il\theta} f(g).$$

Such an f is called a *function of type l*.

The Iwasawa decomposition says that $G = NAK$ so a function of type l is determined by its restriction to (can be regarded as a function on)

$$NA = \left\{ \begin{pmatrix} 1 & \xi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta^{1/2} & 0 \\ 0 & \eta^{-1/2} \end{pmatrix} \right\} = \left\{ \begin{pmatrix} \eta^{1/2} & \eta^{-1/2}\xi \\ 0 & \eta^{-1/2} \end{pmatrix} \right\} \tag{10.178}$$

as in (10.10). This triangular matrix is identified with the point $z = \xi + i\eta$ in the Poincaré upper half-plane.

Remark. *We restrict our considerations to integral l. (We generally drop factors of 2π.) If l is not an integer we have to pass to the universal covering of $G = SL(2,\mathbb{R})$. Topologically $G = \mathbb{R}^2 \times K$.*

Let $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. We want to make the Iwasawa decomposition explicit. We have

$$\begin{aligned} gi &= \frac{ai+b}{ci+d} \\ &= \frac{i + (bd+ac)}{c^2+d^2}. \end{aligned}$$

Since K is the isotropy group of i the NA part of g is, by (10.178),

$$\begin{pmatrix} (c^2+d^2)^{-1/2} & (c^2+d^2)^{1/2}(bd+ac) \\ 0 & (c^2+d^2)^{1/2} \end{pmatrix}.$$

To find the K component we multiply this on the right by

$$k_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

We obtain

$$(c^2+d^2)^{1/2}\sin\theta = c$$
$$(c^2+d^2)^{1/2}\cos\theta = d$$

or

$$\begin{aligned} ci+d &= (c^2+d^2)^{1/2}e^{i\theta} \\ e^{i\theta} &= \frac{ci+d}{(c^2+d^2)^{1/2}} \\ &= \left(\frac{d+ci}{d-ci}\right)^{1/2} \\ &= e^{i\arg(d+ci)}. \end{aligned} \tag{10.179}$$

These formulas define the Iwasawa decomposition of g. If we start with $g_1 = n_1a_1k_1$ then we can compute the Iwasawa decomposition of gg_1 as follows. To evaluate gg_1i we need to compute

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \eta_1^{1/2} & \eta_1^{-1/2}\xi_1 \\ 0 & \eta_1^{-1/2} \end{pmatrix} = \begin{pmatrix} a\eta_1^{1/2} & a\eta_1^{-1/2}\xi_1 + b\eta_1^{-1/2} \\ c\eta_1^{1/2} & c\eta_1^{-1/2}\xi_1 + d\eta_1^{-1/2} \end{pmatrix}. \tag{10.180}$$

The corresponding point in the Poincaré upper half-plane is

$$gg_1i = \frac{a(\xi_1+i\eta_1)+b}{c(\xi_1+i\eta_1)+d} = \frac{az_1+b}{cz_1+d} = gz_1.$$

By (10.179) and (10.180) the K factor of gn_1a_1 is given by

$$\begin{aligned} \theta &= \arg[(ic\eta_1 + (c\xi_1+d)] \\ &= \arg(cz_1+d). \end{aligned}$$

It follows that if f is a function on G which is of type l which is identified with a function of z then

$$gf(z_1) = f(gz_1)e^{il\arg(cz+d)}. \tag{10.181}$$

In order to put the whole of G (rather than G/K) in a geometric setting we have to deal with the four-dimensional representation. We now think of $\mathbb{R}^4$ as the space of 2×2 matrices, not necessarily symmetric, which we write in the form[13]

$$X = \begin{pmatrix} x & t \\ w & z \end{pmatrix}.$$

[13]The reader should be careful to distinguish between z as a matrix coefficient of X and z as a parameter in the Poincaré upper half-plane.

In contrast to the previous sections G now acts on X by left multiplication. When $\det X^0 \neq 0$ the G orbit of X^0 is the set where $\det X = \det X^0 = c$.

We can identify any such orbit with G. When $c = 0$ the matrix X^0 has rank < 2. The set of matrices of rank 1 can be decomposed into the g orbits

$$O_\alpha = \left\{ \begin{pmatrix} x & \alpha x \\ w & \alpha w \end{pmatrix} \right\}$$

and

$$O_\infty = \left\{ \begin{pmatrix} 0 & x \\ 0 & w \end{pmatrix} \right\}.$$

O_α is the orbit of $\left(\begin{smallmatrix} 1 & \alpha \\ 0 & 0 \end{smallmatrix}\right)$. There are analogous orbits of $\left(\begin{smallmatrix} 0 & 0 \\ 1 & \alpha \end{smallmatrix}\right)$.

The isotropy group of $\left(\begin{smallmatrix} 1 & \alpha \\ 0 & 0 \end{smallmatrix}\right)$ is defined by

$$\begin{pmatrix} 1 & \alpha \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & \alpha \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a & a\alpha \\ c & c\alpha \end{pmatrix}.$$

Thus $a = 1, c = 0$ so the isotropy group is N. The isotropy group of $\left(\begin{smallmatrix} 0 & 0 \\ 1 & \alpha \end{smallmatrix}\right)$ is N^t.

We define the inner product by

$$\begin{pmatrix} x & t \\ w & z \end{pmatrix} \cdot \begin{pmatrix} x' & t' \\ w' & z' \end{pmatrix} = xz' + x'z - tw' - t'w \tag{10.182}$$

so

$$\|X\|^2 = X \cdot X = 2 \det X.$$

Moreover the adjoint g' of $g \in G$ is g^{-1}, i.e.

$$\begin{aligned} gX \cdot X &= X \cdot g^{-1} X \\ Xg \cdot X &= X \cdot X g^{-1} \end{aligned} \tag{10.183}$$

because $\det g = 1$ so g preserves the norm. Thus g is orthogonal for this (indefinite) inner product. We use this inner product to define the Fourier transform.

Note that left and right actions of G commute and preserve the norm. Det X is a nondegenerate quadratic form of signature $(2, 2)$ whose orthogonal group is $O(2, 2)$. Since both left and right $G \cong O(1, 2)$ action, i.e. $O(1, 2) \times O(1, 2)$ action, preserve $\det X$,

$$O(2, 2) \supset O(1, 2) \times O(1, 2). \tag{10.184}$$

In fact the inclusion is an equality.

The set $\{\det X = 1\} = \{gI\}$ is exactly G. If we use the Iwasawa decomposition then we can identify the set with NAK. To obtain the remainder of the set $\{\det X > 0\}$ we multiply $\{gI\}$ by $\mathbb{R}^* = \{\left(\begin{smallmatrix} r & 0 \\ 0 & r \end{smallmatrix}\right)\}_{r>0}$. (The set with $\det X < 0$ can be identified with $\mathbb{R}^* G \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.) The change of coordinates from $\mathbb{R}^* G$ to x, z, t, w is given by

$$\begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix} \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a^{1/2} & 0 \\ 0 & a^{-1/2} \end{pmatrix} \begin{pmatrix} c & s \\ -s & c \end{pmatrix} = \begin{pmatrix} x & t \\ w & z \end{pmatrix}.$$

(We have written $a^{1/2}$ instead of a so that $gi = ai + n$; also $c = \cos\theta, s = \sin\theta$.) Thus

$$\begin{aligned} x &= r(a^{1/2}c - na^{-1/2}s) \\ z &= ra^{-1/2}c \\ t &= r(a^{1/2}s + na^{-1/2}c) \\ w &= -ra^{-1/2}s. \end{aligned} \tag{10.185}$$

We want to express the ultrahyperbolic operator $\Box_4$, which is G invariant, in terms of r, a, n, c. We denote the usual euclidean variables in $\mathbb{R}^4$ by

$$\begin{aligned} \alpha &= x + z \\ \beta &= x - z \\ \gamma &= t + w \\ \delta &= t - w \end{aligned}$$

so that

$$\Box_4 = \frac{\partial^2}{\partial\alpha^2} - \frac{\partial^2}{\partial\beta^2} - \frac{\partial^2}{\partial\gamma^2} + \frac{\partial^2}{\partial\delta^2}.$$

Using (10.185) the Jacobian of the transformation is given by[14]

$$J(r, n, a, \theta) = \begin{pmatrix} c\overset{+}{\alpha} - a^{-1/2}sn & \frac{1}{2}a^{-1}(c\overset{-}{\alpha} + a^{-1/2}sn) & -s\overset{+}{\alpha} - a^{-1/2}cn & -a^{-1/2}s \\ c\overset{-}{\alpha} - a^{-1/2}sn & \frac{1}{2}a^{-1}(c\overset{+}{\alpha} + a^{-1/2}sn) & -s\overset{-}{\alpha} - a^{-1/2}cn & -a^{-1/2}s \\ s\overset{-}{\alpha} + a^{-1/2}cn & \frac{1}{2}a^{-1}(s\overset{+}{\alpha} - a^{-1/2}cn) & c\overset{-}{\alpha} - a^{-1/2}sn & a^{-1/2}c \\ s\overset{+}{\alpha} + a^{-1/2}cn & \frac{1}{2}a^{-1}(s\overset{-}{\alpha} - a^{-1/2}cn) & c\overset{+}{\alpha} - a^{-1/2}sn & a^{-1/2}c \end{pmatrix}.$$

Here we have written $\overset{+}{\alpha} = a^{1/2} + a^{-1/2}$ and $\overset{-}{\alpha} = a^{1/2} - a^{-1/2}$ and we have omitted a factor r from the second, third, and fourth columns. The determinant of the Jacobian is $4r^3a^{-2}$.

[14]The reader should not confuse α with $\overset{+}{\alpha}$ or $\overset{-}{\alpha}$.

The matrix of the signed minors of this Jacobian divided by r^3a^{-2} is

$$\begin{pmatrix} -(c\overset{+}{\alpha} - \frac{1}{2}sn) & 2ar^{-1}(c\overset{-}{\alpha} + \frac{1}{2}sn) & 2ar^{-1}\frac{1}{2}s & -2ar^{-1}(s\overset{-}{\alpha} - \frac{1}{2}cn) \\ c\overset{-}{\alpha} - \frac{1}{2}sn & -2ar^{-1}(c\overset{+}{\alpha} + \frac{1}{2}sn) & -2ar^{-1}\frac{1}{2}s & 2ar^{-1}(s\overset{+}{\alpha} - \frac{1}{2}cn) \\ s\overset{-}{\alpha} + \frac{1}{2}cn & -2ar^{-1}(s\overset{+}{\alpha} - \frac{1}{2}cn) & 2ar^{-1}\frac{1}{2}c & -2ar^{-1}(c\overset{+}{\alpha} + \frac{1}{2}sn) \\ -(s\overset{+}{\alpha} + \frac{1}{2}cn) & 2ar^{-1}(s\overset{-}{\alpha} - \frac{1}{2}cn) & -2ar^{-1}\frac{1}{2}c & 2ar^{-1}(c\overset{+}{\alpha} + \frac{1}{2}sn) \end{pmatrix}.$$

This means that

$$\square_4 = \frac{\partial^2}{\partial r^2} + \frac{3}{r}\frac{\partial}{\partial r} - \frac{4a^2}{r^2}\left(\frac{\partial^2}{\partial a^2} + \frac{\partial^2}{\partial n^2}\right) + \frac{2a}{r^2}\frac{\partial^2}{\partial\theta\partial n}. \tag{10.186}$$

As in our study of the three-dimensional representation we are interested in the result of applying $\square_4$ to functions of the form $U = r^s \exp(il\theta)\exp(i\nu n)W(a)$. We find

$$\square_4 U = \left[s(s-1) + 3s - 4a^2\frac{d^2}{da^2} + 4a^2\nu^2 - 2ia\nu l\right]U.$$

In particular if $\square_4 U = 0$ then

$$\left[\frac{d^2}{da^2} - \nu^2 + i\frac{\nu l}{2a} - \frac{s(s+2)}{4a^2}\right]W = 0. \tag{10.187}$$

A comparison with [11, vol. I, p. 248 (4) and p. 264] shows that W is Whittaker's function $M_{\kappa,\mu}(a)$ if we set $\nu = 1/2$, $\kappa = l/4$ and $\mu = (s+1)/2$ (compare [58, p. 110]).

It is important to observe that when $l = 0$ the Whittaker function reduces (essentially) to a Bessel function [11, vol. I, p. 165 (11)].

Let us back-track a little and examine the equation $\square_4 V = 0$ for functions of the form $V = r^s \exp il\theta Y(n,a)$. By (10.186)

$$\square_4 V = \left[s(s+2) - 4a^2\left(\frac{\partial^2}{\partial a^2} + \frac{\partial^2}{\partial n^2}\right) + 2ial\frac{\partial}{\partial n}\right]V.$$

As we have remarked, $z = ai + n$ is the variable in the Poincaré upper half-plane. Hence $\square_4 V = 0$ means

$$\left[4\eta^2\left(\frac{\partial^2}{\partial\xi^2} + \frac{\partial^2}{\partial\eta^2}\right) - 2il\eta\frac{\partial}{\partial\xi}\right]Y = s(s+2)Y. \tag{10.188}$$

The operator in the left of (10.188) is the *Laplacian of weight l* in the upper half-plane. Thus Y is an eigenfunction of the Laplacian of weight l with eigenvalue $s(s+2)$.

These facts are significant to us because our four-dimensional analogs of Eisenstein and Poincaré series are expressed in terms of homogeneous solutions

of $\square_4$ as in the three-dimensional case. Moreover such functions F transform according to $\exp(il\theta)$ when acted on the right by rotation through an angle θ. (Recall that in the four-dimensional representation G acts both left and right.) This means that F, thought of as a function on $\{r\}G$, is of the form

$$F(r, a, \theta, n) = e^{il\theta} r^s H(a, n).$$

Forming the N periods of F is the same as decomposing F further as an integral (sum) of functions of the form

$$F(r, a, \theta, n) = r^s e^{i\nu n} e^{il\theta} W(a). \tag{10.189}$$

By the above W is essentially a Bessel function when $l = 0$ and is a Whittaker function when $l \neq 0$. The case $l = 0$ can easily be reduced to the three-dimensional representation. This clarifies the appearance of Bessel functions in (10.86), (10.100); the Whittaker functions will appear in calculations which we now present.

We are in a position to construct Eisenstein series for the character χ_l. As usual we start with the N invariant point $\hat{\mathbf{p}}^0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. The Fourier transform of $\delta_{\hat{\mathbf{p}}^0}$ is $\exp iz$.

Since G commutes with scalar multiplication (i.e. $\mathbb{R}^*$) and right multiplication by G, in particular by K, we can decompose $\exp iz$ under $\mathbb{R}^* \times$ (right K). According to the ideas set forth in (10.71)ff., this is essentially equivalent to replacing $\delta_{\hat{\mathbf{p}}^0}$ by $\chi_l \hat{r}^s \delta_{\sigma(\hat{\mathbf{p}}^0 k)}$ which is the δ function on $\mathbb{R}^* \hat{\mathbf{p}}^0 K$ with the measure $\chi_l(k)\hat{r}^s \, dk \, d\hat{r}/\hat{r}$ and then taking the Fourier transform.

We want to evaluate this function on the Poincaré half-plane, meaning

$$X = \begin{pmatrix} \eta^{1/2} & \xi\eta^{-1/2} \\ 0 & \eta^{-1/2} \end{pmatrix}.$$

We have

$$\begin{aligned} \hat{\mathbf{p}}^0 k_{\hat{\theta}} &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos\hat{\theta} & \sin\hat{\theta} \\ -\sin\hat{\theta} & \cos\hat{\theta} \end{pmatrix} \\ &= \begin{pmatrix} \cos\hat{\theta} & \sin\hat{\theta} \\ 0 & 0 \end{pmatrix} \end{aligned}$$

so that for X as above

$$X \cdot \hat{r}\hat{\mathbf{p}}^0 k_{\hat{\theta}} = \hat{r}\eta^{-1/2} \cos\hat{\theta}. \tag{10.190}$$

Thus the Fourier transform of $\hat{r}^s \chi_l(k) \delta_{\mathbb{R}^* \hat{\mathbf{p}}^0 k}$ evaluated on the Poincaré half-plane $= \{z\}$ is

$$\begin{aligned} e_l(z,s) &= \iint \hat{r}^s e^{il\hat{\theta}} e^{-i\hat{r}\eta^{-1/2}\cos\hat{\theta}} \frac{d\hat{r}}{\hat{r}}\, d\hat{\theta} \\ &= A\eta^{s/2} \int \cos^{-s}\hat{\theta} e^{il\hat{\theta}}\, d\hat{\theta} \\ &= A(\Im z)^{s/2} \end{aligned} \tag{10.191}$$

where $A = A_l(s)$ is an elementary function of l, s. (A is a generic symbol, i.e. its value may vary from formula to formula. The integral in the second equation is a β function.)

$e_l(g,s)$ is regarded as a function on G which transforms according to χ_l under right action of K and whose value when $g = z$ is given by (10.191). According to (10.181) and (10.191)

$$\sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \gamma e_l(z,s) = A \sum_{(c,d)=1} \frac{\eta^{s/2}}{|cz+d|^s} e^{-il \arg(cz+d)}. \tag{10.192}$$

We denote the last sum by $E_l(z,s)$. As in the case $l = 0$ we use the notation $\mathbf{E}_l(z,s)$ when we sum over all $(c,d) \neq (0,0)$. We shall deal mostly with $E_l(z,s)$.

Even when $s = l$ the Eisenstein series (10.192) is not holomorphic because of the presence of the factor $\eta^{l/2}$. It is closely related to the holomorphic Eisenstein series

$$\sum \left(\frac{d\gamma z}{dz} \right)^{l/2} = \sum \frac{1}{(cz+d)^l}. \tag{10.193}$$

The difference between (10.192) and (10.193) is that in (10.192) Γ acts by fractional linear transformation on $\Im z$ *plus* the action of K via multiplication by χ_l, while (10.193) involves only fractional linear transformation on z.

Our methods are geared to the study of (10.192) which is more general since we do not need $s = l$.

In the computation of parabolic periods of E (so $\eta = \text{const.}$) the factor $\eta^{s/2}$ is irrelevant.

(a) Parabolic periods

We can compute the j-th Fourier series coefficient of the Eisenstein series using the methods of the previous sections. In the upper half-plane the cycle we use is the line $\eta = \text{const.}$; there is no conceptual difficulty in treating general horocycles.

As is usual in our treatment of parabolic cycles we use a modification of the orthogonality formalism used for hyperbolic cycles. We start with the plane

$$L^0 = \left\{ \begin{pmatrix} 0 & t \\ 0 & z \end{pmatrix} \right\}.$$

We set $X = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. The parabolic periods of $E_l(z,s)$ constitute the Radon transform of the restriction of E to G on the spread $\{L^y\}$ where L^y is the two dimensional affine plane

$$L^y = \left\{ \begin{pmatrix} y & t \\ 0 & z \end{pmatrix} \right\}.$$

We have

$$L^y \cap G = \left\{ \begin{pmatrix} y & t \\ 0 & y^{-1} \end{pmatrix} \right\}.$$

In accordance with our identification of the Poincaré half-plane with triangular matrices $L^y \cap G$ corresponds to the line $\eta = y^2$ in the upper half-plane. (We also have $\xi = yt$.)

Since

$$N \begin{pmatrix} y & 0 \\ 0 & y^{-1} \end{pmatrix} = \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y & 0 \\ 0 & y^{-1} \end{pmatrix} \right\} = \left\{ \begin{pmatrix} y & ny^{-1} \\ 0 & y^{-1} \end{pmatrix} \right\} = L^y \cap G$$

we can use the imaginary axis $\left\{ \left(\begin{smallmatrix} y & 0 \\ 0 & y^{-1} \end{smallmatrix}\right) \right\}$ as the angle base ray (line).

The Fourier transform of $\xi_j(n)\delta_{N\left(\begin{smallmatrix} y & 0 \\ 0 & y^{-1} \end{smallmatrix}\right)}$ where $\xi_j(n) = \exp(2\pi ijn)$ is (ignoring factors of 2π)

$$\widehat{\xi_j \delta}_{N\left(\begin{smallmatrix} y & 0 \\ 0 & y^{-1} \end{smallmatrix}\right)} \begin{pmatrix} \hat{x} & \hat{t} \\ \hat{w} & \hat{z} \end{pmatrix} = \int e^{i(y\hat{z} + y^{-1}\hat{x} - ny^{-1}\hat{w} + jn)}\, dn$$

$$= e^{i(y\hat{z} + y^{-1}\hat{x})} \delta_{j - y^{-1}\hat{w}} \tag{10.194}$$

because the exponential is linear in n. We shall assume $j \neq 0$; we leave the easier case $j = 0$ to the reader.

Equation (10.194) can be understood geometrically in the following terms. For fixed y the set $N\left(\begin{smallmatrix} y & 0 \\ 0 & y^{-1} \end{smallmatrix}\right)$ is the affine line which is parallel to $\{\left(\begin{smallmatrix} 0 & t \\ 0 & 0 \end{smallmatrix}\right)\}$ and translated by $\left(\begin{smallmatrix} y & 0 \\ 0 & y^{-1} \end{smallmatrix}\right)$. Thus the Fourier transform of $\delta_{N\left(\begin{smallmatrix} y & 0 \\ 0 & y^{-1} \end{smallmatrix}\right)}$ is the δ function of the orthogonal hyperplane, i.e. $\hat{w} = 0$, times the factor $\exp i(y\hat{z} + y^{-1}\hat{x})$ which compensates for the translation by $\left(\begin{smallmatrix} y & 0 \\ 0 & y^{-1} \end{smallmatrix}\right)$. The Fourier transform of $\exp(ijn)\delta_{N\left(\begin{smallmatrix} y & 0 \\ 0 & y^{-1} \end{smallmatrix}\right)}$ is obtained from this, essentially, by translating the hyperplane by j. We say "essentially" because the hyperplane $\{\left(\begin{smallmatrix} 0 & t \\ 0 & 0 \end{smallmatrix}\right)\}$ is written in the parametric form $\{\left(\begin{smallmatrix} 0 & ny^{-1} \\ 0 & 0 \end{smallmatrix}\right)\}$ and this accounts for the fact that the δ function in (10.194) is $\delta_{j - y^{-1}\hat{w}}$ rather than $\delta_{j - \hat{w}}$.

The set $\mathbb{R}^*\Gamma\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)K$ is two dimensional; it meets the hyperplane $\hat{w} = jy$ in a one-dimensional set over which we perform an integration via Parseval's formula.

We have to evaluate (10.194) on the ray through

$$\gamma\hat{\mathbf{p}}^0 k_\theta = \begin{pmatrix} p & a \\ q & b \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$
$$= \begin{pmatrix} p\cos\theta & p\sin\theta \\ q\cos\theta & q\sin\theta \end{pmatrix}. \tag{10.195}$$

By (10.195) $\hat{w} = q\cos\theta = yj$ so we have to set $\cos\theta = yj/q$. This determines θ. For this value of θ the exponential in (10.194) times χ_l becomes

$$e^{i(l\theta + qy\sin\theta + py^{-1}\cos\theta)}. \tag{10.196}$$

Next we homogenize, which amounts to replacing p, q by $p\hat{r}, q\hat{r}$, multiplying by $\hat{r}^{is}$, and integrating. We obtain from (10.194) and (10.196)

$$\int e^{i(l\theta + q\hat{r}y\sqrt{1-y^2j^2/q^2\hat{r}^2} + jp/q)} \alpha \hat{r}^{is} \frac{d\hat{r}}{\hat{r}}. \tag{10.197}$$

We have included a factor α which accounts for the angle of intersection of $\hat{w} = yj$ and the support of $\mathbb{R}^*\gamma\hat{\mathbf{p}}^0 K$. Except for simple factors, $\alpha = \hat{r}^{-1/2}$.

Note that the factor $\exp(ijp/q)$, which represents the angular contribution as in the case of the three-dimensional representation, is independent of l. $\sum_{(p,q)=1} \exp(ijp/q)$, for a fixed q, is the Ramanujan sum as before. The integral depends on γ only through q which represents the (parabolic) distance in the plane on which N acts. (This will be more clear when we treat hyperbolic cycles.) q is thus the natural choice for grouping the $\{\gamma\}$.

For the transcendental part, suppose first that $l = 0$. The integral is easily reduced via the substitution $\hat{r}q/jy \to \hat{r}$ to the form

$$\int e^{y^2 j\sqrt{\hat{r}^2-1}} \hat{r}^{is-1/2} \frac{d\hat{r}}{\hat{r}} \tag{10.198}$$

which is a Bessel function [11, vol. II, p. 82]. Checking the other factors brings us in agreement with (10.113).

In case $l \neq 0$ there is an additional factor

$$e^{il\theta} = (\cos\theta + i\sin\theta)^l$$
$$= \left(\frac{jy}{\hat{r}q} + i\sqrt{1 - j^2y^2/\hat{r}^2q^2}\right)^l$$
$$= \left(\frac{jy}{\hat{r}q} + i\frac{jy}{\hat{r}q}\sqrt{\frac{\hat{r}^2q^2}{j^2y^2} - 1}\right)^l. \tag{10.199}$$

We make the natural change of variables $\hat{r}q/jy \to \hat{r}$ as in the case $l = 0$ to replace this term by $\hat{r}^{-l}(1 + i\sqrt{\hat{r}^2 - 1})^l$. We then meet an integral of the form

$$\int \hat{r}^{is-1/2-l}(1 + i\sqrt{\hat{r}^2 - 1})^l e^{y^2 j\sqrt{\hat{r}^2-1}} \frac{d\hat{r}}{\hat{r}}. \tag{10.200}$$

There is an additional factor of $(q/jy)^{1/2-is}$.

To evaluate (10.200) we set $t = \sqrt{\hat{r}^2 - 1}$ to obtain

$$2\int (1+t^2)^{\frac{is}{2}-\frac{2l+1}{4}}(1+it)^l e^{y^2 jt} t\frac{dt}{1+t^2} = 2\int (1+it)^{\frac{is}{2}+\frac{l}{2}-\frac{3}{4}}(1-it)^{\frac{is}{2}-\frac{l}{2}-\frac{5}{4}} e^{y^2 jt} t\, dt.$$

Except for the change from t to it and the factor t this is the integral representation for the Whittaker function (see (10.155) and [11, vol. I, p. 273(9)]. The change from t to it is easily handled by change of contour; if we write $t = i(1 - it) - i$ then we can absorb the factor t in other terms.

We have shown

Theorem 10.13 *The parabolic periods of the Eisenstein series split into a factor depending on* Γ *and a transcendental factor depending on* y, s, l, j *and an elementary factor. The arithmetic factor is the same (with* $2s$ *replaced by* $s - \frac{1}{2}$*) as the corresponding series (10.114) in Minkowski space. The transcendental part is the sum of two terms, each of which is a product of an elementary factor with a Whittaker function.*

(b) Parabolic Poincaré series

As in the case of Minkowski space we modify the foundation of Eisenstein series which is $\hat{\mathbf{p}}^0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ to $\exp(i\nu n)$ on the N orbit of $\hat{\mathbf{p}}^1 = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ which is $\{\left(\begin{smallmatrix} 0 & n \\ 0 & 1 \end{smallmatrix}\right)\}$. Applying $\gamma = \begin{pmatrix} p & a \\ q & b \end{pmatrix}$ on the left and K on the right leads to

$$\gamma n \hat{\mathbf{p}}^1 k_\theta = \begin{pmatrix} 0 & p' \\ 0 & q' \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} -p'\sin\theta & p'\cos\theta \\ -q'\sin\theta & q'\cos\theta \end{pmatrix} \tag{10.201}$$

where, as before, $p' = pn + a, q' = qn + b$. The difference between (10.201) and (10.195) involves the interchange of $\cos\theta$ and $\sin\theta$ and the change from p, q to $p'q'$ (and a minus sign in the first column).

Remark. *We shall discuss below our choice of this orbit of* N*.*

As in the case of Minkowski space the group N plays two roles. In the first place it defines the starting orbit $N\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ for Poincaré series. In the second place it defines the parabolic cycle. We write n for the parameter in the first case and $\hat{n}$ in the second case.

The interchange of $\sin\theta$ and $\cos\theta$ does not produce any serious change in our computation starting with (10.195). The main point is the relation

$\sin^2\theta + \cos^2\theta = 1$. The angular term $\sum \exp(ip/q)$ becomes $\sum \exp(ip'/q')$ and we have an additional factor $\exp(in\nu)$ as in the Minkowski theory. We can repeat the argument in the Minkowski theory essentially verbatim. Equation (10.194) is unaltered (except that n becomes $\hat{n}$) and (10.195) is changed by replacing the matrix in the second formula by (10.201). Thus $\hat{w}$ is now $-q'\sin\theta$ so (10.194) implies

$$-q'\sin\theta = jy.$$

Expression (10.196) now reads

$$e^{i(l\theta+q'y\cos\theta-p'y^{-1}\sin\theta+\nu n)} \tag{10.196$'$}$$

which leads to

$$\int e^{i(l\theta+q'\hat{r}y\sqrt{1-j^2y^2/q'^2\hat{r}^2}+jp'/q'+\nu n)}\hat{r}^{is}\frac{d\hat{r}}{\hat{r}}. \tag{10.197$'$}$$

As in the case of the Minkowski space we write

$$\frac{p'}{q'} = \frac{p}{q} - \frac{1}{(qn+b)q}$$

and make the changes of variables $q' = qn + b \to n$, and $\frac{q'\hat{r}}{jy} \to \hat{r}$. The integral becomes

$$A\iint n^{-is}e^{i(l\theta+jy^2\sqrt{\hat{r}^2-1}-j/qn+\nu n/q)}\hat{r}^{is}\frac{d\hat{r}}{\hat{r}}\,dn$$
$$= AK_{1-is}(q^{-1}\sqrt{j\nu})\int e^{il\theta+jy^2\sqrt{\hat{r}^2-1}}\hat{r}^{is}\frac{d\hat{r}}{\hat{r}} = AWK_{1-is}(q^{-1}\sqrt{j\nu})$$

where W is expressed simply in terms of the Whittaker functions as in (10.202)ff. The Bessel function is independent of l and so agrees with the analogous term (10.147) computed in Minkowski space (when $l = 0$). The sum over Γ becomes a Kloosterman sum exactly as in the three-dimensional case (see (10.148)).

We have established

Theorem 10.14 *The Fourier series coefficients of the parabolic Poincaré series are sums of two terms which are elementary factors times (parabolic) Kloosterman sums times $K_{1-is}(q^{-1}\sqrt{j\nu})$ times Whittaker functions which depend on s, j, l, y but not on γ.*

Remark 1 *We find this splitting (modulo elementary factors) into three parts, one depending only on Γ, one depending only on s, j, ν, q, and the third only on s, j, l, y, to be remarkable.*

Remark 2 *We emphasize that the only place l occurs is in the Whittaker factor.*

We defined the Poincaré series using $\delta_{N\left(\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right)}$ as our primary object. In classical theory one starts with the function f which is $\exp(i\nu z)$ on the upper half-plane and transforms like $\chi_l(k)$ under right K multiplication. Precisely

$$f(g) = e^{i\nu(\xi+i\eta)}e^{il\theta} \tag{10.202}$$

for

$$g = \begin{pmatrix} \eta^{\frac{1}{2}} & \eta^{-\frac{1}{2}}\xi \\ 0 & \eta^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

If we examine (10.202) we observe that f is not an eigenfunction of the G component of the ultrahyperbolic operator $\Box_4$ which is the part of (10.186) not involving $\frac{\partial}{\partial r}$ on G which we identify with $\{r = 1\}$

$$\Box_G = -4\eta^2\left(\frac{\partial^2}{\partial\eta^2} + \frac{\partial^2}{\partial\xi^2}\right) + 2\eta\frac{\partial^2}{\partial\xi\partial\theta}. \tag{10.203}$$

Let us replace $f = \exp(i\nu z)\exp(il\theta)$ by $h = \eta^m f$. We find

$$\begin{aligned}\Box_G h &= -4\eta^2[m(m-1)\eta^{m-2}f - 2m\nu\eta^{m-1}f] - 2l\nu\eta^{m+1}f \\ &= -4m(m-1)h + 2\eta\nu(4m-l)h.\end{aligned} \tag{10.204}$$

h is an eigenfunction of $\Box_G$ if $m = l/4$.

Since by (10.186)

$$\Box_4 r^s h = r^{-2}[s(s+2) + \Box_G]r^s h \tag{10.205}$$

$r^s h$ is a solution of $\Box_4 r^s h = 0$ if $s(s+2) = \frac{l}{2}\left(\frac{l}{2} - 2\right)$, i.e.

$$\begin{aligned} s &= -l/2 \\ s &= l/2 - 2. \end{aligned} \tag{10.206}$$

For those values of s we have seen following (10.186) that $r^s h$ is a solution of the ultrahyperbolic equation which is uniquely determined (more precisely, we have seen that there are two choices for s) by its transformation under the groups N and K.

Our construction of the Poincaré series involved the sum of the transforms under $\{\gamma\}$ of the function F which is the Fourier transform of

$$\hat{F} = \hat{r}^s\delta_{\sigma[\hat{r}N\left(\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right)K]}\chi_l(K)e^{i\nu n}.$$

F satisfies the transformation properties of $r^s h$ and F is a solution of the ultrahyperbolic equation because support $\hat{F} \subset \{\hat{\Box} = 0\}$.

This proves that

$$F = c_1(l,\nu)r^{-l/2}h + c_2(l,\nu)r^{l/2-2}h \tag{10.207}$$

for some constants c_j. We leave the computation of the c_j to the reader.

(c) Hyperbolic periods of parabolic Eisenstein series

To understand the computation let us begin with the hyperbolic group $A = \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \right\}$ which does not meet the hypothesis we shall impose but, nevertheless, shows the basic method of computation.

A leaves the linear space L^0 of diagonal matrices invariant, and hence also its orthogonal complement $L^{0\perp}$ of matrices with vanishing diagonal entries. We denote generic points in $L^0, L^{0\perp}$ respectively by $\mathbf{x}, \mathbf{y}$. The orbits of A on L^0 are defined by the values of the A invariant quadratic form

$$\rho(\mathbf{x}) = A(\mathbf{x}) = xz. \tag{10.208}$$

The A invariant quadratic form on $L^{0\perp}$ is

$$\rho(\mathbf{y}) = \rho_A(\mathbf{y}) = tw. \tag{10.209}$$

$\mathbf{y}$ or $\rho(\mathbf{y})$ are the analogs of the quantity y used in the three-dimensional representation.

Remark. *Although A preserves $L^{0\perp}$ it does not preserve the individual points $\mathbf{y} \in L^{0\perp}$. For this reason $L^{\mathbf{y}} = L^0 + \mathbf{y}$ is not A invariant. We shall return to this point below.*

Using the bilinear form $xz - tw$ we identify $\mathbb{R}^4$ with its dual. Then $\rho(\mathbf{x})$ and $\rho(\mathbf{y})$ define quadratic forms on $\hat{L}^0$ and $\hat{L}^{0\perp}$ and dualities between $L^0, \hat{L}^0$ and $L^{0\perp}, \hat{L}^{0\perp}$. Since the (ultra)light cone Γ is defined by $\hat{x}\hat{z} - \hat{t}\hat{y} = 0$, the present situation is comparable to that of Minkowski space, namely Γ is defined by

$$\rho(\hat{\mathbf{y}}) = \rho(\widehat{\mathbf{x}}) \tag{10.210}$$

(see Proposition 10.1). Moreover $\Omega = G$ is defined by

$$\rho^2(\mathbf{x}) - \rho^2(\mathbf{y}) = 1. \tag{10.211}$$

We are interested in the periods of the Eisenstein series (Radon transform) over the spread $\{L^{\mathbf{y}} \cap \Omega\}$. As usual we have to compute the Fourier transform of $\delta_{L^{\mathbf{y}} \cap \Omega}$

$$\hat{\delta}_{L^{\mathbf{y}} \cap \Omega}(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) = e^{i\mathbf{y}\cdot\hat{\mathbf{y}}} \int e^{i\mathbf{x}\cdot\hat{\mathbf{x}}}\, d\theta \tag{10.212}$$

where the integral is taken over $\rho^2(\mathbf{x}) = 1 + \rho^2(\mathbf{y})$.

We introduce cylindrical coordinates so the inner product takes the form

$$(\mathbf{x}, \mathbf{y}) \cdot (\hat{\mathbf{x}}, \hat{\mathbf{y}}) = -\mathbf{y}\cdot\hat{\mathbf{y}} + \rho(\mathbf{x})\rho(\hat{\mathbf{x}}) \cosh[\theta(\mathbf{x}) - \theta(\hat{\mathbf{x}})]. \tag{10.213}$$

$\theta(\mathbf{x})$ is the usual A hyperbolic angle coordinate of $\mathbf{x}$ defined in the L^0 plane.

In cylindrical coordinates (10.212), more generally adding multiplication by a character $\chi(\theta)$, becomes (up to an elementary factor)

$$\begin{aligned}\widehat{\chi_j \delta_{L^{\mathbf{y}} \cap \Omega}}(\hat{x}, \hat{y}) &= e^{i\mathbf{y}\cdot\hat{\mathbf{y}}} \chi_j(\hat{\theta}) \int e^{i[(1+\rho^2(\mathbf{y}))^{\frac{1}{2}} \rho(\hat{\mathbf{x}}) \cosh\theta(\mathbf{x})]} \chi_j(\theta)\, d\theta \\ &= e^{i\mathbf{y}\cdot\hat{\mathbf{y}}} \chi_j(\hat{\theta}) K_j[i(1+\rho^2(\mathbf{y}))^{\frac{1}{2}} \rho(\hat{\mathbf{x}})] \end{aligned} \tag{10.214}$$

which differs little from the three-dimensional result.

To apply this to the Eisenstein series we set

$$\begin{aligned}\mathbf{p}(\phi) &= \begin{pmatrix} p & a \\ q & b \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \\ &= \begin{pmatrix} p\cos\phi & p\sin\phi \\ q\cos\phi & q\sin\phi \end{pmatrix}. \end{aligned} \tag{10.215}$$

The contribution of this point to the Eisenstein series is (essentially)

$$\begin{aligned}&\widehat{\chi_j(\theta)\delta_{L^{\mathbf{y}}\cap\Omega}} \cdot \int \delta_{\sigma(\mathbf{p}(\phi))} \chi_l(\phi)\, d\phi \\ &= \iint e^{i\hat{r}\mathbf{y}\cdot\mathbf{p}(\phi)} \chi_j[\hat{\theta}(\mathbf{p}(\phi))] \chi_l(\phi) K_j[i(1+\rho^2(\mathbf{y}))^{\frac{1}{2}} \rho(\hat{r}\mathbf{p}(\phi))]\, d\phi\, \hat{r}^{is}\, \frac{d\hat{r}}{\hat{r}}. \end{aligned} \tag{10.216}$$

θ is the hyperbolic angle in the $\mathbf{x}$ plane. If we choose the angle base line to be $\left\{ \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \right\}$ then for $\|\mathbf{x}\| \neq 0$ we have

$$e^{\theta} \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} = \begin{pmatrix} x & 0 \\ 0 & z \end{pmatrix}$$

so

$$e^{\theta(\mathbf{x})} = \left(\frac{x}{z}\right)^{1/2}. \tag{10.217}$$

In particular, by (10.215),

$$e^{\hat{\theta}(\mathbf{p}(\phi))} = \left(\frac{p\cos\phi}{q\sin\phi}\right)^{1/2} \tag{10.218}$$

and

$$\rho^2(\mathbf{p}(\phi)) = pq\sin\phi\cos\phi.$$

With the change of variables $\hat{r}\rho(\mathbf{p}(\phi)) \to \hat{r}$, since $\rho(\mathbf{x}) = \rho(\mathbf{y})$ on Γ, the integral takes the form

$$\begin{aligned}&(pq)^{-is/2}(p/q)^{j/2} \iint (\cos\phi\sin\phi)^{-is/2} e^{i\hat{r}\rho(y)\cosh[\hat{\theta}(\mathbf{p}(\phi))-\theta(\mathbf{y})]} \left[\frac{\cos\phi}{\sin\phi}\right]^{j/2} \\ &\quad \times e^{il\phi} K_j[i\hat{r}(1+\rho^2(\mathbf{y})^{1/2})] \hat{r}^{is}\, d\phi\, \frac{d\hat{r}}{\hat{r}}. \end{aligned} \tag{10.219}$$

As in our previous examples we want to write the integral as a product of a function of γ with a transcendental function of j, l, s. γ occurs in (10.219) in the exponential of $\cosh[\hat{\theta}(\mathbf{p}(\phi)) - \theta(y)]$. Since $\exp(\hat{\theta}(\mathbf{p}(\phi))) = (p\cos\phi/q\sin\phi)^{1/2}$ we could change variables to eliminate p/q from the exponential. But γ would then appear in the terms $\exp(i\phi)$ and $(\cos\phi\sin\phi)^{-is/2}$. We have not succeeded in performing the "separation" of γ.

If $\mathbf{y} = 0$, which is the analog of geodesic periods which were mentioned in our discussion of periods in Minkowski space, then separation already exists in (10.219). There are functions $\lambda \not\equiv 1$ which can replace $\exp(i\mathbf{y}\cdot\hat{\mathbf{y}})$ for which separation is possible. One interesting example of such a function λ is

$$\begin{aligned}\lambda_m(\mathbf{y}, \hat{\mathbf{y}}) &= \int \chi_m(\theta(\mathbf{y}))e^{i\hat{r}\mathbf{y}\cdot\hat{\mathbf{y}}}\, d\theta \\ &= \chi_m[\hat{\theta}(\hat{\mathbf{y}})]K_m[i\hat{r}\rho(\mathbf{y})\rho(\hat{\mathbf{y}})] \end{aligned} \tag{10.220}$$

obtained as in the $\mathbf{x}$ integral; that is, integration over the A orbit of $\mathbf{y}$.

If we replace $\exp(i\mathbf{y}\cdot\hat{\mathbf{y}})$ in (10.215) by $\lambda_m(\mathbf{y}, \hat{\mathbf{y}})$ then since $\rho(\hat{\mathbf{y}}) = \rho(\hat{\mathbf{x}})$ when $(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in \Gamma$, (10.219) simplifies to

$$\begin{aligned}(pq)^{-is/2}(p/q)^{(j+m)/2} \iint K_m[i\hat{r}\rho(\mathbf{y})]K_j[i\hat{r}(1+\rho^2(\mathbf{y}))^{1/2}] \\ \times (\sin\phi\cos\phi)^{-is/2}(\tan\phi)^{(j+m)/2}e^{il\phi}\, d\phi\hat{r}^{is}\,\frac{d\hat{r}}{\hat{r}}. \end{aligned} \tag{10.221}$$

As we have seen above the $\hat{r}$ integral can be expressed in terms of the ${}_2F_1$ hypergeometric function. The ϕ integral is a β function. The function (10.221) has a meromorphic extension to the whole plane.

All these calculations were based on the hyperbolic group A. Any other hyperbolic group H is of the form gAg^{-1} for some $g \in G$. The matrix units, which constitute a basis for $\mathbb{R}^4$ formed by eigenvectors of A, become eigenvectors of H. These eigenvectors lie on the light cone. They are not orthogonal but rather they come in pairs e_1, e_2, e_3, e_4 with $e_1 \cdot e_2 = 1$, $e_3 \cdot e_4 = -1$, and all other inner products are 0.

The planes $L^0, L^{0\perp}$ are transformed into planes L_H^0 spanned by e_1, e_2, and $L_H^{0\perp}$ spanned by e_3, e_4. In terms of x, z coordinates the inner product on L_H^0 is a quadratic form $Q_H(x, z) = \rho^2(\mathbf{x})$. Similarly on $L_H^{0\perp}$ we have a quadratic form $Q_H^\perp(t, u) = \rho_H(\mathbf{y})$. Both quadratic forms are indefinite. In conformity with our usual notation we also write $\rho_H^2 = Q_H$, $\rho_{H^\perp}^2 = Q_H^\perp$. In (10.222) pq becomes $Q_H(\mathbf{p})$ and p/q is $\exp(\theta_H(\mathbf{p}))$.

As in our work on Minkowski space there is no problem in deriving from (10.222)

Theorem 10.15 *The hyperbolic periods of the Eisenstein series over the cycle defined by H and the characters $\chi_j(\mathbf{x}), \chi_m(\mathbf{y})$ are given by (compare (10.102))*

$$\zeta[H, (j+m)/2, is/2] \tag{10.222}$$

times the integral in (10.221) where $\rho(\mathbf{y})$ is to be interpreted as $\rho_H(\mathbf{y})$.

Remark. *As we have noted in other cases the γ contribution is the same as for Minkowski space; it does not depend on l.*

(d) Hyperbolic periods of Poincaré series

As usual, going from Eisenstein to Poincaré series entails changing p, q to $p' = pn + a$, $q' = qn + b$, multiplying by $\exp(i\nu n)$ and integrating over n. The $\{\gamma\}$ and N terms appear exactly as in the three-dimensional case so that the γ factor for fixed q is the same hyperbolic Kloosterman sum (10.166) except for the multiplier. The N term is independent of l, θ and so gives rise to the same Whittaker function as in (10.164), with the changes $s \to s/2, j \to (j+m)/2$. We leave to the reader the explicit computation of the transcendental factor when $\mathbf{y} = 0$.

10.7 Higher dimensional groups

To illustrate what can happen in higher dimensions let us replace $SL(2, \mathbb{R})$ by the *symplectic group* $G = Sp(m, \mathbb{R})$. This is the group of 2×2 blocks of $m \times m$ matrices $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ leaving the matrix $J = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ invariant. Thus

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} a' & c' \\ b' & d' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{10.223}$$

which means

$$\begin{aligned} ad' - bc' &= 1 \\ ab' &\quad \text{symmetric} \\ cd' &\quad \text{symmetric.} \end{aligned} \tag{10.224}$$

If $g \in Sp(n)$ so is g'. This follows from the fact that the Lie algebra of G is invariant under transpose since it consists of matrices $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ with $\beta = \beta', \gamma = \gamma'$, $\alpha = -\delta'$; these conditions are clearly invariant under $g \to g'$.

G acts by "fractional linear transformation" on the Siegel upper half-space consisting of symmetric m by m matrices $Z = X + iY$ with Y positive definite:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} Z = (aZ + b)(cZ + d)^{-1}.$$

We distinguish the subgroups

$$\begin{aligned} A &= \left\{ \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \right\} \\ N &= \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \right\} \\ K &= \left\{ \begin{pmatrix} C & S \\ -S & C \end{pmatrix} \right\} = \text{isotropy of } iI; \end{aligned} \tag{10.225}$$

(10.224) and (10.225) imply that n is symmetric and $ad' = 1$.

Let us show that the two expressions for K are the same. It is clear that any $g = \left(\begin{smallmatrix} C & S \\ -S & C \end{smallmatrix}\right)$ lies in the isotropy group of $CC' = iI$. Conversely, let $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ leave iI fixed. Then

$$ai + b = -c + di$$

which yields $a = d, b = -c$, and hence (10.225).

Thus the upper half-plane can be identified with G/K. As in case $m = 1$ it can be identified with the set of matrices

$$\left\{ \begin{pmatrix} \eta^{\frac{1}{2}} & \xi(\eta')^{-\frac{1}{2}} \\ 0 & (\eta')^{-\frac{1}{2}} \end{pmatrix} \right\} = NA \tag{10.226}$$

with ξ symmetric.

For a detailed study of symplectic geometry see [144].

We have

$$\begin{aligned} \dim G &= 2m^2 + m \\ \dim A &= m^2 \\ \dim K &= m^2 \\ \dim N &= \frac{m(m+1)}{2}. \end{aligned} \tag{10.227}$$

(This is seen most simply from the Lie algebra.) We also denote by Γ the *symplectic modular group* which consists of $g \in G$ whose matrix coefficients are integral.

The analog of Minkowski space is the space of $2m \times 2m$ symmetric matrices given in block form $\mathbf{x} = \left(\begin{smallmatrix} x & w \\ w' & z \end{smallmatrix}\right)$ (dimension $2m^2 + m$) with x, z symmetric and with the G action

$$\mathbf{x} \to g\mathbf{x}g'.$$

Since $\det g = 1$ the determinant of $\mathbf{x}$ is preserved. But $\det \mathbf{x}$ is a polynomial of degree $2m$ in the matrix coefficients and so cannot be used to define Fourier

analysis—there is no associated bilinear form. However, we showed (see Remark following (10.3)) that

$$\begin{aligned}\mathbf{x}\cdot\mathbf{x} &= \tfrac{1}{2}\operatorname{tr}[J\mathbf{x}J\mathbf{x}]\\ &= \operatorname{tr}(w^2 - xz)\end{aligned} \tag{10.228}$$

is a nondegenerate G invariant quadratic form. We use this quadratic form to define the Fourier transform.

K is the isotropy group of I so by (10.226) we can identify the upper half-plane with

$$\begin{aligned}\Omega &= \left\{ \begin{pmatrix} \eta^{\frac{1}{2}} & \xi(\eta')^{\frac{-1}{2}} \\ 0 & (\eta')^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} (\eta')^{\frac{1}{2}} & 0 \\ \eta^{-1/2}\xi & \eta^{-\frac{1}{2}} \end{pmatrix} \right\}\\ &= \left\{ \begin{pmatrix} \eta^{\frac{1}{2}}(\eta')^{\frac{1}{2}} + \xi(\eta')^{-\frac{1}{2}}\eta^{-\frac{1}{2}}\xi & \xi(\eta')^{-\frac{1}{2}}\eta^{-\frac{1}{2}} \\ (\eta')^{-\frac{1}{2}}\eta^{-\frac{1}{2}}\xi & (\eta')^{-\frac{1}{2}}\eta^{-\frac{1}{2}} \end{pmatrix} \right\}\end{aligned} \tag{10.229}$$

(compare (10.11)).

The light cone Γ is defined as the orbit of $\left(\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right)$, i.e.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a' & c' \\ b' & d' \end{pmatrix} = \begin{pmatrix} aa' & ac' \\ ca' & cc' \end{pmatrix}. \tag{10.230}$$

This actually corresponds to the positive light cone when $n = 1$ but it is all that we need. It is important to observe that $\|X\| = 0$ for $X \in \Gamma$. Actually the stronger result

$$aa'cc' = ac'ac'$$

is valid because $a'c = c'a$ by (10.224) and our remark that G is invariant under transpose.

In particular

$$\Gamma \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \left\{ \gamma \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right\} = \{\mathbf{p}\}$$

can be thought of as the set of lattice points on the light cone.

We form the symplectic Eisenstein series $E(Z, s)$ in the usual way. That is, we sum the Fourier transforms of $\hat{r}^{is}\delta_{\sigma(\mathbf{p})}$ with the measure $d\hat{r}/\hat{r}$ on each ray and then restrict this sum to G/K. $\Gamma_\infty = \Gamma \cap N$ is the subgroup of Γ leaving $\left(\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right)$ fixed so we sum over Γ/Γ_∞.

To find the periods of E over parabolic cycles we first compute the Fourier transform of E over the cycle defined by $Y = $ const. on the upper half-plane. (Actually, this is nonconvergent; it is defined by a suitable analytic continuation. As usual, we compute the actual periods of E by applying the unfolding lemma

to reduce the calculations to that of integrals over such noncompact cycles.) As in (10.38) this corresponds to the intersection of G/K with the N fixed plane

$$L_X^y = \left\{ \mathbf{x} = \begin{pmatrix} x & w \\ w' & y \end{pmatrix} \right\}$$

with $y = Y = (\eta')^{-1/2}\eta^{-1/2}$ by (10.229). The angle base point on G/K is $\left(\begin{smallmatrix} y^{-1} & 0 \\ 0 & y \end{smallmatrix}\right)$ so the orbit is the set of points

$$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{-1} & 0 \\ 0 & y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix} = \begin{pmatrix} y^{-1} + nyn & ny \\ yn & y \end{pmatrix}.$$

We write $\hat{\mathbf{x}} = \left(\begin{smallmatrix} \hat{y} & \hat{w} \\ \hat{w}' & \hat{x} \end{smallmatrix}\right)$ and use the inner product (10.228). The analog of (10.108) is

$$\hat{\delta}_{L_X^y \cap \Omega}(\mathbf{x}) = \int \exp \operatorname{tr} -[2\hat{w}ny - \hat{y}y - (y^{-1} + nyn)\hat{x}]\, dn. \tag{10.231}$$

The exponential is quadratic in n. In order to bring it into manageable form we "complete the square" as in (10.108)ff. This means we replace n by $n + n_0$ where $n_0 \in \mathbb{N}$ is chosen to remove the linear term in n. Thus n_0 satisfies

$$\operatorname{tr}[-n_0yn\hat{x} - nyn_0\hat{x} + 2\hat{w}ny] = 0 \tag{10.232}$$

for all n. To solve this equation we use the fact that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$. Thus we can replace (10.232) by

$$\operatorname{tr}[-\hat{x}n_0yn - n_0\hat{x}ny + 2\hat{w}ny] = 0. \tag{10.233}$$

We do not know how to solve (10.123) in general. But suppose y is a nonzero scalar multiple of the identity, and hence commutes with all n so we can cancel y. We are left with the equation

$$\operatorname{tr}[\hat{x}n_0n + n_0\hat{x}n - 2\hat{w}n] = 0. \tag{10.234}$$

This equation is to hold for all symmetric n, which means that

$$\hat{x}n_0 + n_0\hat{x} - 2\hat{w} \qquad \text{skew symmetric.} \tag{10.235}$$

In order for the computation of (10.108) to continue we need the analog of the vanishing of $\|\hat{\mathbf{x}}\|$ in (10.109). In (10.109) $\|\hat{\mathbf{x}}\|^2$ arises from the terms $y\hat{y} + n_0^2\hat{x}y - 2n_0y\hat{w}$ since $n_0 = \hat{w}\hat{x}^{-1}$ there. Since we have assumed that y is a nonzero scalar, that vanishing becomes

$$\operatorname{tr}[2n_0\hat{w} - \hat{y} - n_0^2\hat{x}] = 0. \tag{10.236}$$

Since $\hat{x}$ is symmetric we can diagonalize $\hat{x}$. We shall make the calculation assuming $\hat{x}$ is diagonal with entries λ_α; there is no difficulty in dealing with the case of arbitrary symmetric $\hat{x}$. One solution to (10.235) is

$$\begin{aligned}(n_0)_{\alpha\beta} &= \frac{2\hat{w}_{\alpha\beta}}{\lambda_\beta - \lambda_\alpha} \\ &= (\hat{w}_\Lambda)_{\alpha\beta} \end{aligned} \tag{10.237}$$

say. This is meaningful when $\lambda_\alpha - \lambda_\beta \neq 0$ for all α, β. $\hat{w}_\Lambda$ is the only skew-symmetric solution of (10.235). For any other solution would be of the form $\hat{w}_\Lambda + B$ where $\hat{x}B + B\hat{x}$ is skew symmetric which implies that B is skew symmetric.

We complete the square in (10.231) by replacing n by $n + n_0$. In accordance with (10.110) we have to calculate

$$\iint \exp i \operatorname{tr}[2\hat{w}n_0 y\hat{r} - yn^2\hat{x}\hat{r} - yn_0^2\hat{x}\hat{r} - y\hat{y}\hat{r} - y^{-1}\hat{x}\hat{r}]\hat{r}^{is}\, dn\, \frac{d\hat{r}}{\hat{r}}. \tag{10.238}$$

The n integral can be evaluated formally because it is the exponential of a quadratic form (recall $\hat{x}$ and y are diagonal) so we obtain

$$\begin{aligned} & A\frac{1}{\sqrt{\det y\hat{x}}} \int \exp i \operatorname{tr}[(2\hat{w}n_0 y - yn_0^2\hat{x} - y\hat{y} - y^{-1}\hat{x})\hat{r}]\hat{r}^{is-n/2}\, \frac{d\hat{r}}{\hat{r}} \\ &= \frac{A}{\sqrt{\det y\hat{x}}}\{\operatorname{tr}[2\hat{w}n_0 y - yn_0^2\hat{x} - y\hat{y} - y^{-1}\hat{x}]\}^{-is+n/2} \\ &= \frac{A}{\sqrt{\det y\hat{x}}}\{\operatorname{tr}(-y^{-1}\hat{x})\}^{-is+n/2} \end{aligned} \tag{10.239}$$

by (10.236).

A involves the Γ and elementary functions. (Recall that the symbol A may have various connotations.)

The period of E is obtained by summing this over $\Gamma_\infty \setminus \Gamma / \Gamma_\infty$.

Remark. *There are similar formulas when $\lambda_i - \lambda_j = 0$ for some pairs λ_i, λ_j.*

We have calculated the 0 Fourier coefficient of E. For the other Fourier coefficients we "attenuate" the measure dn to $\exp(ij \cdot n)dn$ where j is a symmetric matrix. This has the effect of replacing $\hat{w}$ by $\hat{w}^j = \hat{w} + \frac{1}{2}jy^{-1}$ in (10.231). Equation (10.237) becomes

$$(n_0^j)_{\alpha\beta} = (\hat{w}_\Lambda^j)_{\alpha\beta} = (\hat{w}_\Lambda)_{\alpha\beta} + \frac{1}{2}y^{-1}\frac{j_{\alpha\beta}}{\lambda_\beta - \lambda_\alpha}. \tag{10.240}$$

One essential difference between n_0^0 and n_0^j is that by (10.234) n_0^0 is homogeneous of degree 0 in $\hat{r}$ on $\sigma(\mathbf{p})$ whereas by (10.240) n_0^j contains two degrees of homogeneity (0 and -1) when $j \neq 0$, since $\hat{w}^0$ is homogeneous of degree 1 while

$\hat{w}^j$ is a sum of terms of degrees 1,0. Moreover $\lambda_\alpha, \lambda_\beta$ are the diagonal coefficients of $\hat{x}$ which is homogeneous of degree $+1$; this makes $\hat{w}_\Lambda$ a sum of terms of homogeneity degrees $0, -1$. We shall see that this changes the Γ function, which is the main term in the A of (10.239), into a Bessel function.

We proceed as in the case $j = 0$ to (10.238) with $\hat{w}$ replaced by $\hat{w}^j$ and n_0 by n_0^j. However, the exponential in (10.238) must be modified because $\hat{w}^j$ is a sum of terms of homogeneity 1 and 0 and n_0^j is a sum of terms of degrees $0, -1$. The terms of degree 0 in $\hat{w}$ and in $(n_0^j)^2\hat{x}$ can be taken outside the integral. There remain terms of degree 1 in $\hat{w}$ and $(n_0^j)^2\hat{x}$ (as in case $j = 0$) and a term of degree -1 in $(n_0^j)^2\hat{x}$. This changes the argument of $\exp i\,\mathrm{tr}[\ \ ]$ to be a sum of homogeneous terms of degrees $1, -1$. We call this integral a *symplectic Bessel function* K_s.

To make things precise, we write

$$\hat{w}^j = \hat{w} + \frac{1}{2}jy^{-1}$$
$$n_0^j = n_0 + \tilde{n}_0^j.$$

$\hat{w}$ is homogeneous of degree 1 on $\sigma(\hat{\mathbf{x}})$, jy^{-1} and n_0 are homogeneous of degree 0, and $\tilde{n}_0^j$ is homogeneous of degree -1. On $\sigma(\hat{\mathbf{x}})$ the argument of the exponential in (10.238) is

$$\begin{aligned}&\mathrm{tr}[(2\hat{w}n_0y - yn^2\hat{x} - yn_0^2\hat{x} - y\hat{y} - y^{-1}\hat{x})\hat{r} + (j\tilde{n}_0^j - y(\tilde{n}_0^j)^2\hat{x})\hat{r}^{-1}\\ &\quad + \hat{w}\tilde{n}_0^j + jy^{-1}n_0 - 2yn_0\tilde{n}_0^j\hat{x}]\\ &\qquad = \mathrm{tr}[(-yn^2\hat{x} - y^{-1}\hat{x})\hat{r} + (j\tilde{n}_0^j - y(\tilde{n}_0^j)^2\hat{x})\hat{r}^{-1} + \hat{w}\tilde{n}_0^j + jy^{-1}n_0 - 2yn_0\tilde{n}_0^j\hat{x}]\end{aligned} \tag{10.241}$$

by (10.236). We integrate with respect to n; this leads to $[\det(y\hat{x})]^{-1/2}$ as before. We are left with

$$\begin{aligned}&e^{i\,\mathrm{tr}[\hat{w}\tilde{n}_0^j + jy^{-1}n_0 - 2yn_0\tilde{n}_0^j\hat{x}]}\int e^{i\,\mathrm{tr}[-y^{-1}\hat{x}\hat{r} + (j\tilde{n}_0^j - y(\tilde{n}_0^j)^2\hat{x})]\hat{r}^{-1}}\hat{r}^{is}\frac{d\hat{r}}{\hat{r}}\\ &\qquad = e^{i\,\mathrm{tr}[\hat{w}\tilde{n}_0^j + jy^{-1}n_0 - 2yn_0\tilde{n}_0^j\hat{x}]}K_s[-y^{-1}\hat{x}; j\tilde{n}_0^j - y(\tilde{n}_0^j)^2\hat{x}]\end{aligned} \tag{10.242}$$

where K_s is defined by the formula (10.242).

Equation (10.242) represents the contribution from a single point $\hat{x}$ to the parabolic period of E. To obtain the actual period of E we sum over representative lattice points γ for the double coset $\Gamma_\infty \setminus \Gamma/\Gamma_\infty$ in accordance with the unfolding lemma.

Note that the argument of K_s depends on γ only through $\hat{x}$.

Remark. *We could form Poincaré series. But the formulas do not seem to allow a simple interpretation.*

Let us now pass to hyperbolic periods. There are two natural choices for hyperbolic cycles

(1) *Cartan groups* $\mathcal{C}$. These are conjugates of the subgroup $\mathcal{C} = \left\{ \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix} \right\}$ where α is diagonal.
(2) *Geodesics.* These are the orbits of one-parameter subgroups of conjugates of $\mathcal{C}$ (see [144]).

We shall treat (1) and leave (2), which is somewhat simpler, to the reader.

The polynomial invariants of $\mathcal{C}$ on

$$\left\{ \mathbf{x} = \begin{pmatrix} a & b \\ b' & d \end{pmatrix} \right\},$$

using the $m^2 + m$ matrix coefficients $(a_{ij})_{i \le j}, (b_{ij}), (d_{ij})_{i \le j}$ as coordinates, are generated by

$$\{a_{ij} d_{ij}\}, \{b_{ij} b_{ji}\}_{i \ne j}, \{b_{ii}\}, \{a_{ij} b_{ij} d_{ii}\}_{i \ne j}. \tag{10.243}$$

The number of generators is $2m^2$. It follows that the orbits of $\mathcal{C}$, which are of dimension m, are complete intersections. Thus the harmonic function theory of Chapter 3 applies.

From (10.243) we construct the nondegenerate $\mathcal{C}$ invariant quadratic form

$$\operatorname{tr}(ad' - bb') \tag{10.244}$$

(a and d are symmetric) which we use to define Fourier transform.

We shall compute the periods of E over $L^0 \cap \Omega$ where

$$L^0 = \left\{ \begin{pmatrix} u & 0 \\ 0 & v \end{pmatrix} \right\} \quad u, v \text{ diagonal.}$$

$L^{0\perp}$ consists of all symmetric matrices y with vanishing diagonal elements. As usual L^y is the translate of L^0 by $y \in L^{0\perp}$. Formula (10.229) shows that $\xi = 0$ on $L^0 \cap \Omega$ so $L^0 \cap \Omega = \{ \left(\begin{smallmatrix} u & 0 \\ 0 & u^{-1} \end{smallmatrix} \right) \}$ with $u_j > 0$ for all j.

Distance on L^y is defined by the quadratic form (10.244). For the angle base "line" on L^0 (which is n dimensional) we choose $\{ \left(\begin{smallmatrix} w & 0 \\ 0 & w \end{smallmatrix} \right) \}$. It is clear that the orbit of $\mathcal{C}$ on the angle base line is all of L^0. Moreover for $h \in \mathcal{C}$, $h \ne I$ we cannot have

$$h \begin{pmatrix} w & 0 \\ 0 & w \end{pmatrix} = h \begin{pmatrix} w' & 0 \\ 0 & w' \end{pmatrix}$$

unless $w = w'$ or some $w_j = 0$. Those w with some $w_j = 0$ correspond to degenerate orbits of $\mathcal{C}$; we shall only study generic orbits.

The angle $\theta = \theta_1, \ldots, \theta_n$ of a point $\left(\begin{smallmatrix} u & 0 \\ 0 & v \end{smallmatrix} \right) \in L^0$ is defined by

$$\begin{pmatrix} e^{\theta} & 0 \\ 0 & e^{-\theta} \end{pmatrix} \begin{pmatrix} w & 0 \\ 0 & w \end{pmatrix} \begin{pmatrix} e^{\theta} & 0 \\ 0 & e^{-\theta} \end{pmatrix} = \begin{pmatrix} u & 0 \\ 0 & v \end{pmatrix} \tag{10.245}$$

so that

$$\begin{aligned} w^2 &= uv \\ e^{2\theta} &= uw^{-1}. \end{aligned} \tag{10.246}$$

Here w and $\exp(\theta)$ are diagonal.

For a point

$$\mathbf{p} = \begin{pmatrix} pp' & pq' \\ qp' & qq' \end{pmatrix} \in \Gamma$$

the distance and angle in L^y are defined by subtracting y. Hence they are given by

$$\rho^2(\mathbf{p}) = \rho^2 \begin{pmatrix} (pp')_\alpha & 0 \\ 0 & (qq')_\alpha \end{pmatrix} = \operatorname{tr}[(pp')_\alpha (qq')_\alpha] \tag{10.247}$$

$$\begin{aligned} e^{\theta(\mathbf{p})} &= (pp')_\alpha [(pp')_\alpha (qq')_\alpha]^{-1/2} \\ &= (pp')_\alpha^{1/2} (qq')_\alpha^{-1/2}. \end{aligned} \tag{10.248}$$

We have written U_α for the diagonal of U.

On combining (10.246) and (10.247) we find the inner product of $\mathbf{x} \in L^0$, $\hat{\mathbf{p}} \in \Gamma$

$$\begin{aligned} \mathbf{x} \cdot \hat{\mathbf{p}} &= \begin{pmatrix} u & 0 \\ 0 & v \end{pmatrix} \cdot \begin{pmatrix} pp' & pq' \\ qp' & qq' \end{pmatrix} \\ &= \operatorname{tr}(uqq' + vpp'). \end{aligned} \tag{10.249}$$

One might think it is natural at this point to express the last line of (10.249) in terms of $\rho(u, v), \rho(\hat{\mathbf{p}})$, and an angle between (u, v) and $\hat{\mathbf{p}}$. But such an angle is not readily expressible in terms of $\theta, \hat{\theta}$ and so does not reflect the structure of $\mathcal{C}$.

It seems preferable to express the last sum in terms of components (u_j, v_j) and $\{(pp')_\alpha]_j, [(qq')_\alpha]_j\}$. We then have a product structure so the period of E can be expressed in terms of the m factors. We leave the details to the reader. (This works for $y = 0$; for $y \neq 0$ new ideas are needed.)

Besides the obvious problems of generalizing our results to other groups (in particular to Hilbert modular groups) we are interested in

Problem 10.7 *Find p-adic analogs of E and its periods.*

Some light on this problem is shed in Section 7.3.

We expect that suitable p-adic (and adelic) results would shed new light on many problems involving the arithmetic of quadratic forms.

BIBLIOGRAPHY

[1] M. Agranovsky and J. Globevnik, Analyticity on circles for rational and real-analytic functions of two variables, to appear.

[2] M. Agranovsky and E. T. Quinto, Injectivity of the spherical mean operator and related problems, *Pitman Res. Notes Math.* **347** (1996), 12–36.

[3] M. Agranovsky and E. T. Quinto, Injectivity sets for the Radon transform over circles and complete systems of radial functions, *J. Funct. Anal.* **139** (1996), 384–414.

[4] V. Aguilar, L. Ehrenpreis, and P. Kuchment, Range conditions for the exponential Radon transform, *J. Anal. Math.* **68** (1996), 1–13.

[5] V. Aguilar and P. Kuchment, Range conditions for the multidimensional exponential x-ray transform, *Inverse Prob.* **11** (1995), 5, 977–982.

[6] N. I. Ahiezer and L. I. Ronkin, Separately analytic functions of several variables, and "edge of the wedge" theorems (in Russian), *Usp. Math. Nauk* **28** (1973), 3 (171), 27–42.

[7] W. N. Bailey, *Generalized Hypergeometric Series*, Stechert–Mafner Service Agency, New York and London, 1964.

[8] T. Banh, Holomorphic extendibility of functions via nonlinear Fourier transforms, in *Analysis, geometry, number theory: the mathematics of Leon Ehrenpreis, Contemp. Math. 251*, American Mathematical Society, Providence, RI, 2000.

[9] V. Bargmann, On a Hilbert space of analytic functions and an associated integral transform, *Commun. Pure Appl. Math.* **14** (1961), 187–214; Part II. A family of related function spaces. Application to distribution theory, *Commun. Pure Appl. Math.* **20** (1967), 1–101.

[10] V. Bargmann, Irreducible unitary representations of the Lorentz group, *Ann. Math.* **48** (1947), 2, 568–640.

[11] Bateman Manuscript Project, *Higher Transcendental Functions*, vols. 1, 2, 3, New York, 1953.

[12] C. Berenstein, An inverse spectral theorem and its relation to the Pompeiu problem, *J. Anal. Math.* **37** (1980), 128–144.

[13] C. Berenstein, The Pompeiu problem, what's new?, in *Complex analysis, harmonic analysis and applications (Bordeaux, 1995), Pitman Res. Notes Math. Ser.* **347** (1995), 1–11.

[14] C. Berenstein and E. Casadio-Tarabusi, Inversion formulas for the k-dimensional Radon transform in real hyperbolic spaces, *Duke Math. J.* **62** (1991), 613–631.

[15] C. Berenstein and E. Casadio-Tarabusi, Radon and Riesz transform in real hyperbolic spaces, *Contemp. Math.* **140** (1992), 1–18.

[16] C. Berenstein and L. Zalcman, Pompeiu's problem on spaces of constant curvature, *J. Anal. Math.* **30** (1976), 113–130.

[17] C. Berenstein and L. Zalcman, Pompeiu's problem on symmetric spaces, *Comment. Math. Helv.* **55** (1980), 593–621.

[18] S. Bernstein, *Sur l'ordre de la meilleure approximation des fonctions continues par des polynomes de degree donne*, Bruxelles, 1912.

[19] A. Beurling, Analytic continuation across a linear boundary, *Acta Math.* **128** (1972), 153–182.

[20] A. Besicovitch, *Almost Periodic Functions*, Dover, New York, 1955.

[21] L. C. Biedenharn and J. D. Louck, *The Racah Wigner Algebra in Quantum Theory*, in *Encyclopedia of Mathematics and its Applications* 9, Addison-Wesley, Reading, MA, 1981.

[22] A. Boggess, *CR Manifolds and the Tangential Cauchy–Riemann Complex*, CRC Press, Boca Raton, FL, 1991.

[23] E. D. Bolker, E. Grinberg, and J. Kung, Admissible complexes for the combinatorial Radon transform. A progress report. Integral geometry and tomography (Arcata, CA, 1989), *Contemp. Math.* 113, American Mathematical Society, Providence, RI, 1990, 1–3.

[24] R. Brauer and J. Tate, On the characters of finite groups, *Ann. Math.* **62** (1955), 2, 1–7.

[25] J. Bros and D. Iagolnitzer, Causality and local analyticity: mathematical study, *Ann. Inst. H. Poincaré Sect. A. (NS)* **18** (1973), 147–184.

[26] É. Cartan, *Œuvres Complètes, Partie II. Algèbre, systèmes différentiels et problèmes d'équivalence*, second edition, Editions du Centre National de la Recherche Scientifique (CNRS), Paris, 1984.

[27] S. S. Chern, On integral geometry in Klein spaces, *Ann. Math.* **43** (1942), 178–189.

[28] A. M. Cormack and E. T. Quinto, A Radon transform on spheres through the origin in $\mathbb{R}^n$ and applications to the Darboux equation, *Trans. Am. Math. Soc.* **260** (1980), 575–581.

[29] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, vols. I, II, Interscience, New York, 1953.

[30] J. Delsarte, Théorie des fonctions moyenne-périodiques de deux variables, *Ann. Math.* **72** (1960), 2, 121–178.

[31] P. A. M. Dirac, *The Principles of Quantum Mechanics*, third edition, Clarendon Press, Oxford, 1947.

[32] M. Eastwood, Introduction to Penrose transform. The Penrose transform and analytic cohomology in representation theory (South Hadley, MA,

1992), *Contemp. Math.* **154**, American Mathematical Society, Providence, RI, 1993, 71–75.

[33] L. Ehrenpreis, *Fourier Analysis in Several Complex Variables*, Wiley–Interscience, New York, 1970.

[34] L. Ehrenpreis, Solution of some problems of division I, *Am. J. Math.* **76** (1954), 883–903.

[35] L. Ehrenpreis, Solutions of some problems of division II, *Am. J. Math.* **77** (1955), 286–292.

[36] L. Ehrenpreis, Solutions of some problems of division III, *Am. J. Math.* **78** (1956), 685–715.

[37] L. Ehrenpreis, Solutions of some problems of division IV, *Am. J. Math.* **82** (1960), 522–588.

[38] L. Ehrenpreis, Solutions of some problems of division V, *Am. J. Math.* **84** (1962), 324–348.

[39] L. Ehrenpreis, Mean Periodic Functions I, *Am. J. Math.* **77** (1955), 293–328; also appendix in same journal, 731–733.

[40] L. Ehrenpreis, Theory of infinite derivatives, *Am. J. Math* **81** (1959), 799–845.

[41] L. Ehrenpreis, Analytic functions and the Fourier transform of distributions I, *Ann. Math.* **63** (1956), 945–946.

[42] L. Ehrenpreis, Analytic functions and the Fourier transform of distributions II, *Trans. Am. Math. Soc.* **89** (1958), 450–483.

[43] L. Ehrenpreis, Analytically uniform spaces and some applications, *Trans. Am. Math. Soc.* **101** (1961), 52–74.

[44] L. Ehrenpreis, Cauchy's problem for linear parial differential equations with constant coefficients, *Proc. Natl Acad. Sci. USA* **42** (1956), 642–646.

[45] L. Ehrenpreis, General theory of elliptic equations, *Proc. Natl Acad. Sci. USA* **42** (1956), 39–41.

[46] L. Ehrenpreis, On the theory of kernels of Schwartz, *Proc. Am. Math. Soc.* **7** (1956), 713–718.

[47] L. Ehrenpreis, Sheaves and differential equations, *Proc. Am. Math. Soc.* **7** (1956), 1131–1138.

[48] L. Ehrenpreis, Some applications of the theory of distributions to several complex variables, in *Seminar on Analytic Functions*, vol. 1, Princeton, 1957, 65–79.

[49] L. Ehrenpreis, The fundamental principle for linear constant coefficient partial differential equations, in *Proc. Int. Symp. Linear Spaces, Jerusalem*, 1960, 161–174. Also presented to *Conf. Linear Spaces, Warsaw, 1960*, and summarized in *Stud. Math.*, special series, 1963, 35–36.

[50] L. Ehrenpreis, Lectures at Stanford, 1960 (mimeographed).

[51] L. Ehrenpreis, Conditionally convergent functional integrals and partial differential equations, in *Proc. Int. Congr. Mathematicians, Stockholm*, 1962, 337–338.

[52] L. Ehrenpreis, Function space integration and partial differential equations, Lectures at Harvard University, 1964–1965.

[53] L. Ehrenpreis, A new proof and an extension of Hartogs's theorem, *Bull. Am. Math. Soc.* **67** (1961), 507–509.

[54] L. Ehrenpreis, Complex Fourier transform technique in variable coefficient partial differential equations, *J. d'Anal. (Jerusalem)* **XIX** (1967), 75–95.

[55] L. Ehrenpreis, Reflection, removable singularities and approximation for partial differential equations I, *Ann. Math.* **112** (1980), 1–20.

[56] L. Ehrenpreis, Reflection, removable singularities and approximation for partial differential equations II, *Trans. Am. Math. Soc.* **302** (1987), 1, 1–44.

[57] L. Ehrenpreis, Function theory for Rogers–Ramanujan-like partition identities, *Contemp. Math.* **143** (1993), 259–320.

[58] L. Ehrenpreis, Hypergeometric functions, in *Algebraic Analysis*, vol. II, Academic Press, New York, 1988, 85–128.

[59] L. Ehrenpreis, Fourier analysis on nonconvex sets, *Symp. Mat. Instituto Nazionale de Alte Mat.* **XII** (1971).

[60] L. Ehrenpreis, Conformal geometry, in *Differential Geometry*, Birkhäuser, Boston, 1983, 73–88.

[61] L. Ehrenpreis, Singularities, functional equations, and the circle method, *Contemp. Math.* 166, American Mathematical Society, Providence, RI, 35–80.

[62] L. Ehrenpreis, Fourier analysis, partial differential equations, and automorphic functions, *Proc. Symp. Pure Math.* **49** (1989), 45–100.

[63] L. Ehrenpreis, Fourier analysis on nonconvex sets, *Symp. Math.* **VII**, 413–419. (Convegno sulle Equazioni Ipoellittiche e Spazi Funzionali, INDAM, Rome, 1970) Academic Press, London, 1971.

[64] L. Ehrenpreis, Three problems at Mount Holyoke, *Contemp. Math.* **278**, American Mathematical Society, Providence, RI, 123–130.

[65] L. Ehrenpreis, P. Kuchment, and A. Panchenko, The exponential X-ray transform and Fritz John's equation. I. Range description, in *Analysis, geometry, number theory: the mathematics of Leon Ehrenpreis*, *Contemp. Math.* **251**, American Mathematical Society, Providence, RI, 2000, 173–188.

[66] L. Ehrenpreis and P. Malliavin, Spectral gaps and lacunas, *Bull. Sc. Math.* **105** (1981), 17–28.

[67] L. Ehrenpreis and P. Malliavin, Invertible operators and interpolation in AU spaces, *J. Math. Pures Appl.* **53** (1974), 9, 165–182.

[68] L. Ehrenpreis and F. I. Mautner, Some properties of the Fourier transform on semi-simple Lie groups, I, *Ann. Math.* **61** (1955), 406–439.

[69] L. Ehrenpreis and F. I. Mautner, Some properties of the Fourier transform on semi-simple Lie groups, II, *Trans. Amer. Math. Soc.* **84** (1957), 1–55.

[70] L. Ehrenpreis and F. I. Mautner, Some properties of the Fourier transform on semi-simple Lie groups, III, *Trans. Amer. Math. Soc.* **90** (1959), 3, 431–484.

[71] L. Ehrenpreis and F. I. Mautner, Uniformly bounded representations of groups, *Proc. Natl Acad. Sci. USA* **41** (1955), 231–233.

[72] R. P. Feinerman and D. J. Newman, *Polynomial Approximation*, Williams & Wilkins, Baltimore, MD, 1974.

[73] J. Feldman and C. C. Moore, Ergodic equivalence relations, cohomology, and von Neumann algebras. I, *Trans. Am. Math. Soc.* **234** (1977), 2, 289–324; II, *Trans. Am. Math. Soc.* **234** (1977), 2, 325–359.

[74] L. Flatto, Invariants of finite reflection groups and mean value problems. II, *Am. J. Math.* **92** (1970), 552–561.

[75] L. Flatto, Invariants of finite reflection groups, *Enseign. Math.* **24** (1978), 3–4, 237–292.

[76] L. Flatto and Sister M. M. Wiener, Regular polytopes and harmonic polynomials, *Can. J. Math.* **22** (1970), 7–21.

[77] I. Gelfand, S. Gindikin, and M. Graev, Integral geometry in affine and projective spaces, *J. Sov. Math.* **18** (1980), 39–167.

[78] P. Gilkey, *Invariance theory, the heat equation, and the Atiyah-Singer index theorem*, second edition, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1995.

[79] S. Gindikin, The Radon transform from the cohomological point of view, in *Proc. Conf. 75 Years of Radon Transform, Vienna, 1992*, Lecture Notes Math. Phys., vol. IV, International Press, Cambridge, MA, 1994, 123–128.

[80] F. Gonzalez, Range of Radon transform on Grassmann manifolds, in *Proc. Conf. 75 Years of Radon Transform, Vienna, 1992*, International Press, Hong Kong, 1993.

[81] A. Grothendieck, Sur les espaces (F) et (DF), *Summa Bras. Math.* **3** (1954), 57–123.

[82] A. Grothendieck, Produits tensoriels topologiques et espaces nucléaires, *Mem. Am. Math. Soc.* **16** (1955).

[83] V. Guillemin, On some results of Gel'fand in integral geometry, *Proc. Symp. Pure Math.* **43** (1985), 149–155.

[84] C. Hamaker, K. T. Smith, D. C. Solmon, and S. L. Wagner, The divergent beam X-ray transition, *Rocky Mountain J. Math.* **10** (1980), 253–283.

[85] J. Han, Temple University thesis, 1995.

[86] Harish-Chandra, Representations of semisimple Lie groups, I, *Trans. Am. Math. Soc.* **75** (1953), 185–243.

[87] Harish-Chandra, Spherical functions on a semisimple Lie group, I, *Am. J. Math.* **80** (1958), 241–310.

[88] Harish-Chandra, Spherical functions on a semisimple Lie group, II, *Am. J. Math.* **80** (1958), 553–613.

[89] E. Hecke, Über das Verhalten von $\sum e^{2\pi i\tau\frac{|m^2-2n^2|}{8}}$ und ähnlichen Funktionen bei Modulsubstitutionen, *J. reine angew. Math.* **157** (1927), 159–170.

[90] E. Hecke, *Grundlagen einer Theorie der Integralgruppen und der Integralperioden bei den Normalteilern der Modulgruppe*, Math. Werke, Göttingen (1959), 731–772.

[91] E. Hecke, Über Modulfunktionen und die Dirichletschen Reinen mit Eulersher Produktentwicklung, *Mat. Ann.* **114** (1937), 1–28.

[92] E. Hecke, *Vorlesungen über die Theorie der Algebraischen Zahlen*, Chelsea, New York, 1948.

[93] E. Hecke, Eine neue Art von Zetofunklionesen und ihre Beziehungen zur Verteilung der Primzahlen, Part I, *Math. Z.* **1** (1918), 357–376; Part II, **6** (1920), 11–51.

[94] S. Helgason, *Groups and Geometric Analysis; Integral Geometry, Invariant Differential Operators and Spherical Functions*, Academic Press, New York, 1984.

[95] S. Helgason, *The Radon Transform*, Birkhäuser, Berlin, 1999.

[96] A. Hertle, Continuity of the Radon transform and its inverse in Euclidean space, *Math. Z.* **184** (1983), 165–192.

[97] A. Hertle, On the range of the Radon transform and its dual, *Math. Ann.* **267** (1984), 91–99.

[98] L. Hörmander, *The Analysis of Linear Partial Differential Operators I, II, III, IV*, Springer-Verlag, Berlin and New York, 1983.

[99] F. John, Bestimmung einer Funktion aus ihren Integralen über gewisse Mannigfaltigkeiten, *Math. Ann.* **109** (1934), 488–520.

[100] F. John, *Plane Waves and Spherical Means*, Wiley (Interscience), New York, 1955.

[101] E. Kähler, Innerer und ässerer Differentialkalkül, *Abbh. Deutsch. Akad. Wiss. Berlin Kl. Math. Phys. Tech.* (1960), no. 4.

[102] M. Kashiwara and T. Kawai, Introduction: Professor Mikio Sato and his work, *Algebraic Anal.* **I**, 1–7, Academic Press, Boston, MA, 1988.

[103] M. Kashiwara and T. Kawai, Microlocal analysis, *Publ. Res. Inst. Math. Sci.* **19** (1983), 3, 1003–1032.

[104] M. Knopp, *Modular Functions in Analytic Number Thoery*, Markham, Chicago, 1970.

[105] H. Komatsu *et al.*, *Hyperfunctions and pseudo-differential equations*, Lecture Notes in Mathematics, **287**, Springer-Verlag, Berlin, 1973.

[106] J. Korevaar and J. Wiegerinck, A representation of mixed derivatives with an application to the edge-of-the-wedge theorem, *Nederl. Akad. Wetensch. Indag. Math.* **47** (1985), 1, 77–86.

[107] J. Korevaar and J. Wiegerinck, A lemma on mixed derivatives and results on holomorphic extension, *Nederl. Akad. Wetensch. Indag. Math.* **47** (1985), 3, 351–362.

[108] J. Korevaar and J. Wiegerinck, A lemma on mixed derivatives with applications to edge-of-the-wedge, Radon transformation and a theorem of Forelli, *Int. Symp. Complex Analysis and Applications, Arandjelovac, 1984, Mat. Vestnik* **37** (1985), 145–157.

[109] J.-L. Koszul, Lectures on groups of transformations, Notes by R. R. Simha and R. Sridharan, Tata Institute of Fundamental Research Lectures on Mathematics, 32, Bombay, 1965.

[110] J.-L. Koszul, *Selected papers of J.-L. Koszul* (in French), Series in Pure Mathematics, 17, World Scientific, River Edge, NJ, 1994.

[111] S. Krantz, *Function Theory of Several Complex Variables*, second edition, The Wadsworth and Brooks/Cole Mathematics Series, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1992.

[112] S. S. Kudla and J. J. Millson, Harmonic differentials and closed geodesies on a Riemann surface, *Invent. Math.* **54** (1979), 193–211.

[113] E. Landou, *Vorlesungen über Zahlentheorie*, Chelsea, New York, 1955.

[114] J. Lehner, *Discontinuous Groups and Automorphic Functions*, American Mathematical Society, Providence, RI, 1964.

[115] N. Levinson, *Gap and Devioty Theorems*, American Mathematical Society, New York, 1940.

[116] H. Lewy, On the reflection laws of second order differential equations in two independent variables, *Bull. Am. Math. Soc.* **65** (1959), 37–58.

[117] H. Lewy, An example of a smooth linear partial differential operator without solutions, *Ann. Math.* **66** (1957), 155–158.

[118] H. Lewy, On the local character of the solutions of an atypical linear differential equation in three variables and a related theorem for regular functions of two complex variables, *Ann. Math.* **64** (1956), 2, 514–522.

[119] B. Malgrange, Existence et approximation des solutions des équations aux dérivées partielles et des equations de convolution, *Ann. l'Institut Fourier (Grenoble)*, **6** (1956), 271–355.

[120] R. Meise, B. A. Taylor, and D. Vogt, Indicators of pleurisubbarmonic functions on algebraic varieties and Kaneko's Phragmén–Lindelöf condition, in *Geometrical and algebraical aspects in several complex variables, Sem. Conf.* **8**, 231–250, EditEl, Rende, 1991.

[121] R. Meise, B. A. Taylor, and D. Vogt, Continuous linear right inverses for partial differential opeartors with constant coefficients and Phrogmén–Lindelöf condition, in *Functional analysis*, Lectures Notes in Pure and Applied Mathematics **150**, 357–389, Dekker, New York, 1994.

[122] R. Meise, B. A. Taylor, and D. Vogt, Dietmar continuous linear right inverses for partial differential operators of order 2 and fundamental solutions in half spaces, *Manuscr. Math.* **90** (1996), 449–464.

[123] F. Natterer, On the inversion of the attenuated Radon transform, *Numer. Math.* **32** (1979), 431–438.

[124] F. Natterer, *The Mathematics of Computerized Tomography*, Wiley, New York, 1986.

[125] J. von Neumann, On rings of operators, reduction theory, *Ann. Math.* **41** (1940), 94–161.

[126] R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, New York, 1934.

[127] V. Petersson, Über die Entwicklungskoeffizienten der Automorphen Formen, *Acta Math.* **58** (1932), 168–215.

[128] D. H. Phong and E. M. Stein, Models of degenerate Fourier integral operators and Radon transforms, *Ann. Math.* **140** (1994), 3, 703–722.

[129] D. H. Phong and E. M. Stein, Radon transforms and torsion, *Int. Math. Res. Not.* **4** (1991), 49–60.

[130] V. Pribitkin, The Fourier coefficients of modular forms and Niebur modular integrals having small positive weights, *Acta Arith.* **XCI** (1999), 291–309.

[131] E. T. Quinto, The injectivity of rotation invariant Radon transforms on $\mathbb{C}^n$, *Contemp. Math.* **63** (1987), 245–260.

[132] H. Rademacher, On the expansion of the partition function in series, *Ann. Math.* **44** (1943), 416–422.

[133] J. Radon, Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten, *Ber. Verh. Sächs. Akad. Wiss. Leipzig, Math. Nat. Kl.* **69** (1917), 262–277.

[134] M. Riesz, L'intégrale de Riemann–Liouville et le problème de Cauchy, *Acta Math.* **81** (1949), 1–223.

[135] L. Schwartz, *Théorie des Distributions*, vols. I, II, Paris, 1950–1951.

[136] L. Schwartz, *Etude des Sommes d'Exponentielles Réeles*, Paris, 1943.

[137] L. Schwartz, Théorie générale des fonctions moyenne-périodiques, *Ann. Math.* **48** (1947), 857–929.

[138] L. Schwartz, Jean Delsarte, *Enseign. Math.* **18** (1972), 2, 113.

[139] A. Selberg, Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series, *J. Ind. Math. Soc. (NS)* **20** (1956), 47–87.

[140] H. S. Shapiro, Functions with a spectral gap, *Bull. Am. Math. Soc.* **79** (1973), 355–360.

[141] C. L. Siegel, *Transcendental Numbers*, Princeton University Press, Princeton, NJ, 1949.

[142] C. L. Siegel, *Lectures on Advanced Analytic Number Theory*, Tata Institute, Bombay, 1961.

[143] C. L. Siegel, *Lectures on the Analytical Theory of Quadratic Forms*, The Institute for Advanced Study, Princeton, NJ, 1949.

[144] C. L. Siegel, *Symplectic Geometry*, Academic Press, New York, 1964.

[145] N. Steenrod, *The Topology of Fibre Bundles*, Princeton University Press, Princeton, NJ, 1951.

[146] R. Steinberg, Differential equations invariant under finite reflection groups, *Trans. Am. Math. Soc.* **112** (1964), 392–400.

[147] S. Täcklind, *Sur les classes quasi analytiques des solutions des équations aux dérivées partielles du type parabolique*, Nova Acta Regiae Societatis Scientiarum Upsaliansis, **16** (1936), 1–56.

[148] A. Tumanov, *A Morera type theorem in the strip*, to appear.

[149] B. A. Taylor, Discrete sufficient sets for some spaces of entire functions, *Trans. Am. Math. Soc.* **163** (1972), 207–214.

[150] E. C. Titchmarsh, *The Theory of the Riemann Zeta Function*, Oxford, 1951.

[151] E. C. Titchmarsh, *The Theory of Functions*, Oxford, 1932.

[152] E. C. Titchmarsh, *Introduction to the theory of Fourier integrals*, third edition, Chelsea, New York, 1986.

[153] E. C. Titchmarsh, *Eigen function Expansions*, Part I, second edition, Oxford, 1962.

[154] E. C. Titchmarsh, *Eigen function Expansions*, Part II, Oxford, 1958.

[155] G. N. Watson, *A treatise on the theory of Bessel functions*, Cambridge University Press, Cambridge, 1995.

[156] A. Weil, *Foundations of Algebraic Geometry*, American Mathematical Society, Providence, RI, 1962.

[157] A. Weil, L'oeuvre mathématique de Delsarte, *Enseign. Math.* **18** (1972), 2, 115–134.

[158] H. Weyl, The method of orthogonal projection in potential theory, *Duke Math. J.* **7** (1940), 411–444.

[159] H. Weyl, *The Classical Groups*, Princeton University Press, Princeton, NJ, 1946.

[160] E. T. Whittaker and G. N. Watson, *Modern Analysis*, fourth edition, Cambridge, 1927.

[161] J. Wong, Temple University thesis, 1992.

[162] O. Zariski and P. Samuel, *Commutative Algebra*, vols. I and II, Springer-Verlag, New York and Heidelberg, 1975.

Appendix

SOME PROBLEMS OF INTEGRAL GEOMETRY ARISING IN TOMOGRAPHY

Peter Kuchment
Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA

and

Eric Todd Quinto[1]
Department of Mathematics, Tufts University, Medford, MA 02155, USA

This appendix is designed to coordinate the abstract nature of this book with more concrete and practical aspects of the Radon transform.

A.1 Introduction

Beautiful mathematical theorems in integral geometry form the foundation of tomography. The goal of tomography is to recover the interior structure of a non-transparent body using external measurements. In this appendix, we will outline a few tomographic problems that draw on the integral geometry in this book, including support theorems, inversion formulas, range theorems, and microlocal analysis. We will try to show how these theorems are used in tomography. We do not intend to present here any concise survey of tomography, rather a collection of examples chosen according to their connection to the topics considered in the book and to the authors' interest. The bibliography we provide is also far from being complete, and we refer the reader to books, surveys, and collections of articles [14, 48–52, 67, 72, 73, 80, 124, 126, 138, 139, 150, 161, 170] for details and further references on tomography and related problems of integral geometry.

A.2 X-ray tomography

X-ray computed tomography (CT) is the most basic type of tomography; the model is the classical Radon transform that integrates functions over lines. However, the applied mathematics is deep and the properties of the transform provide insight into all parts of integral geometry. In this section we will discuss some of the basic ideas in X-ray CT and then talk about various types of limited

[1]The second author was partially supported by NSF grants 9877155 and 0200788.

data tomography. Our goal is to show some of the ways integral geometry and microlocal analysis can help one understand these problems.

A.2.1 *Complete data*

The goal of X-ray CT is to get a picture of the internal structure of an object by X-raying the object from many different directions. Let $f : \mathbf{R}^3 \to \mathbf{R}$ be the density function of the object; that is, $f(x)$ is the density of the object at the point $x \in \mathbf{R}^3$. Mathematically, the goal of X-ray CT is to recover the density f from these measurements. As X-rays travel on a line L from the X-ray source through the object to an X-ray detector, they are attenuated by the material on the line they travel along (scatter can be neglected in most X-ray CT problems). If the X-rays are monochromatic, then the attenuation at $x \in L$ is proportional to the density $f(x)$ (with proportionality constant $c > 0$). If $I(x)$ is the number of X-ray photons in the beam when it arrives at x, then the intensity in a small segment of length Δx is decreased by a multiplicative factor of $cf(x)\Delta x$, so

$$\Delta I \approx -\big(cf(x)\Delta x\big)I(x). \tag{A.1}$$

By integrating (A.1), we get:

$$\ln\left[\frac{I(\text{detector})}{I(\text{source})}\right] = c\int_L f(x)dx_L = cRf(L).$$

This integral transform R is exactly the classical Radon transform of f on the line L [71, 124]. Often, X-rays are taken in parallel slices through the object, and the problem becomes a planar problem. We will primarily consider this planar case.

Complete tomographic data is X-ray data over all lines. In practice, this means data is collected on a fairly evenly distributed set of lines throughout the object (the concept of complete data can be made precise using sampling theory as in [124, 126]). In this case, the commonly used reconstruction algorithm is filtered backprojection, which we will now discuss. We start with some standard notation. Let $\omega = (\cos\theta, \sin\theta)$ be a unit vector on the plane and p be a real number. We denote by $L(p,\omega) = \{x \in \mathbf{R}^2 \,|\, x \cdot \omega = p\}$ the line perpendicular to ω and p directed units from the origin. We will also often use $\omega^\perp = (-\sin\theta, \cos\theta)$, the unit vector $\pi/2$ units counterclockwise from ω. We should mention that there are two standard ways of parametrizing lines, the one we use in this section ("hyperplane" or nonparametric notation) and the one used in Section A.3 ("parallel beam" notation). The theorems and ideas in each section are easier to express using the corresponding notation.

The nonparametric Radon transform (A.2) of a function f in the plane can be naturally interpreted as a function of (p,ω) (see Chapter 2):

$$g = Rf(p,\omega) = \int_{x\in L(p,\omega)} f(x)dx_L. \tag{A.2}$$

We define now the backprojection operator, the dual Radon transform, as

$$R^*g(x) = \int_{S^1} g(x \cdot \omega, \omega) d\omega.$$

This is the integral of g over all lines through x, since $L(x \cdot \omega, \omega)$ is the line through x and perpendicular to ω. Let $g \in C_0^\infty(\mathbf{R} \times S^1)$; then the Fourier transform of g in the p variable is

$$\mathcal{F}_{p \to \sigma} g(\sigma, \omega) = \frac{1}{2\pi} \int_{p=-\infty}^{\infty} e^{-ip\sigma} g(p, \omega) dp,$$
$$\mathcal{F}^{-1}_{\sigma \to p} g(p, \omega) = \int_{\sigma=-\infty}^{\infty} e^{ip\sigma} g(\sigma, \omega) d\sigma.$$

For $f \in C_0^\infty(\mathbf{R}^2)$ we define the two-dimensional Fourier transform by

$$\tilde{f}(\xi) = \mathcal{F}_{x \to \xi} f(\xi) = \frac{1}{(2\pi)^2} \int_{x \in \mathbf{R}^2} e^{-ix \cdot \xi} f(x) dx.$$

This allows us to define the Riesz potential I^{-1} as the operator with Fourier multiplier $|\sigma|$:

$$I^{-1} g(p, \omega) = \mathcal{F}^{-1}_{\sigma \to p}(|\sigma| \mathcal{F}_{p \to \sigma} g).$$

A variation of the usual inversion formula for the nonparametric Radon transform (Section 2.1) is

Theorem A.1 [124, 151, 160] *Let $f \in C_0^\infty(\mathbf{R}^2)$. Then $f = \frac{1}{4\pi} R^*(I^{-1} Rf)(x)$.*

The theorem is applied in practice by truncating and smoothing the multiplier $|\sigma|$ in I^{-1} and writing this truncated multiplier as a convolution operator in p [124, 126, 151]. The resulting approximate inversion algorithm becomes $f \approx R^*(\Phi *_p Rf)$ where Φ is the inverse Fourier transform of the truncated multiplier and $*_p$ denotes convolution in the p variable.

Proof The proof uses some of the key elementary formulas for the Radon transform. It begins with the *projection–slice theorem* (which is discussed in Chapter 2 (2.3)ff.):

$$\frac{1}{2\pi} \mathcal{F}_{p \to \sigma} Rf(\sigma, \omega) = \tilde{f}(\sigma\omega). \tag{A.3}$$

Equation (A.3) is proven using Fubini's theorem. It is an elementary application of Radon transforms on spreads: to prove (A.3), we integrate Rf over the spread of lines perpendicular to ω with respect to $\exp(-ip\sigma)dp$ and then set $p = x \cdot \omega$ for $x \in L(p, \omega)$.

To finish the proof, one writes the two-dimensional Fourier inversion formula in polar coordinates and uses (A.3) to get

$$\begin{aligned} f(x) &= \frac{1}{4\pi}\int_{\omega\in S^1}\int_{\sigma\in\mathbf{R}} e^{i\sigma\omega\cdot x}|\sigma|\mathcal{F}_{p\to\sigma}Rf(\sigma,\omega)d\sigma d\omega \\ &= \frac{1}{4\pi}\int_{\omega\in S^1} I^{-1}Rf(\omega\cdot x,\omega)d\omega = \frac{1}{4\pi}R^*I^{-1}Rf(x). \end{aligned}$$

In Section A.2.2 we will use the properties of R as an elliptic Fourier integral operator (FIO) to develop subtle results about singularity detection in tomography. Now, we will use the projection–slice theorem (A.3) to show R is, in fact, an elliptic FIO. Let $J : C^\infty(\mathbf{R}^2) \to C^\infty(\mathbf{R}\times S^1)$ be defined by $Jf(p,\omega) = f(p\omega)$. We take the one-dimensional inverse Fourier transform in (A.3):

$$\begin{aligned} Rf(p,\omega) &= \frac{1}{2\pi}(\mathcal{F}^{-1}_{\sigma\to p}\circ J\circ\mathcal{F}_{x\to\xi})(f)(p,\omega) \\ &= \frac{1}{2\pi}\int_{\sigma\in\mathbf{R}}\int_{x\in\mathbf{R}^2} e^{i(p-\omega\cdot x)\sigma}f(x)dxd\sigma. \end{aligned} \tag{A.4}$$

Of course, this integral does not converge absolutely, although the operation $\mathcal{F}^{-1}_{\sigma\to p}\circ J\circ\mathcal{F}_{x\to\xi}$ is defined for $f\in C_0^\infty(\mathbf{R}^2)$. To make the integrals converge, one does integrations by parts as with FIOs in general to show (A.4) can be defined on distributions of compact support [76,168]. Equation (A.4) shows that R is an elliptic FIO with phase function $\phi(x,p,\omega,\sigma) = (p-\omega\cdot x)\sigma$ and amplitude $1/2\pi$.

A similar exercise using the fact that $R^*Rf(x) = (2/|x|) * f$ shows that

$$R^*Rf(x) = \int_{\xi\in\mathbf{R}^2} e^{ix\cdot\xi}\frac{1}{\pi|\xi|}\tilde{f}(\xi)d\xi. \tag{A.5}$$

This shows R^*R is a classical elliptic pseudo-differential operator with symbol $1/(\pi|\xi|)$. The interplay between microlocal analysis and integral geometry is rich, and its history, beginning with Guillemin's seminal work, will be discussed in detail after Theorem A.3.

A.2.2 *Wave front sets and singularity detection*

Wave front sets provide a subtle and elegant classification of singularities of functions and distributions, and microlocal analysis allows us to understand how FIOs transform wave front sets (compare Section 5.3).

In particular, we will use this methodology to learn about how the FIO R detects singularities and to predict how well reconstruction algorithms will perform when they have limited tomographic data. The wave front set of $f\in\mathcal{D}'(\mathbf{R}^n)$ is defined in Chapter 5. It is useful because it classifies not only singularities of f using points x_0 at which f is not smooth, but also *directions* above each point in which f is not smooth. We also need the concept of Sobolev wave front set. The distribution f is in $H^s(\mathbf{R}^2)$ if and only if its Fourier transform, $\tilde{f} = \mathcal{F}_{x\to\xi}f$,

is in $L^2(\mathbf{R}^2, (1+|\xi|^2)^s)$. This relates global smoothness of f to integrability of its Fourier transform. A local version of this at a point $x_0 \in \mathbf{R}^n$ is obtained by multiplying f by a smooth cut-off function $\psi \in C_0^\infty(\mathbf{R}^n)$ (with $\psi(x_0) \neq 0$) and seeing if $\widetilde{(\psi f)}$ is in this weighted L^2 space. However, this localized Fourier transform $\widetilde{(\psi f)}$ gives even more specific information—microlocal information—namely, the *directions* near which $\widetilde{(\psi f)}$ is in $L^2(\mathbf{R}^2, (1+|\xi|^2)^s)$. The precise definition is (see [135], p. 259)

Definition A.2 *A distribution f is in the Sobolev space H^s locally near $x_0 \in \mathbf{R}^n$ if and only if there is a cut-off function $\psi \in \mathcal{D}(\mathbf{R}^n)$ with $\psi(x_0) \neq 0$ such that the Fourier transform $\widetilde{(\psi f)}(\xi) \in L^2(\mathbf{R}^n, (1+|\xi|^2)^s)$. Let $\xi_0 \in \mathbf{R}^n \setminus 0$. The distribution f is in H^s microlocally near (x_0, ξ_0) if and only if there is a cut-off function $\psi \in \mathcal{D}(\mathbf{R}^n)$ with $\psi(x_0) \neq 0$ and function $u(\xi)$ homogeneous of degree 0 zero and smooth on $\mathbf{R}^n \setminus 0$ and with $u(\xi_0) \neq 0$ such that the product $u(\xi)\widetilde{(\psi f)}(\xi) \in L^2(\mathbf{R}^n, (1+|\xi|^2)^s)$. The H^s wave front set of f, $\mathrm{WF}^s(f)$, is the complement of the set of (x_0, ξ_0) near which f is microlocally in H^s.*

It follows from this definition that, if $(x_0, \xi_0) \notin \mathrm{WF}(f)$, then for all s, f is H^s near (x_0, ξ_0). The wave front set and microlocal Sobolev smoothness are usually defined on $T^*(\mathbf{R}^n) \setminus 0$, the cotangent space of $\mathbf{R}^n$ with its zero section removed because such a definition can be extended invariantly to manifolds using local coordinates. To this end, let $x_0 \in \mathbf{R}^2$. If $y = (y_1, y_2) \in \mathbf{R}^2$, then we let $y\mathbf{dx} = y_1\mathbf{dx}_1 + y_2\mathbf{dx}_2$ be the cotangent vector corresponding to y in $T^*_{x_0}\mathbf{R}^2$. Now let $(p, \omega) \in \mathbf{R} \times S^1$. We let $\mathbf{dp}$ and $\mathbf{d}\omega$ be the standard basis of $T^*_{(p,\omega)}(\mathbf{R} \times S^1)$. Here, θ maps to $\omega = (\cos\theta, \sin\theta)$ provides coordinates on S^1 and the basis covector $\mathbf{d}\omega$ is the dual covector to $\partial/\partial\theta$. The wave front set is extended to distributions on $\mathbf{R} \times S^1$ using these local coordinates, and it is a subset of $T^*(\mathbf{R} \times S^1)$.

The fundamental theorem that gives the relation between Sobolev wave front of a function and its Radon transform is the following.

Theorem A.3 [146] *Let $f \in \mathcal{E}'(\mathbf{R}^2)$, $x_0 \in L(p_0, \omega_0)$, $\eta_0 = \mathbf{dp} - (x_0 \cdot \omega_0^\perp)\mathbf{d}\omega$, and $a \in \mathbf{R}^n \setminus \mathbf{0}$. The correspondence between $\mathrm{WF}(f)$ and $\mathrm{WF}(Rf)$ is*

$$(x_0; a\omega_0\mathbf{dx}) \in \mathrm{WF}(f) \quad \text{if and only if} \quad (p_0, \omega_0; a\eta_0) \in \mathrm{WF}(Rf). \tag{A.6}$$

Given $(p_0, \omega_0; a\eta_0)$, $(x_0; a\omega_0\mathbf{dx})$ is uniquely determined by (A.6). Moreover, $f \in H^s$ microlocally near $(x_0; a\omega_0\mathbf{dx})$ if and only if $Rf \in H^{s+1/2}$ microlocally near $(p_0, \omega_0; a\eta_0)$. Singularities of Rf above (p_0, ω_0) give no information about singularities of f above points not on $L(p_0, \omega_0)$ or at points on this line in directions not conormal to the line.

This theorem is a corollary of fundamental results of Guilleman [59], [61] on the microlocal analysis of R. The general results will first be described and then the relation to Theorem A.3 will be given and finally the proof will be outlined.

Guillemin first developed the microlocal analysis of the Radon transform. For general Radon transforms defined by double fibrations on manifolds with smooth

nowhere-zero measures, he proved that R is an elliptic FIO. He also proved that R^*R is an elliptic pseudo-differential operator under the Bolker assumption on the Lagrangian manifold associated to the FIO R [60]. At that time, he developed a new description of FIO using push forward and pull backs generalizing those used to define Radon transforms using double fibrations [59]. This work was included in [61] (see pp 364ff.).

Quinto [140] described the symbol of R^*R for all generalized Radon transforms satisfying the Bolker assumption in terms of the measures involved into the transform. He also proved more results for the hyperplane transform, including a simple expression for the symbol and a proof of invertibility for translation invariant transforms. Beylkin [19] considered Radon transforms satisfying the Bolker assumption integrating over surfaces in $\mathbf{R}^n$ and analyzed the asymptotics of the symbol of $R^\dagger KR$ where $R^\dagger$ is a backprojection related to R^* and K is a type of filter. We will discuss the generalized Radon transform further in Section A.2.3.3 and Section A.3.

The microlocal correspondence between $\mathrm{WF}(f)$ and $\mathrm{WF}(Rf)$ in Theorem A.3 follows from Guillemin's seminal work [60, 61]. Because R is an FIO corresponding to a specific Lagrangian manifold, (A.6) holds. Sobolev continuity is a basic property for FIOs; any FIO, such as R, of order $\frac{1}{2}$ will map functions in H^s of fixed compact support continuously to functions in $H^{s+1/2}_{\mathrm{loc}}$. If the operator is elliptic, then the function must be $\frac{1}{2}$ order less smooth than its image. Theorem A.3 is a refinement of this observation. Not only does R smooth of order $\frac{1}{2}$ but it maps functions that are in H^s microlocally near a given covector to functions that are in $H^{s+1/2}$ near the covector given by the correspondence (A.6).

The proof outline is as follows. For this classical Radon transform, the key to the correspondence (A.6) is (A.4), that R is an elliptic FIO with phase function $\phi(x, p, \omega, \sigma) = (p - x \cdot \omega)\sigma$ and the explicit Lagrangian manifold associated to this operator [60, 61] (see [140] for details). The fact that $(p_0, \omega_0; a\boldsymbol{\eta_0})$ in (A.6) determines $(x_0, a\omega_0 \mathbf{dx})$ is a simple exercise. The assertion about H^s is given in [146], and, as discussed above, is proven using general Sobolev continuity results for FIO [76, 168].

Theorem A.3 allows one to understand what R does to singularities in a precise and rigorous way and it provides an application of microlocal analysis to tomography. It gives an exact correspondence between singularities of f and those of Rf. Moreover, it states that the singularities of Rf that are detected from the data are of Sobolev order $\frac{1}{2}$ smoother than the corresponding singularities of f. For typical singularities of f (jump singularities in $H^{1/2-\epsilon}$) one can realistically expect the corresponding singularities of Rf not to be masked by noise.

The theorem has the corollary

Corollary A.4 *Data Rf for (p, ω) arbitrarily near (p_0, ω_0) detects singularities of f perpendicular to the line $L(p_0, \omega_0)$ but not in other directions. If Rf is in $H^{s+1/2}$ locally near (p_0, ω_0), then f is microlocally in H^s at all points on*

$L(p_0, \omega_0)$ at all directions conormal to $L(p_0, \omega_0)$. Conversely if a direction above (p_0, ω_0) is in $\mathrm{WF}^{s+1/2}(Rf)$ then f has H^s wave front set conormal to $L(p_0, \omega_0)$ at the point given by (A.6).

Corollary A.4 follows from the correspondence (A.6) for WF and WF^s and the fact that data $\mathrm{WF}^{s+1/2}Rf$ above (p_0, ω_0) provides information only about singularities of f conormal to $L(p_0, \omega_0)$. Palamodov stated a closely related idea in [133]. The "tangent casting" effects of [156] are related to Corollary A.4.

A simple illustration will give an intuitive feeling for the corollary. Let $f : \mathbf{R}^2 \to \mathbf{R}$ be equal to one on the unit disk and zero outside. Then, $Rf(p, \omega) = 2\sqrt{1 - p^2}$ for $|p| \leq 1$ and $Rf(p, \omega) = 0$ for $|p| > 1$. The only lines where Rf is not smooth are those with $|p| = 1$ and these lines are tangent lines to the unit circle. Since $\mathrm{WF}(f)$ is the set of vectors normal to the unit circle, singularities of Rf precisely capture singularities of f.

A.2.3 *Limited data tomography*

Limited tomographic data is tomographic data given on some proper open subset of $\mathbf{R} \times S^1$. Theorem A.3 and Corollary A.4 provide a paradigm to decide which singularities of f are stably visible from limited tomographic data, and we will examine what this predicts for the three common types of limited data: limited angle CT (Section A.2.3.1), exterior CT (Section A.2.3.2), and the interior problem or local CT (Section A.2.3.3).

The basic idea is as follows. If Rf is in $H^{s+1/2}$ near $(p_0, \omega_0) \in \mathbf{R} \times S^1$, then, by Corollary A.4, f must be in H^s at all points of $L(p_0, \omega_0)$ in directions conormal to $L(p_0, \omega_0)$. By Theorem A.3, if Rf is not in $H^{s+1/2}$ microlocally in some direction at (p_0, ω_0), then f cannot be in H^s conormal to $L(p_0, \omega_0)$ at a specific point on this line that is given by (A.6). So, singularities of Rf above (p_0, ω_0) determine specific singularities of f conormal to $L(p_0, \omega_0)$. However, data Rf arbitrarily close to (p_0, ω_0) does not tell anything in a stable way about singularities of f away from $L(p_0, \omega_0)$ or at points of $L(p_0, \omega_0)$ in directions not conormal to this line. By (A.6), these singularities are visible from other data.

A typical tomographic density f is often a piecewise smooth function that is smooth on open sets with well-behaved boundaries. So, singularities of f occur at the boundaries, and the singularities are in $H^{1/2-\epsilon}$ for $\epsilon > 0$. By Theorem A.3, singularities of the Radon data Rf will be in $H^{1-\epsilon}$. A limitation of this analysis is that any discrete data Rf can be considered smooth. However, standard singularities of Rf are going to have large norm in H^1 and so should be visible. Moreover, the paradigm of the preceding paragraph is observed in all typical CT reconstructions from limited data.

Ramm and Zaslavsky [152] have analyzed how the Radon transform itself behaves on functions that are smooth except at smooth boundary surfaces. They give precise asymptotics of Rf at lines tangent to boundaries depending on the curvature of the boundaries, and they have proposed a singularity detection

method using this information on the raw data [152]. This method has been tested on simulated data [84]. A more general method has been proposed [146].

We now apply Theorem A.3 and Corollary A.4 as well as some classical theorems to three common types of limited data tomography problems in the plane. We will use the microlocal results to determine which singularities of f should be visible from limited tomographic data. One can also understand stability of limited data tomography using singular value decompositions [35, 107, 110, 113, 114, 116], and they reflect the principle predicted by the microlocal analysis: singular functions associated with large singular values (which are easy to reconstruct) oscillate in directions in which the wave front is easily detectable, and vice versa.

A.2.3.1 *Limited angle tomography*

Let U be a proper open subset of S^1 such that $U = -U$. *Limited angle data* is tomographic data given on $\mathbf{R} \times U$ only, and the limited angle Radon transform is the Radon transform $R : L_0^1(\mathbf{R}^2) \to L_0^1(\mathbf{R} \times U)$ where L_0^1 is the set of L^1 functions of compact support. The goal is to recover f using data $Rf(p, \omega)$ for $(p, \omega) \in \mathbf{R} \times U$. This problem comes up in electron microscopy [33] and in X-ray airport inspections of luggage traveling on a belt. In both cases, one cannot acquire X-ray CT data all around the object but only in a limited angular range. Alternatively, in electron microscopy, one has a picture that shows many molecules with the same crystalline structure but oriented in random directions. This is like regular CT with complete data, but with no information about the directions [126].

Using the projection–slice theorem (A.3), one can prove this limited data transform is injective for functions of compact support: knowing Rf on $\mathbf{R} \times U$ gives the Fourier transform $\tilde{f}$ on the cone in $\mathbf{R}^2$, $V = \{t\omega \,\big|\, t \in \mathbf{R}, \omega \in U\}$. Since f has compact support, $\tilde{f}$ is real analytic and this Fourier data determines f. Tuy [169] developed a reconstruction algorithm for this problem by extending this Fourier data $\tilde{f}$ on the cone V to $\mathbf{R}^2$ by making $\tilde{f}$ equal to zero off of V, then taking the inverse Fourier transform. This method destroys compact support since the Fourier transforms of compactly supported functions are real analytic. His inversion is really a pseudo-differential operator that is elliptic in directions from V so standard results about wave front sets [76, 168] show that all wave fronts of f in the directions in U will be preserved.

Other appealing and successful inversion methods involve extending the Radon data from $\mathbf{R} \times U$ to $\mathbf{R} \times S^1$ using the range conditions on the Radon transform [103, 107]. The basic idea is to project the data onto the range of the Radon transform with complete data (on $\mathbf{R} \times S^1$) and then invert this consistently completed data using filtered backprojection. Here are the details. Let B be a ball of radius $M > 0$. To implement this, one uses the singular value decomposition for $R : L^2(B) \to L^2([-M, M] \times S^1)$ in appropriate measures [30, 124]. One projects the incomplete data onto this basis of singular functions on $L^2([-M, M] \times S^1)$.

The projection is unstable because it requires an extrapolation: the singular basis is not orthogonal on $[-M, M] \times U$ but on the larger set $[-M, M] \times S^1$. The condition number of the projection matrix increases exponentially as the set U becomes smaller [35, 107]. Alternatively, if one were to construct an orthonormal basis on $L^2([-M, M] \times U)$, then the singular functions would have large norm outside of U.

This instability is intrinsic to the problem, and it is reflected in the fact that some singularities of f are not stably visible from the limited angle data. By Corollary A.4, the only singularities of f that can be detected in a stable way are those with directions in U. To see this, choose $x \in \mathbf{R}^2$ and $\omega \in U$. Any wave front of f at $(x; \omega \mathbf{dx})$ is detected by limited angle data because the line $L(x \cdot \omega, \omega)$ is in this data set. For the same reason, the wave front of f at $(x; \omega \mathbf{dx})$ for $\omega \notin U$ will not be stably detected by this limited angle data.

Moreover, an argument of Finch related to this singularity analysis shows that inversion of limited angle or exterior data is ill-posed in any range of Sobolev norms [40]. The argument for limited angle CT is as follows. We construct a nonsmooth function f that is smooth at all points in $\mathbf{R}^2$ in all directions in U. Then, by (A.6), Rf is smooth on $\mathbf{R} \times U$. So, Rf is in every local Sobolev space above $\mathbf{R} \times U$ but f is not in all Sobolev spaces on $\mathbf{R}^2$ since f is not smooth. Therefore, inversion of the limited angle Radon transform is not continuous in any range of Sobolev norms.

Here is one way to construct such an f. Let $\omega \notin U$, let L be the line through the origin normal to ω, and let H be a half-plane with boundary L. We take a smooth nonzero radial function of compact support, ψ, and let f be the result of setting ψ to zero on H. Then, f is smooth at all points in $\mathbf{R}^2$ in all directions in U but f is not in C^∞. In fact, $\mathrm{WF}(f)$ is a subset of points in L with direction $\omega \mathbf{d}\omega$, i.e. conormals to L.

This instability is also illustrated by the singular functions for the limited angle problem [107]. Those corresponding to large singular values (easy to reconstruct) oscillate generally in directions in U and those corresponding to small singular values (hard to reconstruct) oscillate generally in directions outside of U.

A.2.3.2 *The exterior problem*

Let $M > 1$ and assume *supp* $f \subset \{x \in \mathbf{R}^2 \big| |x| \leq M\}$. In the exterior problem, one has data $Rf(p, \omega)$ for all ω but only for $|p| > 1$. By the support theorem for the Radon transform ([67], see Chapter 2), one can reconstruct $f(x)$ for $|x| > 1$. This problem comes up in studies of the solar corona [6] in which data is total intensities of the corona of the Sun along lines exterior to the core of the Sun that go from the solar corona to the observer on Earth. Exterior data also occurs in industrial tomography of very large objects, for which X-ray data through their centers is too highly attenuated to be usable [148].

Let $|x| > 1$ and $\omega \in S^1$. Then the only singularities of f at x that are reconstructed in a stable manner are those for ω with $L(x \cdot \omega, \omega)$ in the data set, i.e. for $|x \cdot \omega| > 1$. Other singularities of f are not stably detected. Because

some wave front directions are not stably detectable by exterior data, inversion for the exterior problem is highly ill-posed; a translate of Finch's example in Section A.2.3.1 can be used to show inversion of the exterior transform is discontinuous in any range of Sobolev norms. Lissianoi [105] has extended Finch's Sobolev discontinuity result to show that in the exterior problem even recovery of the function in a smaller ring than where the data is given does not help to improve stability. However, logarithmic stability has been proven by Isakov [78].

Lewitt and Bates [103], Louis [108], and Natterer [120] have developed good reconstruction algorithms that use exterior data. Lewitt and Bates' algorithm completes the exterior data by projecting it on the range of the complete Radon transform as discussed in Section A.2.3.1. The projection step is unstable because the singular functions are not orthogonal on the annulus but on a disk. Natterer's algorithm is an effective regularization method. Quinto has developed an exterior reconstruction algorithm which employs a singular value decomposition [134] for the Radon transform on domain $L^2(\{x \in \mathbf{R}^2 \,\big|\, |x| \geq 1\})$ and a priori information about the shape of the object to be reconstructed. Reconstructions for "medical" phantoms are in [142] and those for industrial phantoms are in [143, 144] and real industrial data in [148]. Exactly those singularities that are supposed to be stably reconstructed are clearly imaged in the reconstruction. In Quinto's reconstruction algorithm, singularities that are not "visible" are blurred.

A.2.3.3 *The interior problem and local tomography*

Let $M > 1$ and assume *supp* $f \subset \{x \in \mathbf{R}^2 \big| |x| \leq M\}$. *Interior tomographic data* is data $Rf(p, \omega)$ for all ω but only for $|p| < 1$. Data is missing over lines outside the unit sphere, even though *supp* f can meet the annulus $\{x \in \mathbf{R}^2 \big| 1 \leq |x| \leq M\}$. The goal of interior CT is to reconstruct information about $f(x)$ for $|x| < 1$. This problem comes up whenever scientists want information only about some portion of the object to be reconstructed, not the whole object, or in problems, such as high-resolution tomography of very small parts of objects or electron microscopy, for which it is difficult or impossible to get complete high-resolution CT data.

Simple examples (derived using the range theorem for the Radon transform) show the interior transform, the Radon transform with this limited data, is not injective. However, according to Corollary A.4, one can detect all singularities of f in $|x| < 1$. To see this, choose a point x inside the unit disk and choose a direction $\omega \in S^1$. Then the line through x and normal to ω is in the data set for interior tomography. Therefore, by Theorem A.3, any singularity of f at $(x; \omega \mathbf{dx})$ is stably detected by interior data.

Maaß [114] has developed a singular value decomposition for this problem. See also [110]. The authors of [103] have developed a reconstruction method that projects the interior data onto the range of the Radon transform with complete data and then inverts this completed data. Because of the nonuniqueness of the problem, this projection step is not unique. However, Maaß has shown that singular functions associated to small singular values are fairly constant inside

the region of interest, the unit ball [114]. These singular functions corresponding to small singular values are difficult to reconstruct, but they do not add much detail inside the region since they are relatively constant there. This reflects the fact that all Sobolev singularities inside the region of interest are stably reconstructed.

Even though the interior transform is not invertible, we have shown that all singularities in the region can be stably reconstructed. This explains why singularity detection methods work so well for this problem. Lambda tomography [38, 39, 160, 171] is a well-developed algorithm that finds singularities of a function using interior data over a very small region. One starts with the filtered backprojection inversion formula, $f = 1/4\pi R^* I^{-1} Rf$, and replaces I^{-1} by $I^{-2} = I^{-1} \circ I^{-1} = -d^2/dp^2$. The beauty of this idea is that the formula

$$\sqrt{-\Delta} f = R^* I^{-2} Rf = -R\left(\frac{d^2}{dp^2} Rf\right) \tag{A.7}$$

is a local reconstruction formula. This is true because one needs only data $Rf(p,\omega)$ near $(x\cdot\omega,\omega)$ to calculate $(d^2/dp^2 Rf)(x\cdot\omega,\omega)$ and then to calculate $R^*(d^2/dp^2)Rf(x)$. Moreover, $\sqrt{-\Delta}$ is an elliptic pseudo-differential operator and so $\mathrm{WF}(f) = \mathrm{WF}(\sqrt{-\Delta} f)$ and $\mathrm{WF}^s(f) = \mathrm{WF}^{s-1}(\sqrt{-\Delta} f)$. This reconstruction formula takes singularities of f and makes them more pronounced; any singularity of f in H^s becomes a singularity in H^{s-1} because Λ is an elliptic pseudo-differential operator of order 1. At the same time, there is a cupping effect at the boundaries. Because of this, the developers of lambda CT chose to add a multiple of $R^*Rf = \Lambda^{-1} f$ to the reconstruction to smooth it out; moreover, with a good multiple, the cupping effects are decreased [38, 39].

For the X-ray transform with sources on a curve in $\mathbf{R}^3$, Louis and Maaß [109] have developed a very promising generalization of lambda CT. Greenleaf and Uhlmann [56, 57] completely analyzed the microlocal properties of R^*R for admissible transforms on geodesics, including this example. The microlocal analysis of this three-dimensional lambda CT operator has been investigated by Finch [42], Katsevich [83], and Lan [102]. The operator $R^*\Lambda R$ adds singularities to f because R is not well enough behaved; for the classical line transform in the plane, $R^*\Lambda R$ is an elliptic pseudo-differential operator and therefore preserves singularities. The theorem corresponding to Theorem A.3 for the X-ray transform on lines through a curve is given in [146], and it explains which singularities are stably reconstructed by this transform. The prediction is observed in the reconstructions in [109]. The general analysis in [62] and [56] provides the microlocal results needed to prove this theorem. The microlocal properties of this line transform were also studied in [22] as a basis for support and uniqueness theorems for the X-ray transform.

Limited angle and exterior versions of lambda CT have been developed, and they are promising on simulations [92] and tests on industrial and electron microscope data (E.T. Quinto unpublished).

Another way to deal with interior data is to truncate the data to zero by multiplying by a cut-off function which is equal to one in the region of interest. If one now applies a standard Radon inversion formula to the truncated data, one can check that the result is a pseudo-differential operator with the symbol equal to one in the local region of interest. This means that although the exact values of the function are not recoverable from the local data, its singularities are, by simple usage of truncated data in inversion formulas. One sees, for instance, that the values of jumps of the function are recovered correctly [86, 90]. This idea was further developed by Katsevich and Ramm [85] under the name of pseudo-local tomography. The authors of [38] have provided a different method to calculate the jumps of f at boundaries using lambda CT.

Madych [115] used wavelet analysis to show the strong relationship between lambda CT and pseudo-local CT and regular filtered backprojection. Authors of the papers [36, 153] have also used wavelet techniques for local tomography. They use Radon data to calculate wavelet coefficients of the density to be reconstructed. Although they need some data slightly outside the region of interest, their methods are fairly local.

Candès and Donoho [27] have defined ridgelets, wavelets that are not radially symmetric and are more sensitive to singularities in specified directions. They have developed a tomographic reconstruction algorithm using ridgelets that provides high-quality reconstructions in practice. Their ideas reflect the singularity detection predictions of Theorem A.3.

Technicians currently use the sinogram, the graph of $Rf(p,\omega)$ in rectangular coordinates, to find boundaries, but this method is subjective, and some industrial data is too homogeneous for this to work [149].

Local tomography has been developed for generalized Radon transforms [92], such as the attenuated Radon transform discussed in Section A.3 and in particular in Section A.3.1. If $\nu = \nu(x,\omega)$ is a nowhere-zero smooth weight, we define the generalized Radon transform of $f \in C_0^\infty(\mathbf{R}^2)$ to be

$$R_\nu f(p,\omega) = \int_{x \in L(p,\omega)} f(x)\nu(x,\omega)dx_L, \tag{A.8}$$

the integral of f on the line $L(p,\omega)$ in weight ν (note that the weight ν is determined by x and ω because $x \cdot \omega = p$ when $x \in L(p,\omega)$).

An important generalized Radon transform is the attenuated transform T_μ defined in Section A.3, (A.11). The weight for this transform is

$$\nu(x,\omega) = e^{-\int_{\tau=0}^{\infty} \mu(x-\tau\omega^\perp)d\tau}$$

where $\mu(y)$ is the attenuation coefficient of the body at the point y.

The generalized dual transform of R_ν is

$$R_\nu^* g(x) = \int_{\omega \in S^1} g(x \cdot \omega, \omega)\nu(x,\omega)d\omega$$

for $g \in C^\infty(\mathbf{R} \times S^1)$. Let $\mu(x,\omega)$ and $\nu(x,\omega)$ be nowhere-zero smooth functions on $\mathbf{R}^2 \times S^1$.

We use this to develop a generalized lambda tomography. The calculus of FIOs shows that the generalized lambda operator (compare with (A.7))

$$R_\nu^* \frac{d^2}{dp^2} R_\mu \tag{A.9}$$

is an elliptic pseudo-differential operator of order 1 with principal (top-order) symbol [92]

$$\nu(x, \xi/|\xi|)\mu(x, \xi/|\xi|)|\xi|.$$

The analogous theorem was proven about $R_\nu^* R_\nu$ in [140]. The operator defined by (A.9) is a local operator for the same reasons the lambda operator (A.7) is. Because $R_\nu^* d^2/dp^2 R_\mu$ is an elliptic pseudo-differential operator, it preserves wavefront set and

$$\mathrm{WF}^s(f) = \mathrm{WF}^{s-1}\left(R_\nu^* \frac{d^2}{dp^2} R_\mu f\right).$$

Good reconstructions of simulations using this operator from limited angle, exterior, and interior data are given in [92]. In addition, Katsevich [81] uses properties of pseudo-differential operators to provide an algorithm to detect the values of density jumps at boundaries of regions using this operator (A.9).

In general, the attenuated Radon transform will have a measure that is not smooth on boundaries of regions at which the attenuation changes. This gives rise to pseudo-differential operators $R_\nu^* d^2/dp^2 R_\nu$ that have nonsmooth symbols. Katsevich studied these operators, and he proved that one can detect jumps of f that are not masked by singularities of the measure using this information, if one knows the measure [82]. This work is important even if one does not know the measure ν exactly. Because the operator is an elliptic pseudo-differential operator (albeit with a nonsmooth symbol), the singularities of f that are not masked by singularities of ν will still be visible from the data, even if the measure is unknown. One practical issue would be whether the large amount of noise in SPECT data might create problems for any singularity detection algorithm that "overdifferentiates" the data.

A.3 Attenuated and exponential Radon transforms

Some problems of medical tomography, radiation therapy, and industrial nondestructive testing naturally lead to consideration of a special type of weighted X-ray transforms. We will briefly describe one such model, referring the reader to publications [24, 31, 32, 95, 124, 138, 139, 158] for further details, examples, and references.

The so-called single photon emission computed tomography (SPECT) deals with the situation when a nontransparent body contains a distributed radiation source. The goal is to reconstruct the interior intensity distribution $f(x)$ of the

radiation sources by conducting exterior measurements of the intensity of outgoing radiation. Let us denote by $\mu(x)$ the linear attenuation coefficient (or *absorption*) of the body at the point x. This means that due to the absorption a beam passing through the point x suffers at a small distance Δx the relative intensity loss equal to $\mu(x)\Delta x$. Then assuming that effects of scatter are small and hence can be neglected (which is in fact not always true), the stationary one-velocity transport equation can be written as

$$\nabla_x u \cdot \nu_\phi + \mu u = f. \tag{A.10}$$

Here $u(x, \phi)$ is the density of particles at x moving in the direction of the vector $\nu_\phi = (\cos\phi, \sin\phi)$. In other words,

$$u_{x_1} \cos\phi + u_{x_2} \sin\phi + \mu u = f.$$

Exterior detectors are placed on the boundary. They are collimated in order to detect only the particles following a specific oriented line L. In fact, perfect collimation is impossible, so the detector is affected by all rays from a solid angle around the line L, but we will disregard this circumstance. Assuming that there are no particles entering the region from outside, one can easily solve the transport equation for the intensities measured at the detectors. This gives as the total detected intensity of the beam L the expression

$$T_\mu f(L) = \int_L f(x) e^{-\int_{L_x} \mu(y)dy} dx. \tag{A.11}$$

Here L_x is the segment of the line L between the point x and the detector and dy denotes the standard linear measure on L.

The operator T_μ is said to be the *attenuated X-ray transform* of the function $f(x)$ and $\mu(x)$ is called the attenuation. The formula above is a nonparametric version of this transform. One can also introduce a parametric one. As mentioned in the last section, we will use a slightly different parametrization of lines in this section than we did in Section A.2. The theorems and ideas in each section are easier to express using the respective notation. Namely, let ω be the unit vector parallel to the direction of the oriented line L, $\omega^\perp$ be its orthogonal hyperplane, and $y = L \cap \omega^\perp$. Then the oriented lines L are in one-to-one correspondence with pairs (ω, y), where $y \in \omega^\perp$. (Now, this notation differs from that of the previous chapters; in Chapter 6 we studied attenuations which do not depend on y.) In the present notation we write

$$T_\mu f(\omega, y) = \int_{-\infty}^{\infty} f(y + t\omega) e^{-\int_t^\infty \mu(y+\tau\omega)d\tau} dt.$$

One notices that in contrast with the standard X-ray transform $f \to \int_L f(x)dx$, (A.2), the resulting function depends on the orientation of the line L.

In practice one often averages over the two orientations, thus arriving at a function of nonoriented lines.

In many algorithms one assumes that the attenuation is constant inside the body and zero outside. Let us also assume that the body has a known convex shape. In this case, according to [117], the attenuated transform can be reduced to a simpler one. Indeed, the integral

$$\mathcal{D}\mu(\omega, y + t\omega) = \int_t^\infty \mu(y + \tau\omega)d\tau$$

is the so-called *divergent beam transform* of μ (the integral of μ over the ray in direction ω starting at $y + t\omega$), and it can be evaluated explicitly when μ is constant:

$$\int_t^\infty \mu(y + \tau\omega)d\tau = \mu s(\omega, y) - \mu t.$$

Here $s(\omega, y)$ is the value $\tau \geq t$ for which $y + \tau\omega$ belongs to the boundary of the body. Since the shape of the body is known, $s(\omega, y)$ is a known function. Then the transform T_μ can be rewritten as follows:

$$\begin{aligned} T_\mu f(\omega, y) &= e^{-\mu s(\omega,y)} \int_{-\infty}^\infty f(y + t\omega)e^{\mu t}dt \\ &= e^{-\mu s(\omega,y)} R_\mu f(\omega, y). \end{aligned}$$

Here the function $\exp(-\mu s(\omega, y))$ is known and

$$R_\mu f(\omega, y) = \int_{-\infty}^\infty f(y + t\omega)e^{\mu t}dt$$

is the so-called *exponential X-ray transform* of the function f. (In order to avoid confusion, the reader should notice that we use notations for these transforms different from the ones in [124].) It is necessary to mention that one needs to require that the function $f(x)$ decays at infinity exponentially and sufficiently fast to offset the effect of the exponential weight in the integral.

Among the natural questions to ask about the attenuated and exponential X-ray transforms are:

Injectivity. Can a function of a natural class be reconstructed from its attenuated or exponential transforms? In other words, are these operators injective?
Inversion formulas in cases when injectivity is established.
Stability of inversion.
Range. Are these operators surjective in natural functional classes? Judging by the precedent of the standard Radon transform, the reader certainly should feel that the answer is probably negative, and some nontrivial range conditions should be satisfied. Then the question arises of describing the ranges of these operators.

Simultaneous reconstruction of the sources density f and attenuation μ. In most cases not only the value of the intensity $f(x)$ distribution, but the attenuation coefficient $\mu(x)$ as well is unknown. Is it possible to extract any information about both functions from the values of $T_\mu f$ or $R_\mu f$?

These problems will be briefly addressed below. We would also like to mention papers [37, 95, 131, 132, 157], where some interesting relations between the properties of the exponential transform and complex analysis are discovered that are beyond the scope of this text.

A.3.1 *Uniqueness of reconstruction*

Uniqueness of reconstruction is obviously the first question one should ask. Non-uniqueness would mean that one is unable to recover the interior structure of the object of interest from the exterior information. When we discuss uniqueness we will assume that the function f to be reconstructed has compact support.

The problem of uniqueness for the attenuated X-ray transform in dimension 2 has proven to be hard. The first results on uniqueness were the local ones. Since for smooth μ the operator T_μ (after averaging over opposite directions) is an elliptic FIO [19, 60, 61, 140], one concludes that for each point x there is a neighborhood U of x such that uniqueness holds for functions f with support in U. The idea is that locally the attenuated transform looks almost like the standard X-ray transform. It was shown in [118] that C^2 smoothness of the attenuation μ is sufficient. The paper [66] contains the proof of the following statement. Let us require that the supports of all functions f under consideration belong to a fixed bounded domain Ω. Assume, for instance, that Ω is the unit ball centered at the origin. Then the transformation T_μ is a semi-Fredholm operator between the Sobolev spaces $H_0^s(\Omega)$ and $H^{s+1/2}(\mathbf{S}^1 \times [-1,1])$ for any s (see also [124]). This means that the range of this operator is closed, and its kernel $\mathrm{Ker} T_\mu$ is finite dimensional. In the class of such operators property $\mathrm{Ker} A = 0$ is stable under perturbations of small operator norm. This observation immediately leads to a local uniqueness result: for small Ω, operator T_μ is norm close to the X-ray transform, and hence has zero kernel. The next step was made by Finch [41], who managed to obtain a "semi-local" uniqueness results. Namely, he proved uniqueness under the condition $||\mu||_\infty \, diam\Omega < 5.37$. This result is sufficient for many practical situations. For instance, in medical applications it restricts the diameter of an object to 35.8 cm. The proof is nontrivial and involves energy estimates. A breakthrough came recently in brilliant works [8, 128, 129], where a positive solution of the uniqueness problem under some mild smoothness condition on the attenuation was obtained [8] and an explicit inversion formula found [128, 129] (see also [26, 43, 58, 96, 98, 125] for different derivations and implementations). We will discuss the inversion formulas in a little more detail in the next section.

It was noticed in [41] that a similar uniqueness problem in dimensions 3 and higher is trivial. The reason is that the three-dimensional attenuated X-ray

transform is the collection of the corresponding two-dimensional transforms in all affine planes. If there is a compactly supported function $f(x)$ annihilated by the transform, then we can choose a plane which just barely touches the support of f. In this plane one can use the local uniqueness theorem to conclude that the function is in fact zero there. Proceeding further in this way, one can "eat the support away" and finally conclude that it is empty.

The problem of uniqueness was also considered for transforms with more general positive weights ν, the generalized Radon transform (A.8). Here there are local uniqueness results for fairly arbitrary weights and global uniqueness theorems for some special types of weights (translation invariant, rotation invariant) [100, 140, 141]. In the case of an analytic weight uniqueness in the class of compactly supported functions follows from analytic ellipticity of the corresponding FIO [21]. Analytic ellipticity has been used to prove support theorems and uniqueness in many settings such as for Radon transforms on geodesic spheres in realanalytic manifolds [145] and for circular transforms related to the wave equation (e.g. [2]) (see also [1] and [172] for related results). However, as the famous counterexample by Boman [20] shows, the condition of infinite smoothness of the weight function w alone does not guarantee uniqueness.

Let us move now to the much simpler case of the exponential transform R_μ. One of the most important (albeit simple) properties of the X-ray transform is the Fourier–slice (also called projection–slice) theorem (see Chapter 2 and (A.3)). There is an analog of this property for the exponential transform [121, 124]. Namely, let f be a compactly supported function on the plane. A straightforward calculation shows that

$$\widehat{R_\mu f}(\sigma, \omega) = \sqrt{2\pi}\, \tilde{f}(\sigma\omega + i\mu\omega^\perp), \tag{A.12}$$

where the hat on the left side denotes the one-dimensional Fourier transform with respect to the linear variable s, while the tilde on the right side denotes the two-dimensional Fourier transform. This means that the one-dimensional Fourier transform of the projection data $R_\mu f$ provides the values of the Fourier transform of the function f on a surface in $\mathbf{C}^2$. One can see this as an indication of the possibility to use methods of the theory of functions of several complex variables. Results of [3, 44, 37, 89, 93, 95, 117, 121, 131, 132] show that the relation between two theories is indeed very deep. We will try to show this in our exposition.

Let us turn now to the problem of uniqueness of reconstruction, i.e. injectivity of the operator R_μ. We will follow here considerations of [117] and [74]. Let us assume that function f is either compactly supported or decays as $\exp(-(\mu + \varepsilon)\,|x|)$ for a positive ε. Consider the surface

$$S_\mu = \left\{ z = \sigma\omega + i\mu\omega^\perp | \sigma \in \mathbf{R}, \omega \in \mathbf{S}^1 \right\} \subset \mathbf{C}^2.$$

It is straightforward to check that, except for the points where $\sigma = 0$, this surface is a totally real, smooth two-dimensional submanifold of $\mathbf{C}^2$. This means, in

particular, that it is a uniqueness set for analytic functions. The formula (A.12) shows that knowing $R_\mu f$ one can recover the values of the Fourier transform $\tilde{f}$ of the function f on the surface S_μ. Due to our decay assumption, $\tilde{f}$ is analytic in a tubular neighborhood of $\mathbf{R}^2$ containing S_μ. This implies that $R_\mu f$ determines $\tilde{f}$ and hence f uniquely.

A.3.2 *Inversion formulas*

As soon as injectivity is established, it is natural to look for inversion formulas. By now many inversion formulas have been established for the attenuated Radon transform (see [8,128,129] and also [26,43,58,91,96,98,125] for further discussion and numerical implementation). We will address these below in this section. We would like to start with the much simpler, but still very rich, example of the exponential transform.

A.3.2.1 *Exponential transform*

The first explicit inversion formula for the exponential X-ray transform in the plane was provided by Tretiak and Metz in [167] (see also its discussion in [124]). An inversion procedure was also provided in [12]. Let us introduce the dual exponential X-ray transform (or exponential backprojection) $R_\mu^\#$ as follows: applied to a function $g(s,\omega)$ it produces a function on the plane according to

$$\left(R_\mu^\# g\right)(x) = \int_{\mathbf{S}^1} e^{\mu x\cdot\omega^\perp} g(x\cdot\omega,\omega)d\omega.$$

One can verify the formula

$$R_{-\mu}^\# R_\mu f = \left(2\frac{\cosh\mu|x|}{|x|}\right) * f.$$

Hence, in order to reconstruct the function f from $R_{-\mu}^\# R_\mu f$ one needs to perform a deconvolution. This can be done by using appropriate generalized Riesz-type potentials on the line. Namely, let

$$\zeta_\mu(\sigma) = \begin{cases} |\sigma| & \text{when } |\sigma| > |\mu| \\ 0 & \text{otherwise} \end{cases}$$

and I_μ^{-1} (a generalized Riesz potential) be the Fourier multiplier by $\zeta_\mu(\sigma)$.

Theorem A.5 [167] *Let $f \in C_0^\infty(\mathbf{R}^2)$. Then*

$$f = \frac{1}{4\pi} R_{-\mu}^\# I_\mu^{-1} R_\mu f. \tag{A.13}$$

The original proof of this filtered backprojection-type formula in [167] and the proof provided in [124] are interesting and instructive, although rather

technical. We will now present a different simple approach based on complex analysis [89, 95].

Let us go back to the projection–slice formula (A.12). As we have already discussed, it implies that given $R_\mu f$ one can obtain the values of the Fourier transform of the unknown function f on the surface S_μ in $\mathbf{C}^2$. If instead of S_μ we dealt with $\mathbf{R}^2$, we would just use the Fourier inversion formula to find f. So, the question is whether one can develop a Fourier inversion formula that uses the data on S_μ instead of $\mathbf{R}^2$. The simple idea is to try to use the Cauchy theorem in order to deform the surface of integration from $\mathbf{R}^2$ to S_μ. Let us look at the standard Fourier inversion formula in $\mathbf{R}^2$:

$$f(x) = \int_{\mathbf{R}^2} \tilde{f}(\xi) e^{ix\cdot\xi} d\xi$$

(an additional multiplicative constant may arise depending on normalization of the Fourier transform). Assume that f is compactly supported or decays sufficiently rapidly exponentially, so $\tilde{f}(\xi)$ is analytic in all areas of interest. We consider the holomorphic differential form $\Phi_x = \tilde{f}(z)\exp(ix\cdot z)dz_1 \wedge dz_2$ in $\mathbf{C}^2$. Then the Fourier inversion formula reads

$$f(x) = \int_{\mathbf{R}^2} \Phi_x.$$

Now it is clear that using the Cauchy theorem is a good idea. There are, however, two complications. The first one is minor and can be easily dealt with. The surface S_μ when projected to $\mathbf{R}^2$ by $z \to Re(z)$ covers $\mathbf{R}^2\backslash\{0\}$ twice (since a point $\sigma\omega$ can be also represented as $(-\sigma)(-\omega)$). To treat this problem we can restrict the values of σ in the definition of S_μ to nonnegative numbers only. The second complication is much more serious and interesting. The surface S_μ is not homological to $\mathbf{R}^2$, in particular since it is not simply connected. That is, the points where $\sigma = 0$ constitute the boundary of a circular hole in the surface. This hole is the disk of radius μ in the imaginary plane. The idea to paste this disk to the surface does not work, since the values of $\tilde{f}$ on this disk are not known. The hope is to find a different disk that, on being pasted to the surface, would fix this problem. However, no matter what kind of disk we use, the values of $\tilde{f}$ on this disk will still be unknown. What difference then does the choice of disk make? Assume that we found an analytic disk; then the holomorphic $(2,0)$ form Φ_x would vanish on this disk. Thus, for an analytic disk the integral of the form Φ_x over this disk vanishes independently on the values of $\tilde{f}$. So, we are led to a classical problem of complex analysis about pasting analytic disks to a totally real surface in $\mathbf{C}^2$. A little experimentation shows that such a disk Γ_1 does exist and can be described as follows:

$$\Gamma_1 = \left\{ z = \sigma(\omega + i\omega^\perp) | 0 \le \sigma \le \mu, \omega \in \mathbf{S}^1 \right\}. \tag{A.14}$$

It is contained in the complex line $z_2 = iz_1$ and hence is analytic. We now define the following part of the surface S_μ:

$$\Gamma_2 = \left\{ z = \sigma\omega + i\mu\omega^\perp | \sigma \geq |\mu|, \omega \in \mathbf{S}^1 \right\} \subset S_\mu.$$

Defining $\Gamma = \Gamma_1 \cup \Gamma_2$, applying the Cauchy theorem, and using the equality $\int_{\Gamma_1} \Phi_x = 0$, we get an inversion formula

$$f(x) = \int_{\mathbf{R}^2} \Phi_x = \int_\Gamma \Phi_x = \int_{\Gamma_1} \Phi_x + \int_{\Gamma_2} \Phi_x = \int_{\Gamma_2} \Phi_x.$$

Writing the integral $\int_{\Gamma_2} \Phi_x$ explicitly using the definition of the form and coordinates σ and ω on S_μ, one arrives at the inversion formula (A.13). The details are easily workable and can be found in [89, 95, 158].

The idea of using the Cauchy theorem and Cauchy formula for inverting the exponential X-ray transform was initially used in papers [44, 117, 121], and developed in the presented form in [89, 95]. This approach leads to a wide variety of inversion formulas [95, 158], including cases when the attenuation μ depends on the direction ω [95, 158] and when imperfect detector collimation is taken into account [89]. Invertibility of the exponential X-ray transform was also studied analytically and numerically by different methods in [63–65]. One might also be interested in related inversion formulas available for the cases of half–view [127], three-dimensional limited data acquisition [97], and imperfect collimation [89].

Another interesting relation with complex analysis is the following. The possibility of pasting disks to totally real surfaces in $\mathbf{R}^2$ is a popular topic of complex analysis (e.g. [55]). In particular, the question of rigidity of such a pasted disk has been studied. Rigidity means absence of parametric deformation of the pasted analytic disks. The natural way the disk (A.14) pasted to the surface S_μ appears, suggests that one might expect its rigidity. This happens to be true [112]. Moreover, a general totally real surface in $\mathbf{C}^2$ in a neighborhood of the boundary of a pasted analytic disk can be reduced to a canonical form resembling the definition of S_μ, which in turn leads to the possibility of counting the number of parameters in deformations of this disk [112]. This provides a different approach to the known results [55] concerning this problem.

A.3.2.2 *Attenuated transform*

We follow in this section the technique introduced in [8, 25] with an additional development from [91]. Let as before $f(x)$ be the source's intensity in a planar domain Ω, $u(x, \phi)$ the density of particles at x moving in the direction of $\nu_\phi = (\cos\phi, \sin\phi)$, and $\mu(x)$ the linear attenuation coefficient. When scattering is absent, the transport equation (A.10) can be rewritten as

$$u_{x_1} \frac{1}{2}(e^{i\phi} + e^{-i\phi}) + u_{x_2} \frac{1}{2i}(e^{i\phi} - e^{-i\phi}) + \mu u = f.$$

Using the complex coordinate $z = x_1 - ix_2$ one gets

$$e^{i\phi}\frac{\partial u}{\partial \bar{z}} + e^{-i\phi}\frac{\partial u}{\partial z} + \mu u = f, \tag{A.15}$$

or when the attenuation is zero,

$$e^{i\phi}\frac{\partial u}{\partial \bar{z}} + e^{-i\phi}\frac{\partial u}{\partial z} = f. \tag{A.16}$$

The boundary data $u(x,\phi)|_{x\in\partial\Omega}$ is known. Let us consider the spaces $L_2(S)$ and $H^2 \subset L_2(S)$ (the Hardy space) on the circle and the orthogonal projector $P : L_2(S) \to H^2$. We also denote by $u_k = u_k(x)$ the Fourier coefficients of $u(x,\phi)$. Let R be the right shift in H^2: $(Ru)_k = u_{k-1}$ and $L = R^*$ be the left shift. Then we easily obtain from (A.16) that

$$(u_{-1})_{\bar{z}} + (u_1)_z = f$$

or

$$2Re((u_1)_z) = f \tag{A.17}$$

and

$$(u_k)_{\bar{z}} + (u_{k+2})_z = 0, \quad k = 0, 1, 2, \ldots. \tag{A.18}$$

Let $v = Pu = v_e \oplus v_o$, where $v_e = (v_0, v_2, \ldots)$ and $v_o = (v_1, v_3, \ldots)$. Then (A.18) can be rewritten as

$$v_{\bar{z}} + L^2 v_z = 0, \tag{A.19}$$

which splits into independent equations for v_e and v_o:

$$(v_e)_{\bar{z}} + L^2 (v_e)_z = 0, \quad (v_o)_{\bar{z}} + L^2 (v_o)_z = 0.$$

The boundary values of both vectors v_e and v_o at $\partial\Omega$ are known. According to (A.17), only the odd part v_o plays a role in the recovery of $f(x)$.

Let us introduce a generalized $\bar{\partial}$ operator:

$$D = \frac{\partial}{\partial \bar{z}} + L^2 \frac{\partial}{\partial z}.$$

The equation (A.19) is a generalized $\bar{\partial}$ equation for the function v. If one could prove uniqueness of its solution under given boundary conditions, this would immediately imply uniqueness of recovery of f, and hence uniqueness for the attenuated Radon transform. This would follow, for instance, from a Cauchy-type formula for such "analytic" vector-valued functions (if $A = L^2$, such functions are called A-analytic in [8, 25]). This is exactly what will be outlined in the next subsection.

A.3.2.3 *Cauchy-type formulas and inversion of Radon transform*

Let ω be a planar domain and u be an H^2-valued function. Then analogs of the standard considerations lead to a Cauchy–Green-type formula:

Theorem A.6 *If $u \in C^1(\overline{\omega}, H^2)$ and $z_0 \in \omega$, then*

$$u(z_0) = \frac{1}{2\pi i}\left\{\int_{\partial\omega} (dz - L^2 d\bar{z})\left(z - z_0 - L^2(\bar{z} - \bar{z_0})\right)^{-1} u(z)\right\} + \iint_{\omega} \left(z - z_0 - L^2(\bar{z} - \bar{z_0})\right)^{-1} Du(z) dz \wedge d\bar{z}. \tag{A.20}$$

Corollary A.7 *If $Du = 0$ (i.e. u is A-analytic), then*

$$u(z_0) = \frac{1}{2\pi i}\int_{\partial\omega} (dz - L^2 d\bar{z})\left(z - z_0 - L^2(\bar{z} - \bar{z_0})\right)^{-1} u(z). \tag{A.21}$$

This gives an alternative reconstruction procedure for the standard Radon transform:

(1) Find the boundary values for u_e and u_o.
(2) Find u_e and u_o using (A.21).
(3) Find f from v_o using (A.17).

Remark 1. *As one can read in [8, 25], the resolvent $\left(z - z_0 - L^2(\bar{z} - \bar{z_0})\right)^{-1}$ involved in the above formulas is formally not defined, since the point $\lambda = (z - z_0)/(\bar{z} - \bar{z_0})$ always belongs to the spectrum of the operator L^2. It can, however, be extended from the exterior of the disk $|\lambda| \leq 1$ as a strong limit and hence defined as an unbounded operator.*

Remark 2. *It is interesting to recognize the standard parts of the Radon inversion formula (Hilbert transform, differentiation, and back projection) in the procedure above. The continuation of the resolvent up to the spectrum gives the Hilbert transform, the integration in the Cauchy formula (A.21) corresponds to the back projection, and the differentiation is performed in (A.17).*

A.3.2.4 *Inversion of the attenuated transform*

We have dealt so far with the case of zero attenuation. Let us outline now how one can resolve the situation with the attenuated transform, i.e. when the transport equation has the form

$$Du + \mu u = f.$$

Availability of a Cauchy-type formula would have the same effect of producing the uniqueness result and inversion formulas as above. On a formal level, if we had a "function" h such that

$$Dh - \mu h = 0, \tag{A.22}$$

then Leibnitz's rule would imply that $D(hu) = hf$. This would mean that the attenuation is essentially eliminated, and the above Cauchy-type procedure is applicable to recover hf and hence f. It is easy to find one such function:

$$h(x,\omega) = e^{\mathcal{D}\mu} = e^{\int_0^\infty \mu(x+t\omega)dt}.$$

This, however, is not as simple as it sounds. Indeed, $h(x,\omega)$ for a fixed x acts as the Toeplitz operator $\phi \to P(h\phi)$ in the Hardy space, and hence we are dealing with an operator-valued function of x. On the other hand, the differential operator $D = \bar{\partial} + L^2\partial$ also has operator coefficients. In particular, D and h do not commute in general. This means that Leibnitz's rule does not work. If, however, one could find a solution of (A.22) that for each x has only nonpositive frequencies in its Fourier series with respect to ω, then h and D would commute and the procedure would work. This was implemented (in a not entirely explicit form) in [8], which produced the long-awaited proof of uniqueness for the attenuated Radon transform. However, applying more effort one can show that the procedure can be made explicit [91], which leads to a variety of inversion formulas, including the one obtained in a different way in [128] (see also a different derivation in [125]):

$$f(x) = -\frac{1}{4\pi} Re \operatorname{div} \int_{S^1} \omega e^{(\mathcal{D}\mu)(x,\omega^\perp)} \left(e^{-h} H e^{h} T_\mu f\right)(\omega, x\omega) d\omega, \tag{A.23}$$

where

$$h = \tfrac{1}{2}(Id + iH)R\mu.$$

This formula was implemented numerically in [58, 96, 98, 125].

A.3.3 *Range conditions*

We start with the case of the exponential transform, which is simpler than the one of the general attenuated transform and also leads to interesting analysis.

A.3.3.1 *Exponential transform*

Probably the first appearance of the range conditions for this transform was in the papers [12, 167] devoted to its inversion. We consider a compactly supported smooth function $f(x)$ in the plane and its exponential X-ray (Radon) transform with attenuation μ:

$$g(s,\omega,\mu) = \int_{-\infty}^{\infty} f(s\omega + t\omega^\perp) e^{\mu t} dt.$$

Here $\omega = (\cos\phi, \sin\phi)$ is a unit vector and $\omega^\perp = (-\sin\phi, \cos\phi)$. Now we expand $g(s,\omega,\mu)$ into the Fourier series with respect to the angle ϕ:

$$g(s,\omega,\mu) = \sum_l g_l(s,\mu) e^{il\phi}.$$

We consider the Fourier transform $\hat{g}_l(\sigma, \mu)$ of $g_l(s, \mu)$ with respect to s. It was observed in [12, 167] that the function

$$(\sigma + \mu)^l \hat{g}_l(\sigma, \mu)$$

is even with respect to σ for any $l \in \mathbf{Z}$. Papers [93] and [94] were devoted to finding the complete set of range conditions. It was shown in particular that the set of conditions described above is complete. Namely, the following theorem holds:

Theorem A.8 *A function $g(s, \omega)$ can be represented as $R_\mu f$ for some $f \in C_0^\infty(\mathbf{R}^2)$ if and only if*

(i) $g \in C_0^\infty(\mathbf{R} \times \mathbf{S}^1)$
(ii) function $(\sigma + \mu)^l \hat{g}_l(\sigma, \mu)$ is even with respect to σ for any $l \in \mathbf{Z}$.

An interesting corollary of the range condition provided in this theorem is that the function $\hat{g}_l(\sigma, \mu)$ has a zero of order $|l|$ at the point $\mu \times \operatorname{sgn} l$. Existence of these zeros, however, is obviously not equivalent to the whole set of conditions. One can check directly that (A.30) in the case of constant attenuation leads to the above-mentioned root conditions. This shows in particular that the range conditions (A.30) for the attenuated transform that will be discussed later are incomplete [93].

The range conditions of Theorem A.8 do not have the usual momentum form as they do in the case of the Radon transform. A momentum-type set of conditions was also found in [93] and [94].

Theorem A.9 *A function $g(s, \omega)$ can be represented as $R_\mu f$ for some $f \in C_0^\infty(\mathbf{R}^2)$ if and only if*

(i) $g \in C_0^\infty(\mathbf{R} \times \mathbf{S}^1)$
(ii) the following identity is satisfied for any odd natural n:

$$\sum_{k=0}^{n} \binom{n}{k} \frac{d}{d\phi} \circ \left(\frac{d}{d\phi} - i\right) \circ \cdots \circ \left(\frac{d}{d\phi} - (k-1)i\right) \int_{-\infty}^{\infty} (\mu s)^{n-k} g(s, \omega) ds = 0. \tag{A.24}$$

Here i is the imaginary unit, $\omega = (\cos\phi, \sin\phi)$, and $\circ$ denotes composition of differential operators.

The condition (A.24) is not very intuitive. One way we can try to understand it is the following. Let us assume that g is representable as $R_\mu f$ and plug $R_\mu f$ instead of g into (A.24). In the case of the standard Radon transform, this would lead to a condition that is obviously satisfied. In other words, for the Radon transform, necessity of the momentum conditions (as soon as they are formulated) is simple to observe, and only sufficiency requires a proof. However, even necessity of (A.24) is not obvious. A direct calculation shows that necessity of this condition is equivalent to the following series of identities for the function $\sin\phi$.

For any odd natural n

$$\sum_{k=0}^{n} \binom{n}{k} \left(\frac{d}{d\phi} - \sin\phi\right) \circ \left(\frac{d}{d\phi} - \sin\phi + i\right) \circ \cdots \circ \left(\frac{d}{d\phi} - \sin\phi + (k-1)i\right) \sin^{n-k}\phi = 0. \tag{A.25}$$

The reader might want to try to establish these identities directly. This was done in [93], where it was also shown that an analogous identity can be written in any commutative differential algebra. A similar identity for all natural numbers n holds for $\exp(x)$. Further discussion and reformulation of these conditions can be found in [131].

A better understanding of the meaning of the range conditions for the exponential Radon transform can be achieved using complex analysis. This was done in the series of papers [3, 4, 131, 132]. It was shown that the range description problem essentially reduces to Bernstein–Hartogs' type of theorems on extension of separately analytic functions (see Chapter 5 and further details later on in this chapter).

Consider again formula (A.12):

$$\widehat{R_\mu f}(\sigma, \omega) = \sqrt{2\pi}\tilde{f}(\sigma\omega + i\mu\omega^\perp).$$

If now we have a function $g(s, \omega)$ which we suspect of being in the range of the operator R_μ, then the function defined on S_μ as

$$F(\sigma\omega + i\mu\omega^\perp) = \hat{g}(\sigma, \omega) \tag{A.26}$$

must be extendable from S_μ to the whole $\mathbf{C}^2$ as an entire function of the Paley–Wiener class. If this is correct, then the extension provides the Fourier transform of a compactly supported function f such that $g = R_\mu f$. So, the question arises of extendability of the function defined on the left side of (A.26) to an entire function of the Paley–Wiener class. Such an extendability condition would provide a range description for the operator R_μ. In fact, such a condition can be easily found and has a simple geometric meaning.

Returning to the projection–slice formula (A.12), we notice that although we initially used the real values of σ only, we can in fact use all complex values of σ, provided that the function f is compactly supported. This leads to the determination of values of $\tilde{f}$ on a three-dimensional variety in $\mathbf{C}^2$. Let us denote this variety by Σ_μ:

$$\Sigma_\mu = \left\{ z = \sigma\omega + i\mu\omega^\perp | \sigma \in \mathbf{C}, \omega \in \mathbf{S}^1 \right\} \subset \mathbf{C}^2.$$

This variety is smooth at all points where $Re\ \sigma \neq 0$. At these points it has a complex tangent line. Thus, a boundary $\bar{\partial}$ equation must be satisfied if we want our function to be extendable. One can check, however, that this condition is automatically satisfied due to analyticity of $\hat{g}(\sigma, \omega)$ with respect to σ. So, where

is the extendability condition hidden? It turns out that the right place to look for it is at imaginary values of σ. If $\sigma = ip$ for a real p, then the definition of the function F looks as follows:

$$F(i(p\omega + \mu\omega^{\perp})) = \hat{g}(ip, \omega). \tag{A.27}$$

The line $p \to p\omega + \mu\omega^{\perp}$ is tangent at the point $\mu\omega^{\perp}$ to the circle of radius μ in the plane. Changing ω, we rotate this tangent line, spanning the whole exterior of the disk of radius μ. Let us also notice that the function F is entire with respect to p (i.e. along each tangent line). Since there are two such tangent lines passing through each point in the exterior of this disk, the formula (A.27) might provide a self-contradictory definition of the function F. Let us impose a condition that guarantees that this does not happen. That is, we assume that

$$\hat{g}(ip_1, \omega_1) = \hat{g}(ip_2, \omega_2) \tag{A.28}$$

whenever

$$p_1\omega_1 + \mu\omega_1^{\perp} = p_2\omega_2 + \mu\omega_2^{\perp}.$$

It turns out that this simple geometric condition is sufficient. That is, the following theorem holds.

Theorem A.10 *Let F be a function defined in the exterior of the disk of a positive radius in the plane. If this function is entire along each tangent line to the disk, it is extendable to an entire function of two variables.*

This theorem was proven (although not formulated explicitly in the present form) in [3]. The proof uses an analytic trick, whose role has not yet been well understood. For instance, the question remains whether one can replace tangent lines to a circle by tangent lines to a more general convex algebraic curve. This is certainly a separate analyticity theorem of Bernstein's kind [5, 18], since in a neighborhood of any point in the exterior of the disk the pairs of tangent lines provide a local coordinate system, for which we have separate analyticity. A proof that uses purely complex analytic tools was provided in [132]. It was shown in [136] and [131] that the strange moment conditions (A.24) can be understood as the infinitesimal form of (A.28) at the boundary of the disk.

The last theorem leads to the following range description [3]:

Theorem A.11 *Let $g(s, \omega)$ be a function on $\mathbf{R}\times S^1$, and μ be a positive real. Then $g = R_\mu f$ for some function $f \in C_0^\infty(\mathbf{R}^2)$ if and only if $g \in C_0^\infty(\mathbf{R}\times\mathbf{S}^1)$, and the Fourier transform $\hat{g}(\xi, \phi)$ with respect to the s variable of $g(s, \omega(\phi))$ satisfies the following condition for all real σ:*

$$\hat{g}\left(i\sigma, \phi - \arcsin\frac{\sigma}{\sqrt{\sigma^2+\mu^2}}\right) = \hat{g}\left(-i\sigma, \phi + \arcsin\frac{\sigma}{\sqrt{\sigma^2+\mu^2}}\right). \tag{A.29}$$

Here $\omega(\phi) = (\cos\phi, \sin\phi)$.

The condition (A.29) is a simple restatement of (A.28).

A generalization of this result to the multidimensional case is provided in [4]. A nice discussion can be found in [131] and [132].

Let us mention briefly some applications of these range conditions. It was shown in [137] how the conditions can be effectively used in SPECT in order to detect and correct some data errors arising from hardware imperfection. The paper [111] uses the range conditions in order to treat numerically incomplete data problems of SPECT. It is suggested in paper [32] that the range conditions might be used in some problems of radiation treatment planning (see also [31] and [88]). We will not go into details of these results.

A.3.3.2 *Attenuated transform*

The question about the range descriptions for the attenuated Radon transform turns out to be a tricky one (see also Section A.2). In fact, one could think that there is no chance of getting any explicit range conditions for the attenuated X-ray transform. Despite this natural belief, Natterer [124] found the following infinite set of such range conditions. Let us introduce some notation first. We denote by H the Hilbert transform on the line

$$Hp(x) = \frac{1}{\pi} v.p. \int_{-\infty}^{\infty} \frac{p(y)}{x-y} dy$$

where *v.p.* denotes the principal value integral (the limit as $\epsilon \to 0$ of the integral over $\mathbf{R} \setminus (-\epsilon, \epsilon)$).

Theorem A.12 [124] *Let f and μ belong to the Schwartz space $\mathcal{S}(\mathbf{R}^2)$. Then, for $k > m \geq 0$ integers, we have*

$$\int_{-\infty}^{\infty} \int_{0}^{2\pi} s^m e^{\pm ik\phi + 0.5(I \pm iH)R\mu(\omega,s)} T_\mu f(\omega, s) d\phi ds = 0, \tag{A.30}$$

where $\omega = (\cos\phi, \sin\phi)$, I is the identity operator, H is the Hilbert transform, and $R\mu$ is the Radon transform of μ.

As has already been explained, the set of conditions provided in this theorem is incomplete. Nevertheless, it is of practical importance, since it can be used for the simultaneous recovery of the sources $f(x)$ and attenuation $\mu(x)$ (see [121–123] and further discussion below).

The question had remained on how to write down the missing conditions, until the very recent publication [130], where a complete set of conditions was obtained.

It can be shown [91] that the approach of [8] provides an explanation and natural derivation of (A.30). In order to explain this, let us start with the case of zero attenuation (i.e. standard Radon transform) and look at it in terms of A-analytic functions (see Section A.3.2.2). In the notation of that section, the role of the Radon data is played by a function u on $\partial\Omega$ with values in the Hardy

space. It belongs to the range *iff* it is A-analytically extendable to Ω. Due to an analog of Sokhotsky's formula for such functions, it is easy to show that this happens *iff*

$$\int_{\partial\omega} (dz - L^2 d\bar{z}) \left(z - z_0 - L^2 (\bar{z} - \bar{z_0})\right)^{-1} u(z) = 0 \tag{A.31}$$

for any $z_o \notin \overline{\Omega}$. These conditions, however, do not resemble the standard momentum ones. How do we get those? In standard complex analysis, the boundary values of an analytic function as a measure on $\partial\Omega$ must be orthogonal to the boundary values of any function analytic in Ω and continuous up to the boundary. Any base of such functions would do. For instance, powers of z give

$$\int_{S^1} z^k u(z) dz = 0, \quad k = 0, 1, 2, \dots .$$

Similarly for A-analytic functions: (A.31) means orthogonality on the boundary to the operator functions

$$L_{z_o}(z) = \left(z - z_0 - L^2 (\bar{z} - \bar{z_0})\right)^{-1}.$$

Is there any other convenient basis? What are the analogs of the powers z^n here? The operator-valued function $H(z) = Rz - L\bar{z}$ satisfies $DH = 0$. It is an A-analytic analog of z. Due to the operator nature of H and D, the product Hu is not necessarily A-analytic. Namely,

$$D(Hu) = \left(\frac{\partial}{\partial \bar{z}} + L^2 \frac{\partial}{\partial z}\right) (Rz - L\bar{z})\, u(z) = [L^2, R] \frac{\partial u}{\partial z} z.$$

This expression is not necessarily equal to zero. However, the commutator $[L^2, R]$ equals $-P_1 L$, where P_l is the projector onto the first l terms of Fourier expansion. In particular, $L[L^2, R] = 0$. This implies that

$$L \int_{\partial\omega} (dz - L^2 d\bar{z}) Hu = 0$$

and analogously

$$L^k \int_{\partial\omega} (dz - L^2 d\bar{z}) H^k u = 0, \quad k = 0, 1, \dots .$$

Theorem A.13 [91] *These conditions coincide (modulo evenness) with the standard momentum conditions.*

How can one now include attenuation in this scheme? Using an appropriate "correction" function h as in Section A.3.2.2 and then applying the above conditions to hu one obtains the range conditions. In particular, one can obtain (A.30).

Theorem A.14 [91] *Orthogonality to the suitably understood functions $H^k h$ (i.e. analogs of orthogonality to z^k) gives Natterer's range conditions (A.30).*

We would like to mention again that a full set of range conditions was obtained in [130].

A.3.4 *Recovery of attenuation*

As we have discussed in the beginning of this section, simultaneous recovery of the source's density $f(x)$ and of the attenuation $\mu(x)$ is an important applied issue. At the first glance this problem might look hopeless. Indeed, in the plane case the data $g = T_\mu f$ is a single function of two variables, while we are trying to recover two functions $f(x)$ and $\mu(x)$ of two variables. However, this counting of functions and variables would be persuasive only if the operator T_μ were close to an invertible one (for instance, if it were a Fredholm operator). In fact, the study of the range conditions shows that the operator T_μ has an infinite dimensional cokernel. This means that when μ changes, the range could in principle rotate in such a way that for two different values of the attenuation the two ranges would have zero (or a "very small") intersection. If this were true, then both f and μ would be recoverable or "almost recoverable." For instance, some additional a priori information could help. We will describe now some results that show that something like this does happen, although the understanding of the situation is still very superficial.

Let us start with the simplest case of the exponential X-ray transform. The first attempt to recover both f and μ from $R_\mu f$ was made in [75]. This problem was resolved in [162] (see also [163]). Consider the range conditions in the form of evenness of the function

$$(\sigma + \mu)^l \hat{g}_l(\sigma, \mu).$$

This implies the equality

$$\left(\frac{\sigma + \mu}{-\sigma + \mu}\right)^l = \frac{\hat{g}_l(-\sigma, \mu)}{\hat{g}_l(\sigma, \mu)},$$

whenever $\hat{g}_l(\sigma, \mu)$ is not identically zero. This means that if there is a nonzero harmonic $\hat{g}_l(\sigma, \mu)$ for some $l \neq 0$, then the rational function

$$\frac{\sigma + \mu}{-\sigma + \mu}$$

is uniquely determined. This, in turn, determines μ uniquely as the pole of this function. The only exception arises when $\hat{g}_l(\sigma, \mu) = 0$ for all $l \neq 0$. In this case, however, the function $g = R_\mu f$, and hence f itself, is radial. For radial functions no information about the value of μ can be extracted from $R_\mu f$. Geometrically this means that the intersection of ranges of R_μ for different values of μ consists of radial functions. In other words, we established the following theorem.

Theorem A.15 [162, 163] *Simultaneous recovery of a compactly supported function f and constant attenuation μ from the values of $R_\mu f$ is possible if and only if f is not radial.*

In fact, in the non-radial case one can provide an algorithm of finding the attenuation μ (see [162] and [163]).

An alternative approach to recovery of attenuation was described in [7].

Recovery of a variable attenuation is definitely much more difficult. As in the exponential case, the range theorems are used to this end. Theorem A.12 has been used under some additional restrictions for the sources distribution in order to accomplish this task [121–123]. Papers [8, 128–130] contain some additional indications on how their techniques and results might be employed for that purpose.

A.4 Hyperbolic integral geometry and electrical impedance tomography

Electrical impedance tomography (EIT) is a very promising, inexpensive, and technologically simple (at least in comparison with X-ray, SPECT, PET, and MRI scans) method of medical diagnostics and of industrial nondestructive testing (see, for instance, [9, 10, 16, 17, 23, 28, 29, 47, 54, 77, 87, 104, 150, 154, 155, 164, 165] and references therein). Here is the idea of EIT: one places electrodes around a body, creates some known currents through the electrodes, and then measures the corresponding boundary voltage drops. All of the data is collected on the boundary. After many such measurements one wants to recover the distribution of the electric conductivity inside the body. The information about the electric conductivity is very important for medical diagnostics; it is also vital for some electrical procedures, such as defibrillation; it could also provide a cheap and reliable nondestructive evaluation technology. There are several strong groups in industry and academe that work on the practical and theoretical aspects of this problem using various techniques. We will consider here only one approach, which is related to the integral geometry problems discussed in the book. The reader can refer to the papers quoted above that address other methods that do not use the Radon transform.

Let us describe first the set-up of the inverse conductivity problem, which is the mathematical formulation of EIT. Let $U \subset \mathbf{R}^n$ be a domain with sufficiently smooth boundary Γ. An unknown function $\beta(x)$ (the conductivity) must be recovered from the following data. Given a function ψ on Γ (the current) one solves the Neumann boundary value problem

$$\begin{cases} \nabla \cdot (\beta \nabla u) = 0 \quad \text{in} \quad U \\ \beta \frac{\partial u}{\partial \nu}\big|_\Gamma = \psi, \end{cases}$$

where ν is the unit outer normal vector on Γ. Then one measures the boundary value $\phi = u|_\Gamma$ (the potential). All the pairs (ψ, ϕ) are assumed to be accessible.

In other words, the so-called Dirichlet-to-Neumann operator $\Lambda_\beta : \phi \to \psi$ is known. One needs to recover the conductivity β from this data. The problems is obviously nonlinear.

The inverse conductivity problem is a hard nut to crack both analytically and numerically. The main questions are about uniqueness of determination of the conductivity, stability of the reconstruction, and inversion algorithms. Although the problem of uniqueness can be considered as principally resolved (see [78, 119, 166, 170], and references therein), the other ones are still under thorough investigation. The general understanding is that the problem is highly unstable, so there is no hope of achieving the quality of reconstruction known to other common tomographic techniques, at least without involving additional information about the image to be reconstructed.

The first practical algorithm of Barber and Brown [9–11] deals with the linearized problem, which is still highly unstable. A thorough investigation of this algorithm started by Santosa and Vogelius [155] lead in the works of Berenstein and Casadio Tarabusi [16, 17] to the understanding that the linearized two-dimensional problem can be treated by means of hyperbolic geometry (see Section 7.3). Consider the two-dimensional case when U is the unit disk. It is well known (see for instance [13, 69]) that the unit disk serves as the Poincaré model of the hyperbolic plane $\mathbf{H}^2$. Here are the indications why the hyperbolic geometry might play some role in the (at least linearized) inverse conductivity problem. First of all, the Laplace operator that arises in the linearized problem is invariant with respect to the group of Möbius transformations. Another indication is that if one creates a dipole current through a point on the boundary of U, then the equipotential lines and the current lines form families of geodesics and horocycles in $\mathbf{H}^2$. Following the analysis done in [155] of the algorithm suggested in [9, 10], Berenstein and Casadio Tarabusi [16, 17] discovered that in fact for $n = 2$ the linearized inverse conductivity problem can be reduced to the following integral geometry problem on $\mathbf{H}^2$: the available data enables one to find the function

$$R_G(A * \beta),$$

where R_G is the geodesic Radon transform on $\mathbf{H}^2$, A is an explicitly described radial function on $\mathbf{H}^2$

$$A(r) = \text{const.}(3\cosh^{-4} r - \cosh^{-2} r),$$

and the asterisk $*$ denotes the (non-euclidean) convolution on $\mathbf{H}^2$. Now one can hope to use the known methods of harmonic analysis on $\mathbf{H}^2$ to recover β. That is, one needs to be able to invert the geodesic Radon transform and to deconvolve. Necessary details of harmonic analysis on $\mathbf{H}^2$ can be found in [68, 70]. Several different inversion formulas for the geodesic Radon transform are developed in [15, 69, 106]. The paper [52] provides a unified approach to such formulas. The reader might also be interested in the related papers [99, 101]. The formula obtained in [106] was numerically implemented in [45]. Another part of inversion would be deconvolution. Its numerical implementation can be done by using the

Fourier transform on the hyperbolic plane described in [68] and its inversion. A different approach to the Fourier transform is given in Chapter 7. The Fourier transform is defined as follows:

$$f(z) \to Ff(\lambda, b) = \int_{\mathbf{H}} f(z) e^{(-i\lambda+1)\langle z,b\rangle} dm(z),$$

where $b \in \partial\mathbf{H}^2$, $\lambda \in \mathbf{R}$, $\langle z, b\rangle$ is the (signed) hyperbolic distance from the origin to the horocycle passing through the points z and b, and $dm(z)$ is the invariant measure on $\mathbf{H}^2$. The inverse Fourier transform is

$$g(\lambda, b) \to F^{-1}g(z) = \text{const.} \iint g(\lambda, b) e^{(i\lambda+1)\langle z,b\rangle} \lambda \tanh(\pi\lambda/2) d\lambda db.$$

These operators were numerically implemented in [45]. (By an editorial error, all pictures have been omitted in [45]. They can be found at the URL http://www.math.tamu.edu/kuchment/hypnum.pdf.) Although the problem of numerical implementation of all these transforms looks familiar and similar to the euclidean case, it is in fact much more complex. In particular, even the problem of sampling is not a trivial one. One natural way of sampling would be the following. One chooses a discontinuous group of motions of the hyperbolic disk and uses the orbit of the origin as the grid. Choosing a small mesh, one samples with appropriate accuracy. Unfortunately, this idea does not work, due to the known rigidity of such lattices. Namely, the size of the pixel (i.e. the fundamental domain) for any discrete group acting on $\mathbf{H}^2$ cannot be made arbitrarily small [13]. The way of overcoming this difficulty chosen in [45] was to create a combination of euclidean and hyperbolic grids. A "superlattice" was created using a discrete group of hyperbolic motions, and then a finer euclidean grid was chosen in each of the fundamental domains.

It seemed at first that the hyperbolic geometry approach would not work in dimensions higher than two, due to lack of hyperbolic invariance of the governing equations. It was shown, however, in [46] that a combination of euclidean and hyperbolic integral geometries does the trick.

Bibliography

[1] M. Agranovsky, C. A. Berenstein, and P. Kuchment, Approximation by spherical waves in L^p-spaces, *J. Geom. Anal.* **6** (1996), 3, 365–383.

[2] M. Agranovsky and E. T. Quinto, Injectivity sets for Radon transform over circles and complete systems of radial functions, *J. Funct. Anal.* **139** (1996), 383–414.

[3] V. Aguilar, L. Ehrenpreis, and P. Kuchment, Range conditions for the exponential Radon transform, *J. d'Anal. Math.* **68** (1996), 1–13.

[4] V. Aguilar and P. Kuchment, Range conditions for the multidimensional exponential X-ray transform, *Inverse Probl.* **11** (1995), 977–982.

[5] N. I. Akhiezer and L. I. Ronkin, On separately analytic functions of several variables and theorems on "the thin edge of the wedge," *Russ. Math. Surv.* **28** (1973), 27–42.

[6] M. D. Altschuler, Reconstruction of the global-scale three-dimensional solar corona, in *Image Reconstruction from Projections, Implementation and Applications*, Ed. G.T. Herman, Topics in Applied Physics 32, 105–145, Springer-Verlag, New York/Berlin, 1979.

[7] Yu. E. Anikonov and I. Shneiberg, Radon transform with variable attenuation, *Dok. Akad. Nauk SSSR*, **316** (1991), 1, 93–95; English translation in *Sov. Math. Dokl.*

[8] E. V. Arbuzov, A.L. Bukhgeim, and S.G. Kasantzev, Two-dimensional tomography problems and the theory of A-analytic functions, *Siberian Adv. Math.* **8** (1998), 4, 1–20.

[9] D. C. Barber and B. H. Brown, Applied potential tomography, *J. Phys. E.: Sci. Instrum.* **17** (1984), 723–733.

[10] D. C. Barber and B. H. Brown, Recent developments in applied potential tomography-APT, in *Information Processing in Medical Imaging*, Nijhoff, Amsterdam, 1986, 106–121.

[11] D. C. Barber and B. H. Brown, Progress in electrical impedance tomography, in *Inverse Problems in Partial Differential Equations*, SIAM, Philadelphia, 1990, 151–164.

[12] S. Bellini, M. Piacentini, C. Caffario, and P. Rocca, Compensation of tissue absorption in emission tomography, *IEEE Trans. ASSP* **27** (1979), 213–218.

[13] A. Berdon, *The Geometry of Discrete Groups*, Springer-Verlag, New York–Heidelberg–Berlin, 1983.

[14] C. Berenstein, Radon transforms, wavelets, and applications, in *Integral Geometry, Radon Transforms and Complex Analysis (Venice, 1996)*, 1–33, Lecture Notes in Math. 1684, Springer-Verlag, Berlin, 1998.

[15] C. Berenstein and E. Casadio Tarabusi, Inversion formulas for the k-dimensional Radon transform in real hyperbolic spaces, *Duke Math. J.* **62** (1991), 3, 613–631.

[16] C. Berenstein and E. Casadio Tarabusi, The inverse conductivity problem and the hyperbolic X-ray transform, pp. 39–44 in [53].

[17] C. Berenstein and E. Casadio Tarabusi, Integral geometry in hyperbolic spaces and electrical impedance tomography, *SIAM J. Appl. Math.* **56** (1996), 755–764.

[18] S. Bernstein, Sur l'ordre de la meilleure approximation des fonctions continues par des polynomes de degre donne, Académie Royale de Belgique, Classe des Sciences, Mémoires Collection, **4** (1922), 1–103.

[19] G. Beylkin, The inversion problem and applications of the generalized Radon transform, *Commun. Pure Appl. Math.* **37** (1984), 579–599.

[20] J. Boman, An example of nonuniqueness for a generalized Radon transform. *J. Anal. Math.* **61** (1993), 395–401.

[21] J. Boman and E. T. Quinto, Support theorems for real analytic Radon transforms, *Duke Math. J.* **55** (1987), 4, 943–948.

[22] J. Boman and E. T. Quinto, Support theorems for real analytic Radon transforms on line complexes, *Trans. Am. Math. Soc.* **335** (1993), 877–890.

[23] L. Borcea, J. Berryman, and G. Papanicolaou, High-contrast impedance tomography, *Inverse Probl.* **12** (1996), 835–858.

[24] T. F. Budinger, G. T. Gullberg, and R.H. Huseman, Emission computed tomography, pp. 147–246 in [72].

[25] A. L. Bukhgeim, Inversion formulas in inverse problems, a supplement to M. M. Lavrent'ev and L. Ya. Savel'ev, *Linear Operators and Ill-posed Problems*, Translated from the Russian, Consultants Bureau, New York; Nauka, Moscow, 1995.

[26] A. L. Bukhgeim and S. G. Kazantsev, The attenuated Radon transform inversion formula for divergent beam geometry, Preprint, 2002.

[27] E. J. Candès and D. L. Donoho, Curvelets and reconstruction of images from noisy Radon data, in *Wavelet Applications in Signal and Image Processing VIII*, Eds. A. Aldroubi, A.F. Laine, M.A. Unser, *Proc. SPIE* 4119, 2000.

[28] M. Cheney, D. Isaacson, and J. C. Newell, Electrical impedance tomography, *SIAM Rev.* **41** (1999), 1, 85–101.

[29] B. Cipra, Shocking images from RPI, *SIAM News* July (1994), 14–15.

[30] A. M. Cormack, Representation of a function by its line integrals with some radiological applications, *J. Appl. Phys.* **34** (1963), 2722–2727.

[31] A. Cormack and E. T. Quinto, A problem in radiotherapy: questions of non-negativity, *Int. J. Imaging Syst. Technol.* **1** (1989), 120–124.

[32] A. Cormack and E. T. Quinto, The mathematics and physics of radiation dose planning, *Contemp. Math.* **113** (1990), 41–55.

[33] R. A. Crowther, D. J. DeRosier, and A. Klug, The reconstruction of a three-dimensional structure from projections and its application to electron microscopy, *Proc. R. Soc. London, Ser. A* **317** (1962), 319–340.

[34] R. Courant and D. Hilbert, *Methods of Mathematical Physics, Volume II Partial Differential Equations*, Interscience, New York, 1962.

[35] M. Davison, The ill-conditioned nature of the limited angle tomography problem, *SIAM J. Appl. Math.* **43** (1983), 428–448.

[36] J. De Stefano and T. Olsen, Wavelet localization of the Radon transform in even dimensions, *IEEE Trans. Signal Process* **42** (1994), 2055–2067.

[37] L. Ehrenpreis, P. Kuchment, and A. Panchenko, The exponential X-ray transform and Fritz John's equation. I. Range description, in *Analysis, Geometry, Number Theory: the Mathematics of Leon Ehrenpreis*

(Philadelphia, PA, 1998), 173–188, *Contemp. Math.* 251, American Mathematical Society, Providence, RI, 2000.

[38] A. Faridani, D. V. Finch, E. L. Ritman, and K. Smith, Local tomography II, *SIAM J. Appl. Math.* **57** (1997), 1095–1127.

[39] A. Faridani, E. L. Ritman, and K. T. Smith, Local tomography, *SIAM J. Appl. Math.* **52** (1992), 459–484; Examples of local tomography, *SIAM J. Appl. Math.* **52** (1992), 1193–1198.

[40] D. V. Finch, Cone beam reconstruction with sources on a curve, *SIAM J. Appl. Math.* **45** (1985), 665–673.

[41] D. V. Finch, Uniqueness for the X-ray transform in the physical range, *Inverse Probl.* **2** (1986), 197–203.

[42] D. V. Finch, Talk at Mt. Holyoke' *AMS/IMS/SIAM Summer Conference, June 18–22, 2000.*

[43] D. V. Finch, The attenuated X-ray transform: recent developments, to appear in *Inside–Out: Inverse Problems*, Ed. G. Uhlmann, Cambridge University Press, Cambridge.

[44] D. V. Finch and A. Hertle, The exponential Radon transform, *Contemp. Math.* **63** (1987), 67–74.

[45] B. Fridman, P. Kuchment, K. Lancaster, S. Lissianoi, D. Ma, M. Mogilevsky, V. Papanicolaou, and I. Ponomarev, Numerical harmonic analysis on the hyperbolic plane. *Appl. Anal.* **76** (2000), 3–4, 351–362.

[46] B. Fridman, P. Kuchment, D. Ma, and V. Papanicolaou, Solution of the linearized conductivity problem in the half space via integral geometry, in *Voronezh Winter Mathematical Schools*, Ed. P. Kuchment, 85–95, Amer. Math. Soc. Transl. Ser. 2, 184, American Mathematical Society, Providence, RI, 1998.

[47] A. Friedman and M. Vogelius, Determining cracks by boundary measurements, *Indiana Univ. Math. J.* **38** (1989), 527–556.

[48] I. Gelfand and S. Gindikin (Eds.), *Mathematical Problems of Tomography*, American Mathematical Society, Providence, RI, 1990.

[49] I. Gelfand, S. Gindikin, and M. Graev, Integral geometry in affine and projective spaces, *J. Sov. Math.* **18** (1980), 39–167.

[50] I. Gelfand, M. Graev, and N. Vilenkin, *Generalized Functions, Vol. 5: Integral Geometry and Representation Theory*, Academic Press, New York, 1966.

[51] S. Gindikin (Ed.), *Applied Problems of the Radon Transform*, American Mathematical Society, Providence, RI, 1994.

[52] S. Gindikin, Integral geometry on real quadrics, in *Lie groups and Lie algebras: E.B. Dynkin's Seminar*, 23–31, Amer Math. Soc. Transl. Ser. 2, 169, American Mathematical Society, Providence, RI, 1995.

[53] S. Gindikin and P. Michor (Eds.), *75 Years of the Radon Transform*, International Press, Cambridge, 1994.

[54] D. Gisser, D. Isaacson, and J. Newell, Current Topics in Impedance Imaging, *Clin. Phys. Physiol. Meas.*, **8** (1987), 216–241.

[55] J. Globevnik, Partial indices of analytic discs attached to Lagrangian submanifolds of C^N, *Ann. Inst. Fourier* **46** (1996), 5, 1305–1324.

[56] A. Greenleaf and G. Uhlmann, Non-local inversion formulas for the X-ray transform, *Duke Math. J.*, **58** (1989), 205–240.

[57] A. Greenleaf and G. Uhlmann, Microlocal techniques in integral geometry, *Contemp. Math.* **113** (1990), 149–155.

[58] J.-P. Guillement, F. Jauberteau, L. Kunyansky, R. Novikov, and R. Trebossen, On SPECT imaging based on an exact formula for the nonuniform attenuation correction, Preprint, 2002.

[59] V. Guillemin, Fourier integral operators from the Radon transform point of view, *Proc. Symp. Pure Math.*, **27** (1975) 297–300.

[60] V. Guillemin, Some remarks on Integral geometry, Manuscript, 1975.

[61] V. Guillemin and S. Sternberg, *Geometric Asymptotics*, American Mathematical Society, Providence, RI, 1977.

[62] V. Gullemin, On some results of Gelfand in integral geometry, *Proc. Symp. Pure Math.*, **43** (1985), 149–155.

[63] I. Hazou and D. Solmon, Inversion of the exponential Radon transform I, Analysis, *Math. Methods Appl. Sci.* **10** (1988), 561–574.

[64] I. Hazou and D. Solmon, Inversion of the exponential Radon transform II, Numerics, *Math. Methods Appl. Sci.* **13** (1990), 3, 205–218.

[65] I. Hazou and D. Solmon, Filtered-backprojection and the exponential Radon transform, *Math. Anal. Appl.* **141** (1989), 1, 109–119.

[66] U. Heike, Single-photon emission computed tomography by inverting the attenuated Radon transform with least-squares collocation, *Inverse Probl.* **2** (1986), 307–330.

[67] S. Helgason, *The Radon Transform*, Birkhäuser, Basle, 1980.

[68] S. Helgason, *Groups and Geometric Analysis*, Academic Press, New York, 1984.

[69] S. Helgason, The totally-geodesic Radon transform on constant curvature spaces, *Contemp. Math.* **113** (1990), 141–149.

[70] S. Helgason, *Geometric Analysis on Symmetric Spaces*, American Mathematical Society, Providence, RI, 1994.

[71] S. Helgason, The Radon transform on Euclidean spaces, compact two-point homogeneous spaces and Grassman manifolds, *Acta Math.* **113** (1965), 153–180.

[72] G. Herman (Ed.), *Image Reconstruction from Projections*, Topics in Applied Physics 32, Springer-Verlag, Berlin and New York, 1979.

[73] G. Herman, *Image Reconstruction from Projections*, Academic Press, New York, 1980.

[74] A. Hertle, On the injectivity of the attenuated Radon transform, *Proc. Am. Math. Soc.* **92** (1984), 201–205.

[75] A. Hertle, The identification problem for the constantly attenuated Radon transform, *Math. Z.* **197** (1988), 13–19.

[76] L. Hörmander, *The Analysis of Linear Partial Differential Operators I*, Springer-Verlag, New York, 1983.

[77] D. Isaacson and M. Cheney, Current problems in impedance imaging, in *Inverse Problems in Partial Differential Equations*, SIAM, Philadelphia, 1990, 141–149.

[78] V. Isakov, *Inverse Problems for Partial Differential Equations*, Applied Mathematical Sciences 127, Springer-Verlag, New York, 1998.

[79] F. John, The ultrahyperbolic differential equation with four independent variables, *Duke Math. J.* **4** (1938), 300–322.

[80] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, New York, 1988.

[81] A. Katsevich, Local tomography for the generalized Radon transform, *SIAM J. Appl. Math.* **57** (1997), 1128–1162.

[82] A. Katsevich, Local tomography with nonsmooth attenuation, *Trans. Am. Math. Soc.* **351** (1999), 5, 1947–1974.

[83] A. Katsevich, Cone beam local tomography, *SIAM J. Appl. Math.* **59** (1999), 2224–2246.

[84] A. Katsevich and A. G. Ramm, A method for finding discontinuities of functions from the tomographic data, pp. 115–124 in [150].

[85] A. Katsevich and A. G. Ramm, Pseudolocal tomography, *SIAM J. Appl. Math.* **56** (1996), 167–191.

[86] A. I. Katsevich and A. G. Ramm, A method for finding discontinuities from the tomographic data, pp. 115–123 in [150].

[87] R. Kohn and M. Vogelius, Determining conductivity by boundary measurements, I, *Commun. Pure Appl. Math.*, **37** (1984), 289–298; II, **38** (1985), 643–667.

[88] P. Kuchment, On positivity problems for the Radon transform and some related transforms, *Contemp. Math.*, **140** (1993), 87–95.

[89] P. Kuchment, On inversion and range characterization of one transform arising in emission tomography, pp. 240–248 in [53].

[90] P. Kuchment, Private communication, 1994.

[91] P. Kuchment, Lecture at the AMS Annual Meeting, New Orleans' 2001.

[92] P. Kuchment, K. Lancaster, and L. Mogilevskaya, On local tomography, *Inverse Probl.*, **11** (1995), 571–589.

[93] P. Kuchment and S. Lvin, Paley-Wiener theorem for exponential Radon transform, *Acta Appl. Math.* **18** (1990), 251–260.

[94] P. Kuchment and S. Lvin, The range of the exponential Radon transform, *Sov. Math. Dokl.* **42** (1991), 1, 183–184.

[95] P. Kuchment and I. Shneiberg, Some inversion formulas in the single photon emission tomography, *Appl. Anal.* **53** (1994), 221–231.

[96] L. Kunyansky, A new SPECT reconstruction algorithm based on the Novikov's explicit inversion formula, *Inverse Probl.* **17** (2001), 293–306.

[97] L. Kunyansky, Inversion of the exponential ray transform in 3-D using parallel-beam data measured from closed orbits, Preprint 2002.

[98] L. Kunyansky, Analytic reconstruction algorithms in emission tomography with variable attenuation, to appear in *J. Comput. Math. Sci. Eng.*

[99] A. Kurusa, The Radon transform on hyperbolic space, *Geom. Dedicata* **40** (1991), 2, 325–339.

[100] A. Kurusa, The invertibility of the Radon transform on abstract rotational manifolds of real type, *Math. Scand.* **70** (1992), 112–126.

[101] A. Kurusa, Support theorems for totally geodesic Radon transforms on constant curvature spaces, *Proc. Am. Math. Soc.* **122** (1994), 2, 429–435.

[102] I.-R. Lan, On an operator associated to a restricted x-ray transform, Ph.D. Thesis, Oregon State University, 2000.

[103] R. M. Lewitt and R. H. T. Bates, Image reconstruction from projections. **II:** Projection completion methods (theory), *Optik* **50** (1978), 189–204, **III:** Projection completion methods (computational examples) *Optik* **50** (1978), 269–278.

[104] V. Liepa, F. Santosa, and M. Vogelius, Crack determination from boundary measurements—reconstruction from experimental data, *J. Nondestr. Eval.* **12** (1993), 163–173.

[105] S. Lissianoi, On stability estimates in the exterior problem for the Radon transform, in *Tomography, Impedance Imaging, and Integral Geometry (South Hadley, MA, 1993)*, 143–147, Lectures in Appl. Math. 30, American Mathematical Society, Providence, RI, 1994.

[106] S. Lissianoi and I. Ponomarev, On the inversion of the geodesic Radon transform on the hyperbolic plane, *Inverse Probl.* **13** (1997), 1–10.

[107] A. Louis, Incomplete data problems in X-ray computerized tomography 1. Singular value decomposition of the limited angle transform, *Numer. Math.* **48** (1986), 251–262.

[108] A. Louis, Private communication, 1988.

[109] A. K. Louis and P. Maaß, Contour reconstruction in 3-D X-Ray CT, *IEEE Trans. Med. Imaging*, **12** (1993), 764–769.

[110] A. Louis and A. Rieder, Incomplete data problems in X-ray computerized tomography II. Truncated projections and region-of-interest tomography, *Numer. Math.* **56** (1989), 371–383.

[111] S. Lvin, Data correction and restoration in emission tomography, pp. 149–155 in [150].

[112] D. Ma, Personal communication, 1995.

[113] P. Maaß, 3D Röntgentomographie: Ein Auswahlkriterium für Abtastkurven, *Z. angew. Math. Mech.* **68** (1988), T498–T499.

[114] P. Maaß, The interior Radon transform, *SIAM J. Appl. Math.* **52** (1992), 710–724.

[115] W. Madych, Tomography, approximate reconstruction, and continuous wavelet transforms, *Appl. Comput. Harmon. Anal.* **7** (1999), 54–100.

[116] W. Madych and S. Nelson, Reconstruction from restricted Radon transform data: resolution and ill-conditionedness, *SIAM J. Math. Anal.* **17** (1986), 1447–1453.

[117] A. Markoe, Fourier inversion of the attenuated X-ray transform, *SIAM J. Math. Anal.* **15** (1984), 4, 718–722.

[118] A. Markoe and E.T. Quinto, An elementary proof of local invertibility for generalized and attenuated Radon transforms. *SIAM J. Math. Anal.* **16** (1985), 5, 1114–1119.

[119] A. Nachman, Global uniqueness for a two-dimensional inverse boundary value problem, *Ann. Math.* **143** (1996), 71–96.

[120] F. Natterer, Efficient implementation of "optimal" algorithms in computerized tomography, *Math. Methods Appl. Sci.* **2** (1980), 545–555.

[121] F. Natterer, On the inversion of the attenuated Radon transform, *Numer. Math.* **32** (1979), 431–438.

[122] F. Natterer, Computerized tomography with unknown sources, *SIAM J. Appl. Math.* **43** (1983), 1201–1212.

[123] F. Natterer, Exploiting the range of Radon transform in tomography, in: Eds. P. Deuflhard and E. Hairer, *Numerical Treatment of Inverse Problems in Differential and Integral Equations*, Birkhäuser, Basle, 1983.

[124] F. Natterer, *The Mathematics of Computerized Tomography*, Wiley, New York, 1986.

[125] F. Natterer, Inversion of the attenuated Radon transform, *Inverse Probl.* **17** (2001), 1, 113–119.

[126] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, Monographs on Mathematical Modeling and Computation 5, SIAM, Philadelphia, 2001.

[127] F. Noo and J.-M. Wagner, Image reconstruction in $2D$ SPECT with 180° acquisition, *Inverse Probl.*, **17** (2001), 1357–1371.

[128] R. Novikov, Une formule d'inversion pour la transformation d'un rayonnement X attenue, *CR Acad. Sci. Paris, Ser. I, Math.* **332** (2001), 12, 1059–1063.

[129] R. Novikov, An inversion formula for the attenuated X-ray transformation, *Ark. Math.* **40** (2002), 145–167.

[130] R. Novikov, On the range characterization for the two–dimensional attenuated X-ray transform, *Inverse Probl.* **18** (2002), 677–700.

[131] O. Öktem, Comparing range characterizations of the exponential Radon transform, Research Reports in Mathematics, 17, 1996, Stockholm University, Sweden.

[132] O. Öktem, Extension of separately analytic functions and applications to range characterization of the exponential Radon transform, in *Complex Analysis and Applications (Warsaw, 1997), Ann. Polon. Math.* **70** (1998), 195–213.

[133] V. Palamodov, Nonlinear artifacts in tomography, *Sov. Phys. Dokl.* **31** (1986), 888–890.

[134] R. M. Perry, On reconstruction of a function on the exterior of a disc from its Radon transform, *J. Math. Anal. Appl.* **59** (1977), 324–341.

[135] B. Petersen, *Introduction to the Fourier Transform and Pseudo-Differential Operators*, Pitman, Boston, 1983.

[136] I. Ponomarev, On reconstruction of projection data containing strong displacement distortions in emission gamma-tomography, in *Proc. Conf. on Ill-posed Problems of Mathematical Physics and Analysis, Novosibirsk, 1992*, p. 111.

[137] I. Ponomarev, Correction of emission tomography data. Effects of detector displacement and non-constant sensitivity, *Inverse Probl.*, **10** (1995) 1–8.

[138] N. G. Preobrazhensky and V.V. Pikalov, *Unstable Problems of Plasma Diagnostic*, Nauka, Novosibirsk, 1982.

[139] N. G. Preobrazhensky and V.V. Pikalov, *Reconstructive Tomography in Gas Dynamics and Plasma Physics*, Nauka, Novosibirsk, 1987.

[140] E. T. Quinto, The dependence of the generalized Radon transform on defining measures, *Trans. Am. Math. Soc.* **257** (1980), 331–346.

[141] E. T. Quinto, The invertibility of rotation invariant Radon transforms, *J. Math. Anal. Appl.*, **91** (1983), 510–522.

[142] E. T. Quinto, Tomographic reconstructions from incomplete data—numerical inversion of the exterior Radon transform, *Inverse Probl.* **4** (1988), 867–876.

[143] E. T. Quinto, Limited data tomography in non-destructive evaluation, in *Signal Processing Part II: Control Theory and Applications,* IMA Volumes in Mathematics and its Applications **23**, 347–354, Springer-Verlag, 1990.

[144] E. T. Quinto, Computed tomography and Rockets, in *Mathematical Methods in Tomography, Proceedings, Oberwolfach, 1990*, Lecture Notes In Mathematics 1497, 261–268, Springer-Verlag, Berlin, New York, 1991.

[145] E. T. Quinto, Pompeiu transforms on geodesic spheres in real analytic manifolds, *Isr. J. Math.* **84** (1993), 353–363.

[146] E. T. Quinto, Singularities of the X-ray transform and limited data tomography in $\mathbf{R}^2$ and $\mathbf{R}^3$, *SIAM J. Math. Anal.* **24** (1993), 1215–1225.

[147] E. T. Quinto, Radon transforms satisfying the Bolker assumption, pp. 245–250 in [53].

[148] E. T. Quinto, Exterior and limited angle tomography in non-destructive evaluation, *Inverse Probl.* **14** (1998), 339–353.

[149] E. T. Quinto, Private observation, 1998.

[150] E. T. Quinto, M. Cheney, and P. Kuchment (Eds.), *Tomography, Impedance Imaging, and Integral Geometry*, Lectures in Appl. Math. 30, American Mathematical Society, Providence, RI, 1994.

[151] G. N. Ramachandran and A.V. Lakshminarayanan, Three dimensional reconstruction from radiographs and electron micrographs: applications of convolutions instead of Fourier Transforms, *Proc. Natl Acad. Sci. USA* **68** (1971), 262–277.

[152] A. G. Ramm and A.I. Zaslavsky, Reconstructing singularities of a function given its Radon transform *Math. Comput. Modelling* **18** (1993), 109–138.

[153] F. Rashid-Farroukhi, K.J.R. Liu, C.A. Berenstein, and D.Walnut, Wavelet-based multiresolution local tomography, *IEEE Trans. Image Process.* **6** (1997), 1412–1430.

[154] F. Santosa, Inverse problem holds key to safe, continuous imaging, *SIAM News* July (1994), 1 and 16–18.

[155] F. Santosa and M. Vogelius, A backprojection algorithm for electrical impedance imaging, *SIAM J. Appl. Math.* **50** (1990), 216–241.

[156] L. A. Shepp and S. Srivastava, Computed tomography of PKM and AKM exit cones, *AT&T Tech. J.* **65** (1986), 78–88.

[157] I. Shneiberg, Exponential Radon transform, *Dokl. Akad. Nauk SSSR*, **320** (1991), 3, 567–571; English translation in *Sov. Math. Dokl.*

[158] I. Shneiberg, Exponential Radon transform, pp. 235–246 in [51].

[159] J. Siciak, Separately analytic functions and envelopes of holomorphy of some lower dimensional subsets of $\mathbf{C}^n$, *Ann. Polon. Math.*, **22** (1969), 145–171.

[160] K. Smith and F. Keinert, Mathematical foundations of computed tomography, *Appl. Opt.* **24** (1985), 3950–3957.

[161] K. T. Smith, D.C. Solmon, and S.L. Wagner, Practical and mathematical aspects of the problem of reconstructing objects from radiographs, *Bull. AMS* **83** (1977), 6, 1227–1270.

[162] D. Solmon, Two inverse problems for the exponential Radon transform, in *Inverse Problems in Action*, P.S. Sabatier, Ed. 46–53, Springer-Verlag, Berlin, 1990.

[163] D. Solmon, The identification problem for the exponential Radon transform, *Math. Methods Appl. Sci.* **18** (1995), 687–695.

[164] E. Somersalo, M. Cheney, D. Isaacson, and E. Isaacson, Layer-stripping: a direct numerical method for impedance imaging, *Inverse Probl.*, **7** (1991), 899–926.

[165] J. Sylvester, A convergent layer-stripping algorithm for the radially symmetric impedance tomography problem, *Commun. Partial Differ Equat.*, **17** (1992), 1955–1994.

[166] J. Sylvester and G. Uhlmann, The Dirichlet to Neumann map and its applications, in *Inverse Problems in Partial Differential Equations*, SIAM, Philadelphia, 1990, 101–139.

[167] O. J. Tretiak and C. Metz, The exponential Radon transform, *SIAM J. Appl. Math.* **39** (1980), 341–354.

[168] F. Treves, *Introduction to Pseudodifferential and Fourier integral operators II*, Plenum Press, New York, 1980.

[169] H. K. Tuy, An inversion formula for cone-beam reconstruction, *SIAM J. Appl. Math.* **43** (1983), 546–552.

[170] G. Uhlmann, Inverse boundary value problems and applications, *Asterisque* **207** (1992), 153–211.

[171] E. T. Vainberg, I. A. Kozak, and M. L. Faingois, Backprojection with second derivative filtration in Roentgen computer tomography, in *Linear and nonlinear problems of computer tomography*, Siberian Branch of Academy Sciences, Novosibirsk, Russia, 1985, 16–26.

[172] L. Zalcman, Offbeat integral geometry, *Am. Math. Mon.* **87** (1980), 161–175.

INDEX