Analysis and Mathematical Physics

Björn Gustafsson
Alexander Vasil'ev
Editors

Birkhäuser
Basel · Boston · Berlin

Editors:

Björn Gustafsson
Department of Mathematics
Royal Institute of Technology (KTH)
100 44 Stockholm
Sweden.
e-mail: gbjorn@math.kth.se

Alexander Vasil'ev
Department of Mathematics
University of Bergen
Johannes Brunsgate 12
5008 Bergen
Norway
e-mail: alexander.vasiliev@uib.no

2000 Mathematical Subject Classification: 14H; 28A; 30C,D,E,F; 31B,C; 32A; 34A; 35B,H,Q; 37F; 42A; 45C; 47A; 53C; 70K; 76D; 65E; 81Q; 82C

Library of Congress Control Number: 2009927175

Bibliographic information published by Die Deutsche Bibliothek. Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at http://dnb.ddb.de

ISBN 978-3-7643-9905-4 Birkhäuser Verlag AG, Basel - Boston - Berlin

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. For any kind of use permission of the copyright owner must be obtained.

© 2009 Birkhäuser Verlag AG
Basel · Boston · Berlin
P.O. Box 133, CH-4010 Basel, Switzerland
Part of Springer Science+Business Media
Printed on acid-free paper produced from chlorine-free pulp. TCF ∞
Cover Design: Alexander Faust, Basel, Switzerland
Printed in Germany

ISBN 978-3-7643-9905-4 e-ISBN 978-3-7643-9906-1

9 8 7 6 5 4 3 2 1 www.birkhauser.ch

Contents

Preface

This volume is based on lectures delivered at the international conference "New trends in harmonic and complex analysis", held May 7–12, 2007 in Voss, Norway, and organized by the University of Bergen and the Norwegian University of Science and Technology, Trondheim. It became the kick-off conference of the European Science Foundation Networking Programme "Harmonic and complex analysis and its applications" (2007–2012). The purpose of the Conference was to bring together both experts and novices in analysis with experts in mathematical physics, mechanics and adjacent areas of applied science and numerical analysis. The participants presented their results and discussed further developments of frontier research exploring the bridge between complex, real analysis, potential theory, PDE and modern topics of fluid mechanics and mathematical physics.

Harmonic and Complex Analysis is a well-established area in mathematics. Over the past few years, this area has not only developed in many different directions, it has also evolved in an exciting way at several levels: the exploration of new models in mechanics and mathematical physics and applications has at the same time stimulated a variety of deep mathematical theories.

During the last quarter of the twentieth century the face of mathematical physics changed significantly. One very important aspect has been the increasing degree of cross-fertilization between mathematics and physics with great benefits to both subjects. Whereas the goals and targets in the understanding of fundamental laws governing the structure of matter and energy are shared by physicists and mathematicians alike, the methods used, and even views on the importance and credibility of results, often differ significantly. In many cases, mathematical or theoretical predictions can be made in certain areas, but the physical basis (in particular that of experimental physics) for confirming such predictions remains out of reach, due to natural engineering, technological or economic limitations. Conversely, 'physical' reasoning often provides new insight and suggests approaches that transcend those that may be rigorously treated by purely mathematical analysis; physicists tend to 'jump' over apparent technical obstacles to arrive at conclusions based on physical insight that may form the basis for significant new conjectures. Mathematical analysis in a broad sense has proved to be one of the most useful fields for providing a theoretical basis for mathematical physics. On the other hand, physical insight in domains such as equilibrium problems in potential theory, asymptotics, and boundary value problems often suggests new avenues of approach.

We hope that the present volume will be interesting for specialists and graduate students specializing in mathematics and/or mathematical physics. Many papers in this volume are surveys, whereas others represent original research. We would like to acknowledge all contributors as well as referees for their great service for mathematical society. Special thanks go to Dr. Thomas Hempfling, Birkhäuser, for his kind assistance during preparation of this volume.

Björn Gustafsson
Alexander Vasil'ev
Stockholm-Bergen, 2009

Analysis and Mathematical Physics
Trends in Mathematics, 1–19
© 2009 Birkhäuser Verlag Basel/Switzerland

From $\mathrm{Diff}(S^1)$ to Univalent Functions. Cases of Degeneracy

Hélène Airault

Abstract. We explain in detail how to obtain the Kirillov vector fields $(L_k)_{k\in Z}$ on the space of univalent functions inside the unit disk. Following Kirillov, they can be produced from perturbations by vectors $(e^{ik\theta})_{k\in Z}$ of diffeomorphisms of the circle. We give a second approach to the construction of the vector fields. In our approach, the Lagrange series for the inverse function plays an important part. We relate the polynomial coefficients in these series to the polynomial coefficients in Kirillov vector fields. By investigation of degenerate cases, we look for the functions $f(z) = z + \sum_{n\geq 1} a_n z^{n+1}$ such that $L_k f = L_{-k} f$ for $k \geq 1$. We find that $f(z)$ must satisfy the differential equation:

$$\left[\frac{z^2 w}{1-zw} + \frac{zw}{w-z}\right] f'(z) - f(z) - \frac{w^2 f'(w)^2}{f(w)^2} \times \frac{f(z)^2}{f(w)-f(z)} = 0. \qquad (*)$$

We prove that the only solutions of $(*)$ are Koebe functions. On the other hand, we show that the vector fields $(T_k)_{k\in Z}$ image of the $(L_k)_{k\in Z}$ through the map $g(z) = \frac{1}{f(\frac{1}{z})}$ can be obtained directly as the (L_k) from perturbations of diffeomorphisms of the circle.

Mathematics Subject Classification (2000). Primary 17B68; Secondary 30C35.

Keywords. Reverted series, Koebe function, Kirillov vector fields.

1. Introduction

For $f(z) = z + a_1 z^2 + a_2 z^3 + \cdots$, Schiffer's procedure of elimination of terms in series [16] permits to construct the Kirillov vector fields $L_{-k} f(z)$ for a positive integer k. Let $z^{1-k} f'(z) = f(z)^{1-k}\left[1 + \sum_{j\geq 1} P_j^k f(z)^j\right]$ be the expansion of $z^{1-k} f'(z)$ in powers of $f(z)$, then $L_{-k} f(z) = \sum_{j\geq k+1} P_j^k f(z)^{1+j-k}$. If f is univalent, let $z = f^{-1}(u)$

The author thanks Paul Malliavin for discussions and for having introduced her to the classical book by A.C. Schaeffer and D.C. Spencer, Ref. [15]. Also thanks to Nabil Bedjaoui, Université de Picardie Jules Verne, for his help in the preparation of the manuscript.

in $z^{1-k}f'(z)$, then the expansion in powers of u of $L(u) = \dfrac{f^{-1}(u)^{1-k}}{(f^{-1})'(u)}$ is obtained with the derivative of the Lagrange expansion of $[f^{-1}(u)]^k$. This explains why Laurent expansions for inverse functions are important in the theory of Kirillov vector fields. On the other hand, for positive k, let $L_k f(z) = z^{1+k}f'(z)$ as in [10], [14]. We prove that $L_k f = L_{-k} f$ for any $k \in Z$ if and only if $f(z) = z/(1-\epsilon z)^2$, $\epsilon = 1$ or -1. We exhibit some of the many solutions of $\frac{df_t}{dt} = (L_{-k} - L_k)f_t$. However we can relate these solutions to the Koebe function only when $k = 1$. In sections two and three, we discuss expansions of powers of inverse functions and manipulations on these series. In section four, we relate the inverse series to the Kirillov vector fields and to diffeomorphisms of the circle. In section five, we calculate some of the flows associated to the vector fields (L_k). In section six, we consider the image (T_k) of the vector fields (L_k) under the map $f \to g$ where $g(z) = 1/f(1/z)$. For a univalent function $f(z)$, it is natural to consider $g(z) = \dfrac{1}{f(\frac{1}{z})}$, see for example [17], [5]. Then $z^{1-k}\dfrac{f'(z)}{f(z)} = v^{1+k}\dfrac{g'(v)}{g(v)}$. This leads us to consider expansions of $v^{1+k}g'(v)$ in powers of $g(v)$ for a function $g(v) = v + b_1 + \frac{b_2}{v} + \cdots$. We have $v^{k+1}g'(v) = g(v)^{1+k}\left[1 + \sum_{j\geq 1} V_j^{-k} g(v)^{-j}\right]$. The image vector fields (T_k) are given by $T_k g(z) = \sum_{j \geq k+1} V_j^{-k} g(u)^{-j}$. We compare the two families of vector fields (L_k) and (T_k), they have respectively the generating functions $A(f)$ and $B(g)$ where

$$A(\phi)(u,y) = \frac{\phi'(u)^2\phi(y)^2}{\phi(u)^2(\phi(u)-\phi(y))} = \frac{\phi'(u)^2}{\phi(u)-\phi(y)} - \frac{\phi'(u)^2}{\phi(u)} - \frac{\phi'(u)^2\phi(y)}{\phi(u)^2} \tag{1.1}$$

$$B(\phi)(u,y) = \frac{\phi'(u)^2\phi(y)}{\phi(u)(\phi(u)-\phi(y))} = \frac{\phi'(u)^2}{\phi(u)-\phi(y)} - \frac{\phi'(u)^2}{\phi(u)} \tag{1.2}$$

$$\textit{Let}\quad \psi(z) = \frac{1}{\phi(\frac{1}{z})} \quad \textit{then} \quad \frac{u^2}{\phi(y)^2}\, B(\phi)(u,y) = -\frac{1}{u^2}\, A(\psi)\left(\frac{1}{u},\frac{1}{y}\right). \tag{1.3}$$

In the last section seven, we examine degenerate cases for the vector fields (L_k) and (T_k).

2. Change of variables in series

2.1. Laurent series for $[g^{-1}(z)]^p$ with $g(z) = z + b_1 + \sum_{n\geq 1}\frac{b_{n+1}}{z^n}$ and for $[f^{-1}(z)]^p$ with $f(z) = z + \cdots + b_n z^{n+1} + \cdots$. Derivatives of the Laurent series

For $n \geq 0$, n integer, the Faber polynomial $F_n(z)$ of g is the polynomial part in the Laurent expansion of $[g^{-1}(z)]^n$ where g^{-1} is the inverse function of g, see [6], [5], [9] and [8]. For any $p \in C$, the Laurent expansions of $[g^{-1}(z)]^p$ and of $[f^{-1}(z)]^p$ can be obtained with the method of [2].

Definition 2.1. The homogeneous polynomials K_n^p, F_n and $G_n = K_n^{-1}$, $n \geq 1$, n integer, p a complex number, are defined with

$$\begin{cases} (1 + b_1 z + b_2 z^2 + \cdots)^p = 1 + \sum_{n\geq 1} K_n^p(b_1, b_2, \dots) z^n \\ \log(1 + b_1 z + b_2 z^2 + \cdots) = - \sum_{k\geq 1} \frac{F_k}{k}(b_1, b_2, \dots) z^k \\ \frac{1}{1 + b_1 z + b_2 z^2 + \cdots} = 1 + G_1 z + G_2 z^2 + \cdots \end{cases} . \tag{2.1}$$

Remark that $p = 0$ is a root of K_n^p as a polynomial in p, since for $n \geq 1$, $K_n^0 = 0$.

Lemma 2.2. *See* [2]. *Let* $g(z) = z + b_1 + \frac{b_2}{z} + \cdots + \frac{b_{n+1}}{z^n} + \cdots$ *and let* $p \in Z$*, then*

$$\begin{cases} (\frac{g(z)}{z})^p = 1 + \sum_{j\geq 1} H_j^p \frac{1}{g(z)^j} \\ \frac{zg'(z)}{g(z)} (\frac{g(z)}{z})^p = 1 + \sum_{j\geq 1} H_j^{p-j} \frac{1}{z^j} \end{cases} \quad \text{with the } \underline{\textbf{same}} \text{ coefficients } (H_j^p).$$

If $p \neq 0$, $H_j^{p-j} = (1 - \frac{j}{p}) K_j^p$ *. It extends for any* $p \in C$ *and* $\lim_{p\to 0} \frac{j}{p} K_j^p = -F_j$*. In particular,* $z\frac{g'(z)}{g(z)} = 1 + \sum_{j\geq 1} F_j \frac{1}{z^j}$ *with* $F_j = H_j^{-j}$*.*

Corollary 2.3. *Let* $g(z) = z + b_1 + \frac{b_2}{z} + \cdots + \frac{b_{n+1}}{z^n} + \cdots$ *and* $f(z) = z + b_1 z^2 + b_2 z^3 + \cdots + b_n z^{n+1} + \cdots$*, let* $p \in C$*. With the convention that* $\frac{p}{p-n} K_n^{n-p}$ *is* F_p *if* $p = n$ *and* $\frac{p}{n+p} K_n^{-(n+p)}$ *is equal to* F_p *if* $n + p = 0$*, we have*

$$\begin{cases} [g^{-1}(z)]^p \; = \; z^p \left[1 \; + \; \sum_{n\geq 1} \frac{p}{p-n} K_n^{n-p} \frac{1}{z^n} \right] \\ [f^{-1}(z)]^p \; = \; z^p \left[1 \; + \; \sum_{n\geq 1} \frac{p}{p+n} K_n^{-(n+p)} z^n \right] . \end{cases} \tag{2.2}$$

Corollary 2.3 generalizes: Let $h(z) = 1 + b_1 z + b_2 z^2 + \cdots$. We put $f(z) = zh(z)$ and $g(z) = zh(\frac{1}{z})$. Define the maps $E_p : f(z) \to E_p(f)(z) = z[h(z)]^p$ with $p \neq 0$, $p \in C$ and Inv : $f(z) = z[h(z)] \to \phi(f)(z) = f^{-1}(z)$, the inverse of f and compositions of these maps. Then $E_k \, o \, E_p = E_{kp}$ and Inv o Inv $= Id$.

Lemma 2.4. *Let* $(k_j)_{1\leq j\leq s}$ *be a finite sequence,* $k_j \neq 0$*. then*

$$\phi(f)(z) = [E_{k_s} \, o \text{ Inv } o \, E_{k_{s-1}} \, o \text{ Inv } o \cdots o \, E_{k_1} \, o \text{ Inv } o \, E_{k_0}](f)(z) =$$

$$z \left[1 + \sum_{n\geq 1} \frac{A_n(k_1, k_2, \dots, k_s)}{\beta_n(k_1, k_2, \dots, k_s)} \, K_n^{-k_0 \beta_n(k_1, k_2, \dots, k_s)} \, z^n \right]$$

where $A_n(k_1, k_2, \dots, k_s)$ *and* $\beta_n(k_1, k_2, \dots, k_s)$ *do not depend on the coefficients of* $f(z)$*. If* $\beta_n(k_1, k_2, \dots, k_s) = 0$*, we replace the corresponding term in the expansion by* $(-1)^{s-1} k_0 \times \prod_{j=1}^{s} k_j \times \frac{1}{n} F_n \, z^n$*.*

A similar result holds for $g(z)$ by defining $E_p : g(z) \to E_p(g)(z) = z[h(\frac{1}{z})]^p$ with $p \neq 0$, $p \in C$ and Inv : $g(z) \to \phi(g)(z) = g^{-1}(z)$, the inverse of g. The next expansions will be important. We take the derivative with respect to z in the series of Corollary 2.3, the denominators $(n+p)$ disappear and for any $p \in C$,

$$\begin{cases} \frac{(g^{-1})'(y)}{[g^{-1}(y)]^{1-p}} = \frac{1}{y^{1-p}} \left[1 + \sum_{n\geq 1} K_n^{n-p} \frac{1}{y^n}\right] \\ \frac{(f^{-1})'(y)}{[f^{-1}(y)]^{1-p}} = \frac{1}{y^{1-p}} \left[1 + \sum_{n\geq 1} K_n^{-(n+p)} y^n\right]. \end{cases} \tag{2.3}$$

2.2. Expansions of $z^2(\frac{f'(z)}{f(z)})^2 (\frac{f(z)}{z})^p$, of $\frac{[f^{-1}(y)]^{1-p}}{(f^{-1})'(y)}$ and of $\frac{[g^{-1}(y)]^{1-p}}{(g^{-1})'(y)}$

In [1], see (A.1.2), (A.7.1), the following lemma is used to prove the identity on the polynomials coefficients of the Schwarzian derivative $L_{-k}P_p - L_{-p}P_k = (k-p)P_{p+k}$. With a change of variables in the Cauchy integral, we prove

Lemma 2.5. *See* [1]. *Let* $f(z) = z + \sum_{n\geq 1} c_n z^{n+1}$, *then*

$$\begin{cases} \text{(i)} \quad (\frac{zf'(z)}{f(z)})^2 (\frac{f(z)}{z})^p = 1 + \sum_{n\geq 1} P_n^{n+p} z^n \\ \text{(ii)} \quad \frac{zf'(z)}{f(z)} (\frac{f(z)}{z})^p = 1 + \sum_{n\geq 1} P_n^p f(z)^n \end{cases} \quad \textit{with the } \underline{\textbf{same}} \textit{ polynomials } P_n^p.$$

Lemma 2.6. *See* [2]. *Let* $g(z) = z + b_1 + \frac{b_2}{z} + \frac{b_3}{z^2} + \cdots$, *then*

$$\begin{cases} z\frac{g'(z)}{g(z)} (\frac{g(z)}{z})^k = 1 + \sum_{j\geq 1} V_j^k \frac{1}{g(z)^j} \\ (z\frac{g'(z)}{g(z)})^2 (\frac{g(z)}{z})^k = 1 + \sum_{j\geq 1} V_j^{k-j} \frac{1}{z^j} \end{cases} \quad \textit{with the } \underline{\textbf{same}} \textit{ polynomials } V_j^k.$$

The polynomials P_n^p and V_j^k are homogeneous in the variables (c_j), respectively (b_j), they can be calculated with differential operators as in [1] or with binomial analysis as in [6]. In the following, we relate them to the (K_n^p).

Proposition 2.7. *The polynomials* (P_n^p) *in Lemma* 2.5 *satisfy*

$$\frac{[f^{-1}(y)]^{1-p}}{(f^{-1})'(y)} = y^{1-p} [1 + \sum_{n\geq 1} P_n^p y^n] = \frac{y^{1-p}}{1 + \sum_{n\geq 1} K_n^{-(n+p)} y^n} \tag{2.4}$$

$$P_n^p = G_n(K_1^{-(1+p)}, K_2^{-(2+p)}, \ldots, K_n^{-(n+p)}). \tag{2.5}$$

Proof. We put $z = f^{-1}(y)$ in Lemma 2.5 and we use (2.3). We have

$$G_n(P_1^p, P_2^p, \ldots, P_n^p) = K_n^{-(n+p)} \ \forall n \geq 1. \tag{2.6}$$

Moreover the map G on the manifold of coefficients is involutive (GoG = Identity). □

Proposition 2.8. *The polynomials* (V_j^k) *in Lemma* 2.6 *satisfy*

$$\frac{[g^{-1}(y)]^{1-k}}{y^{1-k}(g^{-1})'(y)} = 1 + \sum_{j\geq 1} V_j^k \frac{1}{y^j} = \frac{1}{1 + \sum_{n\geq 1} K_n^{n-k} \frac{1}{y^n}} \tag{2.7}$$

$$V_n^k = G_n(K_1^{1-k}, K_2^{2-k}, \dots, K_n^{n-k}) \tag{2.8}$$

$$V_n^k(b_1, b_2, \dots, b_n) = P_n^{-k}(G_1(b_1), G_2(b_1, b_2), \dots, G_n(b_1, b_2, \dots, b_n)). \tag{2.9}$$

Moreover

$$\frac{\partial}{\partial b_1} V_{j+1}^p \;=\; (p+j-1)\, V_j^p. \tag{2.10}$$

Proof. Let $z = g^{-1}(y)$ in Lemma 2.6, then $\frac{1}{k}\frac{d}{dy}[(g^{-1})(y)]^k = \frac{y^{k-1}}{1+\sum_{j\geq 1} V_j^k y^{-j}}$. We remark that $\frac{1}{k}\frac{d}{dy}[(g^{-1})(y)]^k = y^{k-1}[1 + \sum_{n\geq 1} K_n^{n-k}\frac{1}{y^n}]$. To prove that $V_n^k = P_n^{-k} oG$, we use $K_n^{k-n} oG = K_n^{n-k}$. To obtain (2.10), we remark that $g'(z)$ does not depend on b_1, then we differentiate with respect to b_1,

$$z^{1+k} g'(z) = \sum_{j\geq 0} V_j^{-k} g(z)^{1+k-j}$$

with $V_0^k = 1$ and we identify equal powers of $g(z)$. □

Writing $g'(z)^2$, we find $V_p^{2-p} = K_n^2(0, -b_2, -2b_3, -3b_4, \dots)$. We have $V_1^p = (p-1)b_1$, $V_2^p = (p-2)b_2 + \frac{(p-1)p}{2} b_1^2 = (p-2)b_2 + p\int V_1^p db_1$, $V_3^p = (p-3)b_3 + (p-2)(p+1)b_1 b_2 + \frac{(p-1)p(p+1)}{3!} b_1^3 = (p-3)b_3 + (p+1)\int V_2^p db_1$, $V_4^p = (p-4)b_4 + \frac{p^2-p-4}{2} b_2^2 + (p+2)\int V_3^p\, db_1$, $V_5^p = (p-5)b_5 + (p^2-p-8)b_2 b_3 + (p+3)\int V_4^p\, db_1$, $V_6^p = (p-6)b_6 + (p^2-p-14)b_2 b_4 + \frac{p^2-p-12}{2} b_3^2 + \frac{(p+4)(p^2-p-6)}{3!} b_2^3 + (p+4)\int V_5^p db_1$.

2.3. The expansion of $z^{k+1}(\frac{f'(z)}{f(z)})^{k+1}(\frac{f(z)}{z})^p$

Lemma 2.9. *Let $f(z) = z + b_1 z^2 + \cdots$. For any k and p, then*

$$\begin{cases} (\frac{zf'(z)}{f(z)})^{k+1}(\frac{f(z)}{z})^p \;=\; 1 + \sum_{n\geq 1} J_n^{n+p}(k) z^n \\ (\frac{zf'(z)}{f(z)})^{k}(\frac{f(z)}{z})^p \;=\; 1 + \sum_{n\geq 1} J_n^{p}(k)\, f(z)^n \end{cases} \quad \textit{with the } \underline{\textbf{same}}\ J_n^p(k).$$

We have $P_n^p = J_n^p(1)$ and $J_n^p(k)$ is a polynomial of the two variables (p,k), its coefficients can be calculated with the method of [6]. The expansion of the Schwarzian derivative $Sf(z) = \frac{f'''(z)}{f'(z)} - \frac{3}{2}(\frac{f''(z)}{f'(z)})^2$ is related to the expansions of Lemma 2.9 since $z^2 S(f^p)(z) = -\frac{(p^2-1)}{2} \times \frac{z^2 f'(z)^2}{f(z)^2} + z^2 Sf(z)$ and $S(f^{-1})(f(z)) = -Sf(z)/f'(z)^2$. For $f(z) = z + b_1 z^2 + \cdots + b_n z^{n+1} + \cdots$, what is the expansion $z^2 Sf(z) = \sum_{n\geq 2} \lambda_n f(z)^n$ in sum of powers of $f(z)$?

3. Taylor series in the Laurent expansions of inverse functions

In this section, we consider Taylor series inside the Lagrange series for inverse functions.

3.1. Polynomials associated to $g(z) = z + b_1 + \frac{b_2}{z} + \cdots$

The inverse function g^{-1}, i.e., $g^{-1}\, o\, g$ = Identity has for expansion

$$g^{-1}(z) = z - b_1 - \sum_{n\geq 1} \frac{1}{n} K_{n+1}^{n} \frac{1}{z^n}$$

and when p is integer, $p \geq 1$,

$$[g^{-1}(z)]^p = z^p + \sum_{1\leq k\leq p-1} \frac{p}{p-k} K_k^{k-p}\, z^{p-k} + F_p - \sum_{n\geq 1} \frac{p}{n} K_{n+p}^{n} \frac{1}{z^n}. \tag{3.1}$$

With the Taylor formula on the polynomial part of $[g^{-1}(z)]^p$, see [5] and [4], we obtain

$$F_p(z) = z^p + \sum_{1\leq k\leq p-1} \frac{p}{p-k} K_k^{k-p} z^{p-k} + F_p \tag{3.2}$$

$$\frac{1}{(p-j)!} \frac{\partial^{p-j}}{\partial z^{p-j}} F_p(z)_{|z=0} = \frac{p}{p-j}\, K_j^{j-p}\ \forall p \geq j. \tag{3.3}$$

Let $z = g(u)$ in (3.1), it gives $u^p = F_p(g(u)) - \sum_{n\geq 1} \frac{p}{n} K_{n+p}^{n} \frac{1}{g(u)^n}$. *We define the Grunsky coefficients* β_{pq} *by* $F_p(g(u)) = u^p + \sum_{q\geq 1} \beta_{pq} \frac{1}{u^q}$. Then, see [2], [7], $\frac{1}{p}\beta_{pq} = \sum_{n=1}^{q-1} \frac{1}{n} K_{n+p}^{n} K_{q-n}^{-n} = \sum_{n=1}^{p-1} \frac{1}{n} K_{n+q}^{n} K_{p-n}^{-n}$.

Proposition 3.1. *The particularity of* $F_p(z)$ *and of the series*

$$S_p(z) = \sum_{n\geq 1} \frac{p}{n} K_{n+p}^{n}(b_1, b_2, \ldots, b_{n+p}) \frac{1}{z^n} \tag{3.4}$$

is that as functions of b_1 *and* z*, they depend only on* $b_1 - z$*,*

$$F_p(z) = F_p(z; b_1, b_2, \ldots, b_p) = F_p(0; b_1 - z, b_2, \ldots, b_p) \tag{3.5}$$

$$S_p(z) = \sum_{n\geq 1} \frac{p}{n} K_{n+p}^{n}(0, b_2, \ldots, b_{n+p}) \frac{1}{(z-b_1)^n}. \tag{3.6}$$

Moreover $K_{p+q}^{q}(b_1, b_2, \ldots, b_{p+q})$ *is a polynomial of degree* $q-1$ *in* b_1*,*

$$K_{p+q}^{q}(b_1, b_2, \ldots, b_{p+q}) = \sum_{1\leq n\leq q} \frac{p}{n} C_{n-1}^{q-1} K_{p+n}^{n}(0, b_2, \ldots, b_{p+n})\, b_1^{q-n}. \tag{3.7}$$

Proof. To prove that $g^{-1}(z)$ as function of (b_1, z) and the other variables is a function of $b_1 - z$ and the other variables, we take the derivative with respect to b_1 in $g^{-1}(g(z)) = z$. Since $\frac{\partial g(z)}{\partial b_1} = 1$, we obtain $(\frac{\partial}{\partial b_1} g^{-1})(g(z)) + (g^{-1})'(g(z)) = 0$. Thus $\frac{\partial}{\partial b_1} g^{-1} + (g^{-1})' = 0$. We expand (3.6) in powers of $\frac{b_1}{z}$ to obtain (3.7) □

3.2. The polynomial part of $\frac{[g^{-1}(y)]^k}{(g^{-1})'(y)}$

If $k \geq 0$, then $\frac{[g^{-1}(y)]^k}{(g^{-1})'(y)} = y^k[1+\sum_{j\geq 1} V_j^{1-k}\,\frac{1}{y^j}]$ has a non zero polynomial part and it depends only on $y - b_1$. In analogy with the Faber polynomials, if $k \geq 0$, let $\phi_k(z)$ be the polynomial part of

$$\frac{[g^{-1}(z)]^k}{(g^{-1})'(z)} = z^k \left[1 + \sum_{j=1}^{k} V_j^{1-k}\,\frac{1}{z^j}\right]. \tag{3.8}$$

Then

$$\phi_k(0) = V_k^{1-k}(b_1, b_2, \ldots, b_k) \tag{3.9}$$

$$\phi_k(z; b_1, b_2, \ldots, b_k) = \phi_k(0; b_1 - z, b_2, \ldots, b_k) = V_k^{1-k}(b_1 - z, b_2, \ldots, b_k) \tag{3.10}$$

$\phi_0(z) = 1$, $\phi_1(z) = z - b_1$, $\phi_2(z) = (z-b_1)^2 - 3b_2$, $\phi_3(z) = (z-b_1)^3 - 4(z-b_1)b_2 - 5b_3$. The generating function of $(\phi_n(z))_n$ is

$$\frac{zg'(z)^2}{g(z) - w} = 1 + \sum_{n\geq 1} \phi_n(w)\frac{1}{z^n}. \tag{3.11}$$

$\phi_n(w)$ is the unique polynomial such that $\phi_n(g(w)) = w^n g'(w) + \sum_{k\geq 1} \gamma_{nk}\frac{1}{w^k}$.

$$\phi_n(z) = F_n(z) - b_2F_{n-2}(z) - 2b_3F_{n-3}(z) - (n-1)b_n \tag{3.12}$$

$$V_{k+p}^{1-k}(b_1, b_2, \ldots) = \sum_{n=0}^{p-1} V_{k+p-n}^{1-k}(0, b_2, \ldots)\frac{(p-1)!}{n!(p-1-n)!}\,b_1^n. \tag{3.13}$$

The subseries $U(z; b_1, b_2, \ldots) = \sum_{j\geq 1} V_{k+j}^{1-k}\,\frac{1}{z^j}$ in the Laurent series of $\frac{[g^{-1}(z)]^k}{(g^{-1})'(z)}$ is a function of $b_1 - z$ and

$$\begin{cases} U(z) = \sum_{j\geq 1} V_{k+j}^{1-k}(b_1, b_2, \ldots)\,\frac{1}{z^j} = \sum_{j\geq 1} V_{k+j}^{1-k}(0, b_2, \ldots)\,\frac{1}{(z-b_1)^j} \\ \qquad = \sum_{j\geq 1, n\geq 0} V_{k+j}^{1-k}(0, b_2, \ldots)\,\frac{(n+j-1)!}{n!(j-1)!}b_1^n\frac{1}{z^{n+j}}. \end{cases} \tag{3.14}$$

3.3. Polynomials related to $f(z) = z + b_1z^2 + \cdots$

When p is a positive integer, $\frac{1}{f^{-1}(z)^p}$ is equal to

$$\frac{1}{z^p} + \sum_{1\leq k\leq p-1} \frac{p}{p-k}K_k^{p-k}\frac{1}{z^{p-k}} - F_p - \sum_{n\geq 1}\frac{p}{n}K_{n+p}^{-n}z^n = H_p(\frac{1}{z}) - \sum_{n\geq 1}\frac{p}{n}K_{n+p}^{-n}z^n$$

where $H_p(u) = u^p + \sum_{1\leq k\leq p-1}\frac{p}{p-k}K_k^{p-k}\,u^{p-k} - F_p$. With $z = f(u)$, it gives

$$H_p(\frac{1}{f(u)}) = \frac{1}{u^p} + \sum_{n\geq 1}\frac{p}{n}K_{n+p}^{-n}f(u)^n = \frac{1}{u^p} + \sum_{q\geq 1}\gamma_{pq}u^q. \tag{3.15}$$

Given k integer, $k \geq 1$, one find a unique sequence of homogeneous polynomials $(Q_p)_{p\geq 1}$ in the variables $b_1, b_2, \ldots$ such that

$$\begin{aligned} &f(z)^k - Q_1 f(z)^{k+1} - Q_2 f(z)^{k+2} - \cdots - Q_p f(z)^{k+p} - z^k \\ &= Q_{p+1} z^{k+p+1} + \text{ higher terms in } z^j \qquad for\ j \geq k+p+2. \end{aligned} \tag{3.16}$$

We have [2, p. 349] $Q_1(b_1) = kb_1,\ Q_2(b_1, b_2) = kb_2 - \frac{k(k+3)}{2} b_1^2,\ Q_3(b_1, b_2, b_3) = kb_3 - k(k+4)b_1 b_2 + \dfrac{k(k+4)(k+5)}{3!} b_1^3, \ldots, Q_n(b_1, b_2, \ldots, b_n) = -\dfrac{k}{k+n} K_n^{-(n+k)}$.

4. Diffeomorphisms of the circle and expansions of inverse functions

4.1. Left invariant and right invariant vector fields on $\mathrm{Diff}(S^1)$

Given $\theta \to \chi(\theta)$, then $\exp(\epsilon\chi)$ is defined by

$$\theta \to \exp(\epsilon\chi(\theta)) = \theta + \epsilon\chi(\theta) + \frac{\epsilon^2}{2} \chi \, o \, \chi(\theta) + \cdots + \frac{\epsilon^n}{n!} \chi \, o \, \chi \ldots o \, \chi(\theta) + \cdots \tag{4.1}$$

Let γ be a diffeomorphism of the circle, $\theta \to \gamma(\theta)$. For small $\epsilon > 0$, we consider

$$\gamma_\epsilon^l(\theta) = \exp(\epsilon\chi) \, o \, \gamma(\theta) = \gamma(\theta) + \epsilon\chi \, o \, \gamma(\theta) + O(\epsilon^2) \quad \text{and } L_\gamma = \frac{d}{d\epsilon}_{|\epsilon=0} \gamma_\epsilon^l = \chi \, o \, \gamma(\theta)$$

$$\gamma_\epsilon^r(\theta) = \gamma \, o \, \exp(\epsilon\chi)(\theta) = \gamma(\theta + \epsilon\chi(\theta) + O(\epsilon^2)) \quad \text{and} \quad R_\gamma = \frac{d}{d\epsilon}_{|\epsilon=0} \gamma_\epsilon^r = \gamma'(\theta)\chi(\theta).$$

Then $\left(\frac{d}{d\epsilon}_{|\epsilon=0} \gamma_\epsilon^l\right)(\gamma^{-1}(u))$ and $\dfrac{1}{\gamma'(u)} \dfrac{d}{d\epsilon} \gamma_\epsilon^r(u)$ are independent of γ.

$$\left(\frac{d}{d\epsilon}_{|\epsilon=0} \gamma_\epsilon\right)(\gamma^{-1}(u)) = -\frac{1}{(\gamma^{-1})'(u)} \frac{d}{d\epsilon}_{|\epsilon=0} \gamma_\epsilon^{-1}(u). \tag{4.2}$$

4.2. From γ to the univalent functions f and g such that $f \, o \, \gamma = g$. The Laurent series L

To γ, we associate g univalent from the exterior of the unit disk and f univalent from the interior of the unit disk such that on the circle $|z| = 1$, we have $f \, o \, \gamma(z) = g(z)$. For γ_ϵ, we have $\gamma_\epsilon = f_\epsilon^{-1} \, o \, g_\epsilon$. Following [10],

$$\frac{d}{d\epsilon} \gamma_\epsilon(z) = \left(\frac{d}{d\epsilon} f_\epsilon^{-1}\right)(g_\epsilon(z)) + (f_\epsilon^{-1})'(g_\epsilon(z)) \times \frac{d}{d\epsilon} g_\epsilon(z). \tag{4.3}$$

Thus $\left(\frac{d}{d\epsilon}\gamma_\epsilon\right)(\gamma_\epsilon^{-1}(z)) = \left(\frac{d}{d\epsilon} f_\epsilon^{-1}\right)(f_\epsilon(z)) + (f_\epsilon^{-1})'(f_\epsilon(z)) \times \left(\frac{d}{d\epsilon} g_\epsilon\right)(g_\epsilon^{-1}(f_\epsilon(z)))$. We divide by $(f_\epsilon^{-1})'(f_\epsilon(z))$ and put $y = f_\epsilon(z)$.

Definition 4.1. Let

$$\begin{cases} \mathbf{L} = \dfrac{1}{(f_\epsilon^{-1})'(y)} \left(\frac{d}{d\epsilon}\gamma_\epsilon\right)(\gamma_\epsilon^{-1}(f_\epsilon^{-1}(y))) \\ \quad = \dfrac{1}{(g_\epsilon^{-1})'(y)} \times \dfrac{1}{\gamma'(g_\epsilon^{-1}(y))} \left(\frac{d}{d\epsilon}\gamma_\epsilon\right)(g_\epsilon^{-1}(y)). \end{cases} \tag{4.4}$$

We have $\mathbf{L} = \dfrac{1}{(f_\epsilon^{-1})'(y)} \times \frac{d}{d\epsilon} f_\epsilon^{-1}(y) + \left(\frac{d}{d\epsilon} g_\epsilon\right)(g_\epsilon^{-1}(y) = \mathbf{L}_f + \mathbf{L}_g$ with

$$\mathbf{L}_f = \frac{1}{(f_\epsilon^{-1})'(y)} \times \frac{d}{d\epsilon} f_\epsilon^{-1}(y) \quad \text{and} \quad \mathbf{L}_g = \left(\frac{d}{d\epsilon} g_\epsilon\right)(g_\epsilon^{-1}(y)). \tag{4.5}$$

Proposition 4.2.

$$\textit{If} \ \left(\frac{d}{d\epsilon}_{|\epsilon=0} \gamma_\epsilon\right)(\gamma^{-1}(u)) = \chi(u) = -u^{1-k}, \ \textit{then} \ \mathbf{L} = -\frac{[f^{-1}(y)]^{1-k}}{(f^{-1})'(y)} \tag{4.6}$$

$$\textit{If} \ \frac{1}{\gamma'(u)} \times \left(\frac{d}{d\epsilon}_{|\epsilon=0} \gamma_\epsilon(u)\right) = \chi(u) = -u^{1-k}, \ \textit{then} \ \mathbf{L} = -\frac{[g^{-1}(y)]^{1-k}}{(g^{-1})'(y)}. \tag{4.7}$$

This shows the importance of the polynomials P_n^p and V_j^k in section 3. As in [12], the vector fields L_f can be related to the variational formulae on univalent functions of Goluzin and Schiffer. See also the related works [3] and [13]. In the following, we shall adopt the point of view of asymptotic expansions. The vector fields do not preserve $Diff(S^1)$ and we shall not study whether the vector fields associated to L_f or L_g as in (4.5) preserve the univalence of f or g. They are simply related to the series of sections 2 and 3. They induce vector fields on the set of functions f and g of the form (**I**) or (**II**) where

$$(\mathbf{I}): \quad \begin{cases} f(z) = z + a_1 z^2 + \cdots + a_n z^{n+1} + \cdots \\ g(z) = c_0 z + c_1 + \frac{c_2}{z} + \cdots + \frac{c_n}{z^{n-1}} + \cdots \end{cases} \tag{4.8}$$

$$(\mathbf{II}): \quad \begin{cases} f(z) = a_0 z + a_1 z^2 + \cdots + a_n z^{n+1} + \cdots \\ g(z) = z + c_1 + \frac{c_2}{z} + \cdots + \frac{c_n}{z^{n-1}} + \cdots . \end{cases} \tag{4.9}$$

Taking the perturbation γ_ϵ of γ, we split differently the Laurent expansion of $(\frac{d}{d\epsilon}\gamma_\epsilon)(\gamma_\epsilon^{-1}(z))$ according to what we take **(I)** or **(II)** as normalization for f and g in the decomposition $f\,o\,\gamma = g$. With **(I)**, we obtain the vector fields (L_k) and the generating function (1.1) and with **(II)**, we obtain the vector fields (T_k) with generating function (1.2). This is explained below.

5. Kirillov vector fields (L_k)

With **(I)**, $f(z) = z + \sum_{n\geq 1} b_n z^{n+1}$. Let $f_\epsilon(z)$ such that

$$f_\epsilon(z) = z + \sum_{n\geq 1} b_n(\epsilon) z^{n+1} \quad \textit{and} \quad f_0(z) = f(z). \tag{5.1}$$

The Taylor series of $\frac{d}{d\epsilon} f_\epsilon(z)$ start with a term in z^2 since $\frac{d}{d\epsilon} f_\epsilon(z) = \sum_{n\geq 2} b_n'(\epsilon) z^n$. The Taylor series $\mathbf{L}_f$ start with a term in z^2. On the other hand, $\mathbf{L}_g$ is of the form

$\alpha z +$ a sum of terms in z^n with $n \leq 0$. In this section, we assume (4.6) and $\chi(u) = -u^{1-k}$. We put $P_0^0 = 1$. **For** $k \geq 0$, $\mathbf{L} = -\frac{[f^{-1}(y)]^{1-k}}{(f^{-1})'(y)}$ is given by

$$\mathbf{L} = \frac{1}{y^{k-1}}\left[1 + \sum_{n\geq 1} P_n^k y^n\right]$$
$$= -\left[\frac{1}{y^{k-1}} + P_1^k \frac{1}{y^{k-2}} + \cdots + P_{k-1}^k + P_k^k\, y + P_{k+1}^k y^2 + \cdots\right].$$

In the decomposition (4.5),

$$\begin{cases} \mathbf{L}_f = \frac{1}{(f_\epsilon^{-1})'(y)}\,\frac{d}{d\epsilon} f_\epsilon^{-1}(y) \;=\; -\sum_{n\geq 1} P_{k+n}^k y^{n+1} \\ \quad\;\; = -\frac{[f^{-1}(y)]^{1-k}}{(f^{-1})'(y)} \;-\; \left(\frac{d}{d\epsilon} g_\epsilon\right)(\, g_\epsilon^{-1}(y)\,) \end{cases} \tag{5.2}$$

$$\mathbf{L}_g \;=\; \left(\frac{d}{d\epsilon} g_\epsilon\right)(\, g_\epsilon^{-1}(y)\,) \;=\; -\,[\,\frac{1}{y^{k-1}} + P_1^k \frac{1}{y^{k-2}} + \cdots + P_{k-1}^k + P_k^k\, y\,]. \tag{5.3}$$

If $\mathbf{k} \leq \mathbf{0}$, we have $\mathbf{L} = \mathbf{L}_f$, we obtain

Proposition 5.1. *If* $\left(\frac{d}{d\epsilon}_{|\epsilon=0}\gamma_\epsilon\right)(\,\gamma^{-1}(u)) \;=\; \chi(u) \;=\; -\,u^{1+k}$ *and* $k \geq 1$*, then*

$$\frac{d}{d\epsilon} f_\epsilon^{-1}(y) \;=\; -\,[f_\epsilon^{-1}(y)]^{k+1}. \tag{5.4}$$

In accordance with (5.2)–(5.3),

Definition 5.2. Let $f(z) = z + b_1 z^2 + b_2 z^3 + \cdots + b_n z^{n+1} + \cdots$. We define the left vector fields $[\,L_k f\,](z)$ for $k \in Z$ with

$$L_{-k} f(z) = z^{1-k} f'(z) - \sum_{j=0}^{k} P_{k-j}^k f(z)^{1-j} = \sum_{j=0}^{+\infty} P_{1+j+k}^k f(z)^{j+2} \quad for \;\; k \geq 0 \tag{5.5}$$

$$L_k f(z) = z^{1+k} f'(z) \quad for \quad k \geq 1. \tag{5.6}$$

5.1. Expression of the vector fields $(L_k)_{k\in Z}$ on the manifold of coefficients

We replace $f(z)^{1-j} = z^{1-j} \sum_{n\geq 0} K_n^{1-j}\, z^n$ in (5.5), it gives

$$[\,L_{-k} f\,](z) \;=\; \sum_{n\geq 1} \,[\sum_{k=1}^{n} P_{k+p}^p K_{n-k}^{k+1}]\; z^{n+1}. \tag{5.7}$$

Since $\frac{\partial}{\partial b_n}[f(z)] = z^{n+1}$, from (5.5)–(5.6), we deduce, see [10], [1],

$$\begin{cases} L_{-p} = \sum_{n\geq 1} A_n^p \frac{\partial}{\partial b_n} \quad with \quad A_n^p = \sum_{k=1}^n P_{k+p}^p K_{n-k}^{k+1} \quad \forall\; p \geq 0 \\ L_p = \frac{\partial}{\partial b_p} + 2b_1 \frac{\partial}{\partial b_{p+1}} + \cdots + (n+1) b_n \frac{\partial}{\partial b_{p+n}} + \cdots \quad for \quad p \geq 1. \end{cases} \tag{5.8}$$

5.2. Integration of $\frac{d}{dt}f_t(z) = L_{-k}f_t(z)$ for positive integer k

We treat the cases $k = 1$ and $k = 2$, the method extends to arbitrary k. If $f(z) = z + b_1z^2 + b_2z^3 + \cdots$, we have $L_0f(z) = zf'(z) - f(z)$,

$$L_1f(z) = z^2f'(z) \quad \textit{and} \qquad L_{-1}f(z) = f'(z) - 1 - 2a_1f(z) \tag{5.9}$$

$$L_2f(z) = z^3f'(z) \quad \textit{and} \qquad L_{-2}f(z) = \frac{f'(z)}{z} - \frac{1}{f(z)} - 3a_1 + (a_1^2 - 4a_2)f(z) \tag{5.10}$$

$$L_{-3}f(z) = \frac{f'(z)}{z^2} - \frac{1}{f(z)^2} - \frac{4a_1}{f(z)} - (a_1^2 + 5a_2) - (6a_3 - 2a_1a_2)f(z) \tag{5.11}$$

$$L_{-4}f(z) = \frac{f'(z)}{z^3} - \frac{1}{f(z)^3} - \frac{5a_1}{f(z)^2} - (4a_1^2 + 6a_2)\frac{1}{f(z)} - (7a_3 + 4a_1a_2 - a_1^3)$$
$$- (8a_4 - 2a_1a_3 - 2a_1^2a_2 + a_1^4)f(z).$$

We see that $f_t(z) = f(\frac{z}{1+tz})$ is solution of $\frac{d}{dt}f_t = L_1f_t$ and the solutions of $\frac{d}{dt}f_t = L_kf_t$ for positive k are given in [10], [1], [11],

$$f_t(z) = \frac{z}{(1 - tkz^k)^{1/k}}. \tag{5.12}$$

On the other hand, for L_{-1}, the function $f_t(z) = \dfrac{f(z+t) - f(t)}{f'(t)}$ is a solution of

$$\frac{d}{dt}f_t(z) = f_t'(z) - 1 - f_t''(0)\, f_t(z). \tag{5.13}$$

The function g_t associated to f_t as in (5.2) is given by $\frac{d}{dt}g_t(z) = -1 - f"_t(0)g_t(z)$. Since $f_t''(0) = \dfrac{f''(t)}{f'(t)}$, we obtain

$$g_t(z) = \frac{g(z) - f(t)}{f'(t)}. \tag{5.14}$$

Let γ be a homographic transformation, $e^{i\gamma(\theta)} = \gamma(e^{i\theta}) = \dfrac{e^{i\theta} + b}{1 + \overline{b}e^{i\theta}} = \gamma(z)$ with $z = e^{i\theta}$, then $fo\gamma = g$ with $f(z) = \dfrac{z}{1 - \overline{b}z}$ and $g(z) = \dfrac{z+b}{1 - b\overline{b}}$. When t varies,

$$f_t(z) = \frac{z(1 - \overline{b}t)}{1 - \overline{b}(t+z)} \tag{5.15}$$

$f_t^{-1}(u) = \dfrac{u(1 - \overline{b}t)}{1 - \overline{b}(t-u)}$, $g_t(z) = \dfrac{(1 - \overline{b}t)^2}{1 - b\overline{b}}\, z + \dfrac{(b-t)(1 - \overline{b}t)}{1 - b\overline{b}}$. We see that $\gamma_t = f_t^{-1}\, o\, g_t$ is the homographic transformation $\gamma_t(z) = \dfrac{(1 - \overline{b}t)z + b - t}{1 + \overline{b}z}$, but γ_t does not transform the unit circle to itself.

The example $k=2$. Since $P_1^2 = 3b_1 = \frac{3}{2} f''(0)$, $P_2^2 = 4b_2 - b_1^2 = \frac{4}{3!} f'''(0) - \frac{f''(0)^2}{4}$, see [1] (A.1.9), we have to integrate

$$\frac{d}{dt} f_t(z) = \frac{f_t'(z)}{z} - \frac{1}{f_t(z)} - \frac{3}{2} f_t''(0) + \left(\frac{f''(0)^2}{4} - \frac{4}{3!} f'''(0) \right) f_t(z)$$

where f_t', f_t'', f_t''' are the derivatives with respect to z of f_t.
For any $k \geq 1$, the $f_t(z)$ defined by

$$f_t(z)^k = \frac{\phi(\tau + z^k) - \phi(\tau)}{\phi'(\tau)} = z^k + \frac{\phi''(\tau)}{2\phi'(\tau)} z^{2k} + \cdots$$

is a solution of $\frac{d}{dt} f_t = L_{-k} f_t$ if $\tau = kt$. It is not difficult to find other solutions of this equation. For example,

Lemma 5.3. *Let* $f_t(z) = z + a_1(t) z^2 + a_2(t) z^3 + \cdots$ *such that*

$$f_t(z)^k - k a_1(t) f_t(z)^{k+1} = z^k. \tag{5.16}$$

If $a_1(t)$ *satisfies the differential equation*

$$\frac{d}{dt} a_1(t) = [(k+1) a_1(t)]^{k+1} \tag{5.17}$$

then $f_t(z)$ *is a solution of* $\frac{d}{dt} f_t(z) = L_{-k} f_t(z)$ *for* $k \geq 1$.

Proof. By identification of the coefficients of equal powers of z, the other coefficients $a_n(t)_{n \geq 2}$ are uniquely determined by $f_t(z)^k - k a_1(t) f_t(z)^{k+1} = z^k$. With (2.1), $K_{n+1}^k = k a_1 K_n^{k+1}$ for $n \geq 1$. Thus $f_t(z)$ depends only on $a_1(t)$. To prove that $f_t(z)$ is a solution of $\frac{d}{dt} f_t(z) = L_{-k} f_t(z)$, it is enough to verify that

$$(k f_t(z)^{k-1} - k(k+1) a_1 f_t(z)^k) \frac{d}{dt} f_t(z) = (k f_t(z)^{k-1} - k(k+1) a_1 f_t(z)^k) L_{-k} f(z). \tag{5.18}$$

We have $(k f_t(z)^{k-1} - k(k+1) a_1 f_t(z)^k) \frac{d}{dt} f_t(z) = k f_t(z)^{k+1} \frac{d}{dt} a_1(t)$ and

$$(k f_t(z)^{k-1} - k(k+1) a_1 f_t(z)^k) f_t'(z) = k z^{k-1}.$$

Thus (5.18) is the same as

$$k f_t(z)^{k+1} \frac{d}{dt} a_1(t) = k - (k f_t(z)^{k-1} - k(k+1) a_1 f_t(z)^k) \sum_{j=0}^{k} P_{k-j}^k f_t(z)^{1-j}.$$

By identifying equal powers of $f(z)$ in this last equation, it is satisfied if

$$(k+1) a_1 P_{k-j}^k = P_{k+1-j}^k \quad for \quad 1 \leq j \leq k. \tag{5.19}$$

This is a consequence of $z^{1-k} f'(z) f(z)^{k-1} = 1 + \sum_{j \geq 1} P_j^k f(z)^j$, see Lemma 2.5. Also $\frac{d}{dt} a_1(t) = (k+1) a_1 P_k^k$. □

Lemma 5.3 extends to a more general class of solutions. According to (3.16), for $p \geq 1$, one can find differential equations for the coefficients $a_1(t)$, $a_2(t)$, $\ldots$, $a_p(t)$ such that a solution of

$$f_t(z)^k + \sum_{j=1}^{p} \frac{k}{k+j} K_j^{-(k+j)} f_t(z)^{k+j} = z^k \tag{5.20}$$

is a solution of $\frac{d}{dt} f_t(z) = L_{-k} f_t(z)$ for $k \geq 1$.

5.3. Integration of $\frac{d}{dt} f_t = (L_{-k} - L_k) f_t$ for $k \in Z$

There exist numerous solutions of $\dfrac{d}{dt} f_t = (L_{-k} - L_k) f_t$. One may think that some of those solutions interact with boundaries of Schlicht regions as studied in [15]. We shall prove in the last section that the Koebe function f satisfies $L_{-k} f = L_k f$ for any integer $k \geq 1$. However we obtain the Koebe function as solution of $f(z)^k = \dfrac{f(u(z)) - f(\tau)}{(1-\tau^2) f'(\tau)}$ only when $k = 1$. Below, we give some examples of solutions.

Lemma 5.4. *The differential equation $\frac{d}{dt} f_t = (L_{-1} - L_1) f_t$ can be written as*

$$\frac{d}{dt} f_t(z) = (1 - z^2) f_t'(z) - 1 - f_t''(0) f_t(z). \tag{5.21}$$

Let $\tau(t)$ such that $\frac{d\tau}{dt} = 1 - \tau^2$ and let

$$h_t(z) = \frac{f(\frac{z+\tau}{1+\tau z}) - f(\tau)}{(1-\tau^2) f'(\tau)}. \tag{5.22}$$

The function $h_t(z)$ satisfies $\dfrac{d}{dt} h_t(z) = (L_{-1} - L_1) h_t(z)$. If $f(z) = \dfrac{z}{(1-z)^2}$, then $h_t(z)$ is equal to $\dfrac{z}{(1-z)^2}$, thus it is independent of t. Conversely if $h_t(z)$ is independent of t, then f is a Koebe function or $f(z) = \frac{1}{2} \ln(\dfrac{1+z}{1-z})$.

Let $k \geq 1$,

$$L_{-k} f(z) = z^{1-k} f'(z) - f(z)^{1-k} - 2k a_k f(z) \ \ if \ \ f(z) = z + a_k z^{k+1} + a_{2k} z^{2k+1} + \cdots . \tag{5.23}$$

Proposition 5.5. *For k integer, $k \geq 1$, $\dfrac{d}{dt} f_t = (L_{-k} - L_k) f_t$ is of the form*

$$\frac{d}{dt} f_t = \frac{1 - z^{2k}}{z^{k-1}} \frac{d}{dz} f_t \; - \; f_t(z)^{1-k} \times \sum_{j=0}^{k} P_j^k f_t(z)^j. \tag{5.24}$$

For $k = 2$, $\frac{d}{dt} f_t = (L_{-2} - L_2) f_t$ is the same as

$$\frac{d}{dt} f_t(z) = \frac{1 - z^4}{z} f_t'(z) - \frac{1}{f_t(z)} - \frac{3}{2} f_t''(0) - \left(\frac{4}{3!} f'''(0) - \frac{f''(0)^2}{4} \right) f_t(z). \tag{5.25}$$

Let $u(z) = \dfrac{z^k + \tau}{1 + \tau z^k}$. *For any function* $\phi(u)$ *having a Taylor expansion at* $u = \tau$ *such that* $\phi'(\tau) \neq 0$, *then* $f_t(z)$ *given by*

$$f_t(z)^k = \frac{\phi(u(z)) - \phi(\tau)}{(1 - \tau^2)\phi'(\tau)} \tag{5.26}$$

is a solution of $\frac{d}{dt} f_t = (L_{-k} - L_k) f_t$ *if* $\frac{d\tau}{dt} = k(1 - \tau^2)$.

For $\phi(u) = \dfrac{u}{(1-u)^2}$, then $f_t(z) = \dfrac{z}{(1 - z^k)^{\frac{2}{k}}}$ and for $\phi(u) = u^2$, $f_t(z)$ is given by $f_t(z)^k = \dfrac{(1+\tau^2)z^{2k} + 2\tau z^k}{2\tau(1 + \tau z^k)^2}, \ldots$.

Proof. We obtain (5.24) since $\frac{d}{dt} f_t = (L_{-k} - L_k) f_t$ is of the form

$$\frac{d}{dt} f_t = \frac{1 - z^{2k}}{z^{k-1}} \frac{d}{dz} f_t \; - \; f_t(z) \times \sum_{j=0}^{k} P^k_{k-j} \frac{1}{f(z)^j}.$$

With (2.4), $\sum_{j=0}^k P_j^k f^j \times \sum_{n \geq 0} K_n^{-(n+k)} f^n = 1 - P_{k+1}^k f^{k+1} +$ *terms in* f^j, $j \geq k+2$. Now, we prove (5.26). We have $\dfrac{du}{dz} = \dfrac{k z^{k-1}(1 - \tau^2)}{(1 + \tau z^k)^2}$ and $\dfrac{du}{d\tau} = \dfrac{1 - z^{2k}}{(1 + \tau z^k)^2}$. The expansion $f_t(z)$ in powers of z is obtained as follows

$$\phi(u) - \phi(\tau) = \phi'(\tau)(u - \tau) + \frac{\phi''(\tau)}{2}(u - \tau)^2 + \cdots$$

$$u = \tau\left(1 + \frac{z^k}{\tau}\right)(1 + \tau z^k)^{-1} = \tau\left(1 + \left(\frac{1}{\tau} - \tau\right) z^k \; + \; (\tau^2 - 1) z^{2k} + \cdots\right)$$

$u - \tau = (1 - \tau^2)(z^k - \tau z^{2k} + \cdots$ thus

$$f_t(z)^k = z^k + \left[-\tau \; + \; \frac{(1 - \tau^2)\phi''(\tau)}{2\phi'(\tau)}\right] z^{2k} + \cdots = z^k[1 + k a_k z^{k+1} + \cdots].$$

This gives $f_t(z) = z + a_k z^{k+1} + \cdots$ with $k\, a_k = -\tau + \dfrac{(1-\tau^2)\phi''(\tau)}{2\phi'(\tau)}$. Taking logarithmic derivatives, we verify that

$$z^{1-k}(1 - z^{2k}) \frac{f_t'(z)}{f_t(z)} - \frac{1}{f_t(z)^k} - 2k a_k = \frac{1}{f_t(z)} \frac{d}{dt} f_t(z). \qquad \square$$

We give another example of solution,

Lemma 5.6. *Let* $f_t(z)$ *be a solution of*

$$f_t(z)^2 - 2a_1(t) f_t(z)^3 = \frac{z^2}{1 - Q_2(t) z^2} \quad \textit{with} \quad Q_2 = 2a_2 - 5a_1^2. \tag{5.27}$$

We assume that $\dfrac{dQ_2(t)}{dt} = -2(1 - Q_2(t)^2)$ *and* $a_1'(t) = 2a_1 Q_2 + 27 a_1^3$, *then* $f_t(z)$ *is a solution of* $\frac{d}{dt} f_t = (L_{-2} - L_2) f_t$.

Proof. We have $f_t(z)^2 - 2a_1(t)f_t(z)^3 = z^2 + Q_2(t)z^4 + \cdots$, compare with (3.16). Let $J = \dfrac{z^2}{1 - Q_2(t)z^2}$, we do a verification by replacing f^3 in terms of J and f^2. □

5.4. $f_t(z) = \dfrac{z}{(1+\tau(t)z^k)^{1/k}}$ **belongs to two different integral manifolds**

Let $f_t(z) = \dfrac{z}{1+\tau(t)(z^k)^{1/k}}$ and k integer, $k \geq 1$, then if $\tau(t) = -kt$, according to (5.12), $f_t(z)$ is solution of $\frac{d}{dt} f_t(z) = L_k f_t(z)$. On the other hand, with Proposition 5.5. taking $\phi(u) = u$, we see that $f_t(z)$ is a solution of $\frac{d}{dt} f_t(z) = (L_{-k} - L_k) f_t(z)$ if $\dfrac{d\tau(t)}{dt} = (1 - \tau^2)$. We explain this as follows: In the infinite-dimensional manifold M of functions f such that $f(z) = z + a_1 z^2 + a_2 z^3 + \cdots + a_k z^{k+1} + \cdots$, we consider the one-dimensional submanifold N_k of functions $f_k(z) = \dfrac{z}{(1+\tau z^k)^{1/k}}$ for fixed $k \geq 1$. This manifold is parametrized by τ. On M, define the functional ϕ_k by $\phi_k(f) = k^2 a_k^2$. Then on N_k, the two vector fields L_{-k} and L_k are proportional, it holds $L_{-k} = \phi_k\, L_k$.

5.5. f **and** g **are like in (I). Generating functions for the Kirillov vector fields**

Let $A(\phi)(u,y)$ as in (1.1) and $\frac{d}{d\epsilon} f_\epsilon(z) = \frac{1}{2i\pi} \int A(f_\epsilon)(u,z) \left(\frac{d}{d\epsilon}\gamma_\epsilon\right)(\gamma_\epsilon^{-1}(u))\, du$ as in (4.4)–(4.5). For $k \in Z$, $|z| < |u|$

$$L_k f(z) = \frac{1}{2i\pi} \int A(f)(u,z) u^{k+1} du \;=\; \frac{f(z)^2}{2i\pi} \int \frac{f'(u)^2}{f(u)^2(f(u)-f(z))}\; u^{k+1} du. \tag{5.28}$$

We deduce $\sum_{k\in Z} (L_k f(z)) u^{-k} \;=\; \dfrac{u^2 f'(u)^2 f(z)^2}{f(u)^2(f(u)-f(z))}$. Since $L_k f(z) = z^{k+1} f'(z)$ for $k \geq 1$,

$$\begin{cases} \sum_{k\geq 1} L_k f(z) u^k = \sum_{k\geq 1} z^{1+k} f'(z) u^k \;=\; \dfrac{z^2 u}{1 - zu}\, f'(z) \\ \sum_{k\geq 1} L_{-k} f(z) u^k = \dfrac{u^2 f'(u)^2}{f(u)^2} \times \dfrac{f(z)^2}{f(u) - f(z)} \;-\; \dfrac{z u f'(z)}{u - z} + f(z). \end{cases} \tag{5.29}$$

6. The vector fields (T_k)

Let $g(z) = z + b_1 + \frac{b_2}{z} + \cdots$ and $g_\epsilon(z) = z + b_1(\epsilon) + \sum_{n\geq 1} \dfrac{b_{n+1}(\epsilon)}{z^n}$. Let $k \geq 0$ and $\dfrac{1}{\gamma'_\epsilon(u)} \times \left(\frac{d}{d\epsilon}\gamma_\epsilon\right)(u) = \chi(u) = u^{1+k}$. Then $\mathbf{L} = \mathbf{L}_f + \mathbf{L}_g$ as in (4.4)–(4.5) and (2.7).

$$\mathbf{L} \;=\; \frac{[g_\epsilon^{-1}(y)]^{1+k}}{(g_\epsilon^{-1})'(y)} \;=\; y^{1+k} \left[1 + \sum_{j\geq 1} V_j^{-k} \frac{1}{y^j} \right] \tag{6.1}$$

$$\mathbf{L}_g = \left(\frac{d}{d\epsilon} g_\epsilon\right) (g_\epsilon^{-1}(y)) = \sum_{j\geq 1} V_{j+k}^{-k}\, y^{1-j} \;=\; \frac{[g_\epsilon^{-1}(y)]^{1+k}}{(g_\epsilon^{-1})'(y)} -\; y^{1+k} - \sum_{j=1}^{k} V_j^{-k}\, y^{1+k-j} \tag{6.2}$$

$$\mathbf{L}_f = \frac{1}{(f_\epsilon^{-1})'(y)} \times \frac{d}{d\epsilon} f_\epsilon^{-1}(y) \;=\; y \times \text{ the polynomial part of } \left(\frac{1}{y} \times \frac{[g_\epsilon^{-1}(y)]^{1+k}}{(g_\epsilon^{-1})'(y)}\right). \tag{6.3}$$

Let $z = g_\epsilon(y)$ in (6.2), then $(\frac{d}{d\epsilon} g_\epsilon)(z) = z^{1+k} g'(z) - g(z)^{1+k} - \sum_{j=1}^{k} V_j^{-k} g(z)^{1+k-j}$. The vector fields induced by $(\frac{d}{d\epsilon} g_\epsilon)(z)$ are different from the right vector fields on g in [10].

Definition 6.1. We put $T_p\, g\,(z) = z^{1+p} g'(z) - \sum_{j=0}^{p} V_j^{-p} g(z)^{p+1-j}$. This defines the vector fields (T_p) with

$$T_p\, g\,(z) = \sum_{j\geq 1} V_{p+j}^{-p}\, g(z)^{1-j} \quad \text{if } p \geq 0 \quad \textit{and} \quad T_{-p}\, g\,(z) = z^{1-p} g'(z) \quad \textit{if } p > 0.$$

Then $T_0 g(z) = z g'(z) - g(z)$, $T_1 g(z) = z^2 g'(z) - g(z)^2 + 2b_1 g(z)$, $T_{-1} g(z) = g'(z)$, $T_2 g(z) = z^3 g'(z) - g(z)^3 + 3b_1 g(z)^2 + (4b_2 - 3b_1^2) g(z)$, $T_{-2} g(z) = \frac{g'(z)}{z}$, ...

6.1. Generating functions for the vector fields $(T_k)_{k\in Z}$

Let $B(\phi)(u,y)$ as in (1.2). We split $\mathbf{L}$ with the normalization **(II)**, see (4.9). In $\mathbf{L}$, the powers y^n, $n \geq 1$ correspond to f and powers y^n, $n \leq 0$ correspond to g. Let $\chi(u) = \frac{1}{\gamma'(u)} \times \big(\frac{d}{d\epsilon}\gamma_\epsilon\big)(u)$, then $\big(\frac{d}{d\epsilon} g_\epsilon\big)(g_\epsilon^{-1}(z)) = \frac{z}{2i\pi} \int \frac{\chi(g^{-1}(u))}{u(u-z)(g^{-1})'(u)} du$. With $z = g(y)$ and $u = g(v)$,

$$\frac{d}{d\epsilon} g_\epsilon(y) \;=\; \frac{1}{2i\pi} \int B(g)(v,y) \chi(v)\, dv. \tag{6.4}$$

When $\chi(u) = u^{1-p}$,

$$(T_p g)(z) = \frac{1}{2i\pi} \int \frac{g'(u)^2 g(z) u^{1-p}}{g(u)(g(u)-g(z))} du \;=\; \frac{z}{2i\pi} \int \frac{[g^{-1}(u)]^{1-p}}{u(u-z)(g^{-1})'(u)} du \tag{6.5}$$

$$\sum_{p\in Z} (T_p g)(z)\, u^p = \frac{u^2 g'(u)^2 g(z)}{g(u)(g(u)-g(z))}. \tag{6.6}$$

We deduce

$$\begin{cases} \sum_{p\geq 1} T_{-p} g(z) u^{-p} = \frac{z}{uz-1}\, g'(z) \\ \sum_{p\geq 1} T_p g(z) u^p = \frac{u^2 g'(u)^2 g(z)}{g(u)(g(u)-g(z))} \;-\; \frac{u z^2 g'(z)}{uz-1} \;+\; g(z). \end{cases} \tag{6.7}$$

6.2. The Schwarzian derivative of g and the operators (T_p)

The Schwarzian derivative of $g(z) = z + b_1 + \frac{b_2}{z} + \cdots$ is given by

$$S_g(z) = \Big(\frac{g''}{g'}\Big)' - \frac{1}{2}\Big(\frac{g''}{g'}\Big)^2 \;=\; \sum_{n\geq 2} \widetilde{Q}_n\, z^{-(n+2)}$$

$$= -6b_2 \frac{1}{z^4} - 24 b_3 \frac{1}{z^5} - 12(b_2^2 + 5b_4)\frac{1}{z^6} - 24(3b_2 b_3 + 5b_5)\frac{1}{z^7} - \cdots. \tag{6.8}$$

On the manifold of coefficients, with $\frac{\partial}{\partial b_n} g(z) = \frac{1}{z^{n-1}}$ and (6.8), we define

$$T_{-p} = -\sum_{n\geq 0}(n-1)b_n \frac{\partial}{\partial b_{n+p}} \quad \forall p > 0. \tag{6.9}$$

Lemma 6.2. $T_{-p}(\widetilde{Q}_n) = -(p^3-p)\delta_{n,p} - (n+p)\widetilde{Q}_{n-p}$

Proof. Since $T_{-p}g(z) = z^{1-p}g'(z)$ if $p > 0$, we have $T_{-p}(S_g) = p(p+1)(1-p)z^{-(p+2)} + z^{1-p}\big(S_g\big)' + 2(1-p)z^{-p}\, S_g$. We identify equal powers of z. □

The coefficients of the Schwarzian derivative S_g in terms of (b_j) are calculated in [7] with the method of [6]. All the polynomials $\widetilde{Q}_n$ have negative coefficients. We can take advantage of that to obtain majorations of $\widetilde{Q}_n$, for all $n \geq 2$. This kind of argument was used in [17] and [5] to obtain majorations of derivatives of the Faber polynomials of $\widetilde{f}(z) = \frac{z^2}{f(z)}$. It differs from the methods in [18].

7. Degeneracy of the vector fields L_k and T_k

7.1. Degeneracy of $(L_k)_{k\in Z}$. The condition $L_k = L_{-k}$

The condition $L_1 f = L_{-1} f$ gives (i) $(1-z^2)f'(z) = 1 + f''(0)\, f(z)$. The condition $L_2 f = L_{-2} f$ gives (ii) $(1-z^4)f'(z) = \frac{z}{f(z)} + 3a_1 z + (4a_2 - a_1^2)\, z\, f(z)$ with $2a_1 = f''(0)$, $6a_2 = f'''(0)$.

Proposition 7.1. *The only solutions f of the system*

$$L_2 f(z) = L_{-2} f(z) \quad \textit{and} \quad L_1 f(z) = L_{-1} f(z) \tag{7.1}$$

are $f(z) = \frac{z}{(1-\epsilon z)^2}$ *with* $\epsilon = 1$ *or* $\epsilon = -1$.

Proof. The solutions of (i) are $f(z) = \frac{1}{2a_1}\,[\,(\frac{1+z}{1-z})^{a_1} - 1\,]$ if $a_1 \neq 0$ and $f(z) = \frac{1}{2}\ln(\frac{1+z}{1-z})$ if $a_1 = 0$. Assume that $a_1 \neq 0$, the condition $f'''(0) = 6a_2$ implies that $3a_2 = 1 + 2a_1^2$. In (ii), we put $u = \frac{1+z}{1-z}$ and $h(u) = f(z) = \frac{1}{2a_1}[u^{a_1} - 1]$, it gives $\frac{4u(u^2+1)}{u^2-1}h'(u) = \frac{1}{h(u)} + 3a_1 + (4a_2 - a_1^2)h(u)$; we replace $h(u)$ by its expression and make $u \to 0$ or $u \to \infty$, we obtain $3a_1^2 = 4a_2$. We deduce from $3a_2 = 1 + 2a_1^2$ and $3a_1^2 = 4a_2$ that $a_2 = 3$ and $a_1 = 2$ or $a_1 = -2$. Assume that $a_1 = 0$ and consider $f(z) = \frac{1}{2}\ln(\frac{1+z}{1-z}) = z + \frac{z^3}{3} + \frac{z^5}{5} + \cdots$. We have $f'''(0) = 2$. In that case, (ii) transforms into $\frac{4u(u^2+1)}{u^2-1}h'(u) = \frac{1}{h(u)} + 8h(u)$, it is immediate that $h(u) = \frac{1}{2}\ln(u)$ is not a solution of this last equation. □

Theorem 7.2. *The function* $f(z) = z + a_1 z^2 + \cdots + a_n z^{n+1} + \cdots$ *is a solution of* $(*)$ *(i.e.,* $L_k f = L_{-k} f$*) if and only if* $f(z) =$ *constant* $\times \frac{z}{(1-\epsilon z)^2}$ *with* $\epsilon = 1$ *or* $\epsilon = -1$*. In particular* $a_3 = 2a_1a_2 - a_1^3$, $4a_1a_2 = 3a_1^3$, $3a_2 = 1 + 2a_1^2$.

Proof. First, to $\chi(u) = u^{k+1} - u^{1-k}$, $k \geq 1$, we associate $L_k f - L_{-k} f$. Then $\sum_{k\geq 1}[L_k f(z) - L_{-k} f(z)]\, w^k$

$$= \left[\frac{z^2 w}{1-zw} + \frac{zw}{w-z}\right] f'(z) - f(z) - \frac{w^2 f'(w)^2}{f(w)^2} \times \frac{f(z)^2}{f(w)-f(z)} \tag{7.2}$$

and the condition $L_k f = L_{-k} f$ for any $k \geq 1$ is the same as $(*)$ or

$$\left[\frac{z(1-z^2)}{(\frac{1}{w}-z)(w-z)}\right] f'(z) - f(z) - \frac{w^2 f'(w)^2}{f(w)^2} \times \frac{f(z)^2}{f(w)-f(z)} = 0. \tag{7.3}$$

If $\dfrac{zf'(z)}{f(z)} = \dfrac{1+z}{1-z}$ or equivalently $f(z) = \mathit{constant} \times \dfrac{z}{(1-z)^2}$, then it is not difficult to verify that f is a solution of $(*)$. Conversely, if f is a solution of $(*)$, then for any $k \geq 1$, we have $L_k f(z) = L_{-k} f(z)$. From Proposition 7.1, we see that $f(z)$ is a Koebe function. □

7.2. Degeneracy of the vectors $(T_k)_{k\in Z}$**,** $(T_k g = T_{-k} g$**,** $\forall k \in Z)$

The condition $T_1 g(z) = T_{-1} g(z)$ gives (iii): $(1-z^2)g'(z) = -g(z)^2 + 2b_1 g(z)$. We put $v(z) = \dfrac{1}{g(\frac{1}{z})} = z - b_1 z^2 + \cdots$. We have $(1-z^2)v'(z) = 1 - 2b_1 v(z)$ and $v''(0) = -2b_1$. This shows that $v(z)$ satisfies (i). The condition $T_2 g(z) = T_{-2} g(z)$ gives (iv): $\dfrac{(1-z^4)}{z} g'(z) = -g(z)^3 + 3b_1 g(z)^2 + (4b_2 - 3b_1^2) g(z)$. We put $v(z) = \dfrac{1}{g(\frac{1}{z})}$ in (iv), it gives $(1-z^4)v'(z) = \dfrac{z}{v(z)} - 3b_1 z - (4b_2 - 3b_1^2) z\, v(z)$. This is the same equation as (ii) where we put $a_1 = G_1(b_1) = -b_1$ and $a_2 = G_2(b_1, b_2) = b_1^2 - b_2$. According to Proposition 7.1, the common solutions of (iii) and (iv) are $v(z) = \frac{z}{(1-\epsilon z)^2}$. It gives $g(z) = z - 2\epsilon + \frac{1}{z}$.

Proposition 7.3. *We have* $T_k g = T_{-k} g$ *for any* $k \in Z$ *if and only if* $g(z) = z - 2\epsilon + \frac{1}{z}$ *with* $\epsilon = 1$ *or* $\epsilon = -1$.

References

[1] H. Airault, P. Malliavin, *Unitarizing probability measures for representations of Virasoro algebra.* J. Math. Pures Appl. **80**, 6, (2001), 627–667.

[2] H. Airault, J. Ren, *An algebra of differential operators and generating functions on the set of univalent functions.* Bull. Sc. Math. **126**, 5, (2002), 343–367.

[3] H. Airault, V. Bogachev, *Realization of Virasoro unitarizing measures on the set of Jordan curves.* C. Rend. Acad Sc. Paris. Série I **336**, (2003), 429–434.

[4] H. Airault, A. Bouali, *Differential calculus on the Faber polynomials.* Bull. Sc. Math. **130**, (2006), 179–222 and H. Airault, A. Bouali, Short communications, Analyse, Madrid 2006.

[5] H. Airault, *Remarks on Faber polynomials*, International Mathematical Forum, **3**, (2008), no. 8, 449–456.

[6] H. Airault and Y. Neretin, *On the action of Virasoro algebra on the space of univalent functions.* Bull. Sc. Math.**132**, (2008), 27–39.

[7] H. Airault, *Symmetric sums associated to the factorization of the Grunsky coefficients. In John Harnad, Pavel Winternitz (Eds.) Groups and symmetries: From the Neolithic Scots to John McKay* (Montreal April 2007), **47**, A.M.S. Proceedings (2009).

[8] C. Badea and S. Grivaux, *Faber hypercyclic operators*, Israel J. Math., **165**, (2008), 43–65.

[9] A. Bouali, *Faber polynomials, Cayley-Hamilton equation and Newton symmetric functions.* Bull. Sc. Math. **130** (2006) 49–70.

[10] A.A. Kirillov, *Geometric approach to discrete series of unireps for Virasoro.* J. Math. Pures Appl. **77** (1998) 735–746.

[11] A.A. Kirillov, D.V. Yuriev, *Kähler geometry of the infinite-dimensional space* $M = \mathrm{Diff}(_+(S^1)/Rot(S^1)$. Functional Analysis Appl. **21**, 4, (1987), 284–294.

[12] I. Markina, D. Prokhorov, A. Vasiliev, *Subriemannian geometry of the coefficients of univalent functions.* J. Funct. Anal. **245** (2007) 475–492.

[13] S. Nag, *Singular Cauchy integrals and conformal welding on Jordan curves.* Annales Acad. Scient. Fennicae Mathematica, **21**, (1996), 81–88.

[14] Yu.A. Neretin, *Representations of Virasoro and affine Lie algebras.* Encyclopedia of Mathematical Sciences, **22**, Springer Verlag (1994) 157–225.

[15] A.C. Schaeffer, D.C. Spencer, *Coefficient regions for Schlicht functions.* Colloquium Publications **35** American Math. Soc. (1950).

[16] M. Schiffer, *Faber polynomials in the theory of univalent functions.* Bull. Amer. Soc., **54** (1948) 503–517.

[17] P.G. Todorov, *On the Faber polynomials of the univalent functions of the class* σ. J. of Math. Anal. and Appl., **162** (1991) 268–276.

[18] S.M. Zemyan, *On the coefficient problem for the Schwarzian derivative.* Ann. Acad. Sci. Fenn., **17** (1992) 221–240.

Hélène Airault
LAMFA CNRS UMR 6140
Université de Picardie Jules Verne
33, Rue Saint Leu, Amiens
Insset, 48 rue Raspail
F-02100 Saint-Quentin (Aisne), France
e-mail: `hairault@insset.u-picardie.fr`

Analysis and Mathematical Physics
Trends in Mathematics, 21–48
© 2009 Birkhäuser Verlag Basel/Switzerland

Poincaré–Steklov Integral Equations and Moduli of Pants

Andrei Bogatyrev

Abstract. Various physical and mathematical settings bring us to a boundary value problem for a harmonic function with spectral parameter in the boundary conditions. One of those problems may be reduced to a singular 1D integral equation with spectral parameter. We present a constructive representation for the eigenvalues and eigenfunctions of this integral equation in terms of moduli of explicitly constructed pants, one of the simplest Riemann surfaces with boundary. Essentially the solution of the integral equation is reduced to the solution of three transcendental equations with three unknown numbers, moduli of pants.

Mathematics Subject Classification (2000). Primary 30Fxx; Secondary 14H15, 30C20, 45C05, 33E30.

Keywords. Spectral parameter, Riemann surface, pair of pants, branched complex projective structure, Dessin d'Enfants.

We introduce the new type of constructive *pictorial* representations for the solutions of the following spectral singular Poincaré–Steklov (PS for brevity) integral equation

$$\lambda\ V.p.\int\limits_I \frac{u(t)}{t-x}\,dt - V.p.\int\limits_I \frac{u(t)\,dR(t)}{R(t)-R(x)} = \text{const}, \quad x \in I := (-1,\ 1), \tag{0.1}$$

where λ is the spectral parameter; $u(t)$ is the unknown function; const is independent of x. The functional parameter $R(t)$ of the equation is a smooth *nondegenerate* change of variable on the interval I:

$$\frac{d}{dt}R(t) \neq 0, \qquad t \in [-1,1]. \tag{0.2}$$

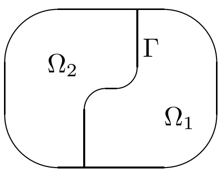
This work was supported by grant MD4798.2007.01.

1. Introduction

H. Poincaré (1896) and V.A. Steklov (1901) were the first who put the spectral parameter to the boundary conditions of the problem for an elliptic operator. Later it became a popular technique for the analysis and optimization in diffraction problems [1], (thermo)conductivity of composite materials, simple 2D model of oil extraction etc.

1.1. Spectral boundary value problem

Let a domain in the plane be subdivided into two simply connected domains Ω_1 and Ω_2 by a smooth simple arc Γ. We are looking for the values of the spectral parameter λ when the following problem has a nonzero solution:



$$\begin{array}{rclll} \Delta U_1 &=& 0 & \text{in } \Omega_1; & U_1 = 0 \text{ on } \partial\Omega_1 \setminus \Gamma; \\ \Delta U_2 &=& 0 & \text{in } \Omega_2; & U_2 = 0 \text{ on } \partial\Omega_2 \setminus \Gamma; \\ U_1 &=& U_2 & \text{on } \Gamma; & \\ -\lambda \dfrac{\partial U_1}{\partial n} &=& \dfrac{\partial U_2}{\partial n} & \text{on } \Gamma, & \end{array} \tag{1.1}$$

Spectral problems of this type naturally arise, e.g., in the justification and optimization of a *domain decomposition method* for the solution of a boundary value problem for Laplace equation. It is easy to show that the eigenfunctions and the eigenvalues of the problem (1.1) are correspondingly the critical points and critical values of the functional (the so-called generalized Rayleigh ratio)

$$F(U) = \frac{\int_{\Omega_2} |\operatorname{grad}\ U_2|^2\, d\Omega_2}{\int_{\Omega_1} |\operatorname{grad}\ U_1|^2\, d\Omega_1}, \qquad U \in H^{1/2}_{oo}(\Gamma), \tag{1.2}$$

where U_s is the harmonic continuation of the function U from the interface Γ to the domain Ω_s, $s = 1, 2$, vanishing at the outer boundary of the domain.

The boundary value problem (1.1) is equivalent to a certain Poincaré–Steklov equation. Indeed, let V_s be the harmonic function conjugate to U_s, $s = 1, 2$. From the Cauchy–Riemann equations and the relations on Γ it follows that the tangent to the interface derivatives of V_1 and V_2 differ by the same factor $-\lambda$. Integrating along Γ we get

$$\lambda V_1(y) + V_2(y) = \text{const}, \quad y \in \Gamma. \tag{1.3}$$

The boundary values of conjugate functions harmonic in the half-plane are related by a Hilbert transformation. To reduce our case to this model we consider a conformal mapping $\omega_s(y)$ from Ω_s to the open upper half-plane $\mathbb{H}$ with normalization $\omega_s(\Gamma) = I$, $s = 1, 2$. Now equation (1.3) may be rewritten as

$$-\frac{\lambda}{\pi}\, V.p. \int_I \frac{U_1(\omega_1^{-1}(t))}{t - \omega_1(y)}\, dt - \frac{1}{\pi}\, V.p. \int_I \frac{U_2(\omega_2^{-1}(t'))}{t' - \omega_2(y)}\, dt' = \text{const}, \quad y \in \Gamma.$$

Introducing the new notation $x := \omega_1(y) \in I$; $R := \omega_2 \circ \omega_1^{-1} : \ I \to \Gamma \to I$; $u(t) := U_1(\omega_1^{-1}(t))$ and the change of variable $t' = R(t)$ in the second integral we arrive at the Poincaré–Steklov equation (0.1). Note that in this context $R(t)$ is the decreasing function on I.

1.2. Some known results

The natural way to study integral equations is operator analysis. This discipline allows to obtain for the *smooth nondegenerate* change of variables $R(x)$ the following results [2]:

- The spectrum is discrete; the eigenvalues are positive and converge to $\lambda = 1$.
- $\sum_{\lambda \in Sp} |\lambda - 1|^2 < \infty$ (a constructive estimate in terms of $R(x)$ is given)
- The eigenfunctions $u(x)$ form a basis in the Sobolev space $H_{oo}^{1/2}(I)$.

1.3. Goal and philosophy of the research

The approach of complex geometry for the same integral equation gives different types of results. For quadratic $R(x) = x + (2C)^{-1}(x^2 - 1)$, $C > 1$, the eigenpairs were found explicitly [3]:

$$u_n(x) = \sin \left[\frac{n\pi}{K'} \int_1^{(C+x)/(C-1)} (s^2 - 1)^{-1/2}(1 - k^2 s^2)^{-1/2} ds \right],$$

$$\lambda_n = 1 + 1/\cosh 2\pi\tau n, \quad n = 1, 2, \ldots,$$

where $\tau = K/K'$ is the ratio of the complete elliptic integrals of modulus $k = (C-1)/(C+1)$. Now we are going to give *constructive representations* for the eigenpairs $\{\lambda, u(x)\}$ of the integral equation with $R(x) = R_3(x)$ being a rational function of degree 3. Equation (0.1) itself will be called PS-3 in this case.

The notion of a constructive representation for the solution should be however specified. Usually this means that we restrict the search for the solution to a certain class of functions such as rational, elementary, abelian, quadratures, the Umemura classical functions, etc. The history of mathematics knows many disappointing results when the solution of the prescribed form does not exist. Say, the diagonal of the square is not commensurable with its side, generic algebraic equations cannot be solved in radicals, linear ordinary differential equations usually cannot be solved by quadratures, Painlevé equations cannot be solved by Umemura functions. Nature always forces us to introduce new types of transcendent objects to enlarge the scope of search. The study of new transcendental functions constitutes the progress of mathematics. This research philosophy goes back to H. Poincaré [4]. From the philosophical point of view our goal is to disclose the nature of emerging transcendental functions in the case of PS-3 integral equations.

1.4. Brief description of the result

The rational function $R_3(x)$ of degree three is explicitly related to a *pair of pants* in Section 2.2. On the other hand, given a spectral parameter λ and two auxiliary real parameters, we explicitly construct in Section 2.3 another pair of pants which additionally depend on two integers. When the above two pants are conformally equivalent, λ is the eigenvalue of the integral equation PS-3 with parameter $R_3(x)$. Essentially, this means that to find the spectrum of the given integral equation (0.1) one has to solve three transcendental equations involving three *moduli of pants*.

Whether this representation of the solutions may be considered as constructive or not is a matter of discussion. Our approach to the notion of a constructive representation is utilitarian: the more we learn about the solution from the given representation the more constructive is the latter. At least we are able to obtain valuable features of the solution: to determine the number of zeroes of the eigenfunction $u(t)$, to find the exact locus for the spectra and to show the discrete mechanism of generating the eigenvalues.

2. Description of the main result

The shape of the two domains Ω_1 and Ω_2 defines the variable change $R(x)$ only up to a certain two-parametric deformation. One can easily check that the *gauge transformation* $R \to L_2 \circ R \circ L_1$, where the linear fractional function $L_s(x)$ keeps the segment $[-1, 1]$, does not affect the spectrum of equation (0.1) and induces only the change of the argument for its eigenfunctions: $u(x) \to u \circ L_1(x)$. For this reason we do not distinguish between two PS equations with their functional parameters $R(x)$ related by the gauge transformation.

The space of equivalence classes of equations PS-3 has real dimension $3 = 7-2-2$ and several components with different topology of the functional parameter R_3. In the present paper we study for brevity only one of the components, the choice is specified in Section 2.1.1.

2.1. Topology of the branched covering

In what follows we consider *rational degree three* functions $R_3(x)$ with *separate real critical values* different from ± 1. The rational function $R_3(x)$ defines a 3-sheeted branched covering of a Riemann sphere by another Riemann sphere. The Riemann–Hurwitz formula suggests that $R_3(x)$ has four separate branch points a_s, $s = 1, \dots, 4$. This means that every value a_s is covered by a critical (double) point b_s, and an ordinary point c_s.

Every point $y \neq a_s$ of the extended real axis $\hat{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ belongs to exactly one of two types. For the type (3:0) the pre-image $R_3^{-1}(y)$ consists of three distinct real points. For the type (1:2) the pre-image consists of a real and two complex conjugate points. The type of the point remains locally constant on the extended real axis and changes when we step over the branch point. Let the branch points

a_s be enumerated in the natural cyclic order of $\hat{\mathbb{R}}$ so that the intervals (a_1, a_2) and (a_3, a_4) are filled with the points of the type (1:2). Later we will specify the way to exclude the relabeling $a_1 \leftrightarrow a_3$, $a_2 \leftrightarrow a_4$ of branch points.

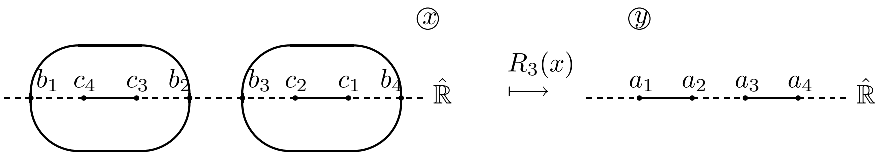


FIGURE 1. The topology of the covering R_3 with real branch points

The total pre-image $R_3^{-1}(\hat{\mathbb{R}})$ consists of the extended real axis and two pairs of complex conjugate arcs intersecting $\hat{\mathbb{R}}$ at the points b_1, b_2, b_3, b_4 as shown at the left side of Fig. 1. The compliment of this pre-image on the Riemann sphere has six components, each of them is mapped 1-1 onto the upper or lower half-plane.

2.1.1. The component in the space of equations. The nondegeneracy condition (0.2) forbids that any of critical points b_s be inside the segment of integration $[-1, 1]$. In what follows we consider the case when the latter segment lies in the annulus bounded by two ovals passing through the critical points b_s. Possibly relabeling the branch points we assume that $[-1, 1] \subset (b_2, b_3)$.

Other components in the space of PS-3 integral equations are treated in [11].

2.2. Pair of pants

For obvious reason *a pair of pants* is the name for the Riemann sphere with three holes in it. Any pair of pants may be conformally mapped to $\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ with three nonintersecting real slots. This mapping is unique up to the real linear-fractional transformation of the sphere. The conformal class of pants with labeled boundary components depend on three real parameters varying in a cell.

Definition To the variable change $R_3(x)$ we associate the pair of pants

$$\mathcal{P}(R_3) := Closure\left(\hat{\mathbb{C}} \setminus \{[-1, 1] \cup [a_1, a_2] \cup [a_3, a_4]\}\right). \tag{2.1}$$

Closure here and below is taken with respect to the intrinsic spherical metrics when every slot acquires two sides. Boundary components of the pair of pants are colored (labeled) in accordance with the palette:

$$\begin{array}{ll} [-1, 1] & - \text{"red"}, \\ [a_1, a_2] & - \text{"blue"}, \\ [a_3, a_4] & - \text{"green"}. \end{array}$$

The conformal class of pants (2.1) depends only on the equivalence class of integral equations. To simplify the statement of our result we assume that infinity lies strictly inside the pants (2.1) which is not a loss of generality – we can always apply a suitable gauge transformation of the parameter $R_3(x)$.

2.2.1. Reconstruction of $R_3(x)$ from the pants. Here we show that the branched covering map $R_3(x)$ with given branch points a_s, $s = 1, \dots, 4$, is essentially unique. A possible ambiguity is due to the conformal motions of the covering Riemann sphere.

Let L_a be the unique linear-fractional map sending the critical values a_1, a_2, a_3, a_4 of R_3 to respectively 0, 1, $a > 1$, ∞. The conformal motion L_b of the covering Riemann sphere sends the critical points b_1, b_2, b_3, b_4 (unknown at the moment) to respectively 0, 1, $b > 1$, ∞. The function $L_a \circ R_3 \circ L_b^{-1}$ with normalized critical points and critical values takes a simple form

$$\widetilde{R_3}(x) = x^2 L(x)$$

with a real linear fractional function $L(x)$ satisfying the restrictions:

$$\begin{array}{ll} L(1) = 1, & L'(1) = -2, \\ L(b) = a/b^2, & L'(b) = -2a/b^3. \end{array}$$

We got four equations for three parameters of $L(x)$ and the unknown b. The first two equations suggest the following expression for $L(x)$

$$L(x) = 1 + 2\frac{(c-1)(x-1)}{x-c}.$$

Another two are solved parametrically in terms of parameter c:

$$b = c\frac{3c-3}{2c-1}; \qquad a = c\frac{(3c-3)^3}{2c-1}.$$

Both functions $b(c)$ and $a(c)$ increase from 1 to ∞ when the argument $c \in (1/3, 1/2)$. So, given $a > 1$ we find the unique c in just specified limits, and therefore the mapping $\widetilde{R_3}(x)$. Now we can restore the linear fractional map L_b. The inverse image $\widetilde{R_3}^{-1}$ of the segment $L_a[-1, 1]$ consists of three disjoint segments. For *our case* we choose the (unique) component of the pre-image belonging to the segment $[1, b]$. The requirement: L_b *maps* $[-1, 1]$ *to the chosen segment* determines $R_3(x)$ up to a gauge transformation.

2.3. Another pair of pants

For real $\lambda \in (1, 2)$ we consider an annulus α depending on λ bounded by $\varepsilon\hat{\mathbb{R}}$, $\varepsilon := \exp(2\pi i/3)$, and the circle

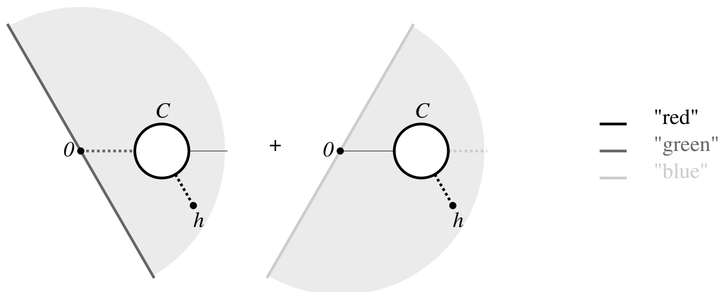
$$C := \{p \in \mathbb{C} : \quad |p - \mu^{-1}|^2 = \mu^{-2} - 1\}, \qquad \mu := \sqrt{\frac{3-\lambda}{2\lambda}} \in (\frac{1}{2}, 1). \tag{2.2}$$

Another annulus bounded by the same circle C and $\varepsilon^2\hat{\mathbb{R}}$ is denoted by $\overline{\alpha}$. Note that for the considered values of λ the circle C does not intersect the lines $\varepsilon^{\pm 1}\mathbb{R}$. We paint the boundaries of our annuli in the following way:

$$\begin{array}{ll} C & \text{– “red”}, \\ \varepsilon\hat{\mathbb{R}} & \text{– “green”}, \\ \varepsilon^2\hat{\mathbb{R}} & \text{– “blue”}. \end{array}$$

Given λ in the specified above limits, real $h_1,\ h_2$ and nonnegative integers m_1, m_2, we define three pairs of pants $\mathcal{P}_s(\lambda, h_1, h_2|m_1, m_2)$ of different fashions $s = 1, 2, 3$.

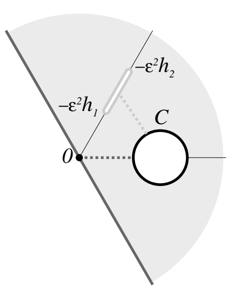
Fashion 1:
$\mathcal{P}_1(\lambda, h_1, h_2|m_1, m_2) :=\ m_1\alpha\ +\ m_2\overline{\alpha} +$

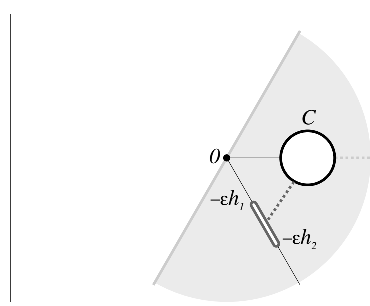


The operations '+' here stand for a certain surgery. First of all take two annuli α and $\overline{\alpha}$ and cut them along the same segment (dashed red line in the figure above) starting at the point $h := h_1 + ih_2$ from the interior of $\alpha \cap \overline{\alpha}$ and ending at the circle C. Now glue the left bank of one cut to the right bank of the other. The resulting two sheeted surface (called *Überlagerungsfläche* in the following) will be the pair of pants $\mathcal{P}_1(\lambda, h_1, h_2|0, 0)$. It is possible to modify the obtained surface sewing several annuli to it. Cut the annulus α contained in the pants and m_1 more copies of this annulus along the same segment (shown by the dashed green line in the figure above) connecting the boundaries of the annulus. The left bank of the cut on every copy of α is identified with the right bank of the cut on another copy so that all copies of the annulus are glued in one piece. A similar procedure may be repeated for the annulus $\overline{\alpha}$ (cut along the dashed blue line). The scheme for sewing together fashion 1 pants from the patches α, $\bar{\alpha}$ when $m_1 = 3$ and $m_2 = 2$ is shown in Fig. 2.

Fashions 2 and 3:
$\mathcal{P}_2(\lambda, h_1, h_2|m_1, m_2) :=$
$m_1\alpha\ +\ m_2\overline{\alpha} +$

$\mathcal{P}_3(\lambda, h_1, h_2|m_1, m_2) :=$
$m_1\alpha\ +\ m_2\overline{\alpha} +$





The pair of pants $\mathcal{P}_2(\lambda, h_1, h_2|0, 0)$ (resp. $\mathcal{P}_3(\lambda, h_1, h_2|0, 0)$) by definition is the annulus α (resp. $\overline{\alpha}$) with removed segment $-\varepsilon^2[h_1, h_2]$ (resp. $-\varepsilon[h_1, h_2]$), $0 < h_1 < h_2 < \infty$. As in the previous case those pants may be modified by sewing in several annuli α, $\overline{\alpha}$. The scheme of cutting and gluing is shown in Fig. 3

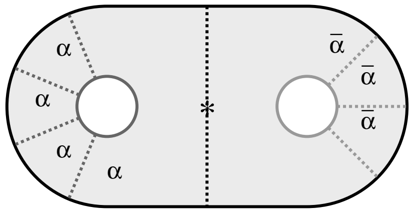


FIGURE 2. The scheme for sewing pants $\mathcal{P}_1(\lambda, h_1, h_2|3, 2)$. Asterisk is the critical point of $p(y)$.

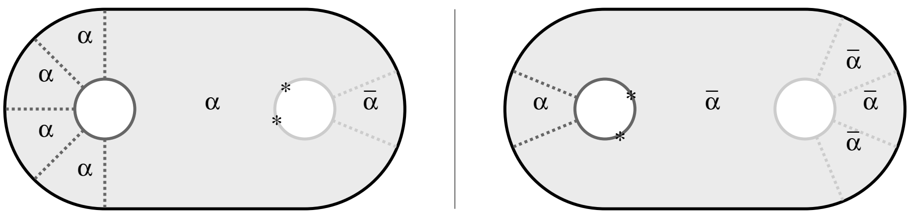


FIGURE 3. The scheme for sewing pants $\mathcal{P}_2(\lambda, h_1, h_2|4, 1)$ (**left**); and $\mathcal{P}_3(\lambda, h_1, h_2|1, 3)$ (**right**). Asterisks are the critical points of the mapping $p(y)$.

2.3.1. Remarks on the constructed pairs of pants. 1. The limiting case of the first fashion of the pants when the branch point $h_1 + ih_2$ tends to $\varepsilon^{\pm 1}\mathbb{R}$ coincides with the limiting cases of the two other fashion pants when $h_1 = h_2 > 0$:

$$\begin{aligned} \mathcal{P}_1(\lambda, -Re(\varepsilon^2 h), -Im(\varepsilon^2 h)|m_1, m_2) &= \quad \mathcal{P}_2(\lambda, h, h|m_1, m_2 + 1), \\ \mathcal{P}_1(\lambda, -Re(\varepsilon h), -Im(\varepsilon h)|m_1, m_2) &= \quad \mathcal{P}_3(\lambda, h, h|m_1 + 1, m_2), \end{aligned} \tag{2.3}$$

where parameter $h > 0$. We denote those intermediate cases as $\mathcal{P}_s(\lambda, h|m_1, m_2)$, $s = 12, 13$ respectively.

2. The surgery procedure of sewing annuli, e.g., to the pants (known as *grafting of projective structures*) was designed by B. Maskit (1969), D. Hejhal (1975) and D. Sullivan-W. Thurston (1983), see also W. Goldman (1987).

3. Every pair of pants $\mathcal{P}_s(\lambda, h_1, h_2|m_1, m_2)$ may be conformally mapped to the sphere with three real slots, i.e., pants of the type (2.1). Let $p(y)$ be the inverse mapping. We observe that $p(y)$ has exactly one critical point in the pants $\mathcal{P}(R_3)$, counting *multiplicity and weight.* For the fashion $s = 1$ this point lies strictly inside the pants and is mapped to $h = h_1 + ih_2$. For the case $s = 2$ (resp.

$s = 3$) there will be two simple critical points of $p(y)$ on the blue (resp. green) boundary component of the pants which are mapped to the points $-\varepsilon^2 h_1$, $-\varepsilon^2 h_2$ (resp.$-\varepsilon h_1$, $-\varepsilon h_2$). Finally, for the intermediate case (see Remark 1) there will be a double critical point on the boundary. The *multiplicity* of the critical point on the boundary should be calculated with respect to the local parameter of the *double* of pants $\mathcal{P}(R_3)$:

$$M := \left\{ w^2 = (y^2 - 1) \prod_{s=1}^{4} (y - a_s) \right\}, \tag{2.4}$$

e.g., at the endpoint $a = \pm 1, a_1, \dots, a_4$ of the slot this local parameter is $\sqrt{y - a}$. We consider the critical points on the boundary with the *weight* $\frac{1}{2}$.

2.4. Main theorem

Later we explain that *real* eigenfunctions of the integral equation PS-3 are split with respect to the reflection symmetry into two groups: the *symmetric* and the *antisymmetric*. In the present paper we consider only the second group of solutions.

Theorem 2.1. *When $\lambda \neq 1, 3$ the antisymmetric eigenfunctions $u(x)$ of the PS-3 integral equation with parameter $R_3(x)$ are in one to one correspondence with the pants $\mathcal{P}_s(\lambda, h_1, h_2 | m_1, m_2)$, $s = 1, 2, 3, 12, 13$, which are conformally equivalent to the pair of pants $\mathcal{P}(R_3)$ with colored boundary components.*

Let $p(y)$ be the conformal map from $\mathcal{P}(R_3)$ to $\mathcal{P}_s(\lambda, h_1, h_2 | m_1, m_2)$, then up to proportionality

$$u(x) = \sqrt{\frac{(y - y_1)(y - y_2)}{p'(y^+)p'(y^-)}} \frac{p(y^+) - p(y^-)}{w(y)}, \tag{2.5}$$

where $x \in [-1, 1]$; $y := R_3(x)$, $y^{\pm} := y \pm i0$. For $s = 1$, $y_1 = \overline{y_2}$ is the critical point of the function $p(y)$; for $s = 2, 3$ the real y_1 and y_2 are critical points of the function $p(y)$.

The *proof* of this theorem will be given in the remaining two sections of the article.

2.5. Corollaries

The representation (2.5) cannot be called explicit in the usual sense, since it comprises a transcendent function $p(y)$. We show that nevertheless the representation is useful as it allows us to understand the following properties of the solutions.

1. The "antisymmetric" part of the spectrum is always a subset of $[1, 2) \cup \{3\}$.

2. Every $\lambda \in (1, 2)$ is the eigenvalue for infinitely many equations PS-3.

Proof. Any of the constructed pants may be transformed to the standard form: a sphere with three real slots. Normalizing the red slot to be $[-1, 1]$, the end points of the two other slots will give the branch points $a_1, \dots, a_4$. We know already how to reconstruct the branched covering $R_3(x)$ from its branch points. □

3. Eigenfunction $u(x)$ related to the pants $\mathcal{P}_s(\dots|m_1, m_2)$ has exactly $m_1 + m_2 + 2$ zeroes on the segment $[-1, 1]$ when $s = 2, 3$ and one more zero when $s = 1$.

Proof. According to the formula (2.5), the number of zeroes of eigenfunction $u(x)$ is equal to the number of points $y \in [-1, 1]$ where $p(y^+) = p(y^-)$. This number in turn is equal to the number of solutions of the inclusion

$$S(y) := \mathrm{Arg}[p(y^-) - \mu^{-1}] - \mathrm{Arg}[p(y^+) - \mu^{-1}] \quad \in 2\pi\mathbb{Z}, \qquad y \in [-1, 1]. \quad (2.6)$$

Let the point $p(y)$ go m times around the circle C when its argument y travels along the two sides of $[-1, 1]$. The integer m is naturally related to the integer parameters of the pants $\mathcal{P}_s(\dots)$. The function $S(y)$ strictly increases from 0 to $2\pi m$ on the segment $[-1, 1]$, therefore the inclusion (2.6) has exactly $m + 1$ solutions on the mentioned segment. □

4. The mechanism for generating the discrete spectrum of the integral equation is explained. Sewing an annulus to the pants $\mathcal{P}_s(\lambda, h_1, h_2|\dots)$ changes the conformal structure of the latter. To return to the conformal structure specified by $\mathcal{P}(R_3)$ we have to change the real parameters of the pants, one of them being the spectral parameter λ.

If we knew how to evaluate the conformal moduli of the pair of pants

$$\mathcal{P}_s(\lambda, h_1, h_2|m_1, m_2)$$

as functions of its real parameters, the solution of the integral equation would be reduced to a system of three transcendental equations for the three numbers λ, h_1, h_2. This solution will depend on the integer parameters s, m_1, m_2.

3. Geometry of integral equation

PS integral equations possess a rich geometrical structure which we disclose in this section. The chain of equivalent transformations of PS-3 equation described here in a somewhat sketchy fashion is given in [10, 11] with more details.

3.1. A nonlocal functional equation

Let us decompose the kernel of the second integral in (0.1) into a sum of elementary fractions:

$$\frac{R_3'(t)}{R_3(t) - R_3(x)} = \frac{d}{dt}\log(R_3(t) - R_3(x)) = \sum_{k=1}^{3} \frac{1}{t - x_k(x)} - \frac{Q'}{Q}(t), \qquad (3.1)$$

where $Q(t)$ is the denominator in an irreducible representation of $R(t)$ as the ratio of two polynomials; $x_1(x) = x,\ x_2(x),\ x_3(x)$ – are all solutions (including multiple and infinite) of the algebraic equation $R_3(x_s) = R_3(x)$. This expansion suggests to rewrite the original equation (0.1) as a certain relationship for the Cauchy-type integral

$$\Phi(x) := \int_I \frac{u(t)}{t - x} dt + \mathrm{const}^*, \quad x \in \hat{\mathbb{C}} \setminus [-1, 1]. \qquad (3.2)$$

The constant term const* in (3.2) is introduced to compensate for the constant terms arising after substitution of expression (3.2) to the equation (0.1).

For a known $\Phi(x)$, the eigenfunction $u(t)$ may be recovered by the Sokhotskii-Plemelj formula:

$$u(t) = (2\pi i)^{-1}\left[\Phi(t+i0) - \Phi(t-i0)\right], \quad t \in I. \tag{3.3}$$

Function $\Phi(x)$ generated by an eigenfunction of PS integral equation satisfies a nonlocal functional equation:

Lemma 3.1. [10] *For $\lambda \neq 1, 3$ the transformations* (3.2) *and* (3.3) *imply a* 1-1 *correspondence between the Hölder eigenfunctions $u(t)$ of the PS-*3 *integral equation and the nontrivial solutions $\Phi(x)$ of the functional equation which are holomorphic on a sphere with the slot* $[-1, 1]$

$$\Phi(x+i0) + \Phi(x-i0) = \delta\Bigg(\Phi(x_2(x)) + \Phi(x_3(x))\Bigg), \quad x \in I, \tag{3.4}$$

$$\delta = 2/(\lambda - 1), \tag{3.5}$$

with Hölder boundary values $\Phi(x \pm i0)$.

3.2. The Riemann monodromy problem

The lifting $R_3^{-1}(\mathcal{P}(R_3))$ of the pants associated to the integral equation consists of three components $\mathcal{O}_s$, $s = 1, 2, 3$. We number them in the following way (see Fig. 1): the segment $[-1, 1]$ lies on the boundary of $\mathcal{O}_1$; the segment $[c_4, c_3]$ is on the boundary of $\mathcal{O}_2$ and the boundary of $\mathcal{O}_3$ comprises the segment $[c_2, c_1]$.

3.2.1. Let the function $\Phi(x)$ be related to the solution $u(x)$ of the integral equation (0.1) as in (3.2). We consider a 3-vector defined in the pair of pants:

$$W(y) := (W_1, W_2, W_3)^t = (\Phi(x_1), \Phi(x_2), \Phi(x_3))^t, \qquad y \in \mathcal{P}(R_3), \tag{3.6}$$

where x_s is the unique solution of the equation $R_3(x_s) = y$ in $\mathcal{O}_s$. Vector $W(y)$ is holomorphic and bounded in the pants $\mathcal{P}(R_3)$ as all three points x_s, $s = 1, 2, 3$, remain in the holomorphicity domain of the function $\Phi(x)$. We claim that *the boundary values of the vector $W(y)$ are related via constant matrices*:

$$W(y+i0) = \mathbf{D}_* W(y-i0), \text{ when } y \in \{slot_*\}. \tag{3.7}$$

The matrix $\mathbf{D}_*$ assigned to the "green" $[a_3, a_4]$, "blue" $[a_1, a_2]$, and "red" $[-1, 1]$ slot respectively is

$$\mathbf{D_2} := \begin{Vmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{Vmatrix}; \qquad \mathbf{D_3} := \begin{Vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{Vmatrix}; \qquad \mathbf{D} := \begin{Vmatrix} -1 & \delta & \delta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix}. \tag{3.8}$$

This in particular means that our vector (3.6) is a solution of a certain Riemann monodromy problem. The monodromy of vector $W(y)$ along the loop crossing only "red", "green" or "blue" slot is given by the matrix $\mathbf{D}$, $\mathbf{D_2}$ or $\mathbf{D_3}$ correspondingly – see Fig. 4.

Indeed, let $y^+ := y + i0$ and $y^- := y - i0$ be two points on opposite sides of the "blue" slot $[a_1, a_2]$. Their inverse images $x_3^+ = x_3^-$, $x_1^\pm = x_2^\mp$ lie outside the cut

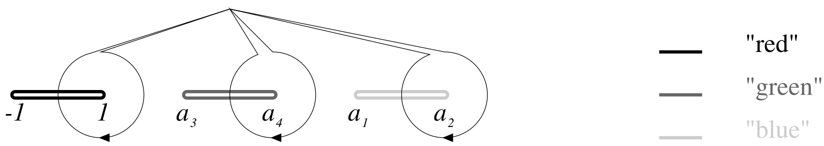


FIGURE 4. Three loops on a sphere with six punctures ± 1, $a_1, \ldots, a_4$

$[-1,1]$. Hence $W(y^+) = \mathbf{D_3}W(y^-)$. For two points $y^\pm$ lying on opposite sides of the "green" slot $[a_3, a_4]$, their inverse images satisfy relations $x_2^+ = x_2^-$, $x_1^\pm = x_3^\mp$, which means $W(y^+) = \mathbf{D_2}W(y^-)$. Finally, let $y^\pm$ lie on both sides of the "red" slot $[-1,1]$. Now two points $x_2^+ = x_2^-$ and $x_3^+ = x_3^-$ lie in the holomorphicity domain of $\Phi(x)$ while x_1^+ and x_1^- appear on the opposite sides of the cut $[-1,1]$. According to the functional equation (3.4),

$$\Phi(x_1^+) = -\Phi(x_1^-) + \delta(\Phi(x_2^\pm) + \Phi(x_3^\pm)),$$

therefore $W(y^+) = \mathbf{D}W(y^-)$ holds on the slot $[-1,1]$.

3.2.2. Conversely, let $W(y)$ be the bounded solution of the Riemann monodromy problem (3.7). We define a piecewise holomorphic function on the Riemann sphere:

$$\Phi(x) := W_s(R_3(x)), \qquad \text{when } x \in \mathcal{O}_s, \qquad s = 1,2,3. \tag{3.9}$$

From the boundary relations for the vector $W(y)$ it immediately follows that the function $\Phi(x)$ has no jumps on the lifted cuts $[a_1, a_2]$, $[a_3, a_4]$, $[-1,1]$ except for the cut $[-1,1]$ from the upper sphere. Say, if the two points $y^\pm$ lie on opposite sides of the cut $[a_1, a_2]$, then $W_3(y^+) = W_3(y^-)$ and $W_1(y^\pm) = W_2(y^\mp)$ which means that the function $\Phi(x)$ has no jump on the components of $R_3^{-1}[a_1, a_2]$. From the boundary relation on the cut $[-1,1]$ it follows that $\Phi(x)$ is the solution for the functional equation (3.4). Therefore it gives a solution of Poincaré–Steklov integral equation with parameter $R_3(x)$. Combining formulae (3.3) with (3.9) we get the reconstruction rule

$$u(x) = (2\pi i)^{-1}\Big(W_1(R_3(x) + i0) - W_1(R_3(x) - i0)\Big), \qquad x \in [-1,1]. \tag{3.10}$$

We have just proved the following

Theorem 3.2. [3] *If $\lambda \neq 1, 3$ then the two formulas* (3.6) *and* (3.10) *imply the one-to-one correspondence between the solutions $u(x)$ of the integral equation* (0.1) *and the bounded solutions $W(y)$ of the Riemann monodromy problem* (3.7) *in the punctured sphere $\hat{\mathbb{C}} \setminus \{-1, 1, a_1, a_2, a_3, a_4\}$.*

3.2.3. Monodromy invariant. The following statement is proved by direct computation.

Lemma 3.3. *All matrixes* (3.8) (i) *are involutive* (*i.e.*, $\mathbf{D}^2 = \mathbf{D_2}^2 = \mathbf{D_3}^2 = \mathbf{1}$) *and* (ii) *conserve the quadratic form*

$$J(W) := \sum_{k=1}^{3} W_k^2 - \delta \sum_{j<s}^{3} W_j W_s. \tag{3.11}$$

The form $J(W)$ is not degenerate unless $-2 \neq \delta \neq 1$, or equivalently $0 \neq \lambda \neq 3$. Since the solution $W(y)$ of our monodromy problem is bounded near the cuts, the value of the form $J(W)$ is independent of the variable y. Therefore the solution takes values either in the smooth quadric $\{W \in \mathbb{C}^3 : \; J(W) = J_0 \neq 0\}$, or the cone $\{W \in \mathbb{C}^3 : \; J(W) = 0\}$.

3.3. Geometry of the quadric surface

The nondegenerate projective quadric $\{J(W) = J_0\}$ contains two families of line elements[1] which for convenience are denoted by the signs '+' and '−'. Two different lines from the same family are disjoint while two lines from different families must intersect. The intersection of those lines with the 'infinitely distant' secant plane gives points on the conic

$$\mathcal{C} := \{(W_1 : W_2 : W_3)^t \in \mathbb{C}P^2 : \quad J(W) = 0\} \tag{3.12}$$

which by means of the stereographic projection p may be identified with the Riemann sphere. Therefore we have introduced two global coordinates $p^{\pm}(W)$ on the quadric, the 'infinite part' of which (= conic $\mathcal{C}$) corresponds to coinciding coordinates: $p^+ = p^-$ (see Fig. 5).

The natural action of them pseudo-orthogonal group $O_3(J)$ in $\mathbb{C}^3$ conserves the quadric, the conic at infinity $\mathcal{C}$, and the families of line elements possibly interchanging their labels '±'. The induced action of the group $O_3(J)$ on the stereographic coordinates $p^{\pm}$ is a linear fractional with a possible change of the superscript '±'.

3.3.1. Stereographic coordinates. To obtain explicit expressions for the coordinate change $W \leftrightarrow p^{\pm}$ on the quadric we bring the quadratic form $J(W)$ to the simpler form $J_\bullet(V) := V_1V_3 - V_2^2$ by means of the linear coordinate change

$$W = \mathbf{K}V \tag{3.13}$$

where

$$\mathbf{K} := (3\delta+6)^{-1/2} \begin{Vmatrix} 1 & 1 & 1 \\ 1 & \varepsilon^2 & \varepsilon \\ 1 & \varepsilon & \varepsilon^2 \end{Vmatrix} \cdot \begin{Vmatrix} 0 & \mu^{-1} & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{Vmatrix}, \tag{3.14}$$

$$\varepsilon := \exp(2\pi i/3), \qquad \mu := \sqrt{\frac{\delta-1}{\delta+2}} = \sqrt{\frac{3-\lambda}{2\lambda}}.$$

[1]This property of quadric is sometimes used in architecture. The line generators of the hyperboloid serve as construction elements, e.g., for the Shukhov tower in Moscow.

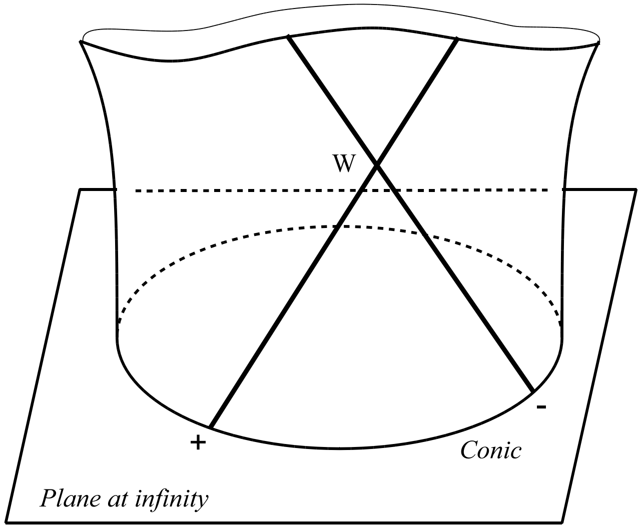


FIGURE 5. Global coordinates p^+ and p^- on quadric

Translating the first paragraph of the current section into formulae we get

$$p^{\pm}(W) := \frac{V_2 \pm i\sqrt{J_0}}{V_1} = \frac{V_3}{V_2 \mp i\sqrt{J_0}}; \tag{3.15}$$

and inverting this dependence,

$$W(p^+, p^-) = \frac{2i\sqrt{J_0}}{p^+ - p^-}\mathbf{K}\begin{pmatrix} 1 \\ (p^+ + p^-)/2 \\ p^+p^- \end{pmatrix}. \tag{3.16}$$

The point $W(p^+, p^-)$ with coordinate p^+ (resp. p^-) being fixed moves on the straight line with the directing vector $\mathbf{K}(1 : p^+ : (p^+)^2)$ (resp. $\mathbf{K}(1 : p^- : (p^-)^2)$) belonging to the conic (3.12).

3.3.2. Action of the pseudo-orthogonal group.

Lemma 3.4. *There exists a (spinor) representation* $\chi : O_3(J) \to PSL_2(\mathbb{C})$ *such that:*

1) *The restriction of* $\chi(\cdot)$ *to* $SO_3(J)$ *is an isomorphism to* $PSL_2(\mathbb{C})$.
2) *For coordinates* $p^{\pm}$ *on the quadric the following transformation rule holds:*

$$\begin{aligned} p^{\pm}(\mathbf{T}W) &= \chi(\mathbf{T})p^{\pm}(W), && \mathbf{T} \in SO_3(J), \\ p^{\pm}(\mathbf{T}W) &= \chi(\mathbf{T})p^{\mp}(W), && \mathbf{T} \notin SO_3(J). \end{aligned} \tag{3.17}$$

3) *The linear-fractional mapping* $\chi p := (ap+b)/(cp+d)$ *is the image of the matrix:*

$$\mathbf{T} := \frac{1}{ad-bc}\,\mathbf{K}\begin{Vmatrix} d^2 & 2cd & c^2 \\ bd & ad+bc & ac \\ b^2 & 2ab & a^2 \end{Vmatrix}\mathbf{K}^{-1} \in SO_3(J). \tag{3.18}$$

4) *The generators of the monodromy group are mapped to the following elements of* PSL_2*:*

$$\begin{aligned} &\chi(\mathbf{D}_s)p = \varepsilon^{1-s}/p, \qquad s = 1,2,3; \\ &\chi(\mathbf{D})p = \frac{\mu p - 1}{p-\mu}. \end{aligned} \tag{3.19}$$

Proof. We define the action of the matrix $\mathbf{A} \in SL_2(\mathbb{C})$ on the vector $V \in \mathbb{C}^3$ by the formula:

$$\mathbf{A} := \begin{Vmatrix} a & b \\ c & d \end{Vmatrix}: \quad \begin{Vmatrix} V_3 & V_2 \\ V_2 & V_1 \end{Vmatrix} \longrightarrow \mathbf{A}\begin{Vmatrix} V_3 & V_2 \\ V_2 & V_1 \end{Vmatrix}\mathbf{A}^t. \tag{3.20}$$

It is easy to check that (3.20) gives the faithful representation of a connected 3-dimensional group $PSL_2(\mathbb{C}) := SL_2(\mathbb{C})/\{\pm\mathbf{1}\}$ into $SO_3(J_\bullet)$ and therefore, an isomorphism. Let us denote by $\chi_\bullet$ the inverse isomorphism $SO_3(J_\bullet) \to PSL_2(\mathbb{C})$ and let $\chi(\pm\mathbf{T}) := \chi_\bullet(\mathbf{K}^{-1}\mathbf{T}\mathbf{K})$ for $\mathbf{T} \in SO_3(J)$. The obtained homomorphism $\chi : O_3(J) \to PSL_2(\mathbb{C})$ will satisfy statement 1) of the lemma.

To prove 2) we replace components of the vector V in the right-hand side of (3.20) with their representation in terms of the stereographic coordinates $p^\pm$:

$$\begin{aligned} &\frac{i\sqrt{J_0}}{p^+-p^-}\mathbf{A}\ \left[(p^+,1)^t\cdot(p^-,1)+(p^-,1)^t\cdot(p^+,1)\right]\ \mathbf{A}^t \\ &= i\sqrt{J_0}\frac{(cp^++d)(cp^-+d)}{p^+-p^-}\left[(\chi p^+,1)^t\cdot(\chi p^-,1)+(\chi p^-,1)^t\cdot(\chi p^+,1)\right] \\ &= \frac{i\sqrt{J_0}}{\chi p^+-\chi p^-}\left[(\chi p^+,1)^t\cdot(\chi p^-,1)+(\chi p^-,1)^t\cdot(\chi p^+,1)\right] \\ &= \begin{Vmatrix} V_3(\chi p^+,\chi p^-) & V_2(\chi p^+,\chi p^-) \\ V_2(\chi p^+,\chi p^-) & V_1(\chi p^+,\chi p^-) \end{Vmatrix}, \end{aligned}$$

where we set $\chi p := (ap+b)/(cp+d)$. Now (3.17) follows immediately for $\mathbf{T} \in SO_3(J)$. It remains to check the transformation rule for any matrix $\mathbf{T}$ from the other component of the group $O_3(J)$, say $\mathbf{T} = -\mathbf{1}$.

Writing the action (3.20) component-wise we arrive at conclusion 3) of the lemma.

And finally, expressions 4) for the generators of monodromy group may be obtained either from analyzing formula (3.18) or from finding the eigenvectors of the matrices $\mathbf{D}_s, \mathbf{D}$ which correspond to the fixed points of linear-fractional transformations. □

For convenience we collect all the introduced objects related to the boundary components of the pair of pants $\mathcal{P}(R_3)$ in Tab. 1

TABLE 1. Slots, their associated colors, matrices and linear-fractional maps

slot	$[-1,1]$	$[a_1,a_2]$	$[a_3,a_4]$
color	"red"	"blue"	"green"
matrix $\mathbf{D}_*$	$\mathbf{D} := \left\lVert\begin{matrix} -1 & \delta & \delta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}\right\rVert$	$\mathbf{D}_3 := \left\lVert\begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{matrix}\right\rVert$	$\mathbf{D}_2 := \left\lVert\begin{matrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{matrix}\right\rVert$
$\chi(\mathbf{D}_*)p$	$\dfrac{\mu p - 1}{p - \mu}$	ε/p	ε^2/p

3.4. Entangled projective structures

Definition 3.5. A branched complex projective structure [5, 6, 8, 9] on a Riemann surface $\mathcal{M}$ is a meromorphic function $p(t)$ on the universal covering $\widetilde{\mathcal{M}}$ which transforms fractionally linear under the cover transformations of $\widetilde{\mathcal{M}}$. The appropriate representation $\chi : \quad \pi_1(\mathcal{M}) \to PSL_2(\mathbb{C})$ is called the *monodromy* of the projective structure. The set of all critical points of $p(t)$ with their multiplicities survives under the cover transformations of $\widetilde{\mathcal{M}}$. The projection of this set to the Riemann surface $\mathcal{M}$ is known as the *branching divisor* $\mathsf{D}(p)$ of projective structure and the branching number of the structure $p(t)$ is $\deg \mathsf{D}(p)$.

Examples. The unbranched projective structures arise in Fuchsian and Schottky uniformizations of the Riemann surface. Any meromorphic function on a Riemann surface is a branched projective structure with trivial monodromy.

3.4.1. Projective structures generated by eigenfunction. Every bounded solution $W(y)$ of the Riemann monodromy problem (3.7) generates two nowhere coinciding meromorphic functions $p^\pm(y)$ in the sphere with three slots. Those functions are stereographic coordinates (3.15) for the vector $W(y)$. The boundary values of functions $p^+(y)$ and $p^-(y)$ on every slot are related by linear-fractional transformations:

$$p^\pm(y + i0) = \chi(\mathbf{D}_*)p^\mp(y - i0), \qquad y \in \{\text{slot}_*\} \tag{3.21}$$

where the matrix $\mathbf{D}_* = \mathbf{D}$, $\mathbf{D}_2$, $\mathbf{D}_3$ stand for the 'red', 'green' and 'blue' slots respectively.

Relations (3.21) allow us to analytically continue both functions $p^+(y)$ and $p^-(y)$ through any slot to the second sheet of the genus 2 Riemann surface

$$M := \left\{ w^2 = (y^2 - 1)\prod_{s=1}^{4}(y - a_s) \right\}, \tag{3.22}$$

and further to its universal covering $\tilde{M}$. Thus obtained functions $p^\pm(t)$, $t \in \tilde{M}$, will be locally single valued on the Riemann surface since all matrices $\mathbf{D}_*$ are involutive. However varying the argument t along the handle of the surface M may result in a linear-fractional transformation of the value $p^\pm(t)$. Say, the continuations of $p^+(y)$

from the pants through the red and green slots will give two different functions on the second sheet related by the linear-fractional mapping $\chi(\mathbf{DD}_2)$.

3.4.2. Branching of the structures $p^{\pm}$. The way we have carried out the continuation of functions $p^{\pm}(y)$ suggests that the branching divisors of the arising projective structures are related via the hyperelliptic involution $H(y,w) := (y,-w)$ of the surface M:

$$\mathsf{D}(p^+) = H\mathsf{D}(p^-). \tag{3.23}$$

The condition $p^+ \neq p^-$ allows to determine the branching numbers of the structures which is done in the next theorem.

Theorem 3.6. [11] *When $\lambda \notin \{0,1,3\}$ the solutions $u(x)$ of the integral equation PS-3 that have invariant $J_0 \neq 0$ are in one-to-one correspondence with the couples of not identically equal functions meromorphic in the pants $\mathcal{P}(R_3)$ $p^{\pm}(y)$ with boundary values satisfying* (3.21) *and two critical points in common. The correspondence $u(x) \to p^{\pm}(y)$ is established by the sequence of formulae* (3.2), (3.6) *and* (3.15)*; the inverse dependence is given by the formula*

$$2\pi u(x) = \sqrt{\frac{(\delta+2)J_0}{3}}\,\frac{p^+(y)p^-(y) - \mu(p^+(y)+p^-(y)) + 1}{p^+(y)-p^-(y)}, \tag{3.24}$$

where $x \in [-1,1]$ and $y := R_3(x) + i0$.

Remark: The number of critical points of the structures in the pants is counted with their *weight and multiplicity* (see Remark 3 on page 28): 1) the branching number of $p^{\pm}(y)$ at the branch point $a \in \{\pm 1, a_1, \dots, a_4\}$ of M is computed with respect to the local parameter $z = \sqrt{y-a}$, 2) every branch point of the projective structure on the boundary of the pants should be considered as a half-point.

Proof. **1.** Let $u(x)$ be an eigenfunction of the integral equation PS-3, then the stereographic coordinates $p^{\pm}(y)$ of the solution of the associated Riemann monodromy problem inherit the boundary relationship (3.21). What remains is to find the branching numbers of the entangled structures $p^{\pm}(y)$. To this end we consider the *Kleinian quadratic differential* on the slit sphere

$$\Omega(y) = \frac{dp^+(y)dp^-(y)}{(p^+(y)-p^-(y))^2}, \qquad y \in \hat{\mathbb{C}}. \tag{3.25}$$

This expression is the infinitesimal form of the cross ratio, hence it remains unchanged after the same linear-fractional transformations of the functions p^+ and p^-. Therefore, (3.25) is a well-defined quadratic differential on the entire sphere. Lifting $\Omega(y)$ to the surface M we get a holomorphic differential. Indeed, $p^+ \neq p^-$ everywhere and applying suitable linear-fractional transformation we assume that $p^+ = 1 + z^{m_+} + \{\text{terms of higher order}\}$ and $p^- = cz^{m_-} + \cdots$ in terms of local parameter z of the surface, $m_\pm \geq 1$, $c \neq 0$. Then $\Omega = cm_+m_-z^{m_++m_--2} + \{\text{terms of higher order}\}$. Therefore

$$\mathsf{D}(p^+) + \mathsf{D}(p^-) = (\Omega).$$

Any holomorphic quadratic differential on a genus 2 surface has 4 zeroes and taking into account the symmetry (3.23) of the branching divisors, we see that each of the structures $p^{\pm}$ has the branching number two on the curve M. It remains to note that the pair of pants $\mathcal{P}(R_3)$ are exactly "one half" of M.

2. Conversely, let $p^+(y)$ and $p^-(y)$ be two not identically equal meromorphic functions on the slit sphere, with boundary conditions (3.21) and total branching number two in the pants (see remark above). We can prove that $p^+ \neq p^-$ everywhere. Indeed, for the meromorphic quadratic differential (3.25) on the Riemann surface M we establish (using a local coordinate on the surface) the inequality

$$\mathsf{D}(p^+) + \mathsf{D}(p^-) \geq (\Omega) \tag{3.26}$$

where the deviation from equality means that there is a point where $p^+ = p^-$. But the degree of the divisor on the left of (3.26) is four and the same number is $\deg(\Omega) = 4g - 4$. Therefore this pair of functions $p^{\pm}$ will give us the holomorphic vector $W(p^+(y), p^-(y))$ in the pants which solves our Riemann monodromy problem. We already know how to convert the latter vector to the eigenfunction of the integral equation PS-3. □

3.4.3. Remark about the non-smooth quadric. It is shown in [11] how to incorporate the exceptional case $J_0 = 0$ into the above scheme. In the latter case the functions $p^{\pm}(y)$ coincide, however the boundary relations (3.21) survive. The total branching number of the function $p^+ = p^-$ in the pair of pants is either zero or one. The solutions to the PS-3 integral equation and the associated Riemann monodromy problem may be recovered up to proportionality from the unified formulae (true whatever J_0)

$$u(x) = \sqrt{\frac{\Omega(y)}{dp^+(y)dp^-(y)}}(p^+(y)p^-(y) - \mu(p^+(y) + p^-(y)) + 1), \tag{3.27}$$

$$W(y) = \sqrt{\frac{\Omega(y)}{dp^+(y)dp^-(y)}}\mathbf{K}(1, (p^+(y) + p^-(y))/2, p^+(y)p^-(y))^t, \tag{3.28}$$

where $\Omega(y) = (y - y_1)(y - y_2)\dfrac{(dy)^2}{w^2(y)}$ is the holomorphic quadratic differential on the Riemann surface M with zeroes at the branching points of the possibly coinciding structures p^+ and p^- [or with two arbitrary double zeroes when the structure $p^+ = p^-$ is unbranched, further analysis however shows that the required unbranched structures do not exist].

3.5. Types of the mirror symmetry of the solution

The eigenvalues of the integral equation are the critical values of the *positive* functional (1.2) – the generalized Rayleigh ratio. So we may consider only *real* eigenfunctions $u(x)$ without loss of generality. Real solutions of the PS-3 equation give rise to exactly two types of *mirror symmetry* for the entangled structures:

$$\begin{array}{ll} \text{Symmetric} & p^{\pm}(\bar{y}) = 1/\overline{p^{\pm}(y)} \\ \text{Antisymmetric} & p^{\pm}(\bar{y}) = 1/\overline{p^{\mp}(y)} \end{array}, \qquad y \in \mathcal{P}(R_3),$$

depending on the sign of the real number $(\delta + 2)J_0$. In what follows we restrict ourselves to the case of *antisymmetric eigenfunctions*. In this case:

$$p^{+}(y \pm i0) = 1/\overline{p^{-}(y \mp i0)} = 1/\overline{\chi(\mathbf{D}_*)p^{+}(y \pm i0)}, \qquad y \in slot_*,$$

and hence we know where the boundary components of the pair of pants $\mathcal{P}(R_3)$ are mapped to. In particular,

$$\begin{array}{ll} \text{“green” boundary} & \to \varepsilon\hat{\mathbb{R}} \\ \text{“blue” boundary} & \to \varepsilon^2\hat{\mathbb{R}} \\ \text{“red” boundary} & \to \left\{ \begin{array}{lll} C & \text{– see (2.2)} & \text{when } 1 < \lambda < 3 \\ \emptyset & & \text{when } \lambda < 1 \textit{ or } 3 < \lambda. \end{array} \right. \end{array} \tag{3.29}$$

We see that the above geometrical analysis of the integral equation gives the universal limits for (the antisymmetric part of) the spectrum.

The branching divisor of the projective structure p^+ has the mirror symmetry: $\mathsf{D}(p^+) = \bar{H}\mathsf{D}(p^+)$ where $\bar{H}(y, w) := (\bar{y}, -\bar{w})$ is the anticonformal involution of the surface M leaving boundary components (ovals) of pair of the pants $\mathcal{P}(R_3)$ intact. Therefore exactly three situations may occur: $p^+(y)$ has one simple critical point strictly inside the pants, or there are two simple critical points on the boundary of pants or there is one double critical point of $p^+(y)$ on the boundary of the pants.

4. Combinatorics of integral equation

For the antisymmetric eigenfunctions we arrive at the essentially combinatorial **Problem** (about putting pants on a sphere). *Find a meromorphic function* $p := p^+$ *defined in the pair of pants* $\mathcal{P}(R_3)$ *mapping boundary ovals to the given circles* (3.29) *and having exactly one critical point* (*counted with weight and multiplicity*) *in the pants.*

The three above-mentioned types of the branching divisor $\mathsf{D}(p)$ will be treated separately in Sections 4.1, 4.2. When the branch point of the structure p is strictly inside the pants we show that the solution of the problem takes the form of the Überlagerungsfläche $\mathcal{P}_1(\dots)$ with certain real and integer parameters. The case of two simple branch points belonging to the boundary gives us the pants $\mathcal{P}_s(\dots)$, $s = 2, 3$ and the unstable intermediate case with double branch point on the boundary brings us to the pants $\mathcal{P}_j(\dots)$, $j = 12,\ 13$ described in (2.3).

Let $p(y)$ be a holomorphic map from a Riemann surface $\mathcal{M}$ with a boundary to the sphere and the selected boundary component $(\partial\mathcal{M})_*$ be mapped to a circle. The reflection principle allows us to holomorphically continue $p(y)$ through this selected component to the double of $\mathcal{M}$. Therefore we can talk of the critical points of $p(y)$ on $(\partial\mathcal{M})_*$. When the argument y passes through a simple critical point, the value $p(y)$ reverses the direction of its movement on the circle. So there should be an even number of critical points (counted with multiplicities) on the selected boundary component.

4.1. The branchpoint is inside a pair of pants

4.1.1. Construction 1. Using otherwise a composition with a suitable linear-fractional map, we suppose that the circle $p((\partial\mathcal{M})_*)$ is the boundary of the unitary disc

$$\mathbb{U} := \{p \in \mathbb{C} : \quad |p| \le 1\}, \tag{4.1}$$

and that a small annular vicinity of the selected boundary component is mapped to the exterior of the unit disc. We define the mapping of a disjoint union $\mathcal{M} \cup \mathbb{U}$ to a sphere

$$\tilde{p}(y) := \begin{cases} p(y), & y \in \mathcal{M}, \\ L(y^d), & y \in \mathbb{U}, \end{cases} \tag{4.2}$$

where the integer $d > 0$ is the degree of the mapping $p : (\partial\mathcal{M})_* \to \partial\mathbb{U}$, and where $L(y)$ is an (at the moment arbitrary) linear fractional mapping keeping the unitary disc (4.1) unchanged. The choice of $L(\cdot)$ will be fixed later to simplify the arising combinatorial analysis.

Now we fill in the hole in $\mathcal{M}$ by the unit disc, identifying the points of $(\partial\mathcal{M})_*$ and the points of $\partial\mathbb{U}$ with the same value of $\tilde{p}$ (there are d ways to do so). The holomorphic mapping $\tilde{p}(y)$ of the new Riemann surface $\mathcal{M} \cup \mathbb{U}$ to the sphere will have exactly one additional critical point of multiplicity $d - 1$ at the center of the glued disc.

4.1.2. Branched covering of a sphere. We return to the function $p(y)$ being the solution of the problem stated in the beginning of section 4. Suppose that the point $p(y)$ completes turns on the corresponding circle d_r, d_g and d_b times when the argument y runs around the 'red', 'green' and 'blue' boundary component of $\mathcal{P}(R_3)$ respectively. We can apply the just introduced *construction* 1 and glue the three discs $\mathbb{U}_r$, $\mathbb{U}_g$, $\mathbb{U}_b$, to the holes of the pants. Essentially, we arrive at a commutative diagram:

$$\begin{array}{ccc} \mathcal{P}(R_3) & \xrightarrow{\text{inclusion}} & \mathbb{C}P^1 \\ & \searrow^{p(y)} & \Big\downarrow \tilde{p} \text{ is branched covering.} \\ & & \mathbb{C}P^1 \end{array} \tag{4.3}$$

Applying the Riemann–Hurwitz formula for the holomorphic mapping $\tilde{p}$ with four ramification points (three of them are in the artificially glued discs and the fourth is inside the pants) we immediately get:

$$d_r + d_g + d_b = 2N, \quad N := \deg \tilde{p}. \tag{4.4}$$

4.1.3. Intersection of circles.

Lemma 4.1. *The circle C does not intersect the two other circles $\varepsilon^{\pm 1}\hat{\mathbb{R}}$. Therefore the spectral parameter $1 < \lambda < 2$ when the projective structure $p(y)$ branch point is inside the pants.*

Proof. We know that the point 0 lies in the intersection of two of our circles: $\varepsilon\hat{\mathbb{R}}$ and $\varepsilon^2\hat{\mathbb{R}}$. The total number $\sharp\{\tilde{p}^{-1}(0)\}$ of the pre-images of this point (counting the multiplicities) is N and cannot be less than $d_b + d_g$ – the number of pre-images on the blue and green boundary components of the pants. Comparing this to (4.4) we get $d_r \geq N$ which is only possible when

$$d_r = d_g + d_b = N. \tag{4.5}$$

Assuming that the circle C intersects any of the circles $\varepsilon^{\pm 1}\hat{\mathbb{R}}$ we repeat the above argument for the intersection point and arrive at the conclusion $d_b = d_r + d_g = N$ or $d_g = d_r + d_b = N$ which is incompatible with the already established equation (4.5). ☐

Remark 4.2. In Section 3.4.3 we promised to show that any meromorphic function p mapping the boundaries of the pants to the circles (3.29) has a *critical point.* Indeed, the inequalities $d_b + d_g \leq N$ and $d_r \leq N$ remain true whatever the branching of the structure p is, while (4.4) originating from the Riemann–Hurwitz formula takes the form $d_b + d_g + d_r = 2N + 1$ for the *unbranched* structure which leads to a contradiction.

4.1.4. Image of the pants. Let us investigate where the artificially glued discs are mapped to. Suppose for instance that the disc $\mathbb{U}_r$ is mapped to the exterior of the circle C. The point 0 will be covered then at least $d_r + d_g + d_b = 2N$ times which is impossible. The discs $\mathbb{U}_g$ and $\mathbb{U}_b$ are mapped to the left of the lines $\varepsilon\mathbb{R}$ and $\varepsilon^2\mathbb{R}$ respectively, otherwise points from the interior of the circle C will be covered more that N times. The image of the pair of pants $\mathcal{P}(R_3)$ is shown on the left of Fig. 6.

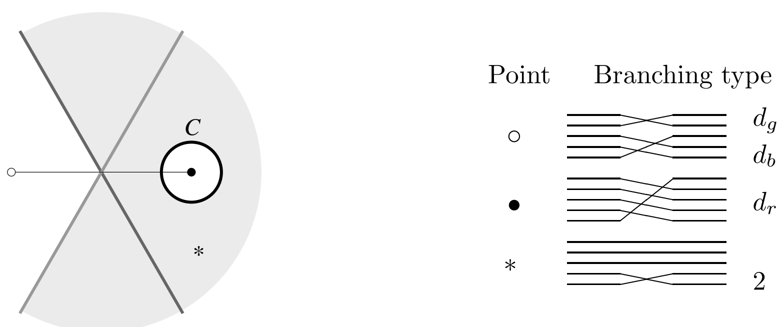
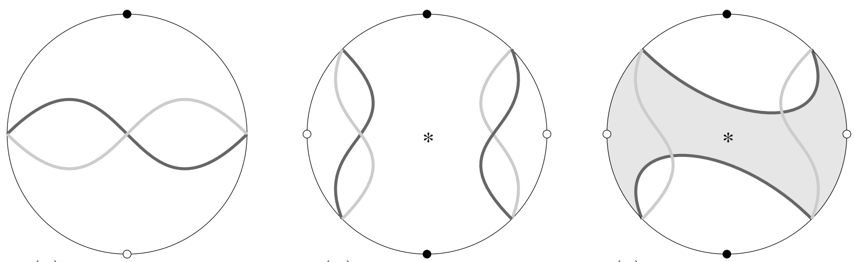


FIGURE 6. **a**) Shaded area is the image of pants (**b**) Branching type of the branch points

We use the ambiguity in the construction of the gluing of the disks to the pants and require that the critical values of $\tilde{p}$ in the discs $\mathbb{U}_g$, $\mathbb{U}_b$ coincide. Now the branched covering $\tilde{p}$ has only three different branch points shown as $\bullet$, $\circ$ and $*$ on Fig. 6a). The branching type at those three points for $d_g = 2, d_b = 3$, $d = N = 5$ is shown on Fig. 6b). The coverings with three branch points are called *Belyi maps* and are described by certain graphs known as Grothendieck's *"Dessins d'Enfants"*. In our case the *dessin* is the lifting of the segment connecting white and black branch points: $\Gamma := \tilde{p}^{-1}[\bullet, \circ]$.

4.1.5. Combinatorial analysis of the dessins. There is exactly one critical point of $\tilde{p}$ over the branch point $*$. Hence, the complement to the graph Γ on the upper sphere of the diagram (4.3) contains exactly one cell mapped $2-1$ to the lower sphere. The rest of the components of the complement are mapped $1-1$. Two types of cells are shown in figures 7 a) and b), the lifting of the red circle is not shown to simplify the pictures. The branch point $*$ should lie in the intersection of the two annuli α and $\overline{\alpha}$, otherwise the discs $\mathbb{U}_g$, $\mathbb{U}_b$ glued to different boundary components of our pants will intersect: the hypothetical case when the branch point of $p(y)$ belongs to one annulus but does not belong to the other is shown in Fig. 7 c).



(**a**) (**b**) (**c**)

FIGURE 7. (**a**) Simple cell (**b**) Double cover (**c**) Impossible double cover

The cells from Fig. 7 a), b) may be assembled in a unique way shown in Fig. 8. The pants are colored in white, three artificially sewed discs are shaded. Essentially this picture shows us how to sew together the patches bounded by our three circles C, $\varepsilon^{\pm 1}\hat{\mathbb{R}}$ to get the pants conformally equivalent to $\mathcal{P}(R_3)$. As a result of the surgery procedure we obtain the pants $\mathcal{P}_1(\lambda, h_1, h_2 | d_g - 1, d_b - 1)$.

4.2. Simple branch points on the boundary of the pants

Our strategy remains the same: to fill in the holes in the pants and to convert $p(y)$ into a branched covering with a simple type of branching.

4.2.1. Construction 2. Let again $p(y)$ be a holomorphic mapping of a bounded Riemann surface $\mathcal{M}$ to the sphere with the selected boundary component $(\partial\mathcal{M})_*$ being mapped to the boundary of the unit disc $\mathbb{U}$. Now the mapping $p(y)$ has two

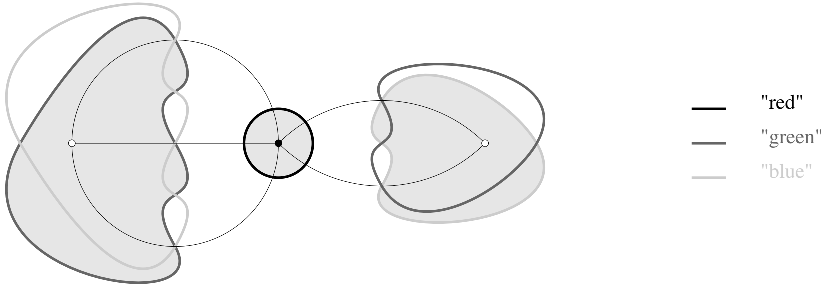


FIGURE 8. Dessin for $d_g = 3, d_b = 2$; the pre-image of the branch point $*$ is at the infinity

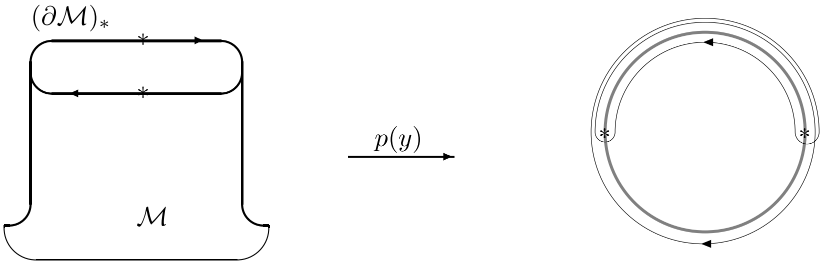


FIGURE 9. Mapping of the boundary component $(\partial\mathcal{M})_*$ with two simple branch points $*$ on it and winding indices $d^+ = 1, d^- = 2$.

simple critical points on the selected boundary component (the case of coinciding critical values is not excluded). Those two points divide the oval $(\partial\mathcal{M})_*$ into two segments: $(\partial\mathcal{M})_*^+$ and $(\partial\mathcal{M})_*^-$. Let the increment of arg $p(y)$ on the segment $(\partial\mathcal{M})_*^+$ be $2\pi d^+ - \phi$, $0 < \phi \le 2\pi$, and the decrement on the segment $(\partial\mathcal{M})_*^-$ be $2\pi d^- - \phi$, the point y moves around the selected oval in the positive direction and $d^\pm$ are positive integers. We are going to fill in the hole in the Riemann surface $\mathcal{M}$ with two copies of the unitary disc (4.1): $\mathbb{U}^+$ and $\mathbb{U}^-$.

We define the mapping from the disjoint union $\mathcal{M}\cup\ \mathbb{U}^+\cup\ \mathbb{U}^-$ to the sphere:

$$\tilde{p}(y) := \begin{cases} p(y), & y \in \mathcal{M}, \\ L^-(y^{d^-}), & y \in \mathbb{U}^-, \\ L^+(y^{-d^+}), & y \in \mathbb{U}^+, \end{cases} \tag{4.6}$$

where $L^\pm(\cdot)$ are the (at the moment arbitrary) linear fractional mappings keeping the unitary disc (4.1) invariant. The choice of $L^\pm(\cdot)$ will be specified later to simplify the combinatorial analysis.

Identifying the points y with the same value of $\tilde{p}(y)$ we glue the segments $(\partial\mathcal{M}*)^\pm$ of the selected boundary oval of $\mathcal{M}$ to the portions of the boundaries of the discs $\mathbb{U}^\pm$ respectively. The remaining parts of the boundaries of $\mathbb{U}^\pm$ are glued to each other as shown in Fig. 10a).

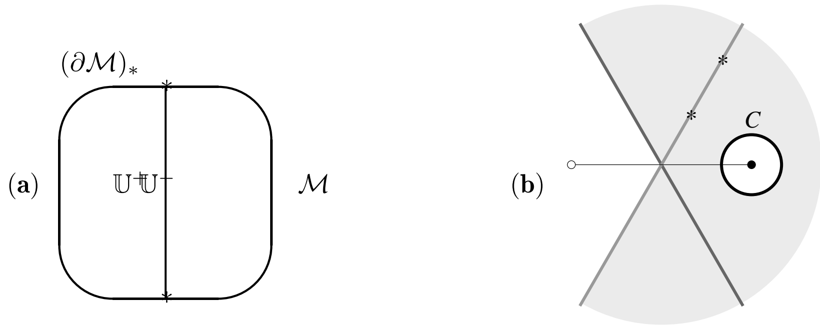


FIGURE 10. (**a**) Filling in the hole bounded by $(\partial\mathcal{M})_*$ (**b**) The shaded area is the image of $\mathcal{P}(R_3)$.

4.2.2. Branched covering of a sphere. At the moment we do not know which of the three boundary ovals of the pants $\mathcal{P}(R_3)$ contains the critical points of $p(y)$. Therefore we introduce the 'nicknames' $\{1,2,3\}$ for the set of colors $\{r,g,b\}$ so that the critical points will be on the oval number 3. The usage of *construction* 2 from Section 4.2.1 allows us to glue two discs $\mathbb{U}_3^{\pm}$ to the latter boundary. The usage of *construction* 1 from Section 4.1.1 fills in the remaining two holes with two discs $\mathbb{U}_1$ and $\mathbb{U}_2$. Positive integers arising in those constructions are denoted by $d_3^{\pm}$, d_1, d_2 respectively.

Again, we split the mapping $p(y)$ from the pants to the sphere as in the diagram (4.3): $p = \tilde{p}\circ$ *inclusion* with the branched covering $\tilde{p}$. The latter mapping has six critical points: two simple ones inherited from the pants and four at the centers of the artificially glued discs and multiplicities $d_3^{\pm}-1$, d_1-1, d_2-1 respectively. The Riemann–Hurwitz formula for this covering gives

$$d_1 + d_2 + d_3^+ + d_3^- = 2N, \qquad N := \deg\tilde{p}. \tag{4.7}$$

Lemma 4.3. *The images of the ovals with numbers* 1 *and* 2 *do not intersect.*

Proof. Suppose the opposite is true and a point Pt lies in the intersection of the images of the first two ovals. Then $N \ge \sharp\tilde{p}^{-1}(Pt) \ge d_1 + d_2$. On the image of the third oval there is a point (e.g., in the right side of Fig. 9 this is a point i) with $d_3^+ + d_3^- \le N$ pre-images. Comparing the last two inequalities to (4.7) we get the equalities

$$d_1 + d_2 = d_3^+ + d_3^- = N$$

and Pt is covered at least $d_1 + d_2 + \min(d_3^+, d_3^-) > N$ times. □

Corollary 4.4. *Two circles* $\varepsilon^{\pm 1}\hat{\mathbb{R}}$ *intersect, therefore the critical points of* $p(y)$ *lie either on the blue or on the green boundary of pants. Moreover, the circle* C *– the image of the red boundary oval – does not intersect the two mentioned circles which may only happen when* $\mu \in (\frac{1}{2}, 1)$*, or equivalently* $\lambda \in (1,2)$.

Convention: We assume that both critical points of p lie on the blue oval. The remaining case when they belong to the green oval is absolutely analogous to the case we consider. Now the notations $\mathbb{U}_b^{\pm}$, $\mathbb{U}_r$, $\mathbb{U}_g$, $d_b^{\pm}$, d_r, d_g have the obvious meaning.

4.2.3. The Image of the Pants. Let us show that the the image of the pants remains the same as in Section 4.1.4.

Lemma 4.5. *The image $p(\mathcal{P}(R_3))$ of the pants lies in the intersection of annuli α and $\bar{\alpha}$ – see Fig.* 10b).

Proof. We refer to the four sectors: $\mathbb{C} \setminus \varepsilon^{\pm 1}\mathbb{R}$ as to 'top', 'down', 'left' and 'right'. It is a matter of notation to say that the disc $\mathbb{U}_b^+$ is mapped to the 'top' and 'left' sectors while the disc $\mathbb{U}_b^-$ is mapped to the 'down' and 'right' sectors.

The disc $\mathbb{U}_g$ covers either the 'top' or the 'left' sector and both are covered by the disc $\mathbb{U}_b^+$. Therefore, $d_g + d_b^+ \leq N$. In a similar way we get $d_r + d_b^- \leq N$. The obtained inequalities and the Riemann–Hurwitz formula (4.7) – which in our notations becomes $d_r + d_g + d_b^+ + d_b^- = 2N$ – give us

$$d_r + d_b^- = d_g + d_b^+ = N.$$

If the disc $\mathbb{U}_r$ is mapped to the exterior of the circle C, then either 'left' or 'top' sector is covered $d_r + d_g + d_b^+ > N$ times. If the disc $\mathbb{U}_g$ is mapped to the right of the line $\varepsilon\mathbb{R}$, then the interior of the circle C is covered $d_r + d_g + d_b^- > N$ times.

We see that the 'left' sector and the interior of the circle C are covered by the artificially inserted discs only. □

Corollary 4.6. *Both critical values of $p(y)$ lie on the ray $-\varepsilon^2(0, \infty)$.*

Corollary 4.7. *The integer d_b^- is equal to* 1*, since the point* 0 *is covered at least $d_g + d_b^+ + d_b^- - 1 \leq N$ times.*

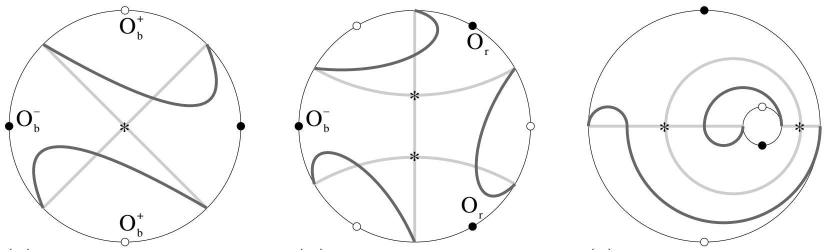
Let us recall that the constructions of attaching discs to the pants allow us to move branch point (= the critical value of $\tilde{p}(y)$ in the inserted disc) within the appropriate circle. In particular, the critical values of $\tilde{p}(y)$ in the discs $\mathbb{U}_g$, $\mathbb{U}_b^+$ may be placed to the same point in the 'left' sector, say to $p = -1$ (point $\circ$ in Fig. 10b) while the critical values in the discs $\mathbb{U}_r$, $\mathbb{U}_b^-$ may be placed to the same point inside C, say to $p = 1$ (point $\bullet$ in Fig. 10b). Now we lift the segment $[\circ, \bullet]$ connecting the branch points to the upper sphere of the diagram (4.3) and analyze the arising graph $\Gamma := \tilde{p}^{-1}([\circ, \bullet])$.

4.2.4. Combinatorial analysis of the graph. The restriction of $\tilde{p}$ to every component F of the compliment $\hat{\mathbb{C}} \setminus \Gamma$ to the graph is naturally continued to the branched coverings over the disc[2] Closure($\hat{\mathbb{C}} \setminus [\circ, \bullet]$). We can list all flat surfaces F covering a

[2] *Closure* here has the same meaning as in the formula (2.1)

TABLE 2. Flat surfaces F covering the disc with the branching number $B \leq 2$

number of sheets	B	surface F	picture
1	0	disc	Fig. 7(**a**)
2	1	disc	Fig. 11(**a**)
3	2	disc	Fig. 11(**b**)
2	2	annulus	Fig. 11(**c**)



(**a**) (**b**) (**c**)

FIGURE 11. Flat surfaces F covering the disc with the branching numbers $B = 1, 2$.

disc with the branching number $B \leq 2$. To this end we use the Riemann–Hurwitz formula for the branched coverings of the bordered surfaces:

$$2 + B = \sharp\{\partial F\} + \deg \tilde{p}|F$$

which relates B – the total branching number of $\tilde{p}$ in the selected flat surface F covering a disc; $\sharp\{\partial F\}$ – the number of its boundary components and $\deg \tilde{p}|F$ – the degree of the restriction of the covering $\tilde{p}$ to the component F. Taking into account that $\sharp\{\partial F\} \leq \deg \tilde{p}|F$ we obtain the list shown in Tab. 2.

The combinatorics of the green and blue circles lifted to the listed covering surfaces F is shown in Fig. 7a) and Fig. 11a–c). Let us denote the centers of the four artificially glued discs $\mathbb{U}_r$, $\mathbb{U}_g$, $\mathbb{U}_b^+$ $\mathbb{U}_b^-$ as respectively O_r (black vertex of graph Γ with valency d_r), O_g, O_b^+ (white vertexes with valencies d_g, d_b^+) and O_b^- (dangling black vertex). Their mutual positions in the graph Γ are subject to the following restriction:

Lemma 4.8. *The vertices O_g and O_b^- are not neighbors in Γ.*

Proof. The disjoint discs $\mathbb{U}_b^-$ and $\mathbb{U}_g$ of the upper sphere in the diagram (4.3) would intersect otherwise – see Fig. 12. □

Corollary 4.9. *The vertices on the border of the triply covering disc F – see Fig. 11b) – appear in the following order: O_g, O_r, O_b^+, O_b^-, O_b^+, O_r.*

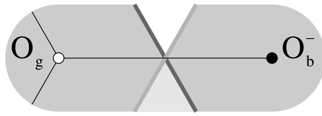


FIGURE 12. If O_g and O_b^- were neighbors, the discs $\mathbb{U}_b^-$ and $\mathbb{U}_g$ would intersect.

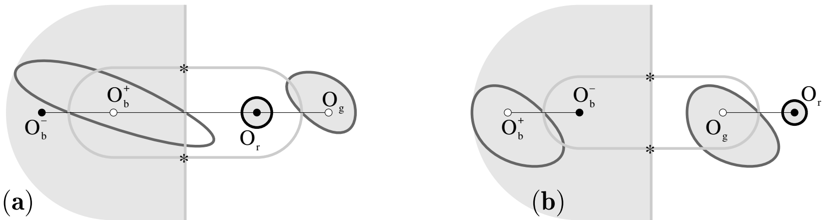


FIGURE 13. Graph Γ for the basic mappings with $d_g = 1$, $d_b^+ = d_r = 2$, $N = 3$ (**a**) and $d_g = d_b^+ = d_r = 1$, $N = 2$ (**b**). Artificially inserted discs are shaded.

They may be uniquely ascribed to the vertices in the picture after the observation: *the blue line divides the vicinity of any critical point $*$ into four quadrants, two of which belong to the pair of pants, one belongs to the disc $\mathbb{U}_b^-$, and the rest is contained in the disc $\mathbb{U}_b^+$.*

Corollary 4.10. *The complement to the graph Γ cannot contain two doubly covering discs F.*

Indeed, the point O_b^- lies on the boundary of one of those discs. Both neighboring vertices on the boundary of the disc F should be O_b^+ according to the lemma. But this contradicts the above *observation*: two quadrants of this covering disc belong to $\mathbb{U}_b^+$ – see Fig. 11a).

4.2.5. Assembly scheme. We see that there remain only two possibilities for the complement to the graph Γ. It consists either of (**a**) one disc mapped 3-1 and $N-3$ *simple* cells mapped 1-1 or (**b**) an annulus mapped 2-1 and $N-2$ simple discs mapped 1-1. The graphs Γ with compliment containing no simple cells are shown in Fig. 13. They correspond to the pants $\mathcal{P}_2(\dots|0,1)$ (**a**) and $\mathcal{P}_2(\dots|0,0)$ (**b**). The graphs with simple cells in the complement are obtained from those two basic pictures as a result of the surgery. We cut the graph along the edge O_rO_g and insert $d_g - 1$ simple discs in the slot as in Fig. 8. The graph on the left side of Fig. 13 admits another surgery: we cut the graph along the edge $O_rO_b^+$ and sew in $d_b^+ - 2$ patches shown in Fig. 7a) in the slot. The arising graph corresponds to the pair of pants $\mathcal{P}_2(\dots, d_g - 1, d_b^+ - 1)$.

4.3. Remaining cases

If the branch points of the projective structure $p := p^+$ belong to the green oval of the pants we arrive at the pair of pants $\mathcal{P}_s$ of fashion $s = 3$. Finally, when the branch points merge the limit variant of construction 2 may be applied for the analysis and we arrive at the pants of intermediate types $s = 12,\ 13$.

5. Conclusion

A similar analysis based on the geometry and combinatorics may be applied to obtain the representations of the solutions of the PS-3 integral equation in all the omitted cases. Much of the techniques used may be helpful for the study of other integral equations with low degree rational kernels.

References

[1] Agranovich M.S., Katzenelenbaum B.Z., Sivov A.N., Voitovich N.N.: *Generalized Method of Eigenoscillations in Diffraction Theory.* Wiley-VCH, Berlin (1999)

[2] Bogatyrev A.B.: *The discrete spectrum of the problem with a pair of Poincaré–Steklov operators.* Doklady RAS, **358**:3, 40–42 (1998)

[3] Bogatyrev A.B.: *A geometric method for solving a series of integral PS equations.* Math. Notes, **63**:3, 302–310 (1998)

[4] Poincaré H.: *Analyse des travaux scientifiques de Henri Poincaré. Acta Math.* **38**, 3–135 (1921)

[5] Gunning R.C.: *Special coordinate coverings of Riemann surfaces. Math. Annalen*, **170**, 67–86 (1967)

[6] Mandelbaum R.: *Branched structures and affine and projective bundles on Riemann surfaces.* Trans. AMS, **183**, 37–58 (1973)

[7] Hejhal D.A.: *Monodromy groups and linearly polymorphic functions.* Acta Math., **135**, 1–55 (1975)

[8] Tyurin A.N.: *On the periods of quadratic differentials.* Russian Math. Surveys, **33**:6, 149–195 (1978)

[9] Gallo D., Kapovich M., Marden A.: *The monodromy groups of Schwarzian equations on closed Riemann surfaces*// Ann. of Math. (2) **151**:2, 625–704 (2000); also arXiv, math.CV/9511213

[10] Bogatyrev A.B.: *Poincaré–Steklov integral equations and the Riemann monodromy problem.* Funct. Anal. Appl. **34**:2, 9–22 (2000)

[11] Bogatyrev A.B.: *PS-*3 *integral equations and projective structures on Riemann surfaces.* Sbornik: Math., **192**:4, 479–514 (2001)

Andrei Bogatyrev
INM RAS, ul. Gubkina, 8
Moscow GSP-1,
119991 Russia
e-mail: `gourmet@inm.ras.ru`

Analysis and Mathematical Physics
Trends in Mathematics, 49–76
© 2009 Birkhäuser Verlag Basel/Switzerland

Generalized Hamilton–Jacobi Equation and Heat Kernel on Step Two Nilpotent Lie Groups

Ovidiu Calin, Der-Chen Chang and Irina Markina

Abstract. We study geometrically invariant formulas for heat kernels of subelliptic differential operators on two step nilpotent Lie groups and for the Grusin operator in $\mathbb{R}^2$. We deduce a general form of the solution to the Hamilton–Jacobi equation and its generalized form in $\mathbb{R}^n \times \mathbb{R}^m$. Using our results, we obtain explicit formulas of the heat kernels for these differential operators.

Mathematics Subject Classification (2000). 53C17, 53C22, 35H20.

Keywords. Sub-Laplacian, Heat operator, H-type groups, action function, volume element.

1. Introduction

Let us start with the Laplace operator on $\mathbb{R}^n$,

$$\Delta = \frac{1}{2}\sum_{j=1}^{n} \frac{\partial^2}{\partial x_j^2}.$$

It is well known that the heat kernel for Δ is the Gaussian:

$$P_t(\mathbf{x}, \mathbf{x}_0) = \frac{1}{(2\pi t)^{\frac{n}{2}}} e^{-\frac{|\mathbf{x}-\mathbf{x}_0|^2}{2t}}.$$

Given a general second-order elliptic operator in n-dimensional Euclidean space,

$$\Delta_X = \frac{1}{2}\sum_{j=1}^{n} X_j^2 + \text{lower-order term},$$

The first author is partially supported by the NSF grant #0631541.
The second author is partially supported by a Hong Kong RGC competitive earmarked research grant #600607, a competitive research grant at Georgetown University.
The third author is supported by NFR grant # 177355/V30, NordForsk Research Network, and ESF Networking Programme HCAA.

where the $\{X_1, \dots, X_n\}$ is a linearly independent set of vector fields, the heat kernel takes the form

$$P_t(\mathbf{x}, \mathbf{x}_0) = \frac{1}{(2\pi t)^{\frac{n}{2}}} e^{-\frac{d^2(\mathbf{x},\mathbf{x}_0)}{2t}} \big(a_0 + a_1 t + a_2 t^2 + \cdots\big).$$

Here $d(\mathbf{x}, \mathbf{x}_0)$ stands for the Riemannian distance between $\mathbf{x}$ and $\mathbf{x}_0$ induced by a metric such that vector fields $X_1, \dots, X_n$ are orthonormal with respect to this metric. The a_j's are functions of $\mathbf{x}$ and $\mathbf{x}_0$. Note that

$$\frac{\partial}{\partial t}\Big(\frac{d^2}{2t}\Big) + \frac{1}{2}\sum_{j=1}^{n}\Big(X_j \frac{d^2}{2t}\Big)^2 = 0,$$

i.e., $\frac{d^2}{2t}$ is a solution of the Hamilton–Jacobi equation.

Now let us move to subelliptic operators. We first consider the famous example: Heisenberg sub-Laplacian on $\mathbb{H}_1$

$$\Delta_X = \frac{1}{2}\Big(\frac{\partial}{\partial x_1} + 2x_2\frac{\partial}{\partial y}\Big)^2 + \frac{1}{2}\Big(\frac{\partial}{\partial x_2} - 2x_1\frac{\partial}{\partial y}\Big)^2. \tag{1.1}$$

We shall try to find a heat kernel in the form

$$\frac{1}{t^q} e^{-\frac{f}{t}} \cdots ,$$

where $h = \frac{f}{t}$ is a solution of the Hamilton–Jacobi equation

$$\frac{\partial h}{\partial t} + \frac{1}{2}\Big(\frac{\partial h}{\partial x_1} + 2x_2\frac{\partial h}{\partial y}\Big)^2 + \frac{1}{2}\Big(\frac{\partial h}{\partial x_2} - 2x_1\frac{\partial h}{\partial y}\Big)^2 = 0.$$

In other words,

$$\frac{\partial h}{\partial t} + H(\mathbf{x}, \nabla h) = 0, \tag{1.2}$$

where

$$H = \frac{1}{2}\Big[\big(\xi_1 + 2x_2\eta\big)^2 + \big(\xi_2 - 2x_1\eta\big)^2\Big] = \frac{1}{2}\big[\zeta_1^2 + \zeta_2^2\big] \tag{1.3}$$

is the Hamilton function associated with the sub-elliptic operator (1.1) and ξ_1, ξ_2 and η are dual variable to x_1, x_2 and y respectively. Using the Lagrange-Chapit method, let us look at the following equation:

$$F(\mathbf{x}, y, t, h, \xi, \eta, \gamma) = \gamma + H(\mathbf{x}, y, \xi, \eta) = 0.$$

We shall find the bicharacteristic curves which are solutions to the following Hamilton system:

$$\begin{aligned}
\dot{x}_1 &= F_{\xi_1} = \xi_1 + 2x_2\eta = \zeta_1, \\
\dot{x}_2 &= F_{\xi_2} = \xi_2 - 2x_1\eta = \zeta_2, \\
\dot{y} &= F_\eta = 2\dot{x}_1 x_2 - 2x_1\dot{x}_2, \\
\dot{t} &= F_\gamma = 1, \\
\dot{\xi}_1 &= -F_{x_1} - \xi_1 F_h = 2\eta\dot{x}_2, \\
\dot{\xi}_2 &= -F_{x_2} - \xi_2 F_h = -2\eta\dot{x}_1,
\end{aligned}$$

$$\dot{\eta} = -F_y - \gamma F_h = 0,$$
$$\dot{\gamma} = -F_t - \gamma F_h = 0,$$
$$\dot{h} = \xi \cdot \nabla_\xi F + \eta F_\eta + \gamma F_\gamma = \xi \cdot \dot{x} + \eta \dot{y} - H$$

since $\dot{t} = 1$ and $\gamma = -H$. With $0 \le s \le t$, one has

$$\begin{aligned} \gamma(s) =& \gamma = \text{constant}, \\ \eta(s) =& \eta = \text{constant}, \\ t(s) =& s. \end{aligned}$$

Here "constant" means "*constant along the bicharacteristic curve*". Furthermore,

$$H = \frac{1}{2}\dot{x}_1^2 + \frac{1}{2}\dot{x}_2^2 = E = \text{energy}.$$

Another way to see that E is constant along the bicharacteristic, note that

$$\begin{aligned} \ddot{x}_1 =& \dot{\xi}_1 + 2\eta\dot{x}_2 = +4\eta\dot{x}_2, \\ \ddot{x}_2 =& \dot{\xi}_2 - 2\eta\dot{x}_1 = -4\eta\dot{x}_1. \end{aligned} \tag{1.4}$$

Therefore, $\ddot{x}_1\dot{x}_1 + \ddot{x}_2\dot{x}_2 = 0$, and E =constant.

We need to find the classical action integral

$$S(t) = \int_0^t \big(\xi \cdot \dot{\mathbf{x}} + \eta\dot{y} - H\big) ds.$$

Let us find ξ and $\mathbf{x}$ from the Hamilton system. We obtain

$$\dddot{x}_1 + 16\eta^2\dot{x}_1 = 0, \quad \dddot{x}_2 + 16\eta^2\dot{x}_2 = 0$$

from (1.4). Hence

$$\begin{aligned} \dot{x}_1(s) =& \dot{x}_1(0)\cos(4\eta s) + \frac{\ddot{x}_1(0)}{4\eta}\sin(4\eta s) \\ =& \dot{x}_1(0)\cos(4\eta s) + \dot{x}_2(0)\sin(4\eta s) \\ =& \zeta_1(0)\cos(4\eta s) + \zeta_2(0)\sin(4\eta s) \end{aligned} \tag{1.5}$$

and

$$\begin{aligned} \dot{x}_2(s) =& \dot{x}_2(0)\cos(4\eta s) + \frac{\ddot{x}_2(0)}{4\eta}\sin(4\eta s) \\ =& \dot{x}_2(0)\cos(4\eta s) - \dot{x}_1(0)\sin(4\eta s) \\ =& -\zeta_1(0)\sin(4\eta s) + \zeta_2(0)\cos(4\eta s), \end{aligned} \tag{1.6}$$

which yields

$$x_1(s) = x_1(0) + \zeta_1(0)\frac{\sin(4\eta s)}{4\eta} + \zeta_2(0)\frac{1-\cos(4\eta s)}{4\eta} \tag{1.7}$$

and

$$x_2(s) = x_2(0) - \zeta_1(0)\frac{1-\cos(4\eta s)}{4\eta} + \zeta_2(0)\frac{\sin(4\eta s)}{4\eta}. \tag{1.8}$$

At $s = t$ one has $x_1(t) = x_1$ and $x_2(t) = x_2$, so

$$\frac{1}{2}\zeta_1(0)\sin(4\eta t) + \frac{1}{2}\zeta_2(0)\big(1-\cos(4\eta t)\big) = 2\eta\big(x_1 - x_1(0)\big),$$
$$-\frac{1}{2}\zeta_1(0)\big(1-\cos(4\eta t)\big) + \frac{1}{2}\zeta_2(0)\sin(4\eta t) = 2\eta\big(x_2 - x_2(0)\big),$$

or,

$$\begin{aligned} +\zeta_1(0)\cos(2\eta t) + \zeta_2(0)\sin(2\eta t) &= \frac{2\eta\big(x_1 - x_1(0)\big)}{\sin(2\eta t)}, \\ -\zeta_1(0)\sin(2\eta t) + \zeta_2(0)\cos(2\eta t) &= \frac{2\eta\big(x_2 - x_2(0)\big)}{\sin(2\eta t)}. \end{aligned} \tag{1.9}$$

Hamilton's equations give

$$\begin{aligned} \xi_2(s) =& -2\eta x_1(s) + \big(\xi_2(0) + 2\eta x_1(0)\big) \\ =& -2\eta x_1(0) - \frac{1}{2}\zeta_1(0)\sin(4\eta s) - \frac{1}{2}\zeta_2(0)\big(1-\cos(4\eta s)\big) + \zeta_2(0) + 4\eta x_1(0) \\ =& 2\eta x_1(0) - \frac{1}{2}\Big[\zeta_1(0)\sin(4\eta s) - \zeta_2(0)\big(1+\cos(4\eta s)\big)\Big], \end{aligned}$$

and

$$\xi_1(s) = -2\eta x_2(0) + \frac{1}{2}\Big[\zeta_1(0)\big(1+\cos(4\eta s)\big) + \zeta_2(0)\sin(4\eta s)\Big].$$

The above calculations imply

$$\begin{aligned} \xi_1\dot{x}_1 + \xi_2\dot{x}_2 =& -2\eta\dot{x}_1(s)x_2(0) + 2\eta x_1(0)\dot{x}_2(s) + \frac{1}{2}\big(\zeta_1^2(0) + \zeta_2^2(0)\big)\big(1+\cos(4\eta s)\big) \\ =& -2\eta\big(\dot{x}_1(s)x_2(0) - x_1(0)\dot{x}_2(s)\big) + \big(1+\cos(4\eta s)\big)E, \end{aligned}$$

and

$$\int_0^t \big(\xi\cdot\dot{\mathbf{x}} + \eta\dot{y} - H\big)ds = \eta\Big[y - y(0) + 2\big(x_1(0)x_2 - x_1x_2(0)\big) + \frac{\sin(4\eta t)}{4\eta^2}E\Big].$$

To find E we square and add the two equations in (1.9),

$$E = \frac{1}{2}\zeta_1^2(0) + \frac{1}{2}\zeta_2^2(0) = 2\eta^2\frac{|\mathbf{x}-\mathbf{x}_0|^2}{\sin^2(2\eta t)}.$$

Hence,

$$\begin{aligned} S(t) =& \int_0^t \big(\xi\cdot\dot{\mathbf{x}} + \eta\dot{y} - H\big)ds \\ =& \eta\Big[y - y(0) + 2\big(x_1(0)x_2 - x_1x_2(0)\big) + |\mathbf{x}-\mathbf{x}_0|^2\cot(2\eta t)\Big]. \end{aligned}$$

We note that

$$\mathbf{x},\ y,\ t,\ \mathbf{x}_0 \quad \text{and} \quad \eta = \eta(0)$$

are free parameters while

$$y(0) = y(0; \mathbf{x}, \mathbf{x}_0, y, \eta; t)$$

is not. Therefore, we need to introduce one more free variable $h(0)$ such that $h(t) = h(0) + S(t)$ is a solution of the Hamilton–Jacobi equation (1.2).

It reduces to find $h(0)$. To find it we shall substitute S into (1.2). Straightforward computation shows that

$$\frac{\partial h}{\partial t} + H(\mathbf{x}, y, \xi(t), \eta(t)) = 0,$$

where

$$h(t) = \eta(0)y(0) + S(t), \quad i.e., \quad h(0) = \eta(0)y(0). \tag{1.10}$$

This yields

$$\frac{\partial h}{\partial t} + H\Big(\mathbf{x}, y, \nabla_{\mathbf{x}} h, \frac{\partial h}{\partial y}\Big) = 0.$$

We have the following theorem.

Theorem 1.1. *We have shown that*

$$\begin{aligned} h =& \eta(0)y(0) + \int_0^t \big(\xi \cdot \dot{\mathbf{x}} + \eta \dot{y} - H\big) ds \\ =& \eta y + 2\eta\big(x_1(0)x_2 - x_1 x_2(0)\big) + \eta|\mathbf{x} - \mathbf{x}_0|^2 \cot(2\eta t) \end{aligned} \tag{1.11}$$

is a "complete integral" of (1.2) *and* (1.3), *i.e., a solution of* (1.2) *and* (1.3) *which depends on* 3 *free parameters* $x_1(0)$, $x_2(0)$ *and* η.

Before we move further, let us consider a more general situation.

2. Generalized Hamilton–Jacobi equations

In this section we study the Hamilton–Jacobi equation which is crucial in the construction of the heat kernel associated with elliptic and sub-elliptic operators. We deduce a general form of the solution to the Hamilton–Jacobi equation and its generalized form. We consider an $(n+m)$-dimensional space $\mathbb{R}^n \times \mathbb{R}^m$. The coordinates are denoted $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$ with dual variables $(\xi_1, \ldots, \xi_n)$ and $(\eta_1, \ldots, \eta_m)$ respectively. The roman indices $i, j, k, \ldots$ will vary from 1 to n and the Greek indices $\alpha, \beta, \ldots$ will vary from 1 to m. As usual, the Hamiltonian function $H(\mathbf{x}, \mathbf{y}, \xi, \eta)$ is a homogeneous polynomial of degree 2 in the variables (ξ, η) and has smooth coefficients in $(\mathbf{x}, \mathbf{y})$.

We have the following nice generalization of a result from [11].

Theorem 2.1. *Set*

$$h(t; \mathbf{x}, \mathbf{y}, \xi, \eta) = \sum_{\alpha=1}^{m} \eta_\alpha(0) y_\alpha(0) + S(t; \mathbf{x}, \mathbf{y}, \xi, \eta), \tag{2.1}$$

where

$$x_j = x_j(s; \mathbf{x}, \mathbf{y}, \xi, \eta; t), \quad j = 1, \ldots, n; \qquad y_\alpha = y_\alpha(s; \mathbf{x}, \mathbf{y}, \xi, \eta; t), \quad \alpha = 1, \ldots, m$$

and

$$S(t;\mathbf{x},\mathbf{y},\xi,\eta)=\int_0^t\Big(\xi(u)\cdot\dot{\mathbf{x}}(u)+\eta(u)\cdot\dot{\mathbf{y}}(u)-H(\mathbf{x}(u),\mathbf{y}(u),\xi(u),\eta(u))\Big)du.$$

Then h satisfies the usual Hamilton–Jacobi equation:

$$\frac{\partial h}{\partial t}+H\Big(\mathbf{x},\mathbf{y},\nabla_{\mathbf{x}}h,\nabla_{\mathbf{y}}h\Big)=0.$$

Proof. In order to prove the theorem, we first calculate the partial derivatives of the function S with respect to all variables explicitly. For $j=1,\dots,n$,

$$\begin{aligned}
&\frac{\partial S}{\partial x_j}(t;\mathbf{x},\mathbf{y},\xi,\eta)\\
&=\int_0^t\Big[\sum_{k=1}^n\Big(\frac{\partial\xi_k}{\partial x_j}\frac{dx_j}{ds}+\xi_k\frac{d}{ds}\frac{\partial x_k(s;\mathbf{x},\mathbf{y},\xi,\eta;t)}{\partial x_j}\Big)\\
&\quad+\sum_{\alpha=1}^m\Big(\frac{\partial\eta_\alpha}{\partial x_j}\frac{dy_\alpha}{ds}+\eta_\alpha\frac{d}{ds}\frac{\partial y_\alpha(s;\mathbf{x},\cdots;t)}{\partial x_j}\Big)-\sum_{k=1}^n\frac{\partial H}{\partial\xi_k}\frac{\partial\xi_k}{\partial x_j}-\sum_{\alpha=1}^m\frac{\partial H}{\partial\eta_\alpha}\frac{\partial\eta_\alpha}{\partial x_j}\\
&\quad-\sum_{k=1}^n\frac{\partial H}{\partial x_k}\frac{\partial x_k(s;\mathbf{x},\mathbf{y},\xi,\eta;t)}{\partial x_j}-\sum_{\alpha=1}^m\frac{\partial H}{\partial y_\alpha}\frac{\partial y_\alpha(s;\mathbf{x},\mathbf{y},\xi,\eta;t)}{\partial x_j}\Big]ds\\
&=\int_0^t\frac{d}{ds}\Big(\sum_{k=1}^n\xi_k\frac{\partial x_k(s;\mathbf{x},\mathbf{y},\xi,\eta;t)}{\partial x_j}+\sum_{\alpha=1}^m\eta_\alpha\frac{\partial y_\alpha(s;\mathbf{x},\mathbf{y},\xi,\eta;t)}{\partial x_j}\Big)ds\\
&=\sum_{k=1}^n\xi_k(s)\frac{\partial x_k(s;\mathbf{x},\mathbf{y},\xi,\eta;t)}{\partial x_j}\Big|_{s=0}^{s=t}+\sum_{\alpha=1}^m\eta_\alpha(s)\frac{\partial y_\alpha(s;\mathbf{x},\mathbf{y},\xi,\eta;t)}{\partial x_j}\Big|_{s=0}^{s=t}.
\end{aligned}$$

It follows that

$$\frac{\partial S}{\partial x_j}(t;\mathbf{x},\mathbf{y},\xi,\eta)=\xi_j(t)-\sum_{\alpha=1}^m\eta_\alpha(0)\frac{\partial y_\alpha(0;\mathbf{x},\mathbf{y},\xi,\eta;t)}{\partial x_j}.$$

Similarly, for $\beta=1,\dots,m$,

$$\frac{\partial S}{\partial y_\beta}(t;\mathbf{x},\mathbf{y},\xi,\eta)=\eta_\beta(t)-\sum_{\alpha=1}^m\eta_\alpha(0)\frac{\partial y_\alpha(0;\mathbf{x},\mathbf{y},\xi,\eta;t)}{\partial y_\beta}.$$

Moreover,

$$\begin{aligned}
\frac{\partial S}{\partial t}(t;\cdots)&=\sum_{k=1}^n\xi_k(t;\cdots)\dot{x}_k(t;\cdots)\\
&\quad+\sum_{\alpha=1}^m\eta_\alpha(t;\cdots)\dot{y}_\alpha(t;\cdots)-H\big(\mathbf{x},\mathbf{y},\xi(t;\cdots),\eta(t;\cdots)\big)\\
&\quad+\sum_{k=1}^n\xi_k(s;\cdots)\frac{\partial x_k(s;\cdots)}{\partial t}\Big|_{s=0}^{s=t}+\sum_{\alpha=1}^m\eta_\alpha(s;\cdots)\frac{\partial y_\alpha(s;\cdots)}{\partial t}\Big|_{s=0}^{s=t}.
\end{aligned}$$

Differentiating $x_1 = x_1(t; \mathbf{x}, \mathbf{y}, \xi, \eta; t)$ yields

$$0 = \frac{d}{dt} x_1(t; \mathbf{x}, \mathbf{y}, \xi, \eta; t) = \dot{x}_1(t; \cdots) + \frac{\partial x_1(s; \mathbf{x}, \mathbf{y}, \xi, \eta; t)}{\partial t}\Big|_{s=t}.$$

On the other hand, one has

$$\xi_k(s; \cdots) \frac{\partial x_k(s; \cdots)}{\partial t}\Big|_{s=0}^{s=t} = -\xi_k(t; \cdots)\dot{x}_k(t; \cdots), \qquad k = 1, \ldots, n,$$

and

$$\eta_\alpha(s; \cdots) \frac{\partial y_\alpha(s; \cdots)}{\partial t}\Big|_{s=0}^{s=t} = -\eta_\alpha(t; \cdots)\dot{y}_\alpha(t; \cdots) - \eta_\alpha(0; \cdots)\frac{\partial y_\alpha(0; \cdots)}{\partial t}, \quad \alpha = 1, \ldots, n,$$

therefore,

$$\frac{\partial S}{\partial t} = -H(t; \cdots) - \sum_{\alpha=1}^{m} \eta_\alpha(0; \cdots)\frac{\partial y_\alpha(0; \cdots)}{\partial t}.$$

It follows that if we set as in the statement of the theorem

$$h(t; \mathbf{x}, \mathbf{y}, \xi, \eta) = \sum_{\alpha=1}^{m} \eta_\alpha(0) y_\alpha(0) + S(t; \mathbf{x}, \mathbf{y}, \xi, \eta),$$

then it satisfies

$$\frac{\partial h}{\partial x_k} = \xi_k(t; \mathbf{x}, \mathbf{y}, \xi, \eta; t), \qquad k = 1, \ldots, n$$
$$\frac{\partial h}{\partial y_\alpha} = \eta_\alpha(t; \mathbf{x}, \mathbf{y}, \xi, \eta; t), \qquad \alpha = 1, \ldots, m,$$

and

$$\frac{\partial h}{\partial t} + H\Big(\mathbf{x}, \mathbf{y}, \xi(t), \eta(t)\Big) = 0 \quad \Rightarrow \quad \frac{\partial h}{\partial t} + H\Big(\mathbf{x}, \mathbf{y}, \nabla_{\mathbf{x}} h, \nabla_{\mathbf{y}} h\Big) = 0.$$

This completes the proof of the theorem. □

We note that the derivation that (2.1) satisfies the Hamilton–Jacobi equation was complete general, not restriction to $H\Big(\mathbf{x}, \mathbf{y}, \nabla_{\mathbf{x}} h, \nabla_{\mathbf{y}} h\Big)$ being (1.3). In particular we did not assume that $\eta_\alpha(s)$ =constant for $\alpha = 1, \ldots, m$. The action integral S is not a solution of the Hamilton–Jacobi equation because some of our free parameters are dual variables $\eta_\alpha(0)$ instead of $y_\alpha(0)$. For the Heisenberg sub-Laplacian or the Grusin operator, $\eta(0) = \eta$ cannot be switched to $y(0)$. As we know, $\dot{y} = 2(\dot{x}_1 x_2 - x_1 \dot{x}_2)$. From (1.5)–(1.8), one has

$$\dot{y} = 2\Big[\dot{x}_1 x_2(0) - x_1(0)\dot{x}_2 + \frac{1}{2}\big(\zeta_1^2(0) + \zeta_2^2(0)\big)\frac{1 - \cos(4\eta s)}{2\eta}\Big],$$

and

$$y(s) = 2\big(x_1(s)x_2(0) - x_1(0)x_2(s)\big) + \frac{E}{4\eta^2}\big(4\eta s - \sin(4\eta s)\big) + C.$$

At $s = t$, one has $x_1(t) = x_1$, $x_2(t) = x_2$ and

$$y = 2\big(x_1 x_2(0) - x_1(0)x_2\big) + \frac{E}{4\eta^2}\big(4\eta t - \sin(4\eta t)\big) + C.$$

Hence, one has

$$\begin{aligned} y(s) = y - 2\Big[\big(x_1 - x_1(s)\big)x_2(0) - x_1(0)\big(x_2 - x_2(s)\big)\Big] \\ - \frac{E}{4\eta^2}\big[4\eta(t-s) - (\sin(4\eta t) - \sin(4\eta s))\big]. \end{aligned}$$

At $s = 0$,

$$y(0) = y + 2\big(x_1(0)x_2 - x_1 x_2(0)\big) + |\mathbf{x} - \mathbf{x}_0|^2 \mu(2\eta t),$$

where we set

$$\mu(\phi) = \frac{\phi}{\sin^2\phi} - \cot\phi.$$

To replace η by $y(0)$, one needs to invert μ,

$$\mu(2\eta t) = \frac{y - y(0) + 2\big(x_1(0)x_2 - x_1 x_2(0)\big)}{|\mathbf{x} - \mathbf{x}_0|^2}.$$

This is impossible since for most of the values on the right-hand side μ^{-1} is a many-valued function [2]. Therefore we must leave η as one of the free parameters which does not permit S to be a solution of the Hamilton–Jacobi equation.

Before we go further, we present a scaling property of the solution to the Hamiltonian system

$$\frac{dx_j}{ds} = \frac{\partial H}{\partial \xi_j}, \quad \frac{dy_\alpha}{ds} = \frac{\partial H}{\partial \eta_\alpha}, \quad \frac{d\xi_j}{ds} = -\frac{\partial H}{\partial x_j}, \quad \frac{d\eta_\alpha}{ds} = -\frac{\partial H}{\partial y_\alpha},$$

$s \in [0, t]$ with the boundary conditions

$$\mathbf{x}(0) = \mathbf{x}_0, \quad \mathbf{x}(t) = \mathbf{x}, \quad \mathbf{y}(t) = \mathbf{y}, \quad \eta(0) = \eta(0).$$

Lemma 2.1. *One has the following scaling property*

$$\begin{aligned} x_j(s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \eta(0); t) &= x_j\Big(\lambda s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\Big), \quad j = 1, \dots, n \\ y_\alpha(s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \eta(0); t) &= y_\alpha\Big(\lambda s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\Big), \quad \alpha = 1, \dots, m \\ \xi_j(s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \eta(0); t) &= \lambda\xi_j\Big(\lambda s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\Big), \quad j = 1, \dots, n \\ \eta_\alpha(s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \eta(0); t) &= \lambda\eta_\alpha\Big(\lambda s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\Big), \quad \alpha = 1, \dots, m \end{aligned} \tag{2.2}$$

for $\lambda > 0$, if the two sides of (2.2) *stays in the domain of unique solvability of the Hamiltonian system.*

Proof. Denote the curve on the right-hand side of (2.2) by $\{\tilde{\mathbf{x}}(s), \tilde{\mathbf{y}}(s), \tilde{\xi}(s), \tilde{\eta}(s)\}$. Note that $s \in (0, t)$. Then for $j = 1, \ldots, n$

$$\begin{aligned}\frac{\partial \tilde{x}_j}{\partial s} &= \lambda \dot{x}_j\Big(\lambda s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\Big) \\ &= \lambda \frac{\partial H}{\partial \xi_j}\Big(x_1\Big(\lambda s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\Big), x_2\Big(\lambda s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\Big), \ldots\Big) \\ &= \frac{\partial H}{\partial \xi_j}\big(\tilde{\mathbf{x}}(s), \tilde{\mathbf{y}}(s), \tilde{\xi}(s), \tilde{\eta}(s)\big),\end{aligned}$$

since $\frac{\partial H}{\partial \xi_j}$, $j = 1, \ldots, n$, are homogeneous of degree 1 in $\xi_1, \ldots, \xi_n$ and $\eta_1, \ldots, \eta_m$. Similar calculations and homogeneity of degree 2 of $\frac{\partial H}{\partial x_j}$ and $\frac{\partial H}{\partial y_\alpha}$ in $\xi_1, \ldots, \xi_n$ and $\eta_1, \ldots, \eta_m$ yield

$$\frac{\partial \tilde{y}_\alpha}{\partial s} = \frac{\partial H}{\partial \eta_\alpha}, \qquad \frac{\partial \tilde{\xi}_j}{\partial s} = -\frac{\partial H}{\partial x_j}, \qquad \frac{\partial \tilde{\eta}_\alpha}{\partial s} = -\frac{\partial H}{\partial y_\alpha}.$$

Clearly,

$$\begin{aligned}\tilde{x}_j(0) &= x_j\Big(0; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\Big) = x_j(0), \\ \tilde{x}_j(t) &= x_j\Big(\lambda t; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\Big) = x_j,\end{aligned}$$

for $j = 1, \ldots, n$ and

$$\begin{aligned}\tilde{y}_\alpha(t) &= y_\alpha\left(\lambda t; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\right) = y_\alpha, \\ \tilde{\eta}_\alpha(0) &= \lambda \eta_\alpha\left(0; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\right) = \lambda \frac{\eta_\alpha(0)}{\lambda} = \eta_\alpha(0)\end{aligned}$$

for $\alpha = 1, \ldots, m$. The bicharacteristic curves are unique, so the two sides of (2.2) agree. □

Corollary 2.2. *One has*

$$h(\mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \eta(0); t) = \lambda h\left(\mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\right).$$

Proof. In the case of the Heisenberg group, the corollary is a direct consequence of the explicit formula (1.11) and in this case, $\eta(0) = \eta$ is a constant. Here we would like to give a proof which applies in more general case. We know that for $j = 1, \ldots, m$,

$$\begin{aligned}\dot{x}_j(s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \eta(0); t) &= \frac{dx_j}{ds}(s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \eta(0); t) \\ &= \frac{dx_j}{ds}\left(\lambda s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\right) \\ &= \lambda \dot{x}_j\left(\lambda s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\right).\end{aligned}$$

Similar result holds for $\dot y_\alpha$ for $\alpha = 1, \dots, m$. Therefore,

$$\begin{aligned}
&\int_0^t \big[\xi(s)\cdot\dot{\mathbf{x}}(s) + \eta(s)\cdot\dot{\mathbf{y}}(s) - H(\mathbf{x}(s;\dots),\dots)\big]\,ds \\
&= \int_0^t \Big[\lambda\xi\left(\lambda s; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\right)\cdot\lambda\dot{\mathbf{x}}(\lambda s;\dots) + \sum_{\alpha=1}^m \lambda\eta_\alpha(\lambda s;\dots)\cdot\lambda\dot y_\alpha(\lambda s;\dots) \\
&\quad - \lambda^2 H(\mathbf{x}(\lambda s;\dots),\dots)\Big]\,ds \\
&= \frac{1}{\lambda}\int_0^t \Big[\lambda^2\sum_{k=1}^n \xi_k(\lambda s;\dots)\dot x_k(\lambda s;\dots) + \lambda^2\sum_{\alpha=1}^m \eta_\alpha(\lambda s;\dots)\dot y_\alpha(\lambda s;\dots) \\
&\quad - \lambda^2 H(\mathbf{x}(\lambda s;\dots),\dots)\Big]\,d(\lambda s) \\
&= \lambda\int_0^t \Big[\xi\Big(s'; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \frac{\eta(0)}{\lambda}, \lambda t\Big)\cdot\dot{\mathbf{x}}(s';\dots) + \eta(s';\dots)\cdot\dot{\mathbf{y}}(s';\dots) \\
&\quad - H(\mathbf{x}(s';\dots),\dots)\Big]\,ds' \\
&= \lambda S\Big(\mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}, \lambda t\Big).
\end{aligned}$$

Also,

$$\sum_{\alpha=1}^m \eta_\alpha(0) y_\alpha(0; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \eta(0); t) = \lambda\sum_{\alpha=1}^m \frac{\eta_\alpha(0)}{\lambda} y_\alpha\Big(0; \mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \frac{\eta(0)}{\lambda}; \lambda t\Big)$$

and the proof of the corollary is therefore complete. □

Set

$$f(\mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \eta(0)) = h(\mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \eta(0), t)\Big|_{t=1}.$$

Then

Theorem 2.2. *f is a solution of the generalized Hamilton–Jacobi equation*

$$\sum_{\alpha=1}^m \eta_\alpha(0)\frac{\partial f}{\partial\eta_\alpha(0)} + H\Big(\mathbf{x}, \mathbf{y}, \nabla_{\mathbf{x}} f, \nabla_{\mathbf{y}} f\Big) = f. \tag{2.3}$$

Proof. By homogeneity property of the function h, one has

$$h(\mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, \eta(0), t) = \frac{1}{t} h(\mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, t\eta(0), 1) = \frac{1}{t} f(\mathbf{x}, \mathbf{x}_0, \mathbf{y}, \xi, t\eta(0)),$$

so,

$$\frac{\partial h}{\partial t} = -\frac{1}{t^2} f + \frac{1}{t}\sum_{\alpha=1}^m \eta_\alpha(0)\frac{\partial f}{\partial\eta_\alpha(0)} \tag{2.4}$$

on one hand. On the other hand,

$$\frac{\partial h}{\partial t} = -H\big(\mathbf{x}, \mathbf{y}, \nabla_{\mathbf{x}} h, \nabla_{\mathbf{y}} h\big) \tag{2.5}$$

from Theorem 2.1. Since (2.4) agrees with (2.5) for all t so we may set $t = 1$ which yields the proposition. □

At the rest of the section we present some examples that reveal the geometrical nature of functions h and f.

2.3. Laplace operator

We start from the Laplace operator $\Delta = \frac{1}{2}\sum_{k=1}^{n}\frac{\partial^2}{\partial x_k^2}$ in $\mathbb{R}^n$. The Hamiltonian function $H(\xi)$ is

$$H(\xi) = \frac{1}{2}\sum_{k=1}^{n}\xi_k^2$$

and hence we need to deal with $F(\xi,\gamma) = H + \gamma = 0$. The Hamilton's system is

$$\dot{\mathbf{x}} = \xi, \qquad \dot{\xi} = 0, \qquad \dot{\gamma} = 0.$$

with initial-boundary conditions $\mathbf{x}(0) = \mathbf{x}_0$, $\mathbf{x}(t) = \mathbf{x}$. Since $\dot{\xi} = 0$, it follows that $\xi(s) = \xi(0) =$ constants, is a constant vector. Then

$$\ddot{\mathbf{x}} = \dot{\xi} = 0 \quad \Rightarrow \quad \mathbf{x}(s) = \xi(0)s + \mathbf{x}_0.$$

Moreover,

$$\mathbf{x} = \mathbf{x}(t) = \xi(0)t + \mathbf{x}_0 \quad \Rightarrow \quad \xi(0) = \frac{\mathbf{x} - \mathbf{x}_0}{t}$$

and

$$\frac{\partial h}{\partial t} = \frac{1}{2}\sum_{k=1}^{n}\xi_k^2 = \sum_{k=1}^{n}\frac{(x_k - x_k^{(0)})^2}{2t^2} = \frac{|\mathbf{x} - \mathbf{x}_0|^2}{2t^2}$$

or,

$$h(\mathbf{x},\mathbf{x}_0,t) = h(0) + \frac{|\mathbf{x} - \mathbf{x}_0|^2}{2t^2}t = h(0) + \frac{|\mathbf{x} - \mathbf{x}_0|^2}{2t}.$$

Since this is a translation invariant case, we may assume that $h(0) = 0$. Therefore,

$$f(\mathbf{x},\mathbf{x}_0) = h(\mathbf{x},\mathbf{x}_0,t)\Big|_{t=1} = \frac{|\mathbf{x} - \mathbf{x}_0|^2}{2}$$

gives us the Euclidean action function.

2.4. Grusin operator

We are in $\mathbb{R}^2$ now and the horizontal vector fields X_1, X_2 are given by

$$X_1 = \frac{\partial}{\partial x}, \quad \text{and} \quad X_2 = x\frac{\partial}{\partial y}.$$

The Grusin operator is given as follows: $\Delta_X = \frac{1}{2}\left(\frac{\partial}{\partial x}\right)^2 + \frac{1}{2}x^2\left(\frac{\partial}{\partial y}\right)^2$. It is obvious that Δ_X is elliptic away from the y-axis but degenerate on the y-axis. Since $[X_1, X_2] = \frac{\partial}{\partial y}$, hence $\{X_1, X_2, [X_1, X_2]\}$ spanned the tangent bundle of $\mathbb{R}^2$ everywhere. By Hörmander's theorem [12], Δ_X is hypoelliptic.

The Hamiltonian function H for the Δ_X is

$$H(x,y,\xi,\eta) = \frac{1}{2}\xi^2 + \frac{1}{2}x^2\eta^2. \tag{2.6}$$

The Hamilton system can be obtained as follows;

$$\begin{aligned}\dot{x} =& H_\xi = \xi,\\ \dot{y} =& H_\eta = \eta x^2,\\ \dot{\xi} =& - H_x = -\eta^2 x,\\ \dot{\eta} =& - H_y = 0,\\ \dot{S} =& \xi\dot{x} + \eta\dot{y} - H.\end{aligned}$$

With $0 \le s \le t$,

$$\eta(s) = \eta(0) = \eta_0 = \text{constant},$$

here "constant" means "*constant along the bicharacteristic curve*". Next,

$$\ddot{x} = \dot{\xi} = -x\eta^2, \qquad \text{so} \qquad \ddot{x} + \eta^2 x = 0.$$

It follows that

$$\begin{aligned}x(s) = A\cos(\eta s) + B\sin(\eta s) &= x(0)\cos(\eta s) + \frac{\xi(0)}{\eta}\sin(\eta s)\\ &= x_0\cos(\eta s) + \frac{\xi(0)}{\eta}\sin(\eta s).\end{aligned}$$

Hence,

$$\xi(s) = \dot{x}(s)$$

yields

$$\xi(s) = \xi(0)\cos(\eta s) - \eta x_0\sin(\eta s).$$

We also have

$$x = x(t) = x_0\cos(\eta t) + \frac{\xi(0)}{\eta}\sin(\eta t),$$

and

$$\frac{\xi(0)}{\eta} = \frac{x - x_0\cos(\eta t)}{\sin(\eta t)}. \tag{2.7}$$

Consequently,

$$x(s) = x(0)\cos(\eta s) + \frac{x - x_0\cos(\eta t)}{\sin(\eta t)}\sin(\eta s).$$

The singularities occur at $\eta = \eta_0 = \frac{k\pi}{t}$ when $x = \pm x_0$; they are $\eta = \frac{(2k+1)\pi}{t}$ if $x = x_0$ and $\eta_0 = \frac{2k\pi}{t}$ if $x = -x_0$. Next,

$$\begin{aligned}&\dot{y}(s) = \eta x^2(s)\\ &= \eta\Big[x_0\Big(\frac{1}{2} + \frac{1}{2}\cos(2\eta s)\Big) + 2x_0\frac{\xi(0)}{\eta}\sin(\eta s)\cos(\eta s) + \Big(\frac{\xi(0)}{\eta}\Big)^2\Big(\frac{1}{2} - \frac{1}{2}\cos(2\eta s)\Big)\Big]\\ &= \frac{d}{ds}\Big\{\eta\Big[\frac{x_0^2}{2}\Big(s + \frac{\sin(2\eta s)}{2\eta}\Big) + \frac{x_0\xi(0)}{\eta^2}\sin^2(\eta s) + \frac{1}{2}\Big(\frac{\xi(0)}{\eta}\Big)^2\big(s - \frac{\sin(2\eta s)}{2\eta}\big)\Big]\Big\}\\ &= \frac{d}{ds}\Big\{\frac{\eta}{2}\Big[x_0^2 + \big(\frac{\xi(0)}{\eta}\big)^2\Big]s + \frac{1}{4}\Big[x_0^2 - \big(\frac{\xi(0)}{\eta}\big)^2\Big]\sin(2\eta s) + \frac{x_0}{2}\frac{\xi(0)}{\eta}\big(1 - \cos(2\eta s)\big)\Big\}.\end{aligned}$$

We replace $\frac{\xi(0)}{\eta}$ by (2.7) and collect terms with x_0^2:

$$\begin{aligned}
&\frac{x_0^2}{2}\Big\{\eta s+\frac{1}{2}\sin(2\eta s)+\eta s\frac{\cos^2(\eta t)}{\sin^2(\eta t)}-\frac{1}{2}\frac{\cos^2(\eta t)}{\sin^2(\eta t)}\sin(2\eta t)-\frac{\cos(\eta t)}{\sin(\eta t)}\big(1-\cos(2\eta s)\big)\Big\}\\
&=\frac{x_0^2}{2}\Big\{\frac{\eta s}{\sin^2(\eta t)}-\frac{1}{2}\frac{\cos^2(\eta t)-\sin^2(\eta t)}{\sin^2(\eta t)}\sin(2\eta s)-\frac{\cos(\eta t)}{\sin(\eta t)}\big(1-\cos(2\eta s)\big)\Big\}\\
&=\frac{x_0^2}{2\sin^2(\eta t)}\Big\{\eta s-\frac{1}{2}\big[\cos(2\eta t)\sin(2\eta s)+\sin(2\eta t)\big(1-\cos(2\eta s)\big)\big]\Big\}\\
&=\frac{x_0^2}{4\sin^2(\eta t)}\Big\{2\eta s-\big[\sin(2\eta t)-\sin\big(2\eta(t-s)\big)\big]\Big\}.
\end{aligned}$$

The terms containing x^2 are:

$$\frac{1}{4}\frac{x^2}{\sin^2(\eta t)}\big(2\eta s-\sin(2\eta s)\big),$$

and the terms with x_0x are the following:

$$\frac{1}{2}\frac{2xx_0}{\sin^2(\eta t)}\Big\{\frac{1}{2}\big[\sin\big(\eta(2s-t)\big)+\sin(\eta t)\big]-\eta\, s\cos(\eta t)\Big\}.$$

So,

$$\begin{aligned}
\dot y(s)=\frac{d}{ds}\Big\{&\frac{x_0^2}{4\sin^2(\eta t)}\Big[2\eta s-\big[(\sin(2\eta t)-\sin\big(2\eta(t-s)\big)\big])\Big]\\
&+\frac{x^2}{4\sin^2(\eta t)}\big(2\eta s-\sin(2\eta s)\big)\\
&+\frac{2xx_0}{4\sin^2(\eta t)}\Big[\frac{1}{2}\big(\sin\big(\eta(2s-t)\big)+\sin(\eta t)\big)-\eta\, s\cos(\eta t)\Big]\Big\}.
\end{aligned}$$

The action function has the form

$$S=\int_0^t(\xi\dot x+\eta\dot y-H)ds=\eta(y-y(0))+\int_0^t(\xi^2-H)ds.$$

We find ξ^2 as follows

$$\begin{aligned}
\xi^2(s)&=\frac{\xi^2(0)}{2}\big(1+\cos(2\eta s)\big)-\xi(0)\eta x_0\sin(2\eta s)+\frac{1}{2}\eta^2x_0^2\big(1-\cos(2\eta s)\big)\\
&=\underbrace{\frac{1}{2}\big[\xi^2(0)+\eta^2x_0^2\big]}_{=H(0)}+\frac{1}{2}\big[\xi^2(0)-\eta^2x_0^2\big]\cos(2\eta s)-\eta x_0\xi(0)\sin(2\eta s).
\end{aligned}$$

Since H is constant along the bicharacteristic, one has

$$H=H(0)=\frac{1}{2}\big[\xi^2(0)+\eta^2x_0^2\big].$$

Continuing, we obtain the action function

$$
\begin{aligned}
S =& \eta(y - y(0)) + \int_0^t \Big[(\xi^2(0) - \eta^2 x_0^2)\frac{\cos(2\eta s)}{2} - \eta x_0 \xi(0)\sin(2\eta s)\Big] ds \\
=& \eta(y - y(0)) + \frac{1}{2}\big(\xi^2(0) - \eta^2 x_0^2\big)\frac{\sin(2\eta t)}{2\eta} + \eta x_0 \xi(0)\frac{\cos(2\eta t) - 1}{2\eta}.
\end{aligned}
$$

We simplify this

$$
\begin{aligned}
S - \eta(y - y(0)) =& \frac{\eta^2}{2}\Big(\frac{x - x_0\cos(\eta t)}{\sin(\eta t)}\Big)^2 \frac{\sin(2\eta t)}{2\eta} - \frac{1}{2}\eta^2 x_0^2 \frac{\sin(2\eta t)}{2\eta} \\
&+ \eta^2 x_0 \frac{x - x_0\cos(\eta t)}{\sin(\eta t)}\frac{\cos(2\eta t) - 1}{2\eta} \\
=& \frac{\eta}{4}\Big\{\Big(\frac{x - x_0\cos(\eta t)}{\sin(\eta t)}\Big)^2 \sin(2\eta t) - x_0^2 \sin(2\eta t) \\
&- 2x_0 \frac{x - x_0\cos(\eta t)}{\sin(\eta t)}\big(1 - \cos(2\eta t)\big)\Big\}.
\end{aligned} \tag{2.8}
$$

In the bracket $\{\cdots\}$ of (2.8), terms involved x_0^2 are

$$
\begin{aligned}
x_0^2 &\Big[\Big(\frac{\cos^2(\eta t)}{\sin^2(\eta t)} - 1\Big)\sin(2\eta t) + 2\frac{\cos(\eta t)}{\sin(\eta t)}(1 - \cos(2\eta t))\Big] \\
&= x_0^2\Big(\frac{\cos(2\eta t)\sin(2\eta t)}{\sin^2(\eta t)} + 2\frac{\cos(\eta t)}{\sin(\eta t)} - \frac{\cos(2\eta t)\sin(2\eta t)}{\sin^2(\eta t)}\Big) = 2x_0^2\cot(\eta t),
\end{aligned}
$$

terms involved x^2 are

$$
x^2 \frac{\sin(2\eta t)}{\sin^2(\eta t)} = 2x^2\cot(\eta t),
$$

and terms containing $x_0 x$ are

$$
\begin{aligned}
2xx_0\Big(-\frac{\cos(\eta t)}{\sin^2(\eta t)}\sin(2\eta t) - \frac{1 - \cos(2\eta t)}{\sin(\eta t)}\Big) &= -2xx_0\Big(\frac{2\cos^2(\eta t)}{\sin(\eta t)} + 2\sin(\eta t)\Big) \\
&= -\frac{4xx_0}{\sin(\eta t)}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\{\cdots\} =& 2(x^2 + x_0^2)\cot(\eta t) - \frac{4xx_0}{\sin(\eta t)} \\
=& \big[(x + x_0)^2 + (x - x_0)^2\big]\cot(\eta t) - \frac{(x + x_0)^2 - (x - x_0)^2}{\sin(\eta t)} \\
=& (x + x_0)^2\Big(\cot(\eta t) - \frac{1}{\sin(\eta t)}\Big) + (x - x_0)^2\Big(\cot(\eta t) + \frac{1}{\sin(\eta t)}\Big) \\
=& (x + x_0)^2\frac{\cos(\eta t) - 1}{\sin(\eta t)} + (x - x_0)^2\frac{\cos(\eta t) + 1}{\sin(\eta t)} \\
=& -(x + x_0)^2\tan\big(\frac{\eta t}{2}\big) + (x - x_0)^2\cot\big(\frac{\eta t}{2}\big).
\end{aligned}
$$

Thus S has the following form:

$$S = \eta(y - y(0)) - \frac{\eta}{4}\Big[(x + x_0)^2 \tan\Big(\frac{\eta t}{2}\Big) - (x - x_0)^2 \cot\Big(\frac{\eta t}{2}\Big)\Big].$$

By Theorem 2.1, we know that

$$\begin{aligned} h(t; x, x_0, y, \eta) &= \eta y(0) + S(t; x, y, \eta) \\ &= \eta y(0) + \eta(y - y(0)) - \frac{\eta}{4}\Big[(x + x_0)^2 \tan\Big(\frac{\eta}{2}\Big) - (x - x_0)^2 \cot\Big(\frac{\eta}{2}\Big)\Big] \\ &= \eta y - \frac{\eta}{4}\Big[A^2 \tan\big(\frac{\eta t}{2}\big) - B^2 \cot\Big(\frac{\eta t}{2}\Big)\Big] \end{aligned}$$

is a solution of the Hamilton–Jacobi equation. Here $A = x + x_0$ and $B = x - x_0$. Now by Theorem 2.2, the function

$$f(x, x_0, y, \eta) = h(t; x, x_0, y, \eta)\Big|_{t=1} = \frac{1}{2}\frac{\eta}{2}\Big\{4y - A^2 \tan\Big(\frac{\eta}{2}\Big) + B^2 \cot\Big(\frac{\eta}{2}\Big)\Big\}$$

is a solution of the generalized Hamilton–Jacobi equation

$$\eta\frac{\partial f}{\partial \eta} + H\Big(x, x_0, y, \partial_x f, \partial_y f\Big) = f.$$

We set

$$\frac{\eta}{2} = \widetilde{\eta}\tau,$$

where $\tau \in \mathbb{R}$ yields the domain of integration and $\widetilde{\eta}$ is a fixed complex number.

Lemma 2.5. *Suppose f is a smooth function of $\tau \in \mathbb{R}$ and*

$$\lim_{\tau \to \pm\infty} Re(f)(\tau) = \infty$$

off the canonical curve $x_0^2 + x^2 = 0$. Then $\widetilde{\eta}$ is pure imaginary.

Proof. Let $\widetilde{\eta} = \eta_1 + i\eta_2$. An elementary calculation yields

$$\begin{aligned} f = \frac{1}{2}\big(\eta_1 + i\eta_2\big)\tau\Big\{4y &+ \frac{\sin(2\eta_1\tau)\big[(B^2 - A^2)\cosh(2\eta_2\tau) + (B^2 + A^2)\cos(2\eta_1\tau)\big]}{\cosh^2(2\eta_2\tau) - \cos^2(2\eta_1\tau)} \\ &- i\frac{\sinh(2\eta_2\tau)\big[(B^2 + A^2)\cosh(2\eta_2\tau) + (B^2 - A^2)\cos(2\eta_1\tau)\big]}{\cosh^2(2\eta_2\tau) - \cos^2(2\eta_1\tau)}\Big\} \end{aligned}$$

(i) $\eta_1 = 0$, *i.e.*, $\eta \in i\mathbb{R}$. When $\tau \approx \pm\infty$,

$$f \approx \frac{1}{2}i\eta_2\tau\Big\{4y - i2(x_0^2 + x^2)\tanh(2\eta_2\tau)\Big\},$$

and

$$\mathrm{Re}(f) \approx \frac{1}{4}(x_0^2 + x^2)2\eta_2\tau\tanh(2\eta_2\tau) \;\to\; \pm\infty$$

as $\tau \to \pm\infty$ as long as $x_0^2 + x^2 \neq 0$.

(ii) $\eta_2 = 0$, that is $\eta \in \mathbb{R}$. Then

$$f = 2\eta_1\tau y + \frac{1}{4}\frac{2\eta_1\tau}{\sin(2\eta_1\tau)}\Big[B^2 - A^2 + (B^2 + A^2)\cos(2\eta_1\tau)\Big]$$

is singular in $\tau \in \mathbb{R}$ when $x_0^2 + x^2 \neq 0$, otherwise

$$\mathrm{Re}(f) = f = 2\eta_1 \tau y \underbrace{\longrightarrow}_{\tau \to \pm\infty} \pm(\mathrm{sgn}(y))\infty.$$

(iii) $\eta_1 \neq 0$, $\eta_2 \neq 0$. Here

$$f \approx \frac{1}{2}(\eta_1 + i\eta_2)\tau \Big\{ 4y - i(A^2 + B^2)\tanh(2\eta_2 \tau) \Big\}$$

as $\tau \to \pm\infty$, and

$$\begin{aligned} \mathrm{Re}(f) &\approx 2\eta_1 \tau y + (x_0^2 + x^2)\big|\eta_2 \tau\big| \\ &= |\tau|\big[2(\mathrm{sgn}(\tau))\eta_1 y + (x_0^2 + x^2)|\eta_2|\big] \end{aligned}$$

and choosing x_0, x, y so that

$$2\eta_1 y > (x_0^2 + x^2)|\eta_2|$$

we have

$$\lim_{\tau \to \pm\infty} \mathrm{Re}(f) = \pm\infty$$

which we do not want. This complete the proof of Lemma (2.5). □

Following the tradition, we shall choose

$$\widetilde{\eta} = -\frac{i}{2}.$$

Then

$$f = -i\tau y + \frac{1}{2}(x_0^2 + x^2)\tau \coth \tau - \frac{\tau x_0 x}{\sinh \tau}.$$

2.6. Sub-Laplace operator on step 2 nilpotent Lie groups

Let $\mathcal{M}$ be a simply connected 2-step nilpotent Lie group $\mathbb{G}$ equipped with a left invariant metric. Let $\mathcal{G}$ be its Lie algebra and it is identified with the group $\mathbb{G}$ by the exponential map:

$$\exp : \mathcal{G} \to \mathbb{G}.$$

We assume

$$\mathcal{G} = [\mathcal{G}, \mathcal{G}] \oplus [\mathcal{G}, \mathcal{G}]^{\perp} = \mathcal{C} \oplus [\mathcal{G}, \mathcal{G}]^{\perp} = \mathcal{C} \oplus \mathcal{H},$$

where $\mathcal{H}$ and $\mathcal{C}$ are vector spaces over $\mathbb{R}$ with an skew-symmetric bilinear form

$$B : \mathcal{H} \times \mathcal{H} \to \mathcal{C}$$

such that $B(\mathcal{H}, \mathcal{H}) = \mathcal{C}$. The group law is given by

$$(\mathcal{H} \oplus \mathcal{C}) \times (\mathcal{H} \oplus \mathcal{C}) \to \mathcal{H} \oplus \mathcal{C}$$

with

$$(\mathbf{x}, \mathbf{y}) * (\mathbf{x}', \mathbf{y}') = \big(\mathbf{x} + \mathbf{x}', \mathbf{y} + \mathbf{y}' + \frac{1}{2}B(\mathbf{x}, \mathbf{x}')\big)$$

and then the exponential map is the identity map. Let $\{X_1, \ldots, X_n\}$ be a basis of $\mathcal{H}$ and let $\{Y_1, \ldots, Y_m\}$ be a basis of the center $[\mathcal{G}, \mathcal{G}] = \mathcal{C}$. We assume $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_m\}$ are orthonormal, and introduce a left invariant Riemannian metric on the group $\mathbb{G}$ in an obvious way.

We write the vector fields X_j, $j = 1, \dots, n$ by:

$$X_j = \frac{\partial}{\partial x_j} + \sum_{k=1}^{n} \sum_{\alpha=1}^{m} a_{jk}^{\alpha} x_k \frac{\partial}{\partial y_\alpha}$$

where the a_{jk}^{α} are real numbers and form skew-symmetric matrices $\big[a_{jk}^{\alpha}\big]_{j,k}$, *i.e.*, $a_{jk}^{\alpha} = -a_{kj}^{\alpha}$. We are interested in the sub-Laplacian Δ_X which can be defined as follows:

$$\Delta_X = \frac{1}{2} \sum_{j=1}^{n} X_j^2$$

It is easy to see that

$$\big[X_j, X_k\big] = 2 \sum_{k=1}^{n} \sum_{\alpha=1}^{m} a_{jk}^{\alpha} \frac{\partial}{\partial y_\alpha}. \tag{2.9}$$

Lemma 2.7. *The operator Δ_X is hypoelliptic if and only if the rectangular matrix of order $\frac{n(n-1)}{2} \times m$ with element $\big[a_{jk}^{\alpha}\big]_{\{(j<k),\alpha\}}$ is of rank m (which implies that $m \leq \frac{n(n-1)}{2}$).*

Proof. The operator Δ_X is hypoelliptic when the vector fields $\{X_j\}_{j=1}^{n}$ satisfy the "first" bracket generating condition. This implies that we can recover all the $\frac{\partial}{\partial y_\alpha}$ from the $\frac{n(n-1)}{2}$ relations (2.9). If we consider $\big[a_{jk}^{\alpha}\big]$ as a matrix with indices $\alpha = 1, \dots, m$ and the couples (j, k) where $j < k$, this means that this matrix should have rank m. □

We may define a Lie group structure on $\mathbb{R}^n \times \mathbb{R}^m$ with the following group law:

$$(\mathbf{x}, \mathbf{y}) \circ (\mathbf{x}', \mathbf{y}') \tag{2.10}$$
$$= \Big(x_1 + x_1', \dots, x_n + x_n', y_1 + y_1' + \sum_{j,k=1}^{n} a_{jk}^{1} x_j' x_k, \dots, y_m + y_m' + \sum_{j,k=1}^{n} a_{jk}^{m} x_j' x_k\Big).$$

It is easy to see that the X_j are left invariant vector fields such that

$$\big(X_j f\big)(\mathbf{x}, \mathbf{y}) = \frac{\partial}{\partial x_j'} \big(f \circ \mathcal{L}_{(\mathbf{x},\mathbf{y})}\big)(\mathbf{x}', \mathbf{y}')\Big|_{\mathbf{x}'=0, \mathbf{y}'=0}$$

where

$$\mathcal{L}_{(\mathbf{x},\mathbf{y})}(\mathbf{x}', \mathbf{y}') = (\mathbf{x}, \mathbf{y}) \circ (\mathbf{x}', \mathbf{y}')$$

is the left translation by the element $(\mathbf{x}, \mathbf{y})$. In particular, Δ_X is a left invariant operator for this group structure (see [1] and [16]).

Let $\xi_1, \dots, \xi_n$ be the dual variables of $\mathbf{x}$ and $\eta_1, \dots, \eta_m$ be the dual variables of $\mathbf{y}$. We define the symbols ζ_j of the vector field X_j by

$$\zeta_j = \xi_j + \sum_{k=1}^{n} \sum_{\alpha=1}^{m} a_{jk}^{\alpha} x_k \eta_\alpha.$$

We shall try to find a solution of the following equation:

$$\frac{\partial h}{\partial t} + \frac{1}{2}\sum_{j=1}^{n}\Big(\frac{\partial h}{\partial x_j} + \sum_{k=1}^{n}\sum_{\alpha=1}^{m} a_{jk}^{\alpha} x_k \frac{\partial h}{\partial y_\alpha}\Big)^2 = 0.$$

Thus we start with

$$\frac{\partial z}{\partial t} + H(\nabla z) = 0, \tag{2.11}$$

where $H(\mathbf{x}, \mathbf{y}; \xi, \eta)$ is the Hamiltonian function as the full symbol of Δ_X,

$$H(\mathbf{x}, \mathbf{y}; \xi, \eta) = \frac{1}{2}\sum_{j=1}^{n}\Big(\xi_j + \sum_{k=1}^{n}\sum_{\alpha=1}^{m} a_{jk}^{\alpha} x_k \eta_\alpha\Big)^2 = \frac{1}{2}\sum_{j=1}^{n}\Big(\xi_j + \sum_{k=1}^{n}\mathcal{A}_{kj}(\eta)\cdot x_k\Big)^2. \tag{2.12}$$

Here

$$\mathcal{A}_{kj}(\eta) = \sum_{\alpha=1}^{m} a_{kj}^{\alpha}\eta_\alpha.$$

We shall find the bicharacteristic curves which are solutions to the corresponding Hamilton's system. The solutions define a one parameter family of symplectic isomorphism of the (punctures) cotangent bundle $T^*(\mathbb{R}^n \times \mathbb{R}^m)\setminus\{\mathbf{0}\}$. Since $\mathcal{A}^t(\eta) = -\mathcal{A}(\eta)$, the Hamilton's system can be written explicitly as follows:

$$\begin{aligned}
\dot{x}_j &= H_{\xi_j} = \xi_j - \sum_{k=1}^{n}\mathcal{A}_{jk}(\eta)\cdot x_k = \zeta_j, \quad \text{for } j = 1, \dots, n\\
\dot{y}_\alpha &= H_{\eta_\alpha} = \sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}^{\alpha} x_k \zeta_j, \quad \text{for } \alpha = 1, \dots, m\\
\dot{\xi}_j &= -H_{x_j} = -\sum_{k=1}^{n}\mathcal{A}_{jk}(\eta)\cdot\zeta_k = \sum_{k=1}^{n}\mathcal{A}_{kj}(\eta)\cdot\zeta_k, \quad \text{for } j = 1, \dots, n\\
\dot{\eta}_\alpha &= -H_{y_\alpha} = 0, \quad \text{for } \alpha = 1, \dots, m
\end{aligned} \tag{2.13}$$

with the initial-boundary conditions such that

$$\begin{cases} \mathbf{x}(0) = 0\\ \mathbf{x}(t) = \mathbf{x} = (x_1, \dots, x_n)\\ \mathbf{y}(t) = \mathbf{y} = (y_1, \dots, y_m)\\ \eta(0) = i\tau = i(\tau_1, \dots, \tau_m), \end{cases} \tag{2.14}$$

where $t \in \mathbb{R}$, $\mathbf{x}$ and $\mathbf{y}$ are arbitrarily given. With $0 \le s \le t$,

$$\eta_\alpha(s) = \eta_\alpha = \text{constant}, \quad \text{for } \alpha = 1, \dots, m.$$

Again, "constant" means "constant along the bicharacteristic curve". Also

$$H = \frac{1}{2}\sum_{j=1}^{n}\dot{x}_j^2 = \frac{1}{2}\sum_{j=1}^{n}\zeta_j^2 = E = \text{energy}.$$

Another way to see that E is constant along the bicharacteristic, note that

$$\begin{aligned}\ddot{x}_j = \dot{\zeta}_j = \dot{\xi}_j - \sum_{k=1}^{n} \mathcal{A}_{jk}(\eta)\cdot \dot{x}_k &= -\sum_{k=1}^{n} \mathcal{A}_{jk}(\eta)\cdot \zeta_k - \sum_{k=1}^{n} \mathcal{A}_{jk}(\eta)\cdot \zeta_k \\ &= -2\sum_{k=1}^{n} \mathcal{A}_{jk}(\eta)\cdot \zeta_k\end{aligned} \tag{2.15}$$

for $j = 1, \ldots, n$. Hence

$$\ddot{\mathbf{x}} = \dot{\zeta} = \dot{\xi} + \mathcal{A}(\eta)\dot{\mathbf{x}} = -2\mathcal{A}(\eta)\zeta. \tag{2.16}$$

Therefore,

$$\ddot{\mathbf{x}}\cdot\dot{\mathbf{x}} = -2\mathcal{A}(\eta)\zeta\cdot\zeta = 0$$

since $\mathcal{A}$ is skew-symmetric. It follows that

$$\frac{1}{2}\sum_{j=1}^{n} \dot{x}_j^2 = \frac{1}{2}\dot{\mathbf{x}}\cdot\dot{\mathbf{x}} = E = \text{energy}.$$

Since $\dot{\mathbf{x}}(s) = e^{-2s\mathcal{A}(\eta)}\xi(0)$, by integrating the equation

$$\mathcal{A}(\eta)\dot{\mathbf{x}}(s) = \mathcal{A}(\eta)e^{-2s\mathcal{A}(\eta)}\xi(0),$$

one has

$$\mathcal{A}(\eta)\mathbf{x}(s) = -\frac{1}{2}\Big(e^{-2s\mathcal{A}(\eta)} - I\Big)\xi(0)$$

where I is the $n\times n$ identity matrix. Since $\eta_\alpha = \eta(0) = i\tau_\alpha$ is pure imaginary, the matrix $i\mathcal{A}(\tau)$ is self-adjoint. It follows that the matrix

$$\frac{is\mathcal{A}(\tau)}{\sinh(it\mathcal{A}(\tau))} = \frac{1}{2\pi i}\int_\gamma \frac{\lambda}{\sinh(\lambda)}\Big(\lambda - it\mathcal{A}(\tau)\Big)^{-1} d\lambda$$

is well defined and invertible for any $t\in\mathbb{R}$ and $\tau\in\mathbb{R}^m$. Here γ is a suitable contour surrounding the spectrum of the matrix $it\mathcal{A}(\tau)$. The matrix

$$\frac{1}{2\pi i}\int_\gamma \frac{\lambda}{\sinh(\lambda)}\Big(\lambda - it\mathcal{A}(\tau)\Big)^{-1} d\lambda$$

has an inverse:

$$\frac{1}{2\pi i}\int_\gamma \frac{\sinh(\lambda)}{\lambda}\Big(\lambda - it\mathcal{A}(\tau)\Big) d\lambda.$$

We write it as

$$\frac{\sinh(i\mathcal{A}(\tau))}{i\mathcal{A}(\tau)} = \sum_{k=0}^{\infty} \frac{(i\mathcal{A}(\tau))^{2k}}{(2k+1)!}.$$

Then for any fixed $t\in\mathbb{R}$, we have one-to-one correspondence between the initial condition $\xi(0)$ and boundary condition $\mathbf{x}$:

$$\xi(0) = e^{it\mathcal{A}(\tau)}\cdot\frac{i\mathcal{A}(\tau)}{\sinh(it\mathcal{A}(\tau))}\cdot\mathbf{x}, \qquad t\neq 0.$$

Now we may solve the initial value problem:

$$\begin{cases} \dot{x}_j(s) = & \frac{\partial H}{\partial \xi_j} = \xi_j + i\sum_{k=1}^n \sum_{\alpha=1}^m a_{jk}^\alpha x_k \tau_\alpha = \xi_j + i\sum_{k=1} \mathcal{A}_{kj}(\tau) x_k, \\ \dot{\xi}_j(s) = & -\frac{\partial H}{\partial x_j} = -i\sum_{k=1}^n \Big(\xi_k + i\sum_{\ell=1}^n \mathcal{A}_{\ell k}(\tau) x_\ell\Big) \cdot \mathcal{A}_{jk}(\tau) \end{cases}$$

with the initial conditions

$$\begin{cases} \mathbf{x}(0) = 0 \\ \xi(0) = e^{it\mathcal{A}(\tau)} \cdot \frac{i\mathcal{A}(\tau)}{\sinh(it\mathcal{A}(\tau))}\mathbf{x}. \end{cases}$$

Straightforward computations show that

$$\begin{aligned} \mathbf{x}(s) =& \mathbf{x}(s;\mathbf{x},\tau,t) = e^{i(t-s)\mathcal{A}(\tau)} \frac{\sinh(is\mathcal{A}(\tau))}{\sinh(it\mathcal{A}(\tau))} \cdot \mathbf{x} \\ \xi(s) =& \xi(s;\mathbf{x},\tau,t) \\ =& \frac{i\mathcal{A}(\tau)}{\sinh(it\mathcal{A}(\tau))} \cdot e^{it\mathcal{A}(\tau)} \Big(I - e^{-is\mathcal{A}(\tau)} \sinh(is\mathcal{A}(\tau))\Big) \cdot \mathbf{x} \\ =& \Big(e^{-is\mathcal{A}(\tau)} \cosh(is\mathcal{A}(\tau))\Big) \cdot \Big(e^{it\mathcal{A}(\tau)} \frac{i\mathcal{A}(\tau)}{\sinh(it\mathcal{A}(\tau))}\Big)\mathbf{x} \\ =& \Big(e^{-is\mathcal{A}(\tau)} \cosh(is\mathcal{A}(\tau))\Big) \cdot \xi(0). \end{aligned}$$

Hence we obtain solutions for the initial-boundary problem (2.13) under the condition (2.14). We also have the following solutions for $\mathbf{y}(s)$:

$$y_\alpha(s) = y_\alpha(0) + \int_0^s \sum_{k=1}^n \Big(\big(e^{-2iu\mathcal{A}(\tau)}\xi(0)\big)_k \cdot \sum_{\ell=1}^n a_{\ell k}^\alpha x_\ell(u)\Big) du, \quad \alpha = 1,\ldots,m.$$

Again by Theorem 2.2, the function

$$f(\mathbf{x},\mathbf{y},\tau) = h(\mathbf{x},\mathbf{y},\tau,t)\Big|_{t=1}$$

is a solution of the generalized Hamilton–Jacobi equation. In our case, the function f can be calculated explicitly.

$$\begin{aligned} f(\mathbf{x},\mathbf{y},\tau) =& h(\mathbf{x},\mathbf{y},\tau,t)\Big|_{t=1} \\ =& \sum_{\alpha=1}^m \eta_\alpha(0) y_\alpha(0) + \int_0^1 \big(\xi \cdot \dot{\mathbf{x}} + \eta \cdot \dot{\mathbf{y}} - H\big) ds \\ =& \eta_0 \sum_{\alpha=1}^m \tau_\alpha y_\alpha + \int_0^1 \big(\xi \cdot \dot{\mathbf{x}} - H\big) ds. \end{aligned}$$

Here η_0 is a pure imaginary number. This choice can be motivated by Lemma 2.5.

Since

$$\xi \cdot \dot{\mathbf{x}} - H = \frac{1}{2}\langle \zeta, \zeta\rangle - \langle \zeta, \mathcal{A}\mathbf{x}\rangle,$$

then

$$\langle \zeta, \mathcal{A}\mathbf{x}\rangle = \Big\langle \zeta, \frac{\mathcal{A}(\tau)e^{2s\mathcal{A}(\tau)}}{e^{2\mathcal{A}(\tau)} - I}\mathbf{x}\Big\rangle - \Big\langle \zeta, \frac{\mathcal{A}(\tau)}{e^{2\mathcal{A}(\tau)} - I}\mathbf{x}\Big\rangle$$
$$= \frac{1}{2}\langle \zeta, \zeta\rangle - \Big\langle \frac{2\mathcal{A}(\tau)e^{2s\mathcal{A}(\tau)}}{e^{2\mathcal{A}(\tau)} - I}\mathbf{x}, \frac{\mathcal{A}(\tau)}{e^{2\mathcal{A}(\tau)} - I}\mathbf{x}\Big\rangle.$$

It follows that

$$\xi\cdot\dot{\mathbf{x}} - H = \Big\langle \frac{2\mathcal{A}(\tau)e^{2s\mathcal{A}(\tau)}}{e^{2\mathcal{A}(\tau)} - I}\mathbf{x}, \frac{\mathcal{A}(\tau)}{e^{2\mathcal{A}(\tau)} - I}\mathbf{x}\Big\rangle = \Big\langle \frac{2\mathcal{A}(\tau)\cosh(2s\mathcal{A}(\tau))}{e^{2\mathcal{A}(\tau)} - I}\mathbf{x}, \frac{\mathcal{A}(\tau)}{e^{2\mathcal{A}(\tau)} - I}\mathbf{x}\Big\rangle.$$

The second equality due to $\mathcal{A}$ is skew-symmetric. Now we can integrate from $s = 0$ to $s = 1$ to obtain

$$\int_0^1 \big(\xi\cdot\dot{\mathbf{x}} - H\big)ds = \frac{1}{2}\Big\langle \big(\mathcal{A}(\tau)\coth(\mathcal{A}(\tau))\big)\mathbf{x}, \mathbf{x}\Big\rangle.$$

It follows that

$$f(\mathbf{x}, \mathbf{y}, \tau) = -i\sum_{\alpha=1}^{m} \tau_\alpha y_\alpha + \frac{1}{2}\Big\langle \big(\mathcal{A}(\tau)\coth(\mathcal{A}(\tau))\big)\mathbf{x}, \mathbf{x}\Big\rangle. \tag{2.17}$$

Using equation (2.17), we may complete the discussion in Section 2.

Example 2.8. When

$$\mathcal{A} = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & a_{2n} \end{bmatrix} \in M_{2n\times 2n}, \quad \text{with } a_j = a_{j+n}, \quad j = 1, \ldots, n,$$

i.e., the group is an anisotropic Heisenberg group. In this case, $m = 1$ and

$$f(\mathbf{x}, y, \tau) = -i\tau y + \tau\sum_{k=1}^{n} a_k \coth(2a_k\tau)\big(x_k^2 + x_{n+k}^2\big).$$

Example 2.9. In $\mathbb{R}^4$, the basis of quaternion numbers $\mathbb{H} = \{a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} : a, b, c, d \in \mathbb{R}\}$ can be given by real matrices

$$M_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad M_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix},$$

$$M_2 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \qquad M_3 = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}.$$

We have

$$q = \begin{bmatrix} a & b & -d & -c \\ -b & a & -c & d \\ d & c & a & b \\ c & -d & -b & a \end{bmatrix} = aM_0 + bM_1 + cM_2 + dM_3.$$

The number a is called the *real part* and denoted by $a = \mathrm{Re}(q)$. The vector $\mathbf{u} = (b, c, d)$ is the *imaginary part* of q. We use the notations

$$b = \mathrm{Im}_1(q), \quad c = \mathrm{Im}_2(q), \quad d = \mathrm{Im}_3(q), \quad \text{and} \quad \mathrm{Im}(q) = \mathbf{u} = (b, c, d).$$

We introduce the quaternionic H-type group denoted by $\mathcal{Q}$. This group consists of the set

$$\mathbb{H} \times \mathbb{R}^3 = \{[\mathbf{x}, \mathbf{y}] : \ \mathbf{x} \in \mathbb{H}, \ \mathbf{y} = (y_1, y_2, y_3) \in \mathbb{R}^3\}$$

with the multiplication law defined in (2.10) with $[a^{\alpha}_{jk}] = M_\alpha$, $\alpha = 1, 2, 3$. The horizontal vector fields $X = (X_1, X_2, X_3, X_4)$ of the group $\mathcal{Q}$ can be written as follows:

$$X = \nabla_{\mathbf{x}} + \frac{1}{2}\Big(M_1\mathbf{x}\frac{\partial}{\partial y_1} + M_2\mathbf{x}\frac{\partial}{\partial y_2} + M_3\mathbf{x}\frac{\partial}{\partial y_3}\Big),$$

with $\mathbf{x} = (x_1, x_2, x_3, x_4)$ and

$$\nabla_{\mathbf{x}} = \Big(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}, \frac{\partial}{\partial x_4}\Big).$$

In this case, the solution for the generalized Hamilton–Jacobi equation is

$$f(\mathbf{x}, y_1, y_2, y_3, \tau_1, \tau_2, \tau_3) = -i\sum_{\alpha=1}^{3} \tau_\alpha y_\alpha + \frac{|\mathbf{x}|^2}{2}|\tau| \coth(2|\tau|)$$

See details in [6]. In general multidimensional case, the matrix $\mathcal{A}$ can be defined as follows:

$$\mathcal{A} = \begin{bmatrix} \sum_{\alpha=1}^{3} a^{\alpha}_1 M_\alpha & 0 & \dots & 0 \\ 0 & \sum_{\alpha=1}^{3} a^{\alpha}_2 M_\alpha & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \sum_{\alpha=1}^{3} a^{\alpha}_n M_\alpha \end{bmatrix}.$$

In this case we obtain the so-called anisotropic quaternion Carnot group considered in [7]. The complex action is given by

$$f(x, y, \tau) = -i\sum_{\alpha} \tau_\alpha y_\alpha + \frac{1}{2}\sum_{l=1}^{n} |x_l|^2 |\tau|_l \coth(2|\tau|_l),$$

where $|x_l|^2 = \sum_{j=0}^{3} x^2_{4l-j}$, $|\tau|_l = \big(\sum_{\alpha=1}^{3}(a^{\alpha}_l)^2\tau^2_\alpha\big)^{1/2}$. If all a^{α}_l, $l = 1, \dots, n$ are equal, we get the example of multidimensional quaternion H-type group. More information about H-type groups can be found in [5, 13, 14, 15].

3. Heat kernel and transport equation

Let us return to the heat kernel. We consider the sub-Laplacian

$$\Delta_X = \frac{1}{2}\sum_{k=1}^{n} X_k^2 \quad \text{with} \quad X_k = \frac{\partial}{\partial x_k} + \sum_{j=1}^{n}\sum_{\alpha=1}^{m} a_{kj}^{\alpha} x_j \frac{\partial}{\partial y_\alpha}.$$

Assume that $\{X_1, \ldots, X_n\}$ is an orthonormal basis of the "*horizontal subbundle*" on a simply connected nilpotent 2 step Lie group. The Hamiltonian of the operator Δ_X is

$$H(\mathbf{x}, \mathbf{y}, \xi, \eta) = \frac{1}{2}\sum_{k=1}^{n}\Big(\xi_k + \sum_{j=1}^{n}\sum_{\alpha=1}^{m} a_{kj}^{\alpha} x_j \eta_\alpha\Big)^2.$$

By Theorem 2.2, the function f associated with H is a solution of the generalized Hamilton–Jacobi equation:

$$H(\mathbf{x}, \mathbf{y}, \nabla_{\mathbf{x}} f, \nabla_{\mathbf{y}} f) + \sum_{\alpha=1}^{m} \tau_\alpha \frac{\partial f}{\partial \tau_\alpha} = f(\mathbf{x}, \mathbf{y}; \eta_1, \ldots, \eta_m).$$

As we know, the function f depends on free variables η_α, $\alpha = 1, \ldots, m$. To this end we shall sum over η_α, or for convenience $\tau_\alpha = t\eta_\alpha$, $\alpha = 1, \ldots, m$; an extra t can always be absorbed in the power q which can be determined after we solve the generalized Hamilton–Jacobi equation. Thus we write heat kernel of $\Delta_X - \frac{\partial}{\partial t}$ as following

$$K(\mathbf{x}, \mathbf{y}; t) = K_t(\mathbf{x}, \mathbf{y}) = \frac{1}{t^q}\int_{\mathbb{R}^m} e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}} V(\tau)d\tau. \tag{3.1}$$

Here V is the volume element. To see whether (3.1) is a representation of the heat kernel we apply the heat operator to K and take it across the integral.

$$\Big(\Delta_X - \frac{\partial}{\partial t}\Big)\frac{e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}}}{t^q} = \frac{e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}}}{t^{q+2}}\big(H(\mathbf{x}, \mathbf{y}, \nabla_{\mathbf{x}} f, \nabla_{\mathbf{y}} f) - f\big) - \frac{e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}}}{t^{q+1}}\big(\Delta_X(f) - q\big),$$

and the eiconal equation (2.3) implies that

$$\begin{aligned}
&\Big(\Delta_X - \frac{\partial}{\partial t}\Big)\frac{e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}} V(\tau)}{t^q} \\
&= \frac{e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}}}{t^{q+1}}\sum_{\alpha=1}^{m} \tau_\alpha\Big(-\frac{1}{t}\frac{\partial f}{\partial \tau_\alpha}\Big)V(\tau) - \frac{e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}}}{t^{q+1}}\big(\Delta_X f - q\big)V(\tau) \\
&= -\frac{e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}}}{t^{q+1}}\Big[\sum_{\alpha=1}^{m} \tau_\alpha \frac{\partial V}{\partial \tau_\alpha} \\
&\qquad + \big(\Delta_X f - q + m\big)V(\tau)\Big] + \sum_{\alpha=1}^{m}\frac{\partial}{\partial \tau_\alpha}\Big(\frac{e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}} \tau_\alpha V(\tau)}{t^{q+1}}\Big).
\end{aligned}$$

Assuming

$$\frac{e^{-\frac{f(u)}{t}} \tau_\alpha V(\tau)}{t^{q+1}} \to 0$$

as τ_α tends to the ends of an appropriate contour Γ_α for $\alpha = 1, \dots, m$, one has

$$\begin{aligned}
&\Big(\Delta_X - \frac{\partial}{\partial t}\Big) K_t(\mathbf{x}, \mathbf{y}) \\
&= \Big(\Delta_X - \frac{\partial}{\partial t}\Big)\Big\{\frac{1}{t^q}\int_{\cup_{\alpha=1}^m \Gamma_\alpha} e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}} V(\tau) d\tau\Big\} \\
&= -\frac{1}{t^{q+1}}\int_{\cup_{\alpha=1}^m \Gamma_\alpha} e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}} \Big[\sum_{\alpha=1}^m \tau_\alpha \frac{\partial V}{\partial \tau_\alpha} + \big(\Delta_X f - q + m\big) V(\tau)\Big] d\tau = 0
\end{aligned}$$

if $t \neq 0$ and

$$\sum_{\alpha=1}^m \tau_\alpha \frac{\partial V}{\partial \tau_\alpha}(\tau) + \big(\Delta_X f - q + m\big) V(\tau) = 0. \tag{3.2}$$

Equation (3.2) is called the first-order transport equation.

Remark 3.1. Here we have made a crucial assumption on the volume element, *i.e.,* V does not depend on the space variables $\mathbf{x}$ and $\mathbf{y}$. That simplify the transport equation significantly. Under a more general situation a function V will found among co-dimension one form

$$V d\tau = \sum_{\ell=1}^m (-1)^{\ell-1} V_\ell d\tau_1 \wedge \cdots \wedge \widehat{d\tau_\ell} \wedge \cdots \wedge d\tau_m$$

which satisfies a so-called "generalized transport equation":

$$df \wedge \Delta_X(V) + \sum_{\ell=1}^n X_\ell(f) X_\ell(dV) + \mathcal{D}(dV) - \big(\Delta_X f + n - m - 1\big) dV = 0,$$

where $\mathcal{D}(V)$ is defined by

$$\mathcal{D}(V) = \sum_{k=1}^m \tau_k \frac{\partial}{\partial \tau_k}(V) = \sum_{k=1}^m \sum_{\ell=1}^m (-1)^{\ell-1} \tau_k \frac{\partial V_\ell}{\partial \tau_k} d\tau_1 \wedge \cdots \wedge \widehat{d\tau_\ell} \wedge \cdots \wedge d\tau_m.$$

Detailed discussion can be found in Furutani [9] and Greiner [11].

With f given by (2.17), one has

$$\Delta_X f = \frac{1}{2}\mathrm{tr}\big(\mathcal{A}(\tau)\coth(\mathcal{A}(\tau))\big) = \frac{1}{2}\mathrm{tr}\Big(\frac{1}{2\pi i}\int_{\mathcal{C}} \lambda \frac{\cosh(\lambda)}{\sinh(\lambda)}\big(\lambda - i\mathcal{A}(\tau)\big)^{-1} d\lambda\Big). \tag{3.3}$$

Then (3.2) becomes

$$\begin{aligned}
&\sum_{\alpha=1}^m \tau_\alpha \frac{\partial V}{\partial \tau_\alpha}(\tau) + \big(\Delta_X f - q + m\big) V(\tau) = 0 \iff \\
&\sum_{\alpha=1}^m \tau_\alpha \frac{\partial V}{\partial \tau_\alpha}(\tau) = \Big(q - m - \frac{1}{2}\mathrm{tr}\big(\mathcal{A}(\tau)\coth(\mathcal{A}(\tau))\big)\Big) V.
\end{aligned} \tag{3.4}$$

Fix τ and define for $0 \le \lambda \le 1$

$$W(\lambda) = V(\lambda\tau).$$

Hence, (3.4) reduces to

$$\lambda\frac{dW}{d\lambda} = \Big[q - m - \frac{1}{2}\mathrm{tr}\big(\lambda\mathcal{A}(\tau)\coth(\lambda\mathcal{A}(\tau))\big)\Big]W.$$

Here we are using the fact that $\mathcal{A}(\tau)$ is linear in τ. It follows that

$$\frac{dW}{W} = \Big(\frac{q-m}{\lambda} - \frac{1}{2}\mathrm{tr}\big(\mathcal{A}(\tau)\coth(\lambda\mathcal{A}(\tau))\big)\Big)d\lambda.$$

Hence,

$$\log W = (q-m)\big(\log\lambda + \log C\big) - \frac{1}{2}\log(\sinh(\mathcal{A}(\lambda\tau))).$$

Therefore,

$$V(\tau) = \frac{(\det\mathcal{A}(\tau))^{q-m}}{\sqrt{\big(\det\sinh(\mathcal{A}(\tau))\big)}}.$$

If we propose the volume element V is real analytic and non-vanish at 0, then we have $q = \frac{n}{2} + m$.

Consequently,

$$P = \frac{A}{(2\pi t)^q}\int_{\mathbb{R}^m} e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}}V(\tau)d\tau, \tag{3.5}$$

where f is given by (2.17) and

$$V(\tau) = \frac{(\det\mathcal{A}(\tau))^{\frac{n}{2}}}{\sqrt{\big(\det\sinh(\mathcal{A}(\tau))\big)}}\ ,$$

where the branch is taken to be $V(0) = 1$. Finally we can write down the second main results on this paper.

Theorem 3.2. *The equation*

$$P_t(\mathbf{x},\mathbf{y}) = \frac{A}{(2\pi t)^q}\int_{\mathbb{R}^m} e^{-\frac{f(\mathbf{x},\mathbf{y},\tau)}{t}}V(\tau)d\tau,$$

represents the heat kernel for Δ_X if and only if $q = \frac{n}{2} + m$, in which case $A = 1$.

We clearly have

$$\frac{\partial P}{\partial t} - \Delta_X P = 0, \qquad t > 0$$

and

$$\lim_{t\to 0} P(\mathbf{x},\mathbf{y},t) = \delta(\mathbf{x})\delta(\mathbf{y}).$$

The calculation is long but straightforward. Readers can find the proof of this theorem in many places, see, *e.g.,* [1, 2, 3, 4, 8, 10]. We skip the proof here. Instead, we list some examples.

Example 3.1. The Heisenberg sub-Laplacian: $\Delta_X = \frac{1}{2}\Big(\frac{\partial}{\partial x_1} + 2x_2\frac{\partial}{\partial y}\Big)^2 + \frac{1}{2}\Big(\frac{\partial}{\partial x_2} - 2x_1\frac{\partial}{\partial y}\Big)^2$ which is defined as (1.1). The action function is $f(\mathbf{x}, y) = -i\tau y + (x_1^2 + x_2^2)\tau\coth(2\tau)$. The volume is $V(\tau) = \frac{2\tau}{\sinh(2\tau)}$. In this case $n = 2$, $m = 1$ and (3.5) has the following expression:

$$P_t(\mathbf{x}, y) = \frac{2}{(2\pi t)^2}\int_{-\infty}^{+\infty} e^{-\frac{f(\mathbf{x},y,\tau)}{t}}\frac{\tau}{\sinh(2\tau)}d\tau. \tag{3.6}$$

Example 3.2. The Grusin operator: $\Delta_G = \frac{1}{2}\Big(\frac{\partial}{\partial x}\Big)^2 + \frac{1}{2}x^2\Big(\frac{\partial}{\partial y}\Big)^2$. There is no group structure in this case. However, this operator has connection with the Heisenberg sub-Laplacian. Let $\mathbf{H}_1$ be the Heisenberg group whose Lie algebra has a basis $\{X_1, X_2, T\}$ with the bracket relation $[X_1, X_2] = -4T$. As in (1.1),

$$\Delta_X = -\frac{1}{2}\big(X_1^2 + X_2^2\big)$$

is the sub-Laplacian on $\mathbf{H}_1$. Let $\mathbf{N}_{X_2} = \langle X_2\rangle = [\{aX_2\}_{a\in\mathbb{R}}]$ be a subgroup generated by the element X_2. The map $\rho : \mathbf{H}_1 \to \mathbb{R}^2$ defined by

$$\begin{aligned}\rho : \mathbf{H}_1 \to \mathbb{R}^2 \cong \mathbf{h} \ni g =& x_1X_1 + x_2X_2 + zZ\\ =&(x_1, x_2, z) \mapsto (u, v) \in \mathbb{R}^2\end{aligned}$$

where

$$u = x_1, \qquad v = z + \frac{1}{2}x_1x_2$$

realizes the projection map

$$\mathbf{H}_1 \cong \mathbb{R}^3 \ \to\ \mathbf{N}_{X_2} \setminus \mathbf{H}_1 \cong \mathbb{R}^2.$$

In fact, this is a principal bundle and the trivialization is given by the map

$$\mathbf{N}_{X_2} \times (\mathbf{N}_{X_2} \setminus \mathbf{H}_1) \cong \mathbb{R} \times \mathbb{R}^2 \ni (a; u, v) \mapsto (x_1, x_2, z) \in \mathbb{R}^3 \cong \mathbf{H}_1$$

where

$$(a; u, v) \ \mapsto\ \big(u, a, v - \frac{1}{2}au\big).$$

So the sub-Laplacian Δ_X on $\mathbf{H}_1$ and Grusin operator Δ_G commutes each other through the map ρ:

$$\Delta_H \circ \rho^* = \rho^* \circ \Delta_G.$$

The heat kernel $P_t(\mathbf{x}, y) \in C^\infty(\mathbb{R}_+ \times \mathbf{H}_1)$ is given by (3.6). Hence,

$$\int_{-\infty}^{+\infty} P_t\Big((x_1, x_2, y), \Big(u, a, v - \frac{1}{2}ua\Big)\Big) = P_t^G\Big(\Big(x_1 + y + \frac{1}{2}x_1x_2\Big), (u, v)\Big)$$

that is, the fiber integration of the function $P_t(g, h)$ along the fiber of the map ρ gives the heat kernel of the Grusin operator.

$$P_t^G((x_0, 0), (x, y)) = \frac{1}{(2\pi t)^{\frac{3}{2}}}\int_{-\infty}^{+\infty} e^{-\frac{f(x,x_0,y,\tau)}{t}}\sqrt{\frac{|\tau|}{\sinh|\tau|}}d\tau$$

Example 3.3. Step 2 nilpotent Lie group: $\Delta_X = -\frac{1}{2}\sum_{j=1}^n X_j^2$ where

$$X_j = \frac{\partial}{\partial x_j} + \sum_{k=1}^n \Big(\sum_{\alpha=1}^m a_{jk}^\alpha x_k\Big)\frac{\partial}{\partial y_\alpha}$$

with $\mathcal{A}_{jk}^{(\alpha)} = \left[a_{jk}^\alpha\right]_{j,k}$ is a skew-symmetric and orthogonal matrix. The heat kernel is

$$P_t(\mathbf{x},\mathbf{y}) = \frac{1}{(2\pi t)^{\frac{n}{2}+m}}\int_{\mathbb{R}^m} e^{-\frac{2i\mathbf{y}\cdot\tau - \langle\mathcal{A}(\tau)\coth(\mathcal{A}(\tau))\mathbf{x},\mathbf{x}\rangle}{2t}}\sqrt{\det\frac{\mathcal{A}(\tau)}{\sinh(\mathcal{A}(\tau))}}d\tau.$$

Here $\mathcal{A}_{jk}(\tau) = \sum_{\alpha=1}^m a_{jk}^\alpha \tau_\alpha$. In particular, if $\mathcal{A}_{jk}^{(\alpha)}$ satisfies further assumption: $\mathcal{A}_{jk}^{(\alpha)}\mathcal{A}_{jk}^{(\gamma)} + \mathcal{A}_{jk}^{(\gamma)}\mathcal{A}_{jk}^{(\alpha)} = 0$, *i.e.,* the group in a H-type group. Then the heat kernel has the following form:

$$P_t(\mathbf{x},\mathbf{y}) = \frac{1}{(2\pi t)^{\frac{n}{2}+m}}\int_{\mathbb{R}^m} e^{-\frac{2i\mathbf{y}\cdot\tau + |\mathbf{x}|^2|\tau|\coth(|\tau|)}{2t}}\Big(\frac{|\tau|}{\sinh|\tau|}\Big)^{\frac{n}{2}}d\tau.$$

Acknowledgement

Part of this article is based on a lecture presented by the first author during the International Conference on “New Trends in Complex and Harmonic Analysis” which was held on May 7–11, 2007 at Voss, Norway. The first author thanks the organizing committee, especially Professor Alexander Vasiliev for his invitation. He would also like to thank all the colleagues at the Mathematics Department of University of Bergen for the warm hospitality during his visit to Norway. We also would like to thank Professor Peter Griener and Professor Kenro Furutani for many inspired conversations on this project.

References

[1] Beals R., Gaveau B., and Greiner P.C. *The Green function of model step two hypoelliptic operators and the analysis of certain tangential Cauchy-Rieman complexes.* Advances in Math. **121** (1996), 288–345.

[2] Beals R., Gaveau B., and Greiner P.C. *Hamilton–Jacobi theory and the heat kernel on Heisenberg groups.* J. Math. Pures Appl. **79** (2000), no. 7, 633–689.

[3] Calin O., Chang D.C., and Greiner P.C. *On a step* $2(k+1)$ *sub-Riemannian manifold.* J. Geom. Anal., **14** (2004), no. 1, 1–18

[4] Calin O., Chang D.C., and Greiner P.C. *Geometric Analysis on the Heisenberg Group and Its Generalizations*, to be published in AMS/IP series in advanced mathematics, International Press, Cambridge, Massachusetts, 2007. 244 pp.

[5] Cowling M., Dooley A.H., Korányi A., and Ricci F. *H-type groups and Iwasawa decompositions.* Adv. Math. **87** (1991), no. 1, 1–41.

[6] Chang D.C., Markina I. *Geometric analysis on quaternion H-type groups.* J. Geom. Anal. **16** (2006), no. 2, 266–294.

[7] Chang D.C., Markina I. *Anisotropic quaternion Carnot groups: geometric analysis and Green's function*, to appear in Advanced in Applied Math., (2007).

[8] Chang D.C., Markina I. *Quaternion H-type group and differential operator Δ_λ*, to appear in Science in China, Series A: Mathematics, (2007).

[9] Furutani K. *Heat kernels of the sub-Laplacian and the Laplacian on nilpotent Lie groups*, Analysis, Geometry and Topology of Elliptic Operators, World Scientific (2006), 185–226.

[10] Gaveau B. *Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents*, Acta Math. **139** (1977), no. 1–2, 95–153.

[11] Greiner P. *On Hörmander operators and non-holonomic geometry*, Fields Institute Communications, **52** (2008).

[12] Hörmander L. *Hypoelliptic second-order differential equations.* Acta Math. **119** (1967) 147–171.

[13] Kaplan A. *Fundamental solutions for a class of hypoelliptic PDE generated by composition of quadratics forms.* Trans. Amer. Math. Soc. **258** (1980), no. 1, 147–153.

[14] Kaplan A. *On the geometry of groups of Heisenberg type.* Bull. London Math. Soc. **15** (1983), no. 1, 35–42.

[15] Korányi A. *Geometric properties of Heisenberg-type groups.* Adv. in Math. **56** (1985), no. 1, 28–38.

[16] Nagel A., Ricci F., and Stein E.M. *Singular integrals with flag kernels and analysis on quadratic CR manifolds* Jour. Func. Anal. **181** (2001), 29–181.

Ovidiu Calin
Department of Mathematics
Eastern Michigan University
Ypsilanti, MI, 48197, USA
e-mail: `ocalin@emunix.emich.edu`

Der-Chen Chang
Department of Mathematics
Georgetown University
Washington D.C. 20057, USA

and

National Center for Theoretical Sciences
Mathematics Division
National Tsing Hua University
Hsinchu, Taiwan 30013, ROC
e-mail: `chang@georgetown.edu`

Irina Markina
Department of Mathematics
University of Bergen
Johannes Brunsgate 12
Bergen 5008, Norway
e-mail: `irina.markina@uib.no`

Analysis and Mathematical Physics
Trends in Mathematics, 77–106
© 2009 Birkhäuser Verlag Basel/Switzerland

Bergen Lecture on $\bar{\partial}$-Neumann Problem

Der-Chen Chang

Abstract. Let $\Omega \subset\subset \mathbb{C}^{n+1}$ be a bounded, pseudoconvex domain of finite type with smooth boundary. We assume further that the Levi form of $\partial\Omega$ is diagonalizable. In this article, we give detailed discussion of recent progress of the $\bar{\partial}$-Neumann problem. Using this result, we obtain solving operator for inhomogeneous Cauchy-Riemann equation $\bar{\partial}U = f$ in Ω. Here $f = \sum_{j=1}^{n+1} f_j\bar{\omega}$ is a given $(0,1)$-form. Then we discuss the "possible" optimal estimates of the solution.

Mathematics Subject Classification (2000). 32A20.

Keywords. Cauchy-Riemann equation, $\bar{\partial}$-Neumann problem, Heisenberg group, Carnot-Carathéodory distance, Calderón operator, Cauchy-Szegö projection, pseudo-differential operators, singular integral operators.

1. Introduction

This article is based on a series of lectures presented by the author at the Matematisk Institutt, Universitetet i Bergen, May 2007. The purpose of this article is to give an exposition of some recent progress in the $\bar{\partial}$-Neumann problem.

Given a bounded domain $\Omega \subset\subset \mathbb{C}^{n+1}$ with smooth boundary $\partial\Omega$, *i.e.,* there exists a real-valued function $\rho \in C^\infty(\bar{\Omega})$ such that

$$\partial\Omega = \{z \in \mathbb{C}^{n+1} : \ \rho(z) = 0\}$$

with $d\rho(z) \neq 0$, $\forall\, z \in \partial\Omega$. One of the basic problems in several complex variables is to solve the inhomogeneous Cauchy-Riemann equation

$$\bar{\partial}U = f \qquad \text{in} \quad \Omega \tag{1.1}$$

with "good" bounds on Ω, where f is a given $(0,1)$-form $f = \sum_{j=1}^{n+1} f_j\bar{\omega}_j$. Obvious, the right-hand side of (1.1) has $n+1$ data but the left-hand side of (1.1) has only one function. Therefore, the equation (1.1) is over-determined. It follows that

The author is partially supported by European Science Foundation Networking Program HCAA, a Hong Kong RGC competitive earmarked research grant #600607 and a competitive research grant at Georgetown University.

the equation (1.1) is solvable only when f satisfies a consistence condition, *i.e.,* $\bar\partial f = 0$. Moreover, solution for the equation (1.1) is highly non-unique. Suppose U is a solution of (1.1), then $U+F$ is also a solution whenever $F \in \mathcal{H}(\Omega)$ where $\mathcal{H}(\Omega)$ is the set of all holomorphic functions defined on Ω. Denote $\mathcal{A}^2(\Omega) = L^2(\Omega)\cap\mathcal{H}(\Omega)$ the Bergman space. Then we can find a "canonical solution" which satisfies an extra condition:

$$U \perp \mathcal{H}(\Omega) \quad \text{in} \quad \mathcal{A}^2(\Omega).$$

It is also called the Kohn solution of (1.1) which minimizes the L^2-norm among all solutions. In order to find the Kohn solution, let us consider a first-order differential operator D and μ, $\nu \in C^\infty(\bar\Omega)$, then the formal adjoint D^* of D can be defined as follows

$$\int_\Omega (D\mu)\bar\nu dV = \int_\Omega \mu\overline{(D^*\nu)}dV + \int_{\partial\Omega} \mu\overline{(A^\sharp\nu)}d\sigma,$$

where $A^\sharp$ is a 0th-order operator defined on $\partial\Omega$. In our case $D = \bar\partial$ is the Cauchy-Riemann operator. Hence,

$$\mathrm{dom}(\bar\partial^*) = \big\{\nu \in \mathcal{C}^\infty(\bar\Omega) : \, A^\sharp\nu = 0 \text{ on } \partial\Omega\big\}.$$

Note that with $U = \bar\partial^* u$, then

$$\langle\bar\partial^* u, F\rangle = \langle u, \bar\partial F\rangle = 0, \qquad \text{for all} \quad F \in \mathcal{H}(\Omega).$$

This means that if we solve the equation

$$\bar\partial\bar\partial^* u = f, \quad u \in \mathrm{dom}(\bar\partial^*), \tag{1.2}$$

then we solve (1.1) with a canonical solution.

In fact, problem (1.2) is equivalent to the case $\bar\partial f = 0$ of the system

$$\begin{aligned} &\Box u = \big(\bar\partial\bar\partial^* + \bar\partial^*\bar\partial\big)u = f, \\ &u \in \mathrm{dom}(\bar\partial^*), \;\; \bar\partial u \in \mathrm{dom}(\bar\partial^*). \end{aligned} \tag{1.3}$$

To see that,

$$0 = \bar\partial f = \bar\partial\big(\bar\partial\bar\partial^* + \bar\partial^*\bar\partial\big)u = \bar\partial\bar\partial^*\bar\partial u,$$

and so

$$0 = \langle\bar\partial u, \bar\partial\bar\partial^*\bar\partial u\rangle = \langle\bar\partial^*\bar\partial u, \bar\partial^*\bar\partial u\rangle \;\Rightarrow\; \bar\partial^*\bar\partial u = 0.$$

For general $u \in B^{(0,q)}(\Omega)$, the system (1.3) is called that "$\bar\partial$-Neumann problem". Here $B^{(0,q)}(\Omega)$ is the collection of all $(0,q)$ forms defined on Ω. The formalism of the $\bar\partial$-Neumann problem was introducing by D.C. Spencer in the early 1950's. The first result was obtained by Kohn (see [**25**], [**26**]):

$$\|u\|_{L^2_{k+1}(\Omega)} \le C\big(\|f\|_{L^2_k(\Omega)} + \|u\|_{L^2(\Omega)}\big), \quad k \in \mathbb{Z}_+.$$

This estimate is sharp in L^2. Unlike the elliptic case, the solution u does not gain two in all directions. Therefore, the system (1.3) has great interests from the point of view of partial differential equations.

There are essentially three aspects to this problem:

1. Existence of solutions;
2. Find the solving operator $\mathbf{N}$ (and hence $\bar{\partial}^*\mathbf{N}$) for the system (1.3) (and hence the system (1.2));
3. Sharp estimates for $\mathbf{N}$ and $\bar{\partial}^*\mathbf{N}$.

It is well known that although the given $(0,1)$-form f satisfies the consist condition, the equation $\bar{\partial}u = f$ is not necessary solvable. The solvability of inhomogeneous Cauchy-Riemann equation heavily rely on the geometry of Ω. This is a significant difference between analysis in one and several complex variables. Before we go further, let us recall some basic tools in several complex variables (see *e.g.,* Folland-Kohn [**20**] and Fefferman-Kohn [**19**]).

Definition 1.1. Let $\mathcal{M}$ be a $(2n+1)$-dimensional manifold. Then a CR structure on $\mathcal{M}$ is given by a subbundle $T^{1,0}(\mathcal{M})$ of the complex tangent bundle $\mathbb{C}T(\mathcal{M})$ satisfying the following properties:

(1) $T^{1,0}(\mathcal{M}) \cap \overline{T^{1,0}(\mathcal{M})} = \{0\}$;
(2) The fiber dimension of $T^{1,0}(\mathcal{M})$ is n;
(3) If Z and Z' are local vector fields with values in $T^{1,0}(\mathcal{M})$, then the commutator

$$[Z, Z'] = ZZ' - Z'Z$$

also has values in $T^{1,0}(\mathcal{M})$.

In general, a manifold $\mathcal{M}$ with a fixed CR structure is called a CR manifold.

Definition 1.2. Let $Z_1, \dots, Z_n$ be C^∞ vector fields on an open set $U \subset \mathcal{M}$ which are a local basis of sections of $T^{1,0}(\mathcal{M})$ on U. Let T be a local vector field on U such that $\{Z_j, \bar{Z}_j, T\}_{j=1,\dots,n}$ forms a basis of the complex tangent bundle. The vector fields $[Z_j, \bar{Z}_k]$ in terms of this basis is given by

$$[Z_j, \bar{Z}_k] = c_{jk}\sqrt{-1}T + \sum_{\ell=1}^{n} a_{jk}^{\ell} Z_\ell + \sum_{\ell=1}^{n} b_{jk}^{\ell} \bar{Z}_\ell.$$

The Hermitian form (c_{jk}) is called the Levi form. $\mathcal{M}$ is called pseudoconvex if each point of $\mathcal{M}$ has a neighborhood on which the vector field T can be chosen so that $(c_{jk}) \geq 0$. The Levi form is said to be diagonalizable on U if the local basis $Z_1, \dots, Z_n$ can be chosen so that

$$c_{jk} = \delta_{jk}\lambda_j \quad \text{on} \quad U.$$

Remark 1.1.

1. $\mathcal{M}$ is strongly pseudoconvex if the Levi form (c_{jk}) is positive definite.
2. If $\mathcal{M}$ is a hypersurface in $\mathbb{C}^{n+1}$ then it has the CR structure induces by $\mathbb{C}^{n+1}$ where

$$T^{1,0}(\mathcal{M}) = T^{1,0}(\mathbb{C}^{n+1}) \cap \mathbb{C}T(\mathcal{M});$$

that is, the fiber $T_p^{1,0}(\mathcal{M})$ consists of vectors of the form

$$\sum_{j=1}^{n+1} a_j \frac{\partial}{\partial z_j}$$

which are tangent to $\mathcal{M}$ at p.

3. When the Levi form (c_{jk}) of $\mathcal{M}$ is diagonalizable, "finite type" means the horizontal subbundle of the tangent bundle $T(\mathcal{M})$ satisfies the Chow's condition [**16**], *i.e.,* the vector fields $\{\mathrm{Re}(Z_1), \mathrm{Im}(Z_1), \ldots, \mathrm{Re}(Z_n), \mathrm{Im}(Z_n)\}$ together with their brackets generate the tangent bundle of $\mathcal{M}$.

Let g be a smooth Hermitian metric on $\mathbb{C}^{n+1}$. Then there is an open neighborhood U of $\partial\Omega$ such that if ρ denotes a signed geodesic distance in the metric g to $\partial\Omega$, then

$$\Omega^+ = \Omega \cap U = \{z \in U : \rho(z) > 0\};$$
$$\bigtriangledown\rho(z) \neq 0 \quad \text{for all} \quad z \in U.$$

We choose a smooth orthogonal basis for $(0,1)$-form on U, given by $\bar{\omega}_1, \ldots, \bar{\omega}_{n+1}$ with

$$\bar{\omega}_{n+1} = \sqrt{2}\bar{\partial}\rho.$$

We let $\bar{Z}_1, \ldots, \bar{Z}_{n+1}$ be the dual basis of antiholomorphic vector fields on U. Then

$$Z_1, \ldots, Z_n, \bar{Z}_1, \ldots, \bar{Z}_n$$

are tangential to $\partial\Omega$. If $\frac{\partial}{\partial\rho}$ is the vector field dual to the one form $d\rho$, then

$$Z_{n+1} = \frac{1}{\sqrt{2}}\frac{\partial}{\partial\rho} + iT$$

is the complex normal.

Because the vector fields split into "tangential" and "normal" part, we may consider a $(0,1)$-form u as follows:

$$u = \sum_{j=1}^{n+1} u_j \bar{\omega}_j = \sum_{j=1}^{n} u_j \bar{\omega}_j + u_{n+1}\bar{\omega}_{n+1}.$$

Then the $\bar{\partial}$-Neumann problem is the following boundary value problem:

$$\begin{aligned} \Box u = f \qquad & \text{in} \qquad \Omega \\ u_{n+1} = 0 \qquad & \text{on} \qquad \partial\Omega \\ \bar{Z}_{n+1}(u_j) - [S(u)]_{j,n+1} = 0 \qquad & \text{on} \qquad \partial\Omega \end{aligned}$$

for $j = 1, \ldots, n$. Here

$$[S(u)]_{j,n+1} = \sum_{\ell=1}^{n} \bar{s}^{\ell}_{j,n+1} u_j, \qquad j = 1, \ldots, n$$

and the matrix S is defined by the equations

$$\bar\partial\bar\omega_\ell = \sum_{j<k} \bar s^\ell_{jk}\bar\omega_j \wedge \bar\omega_k.$$

Then

$$\Box u = \begin{bmatrix} \Box_1 & 0 & \cdots & 0 & 0 \\ 0 & \Box_2 & \cdots & 0 & 0 \\ \cdots & & & & \\ 0 & 0 & \cdots & \Box_n & 0 \\ 0 & 0 & \cdots & 0 & \Box_{n+1} \end{bmatrix} u$$
$$+ (h_{n+1}\mathbf{I}_{n+1} + S^t)(\bar Z_{n+1}u) - (\bar S(Z_{n+1}u)) + \varepsilon(Z,\bar Z)u + \varepsilon(u)$$

where h_{n+1} is a smooth function which comes from the volume element. $\varepsilon(Z,\bar Z)u$ represents terms of first derivatives of u along horizontal directions and $\varepsilon(u)$ represents terms of u multiplying by smooth functions. Here

$$\Box_\ell = -\frac{1}{2}\sum_{j=1}^n (Z_j\bar Z_j + \bar Z_j Z_j) - Z_{n+1}\bar Z_{n+1} + \Big(\sum_{j=1}^n \lambda_j - 2\lambda_\ell\Big) iT$$

for $\ell = 1, 2, \ldots, n$ and

$$\Box_{n+1} = -\frac{1}{2}\sum_{j=1}^n (Z_j\bar Z_j + \bar Z_j Z_j) - Z_{n+1}\bar Z_{n+1} + \Big(\sum_{j=1}^n \lambda_j\Big) iT$$

We shall construct a solving operator for the $\bar\partial$-Neumann problem according to the geometry of $\partial\Omega$, *i.e.*, construct $\mathbf{N} = (\mathbf{N}_1, \ldots, \mathbf{N}_{n+1})$ such that modulo smooth error $u_j = \mathbf{N}_j(f_j)$ for $j = 1, \ldots, n+1$.

2. $n \geq 2$ and Ω is strongly pseudoconvex

When Ω is strongly pseudoconvex, the Levi form is positive definite. In this case, Ω has a foliation by a result of Chern and Moser [**15**]. Locally, Ω can be approximated by the "upper half-space" in $\mathbb{C}^{n+1}$, holomorphically equivalent to the unit ball. Readers can consult Stein's book [**35**] for background of Heisenberg group and its connection with analysis in several complex variables. The domain consists of all $z \in \mathbb{C}^{n+1}$, $n \geq 1$, so that

$$\Omega = \Big\{(z_1, \ldots, z_{n+1}) : \ \operatorname{Im}(z_{n+1}) > \sum_{j=1}^n |z_j|^2\Big\}.$$

Its boundary is the "paraboloid"

$$\partial\Omega = \Big\{ \operatorname{Im}(z_{n+1}) = \sum_{j=1}^n |z_j|^2\Big\}. \tag{2.1}$$

The domain Ω is bi-holomorphic to the unit ball in $\mathbb{C}^{n+1}$:

$$\mathbb{B} = \Big\{(w_1, \ldots, w_{n+1}) : \ \sum_{j=1}^{n+1} |w_j|^2 < 1\Big\}$$

via the generalized Cayley transform

$$w_{n+1} = \frac{i - z_{n+1}}{i + z_{n+1}}, \ w_k = \frac{2iz_k}{i + z_{n+1}}, \ k = 1, \ldots, n.$$

In this correspondence the boundary (2.1), together with the "point at ∞" maps onto the unit sphere in $\mathbb{C}^{n+1}$. The Heisenberg group $\mathbb{H}_n$ gives the translation of the domain Ω:

$$\mathbb{H}_n = \mathbb{C}^n \times \mathbb{R} = \{(\zeta, \xi) : \ \zeta \in \mathbb{C}^n, \ \xi \in \mathbb{R}\}.$$

The action of an element (ζ, ξ) on Ω is given by

$$(z', z_{n+1}) \ \rightarrow \ \big(z' + \zeta, z_{n+1} + \xi + 2iz' \cdot \bar{\zeta} + i|\zeta|^2\big)$$

where $z' = (z_1, \ldots, z_n)$.

Multiplication on $\mathbb{H}_n$ is given by

$$(\zeta_1, \xi_1) \cdot (\zeta_2, \xi_2) = \big(\zeta_1 + \zeta_2, \xi_1 + \xi_2 + 2\,\mathrm{Im}(\zeta_1 \cdot \bar{\zeta}_2)\big).$$

The group $\mathbb{H}_n$ acts on $\partial\Omega$ in a simply transitive manner. The mapping of $\mathbb{H}_n$ to $\mathbb{C}^{n+1}$

$$(\zeta, \xi) \ \rightarrow \ \big(\zeta, \xi + i|\zeta|^2\big) \tag{2.2}$$

identifies the group $\mathbb{H}_n$ with $\partial\Omega$. In this case, the vector fields Z_j can be written as

$$Z_j = \frac{\partial}{\partial z_j} + i\bar{z}_j \frac{\partial}{\partial t}, \qquad j = 1, \ldots, n$$

and the complex normal Z_{n+1} is

$$Z_{n+1} = \frac{1}{\sqrt{2}}\Big(\frac{\partial}{\partial \rho} + iT\Big) = \frac{1}{\sqrt{2}}\Big(\frac{\partial}{\partial \rho} + i\frac{\partial}{\partial t}\Big)$$

where $\rho = \mathrm{Im}(z_{n+1}) - \sum_{j=1}^{n} |z_j|^2$ is the "height" function and $T = \frac{\partial}{\partial t}$ is the "missing direction". On the model domain, the system is split in an obvious way. The "normal" component u_{n+1} of u is the solution of a Dirichlet problem of the complex Laplacian:

$$\begin{aligned} \Box^\sharp u_{n+1} &= f_{n+1} \quad \text{in } \ \Omega \\ u_{n+1} &= 0 \qquad \text{on } \ \partial\Omega \end{aligned}$$

Recall that

$$\Box^\sharp = \Big\{ -\frac{1}{2}\sum_{j=1}^{n} \big(Z_j\bar{Z}_j + \bar{Z}_j Z_j\big) - \frac{1}{2}\frac{\partial^2}{\partial t^2} - inT\Big\} - \frac{1}{2}\frac{\partial^2}{\partial \rho^2}.$$

Denote

$$A_\alpha = 2\Big\{ -\frac{1}{2}\sum_{j=1}^{n} \big(Z_j\bar{Z}_j + \bar{Z}_j Z_j\big) - \frac{1}{2}\frac{\partial^2}{\partial t^2} - i\alpha T\Big\}.$$

So,

$$\Box^\sharp = \frac{1}{2}A_\alpha - \frac{1}{2}\frac{\partial^2}{\partial\rho^2}, \qquad \alpha = n.$$

Taking the partial Fourier transform in ρ-variable, one has

$$(\mathcal{F}_{\rho\to\varsigma}u)(z,t;\varsigma) = \frac{2\tilde{f}(z,t;\varsigma)}{A_\alpha + \varsigma^2}.$$

It follows that

$$u(z,t;\rho) = \frac{1}{2\pi}\int\int \frac{e^{i(\rho-\tilde{\rho})\varsigma}}{A_\alpha+\varsigma^2} f(z,t;\tilde{\rho})d\tilde{\rho}d\varsigma = \int \frac{e^{i|\rho-\tilde{\rho}|\sqrt{A_\alpha}}}{\sqrt{A_\alpha}} f(z,t;\tilde{\rho})d\tilde{\rho}.$$

Plugging in the boundary condition, one has

$$u(z,t;\rho) = \int_0^\infty \frac{e^{-|\rho+\tilde{\rho}|\sqrt{A_\alpha}}}{\sqrt{A_\alpha}} f(z,t;\tilde{\rho})d\tilde{\rho} - \int_0^\infty \frac{e^{-(\rho-\tilde{\rho})\sqrt{A_\alpha}}}{\sqrt{A_\alpha}} f(z,t;\tilde{\rho})d\tilde{\rho}.$$

Since we want to find the kernel $G^\sharp$ of the Green's function for the operator $\Box^\sharp$, we let

$$f(z,t;\tilde{\rho}) = \delta_z(w)\otimes\delta_t(s)\otimes\delta_{\tilde{\rho}}(\varrho).$$

Then

$$G^\sharp(z,t;\rho) = \frac{e^{-|\rho-\varrho|\sqrt{A_\alpha}}}{\sqrt{A_\alpha}}\big(\delta_z(w)\otimes\delta_t(s)\big) - \frac{e^{-(\rho+\varrho)\sqrt{A_\alpha}}}{\sqrt{A_\alpha}}\big(\delta_z(w)\otimes\delta_t(s)\big).$$

Let $\sigma(A_\alpha)$ be the symbol of A_α. Then it is easy to see that

$$\sigma(A_\alpha) = \Delta^2 + 2\alpha\tau = \sqrt{2\sum_{j=1}^n |\sigma(Z_j)|^2 + \tau^2} + 2\alpha\tau$$

where $\tau = \sigma\big(\frac{\partial}{i\partial t}\big)$. Hence we have

$$\begin{aligned}
&\frac{e^{-\rho\sqrt{A_\alpha}}}{\sqrt{A_\alpha}}\big(\delta_z(w)\otimes\delta_t(s)\big)\\
&= \frac{1}{(2\pi)^{n+1}}\int_{\mathbb{R}^{2n+1}} \sigma\Big(\frac{e^{-\rho\sqrt{A_\alpha}}}{\sqrt{A_\alpha}}\Big)\mathcal{F}\big(\delta_z(w)\otimes\delta_t(s)\big)e^{i(\sum_{j=1}^{2n}x_j\xi_j+t\tau)}d\xi d\tau\\
&= \frac{1}{(2\pi)^{n+1}}\int_{\mathbb{R}^{2n+1}} \frac{e^{-\rho\Delta}}{\Delta}e^{i(\sum_{j=1}^{2n}(x_j-y_j)\xi_j+(t-s)\tau)}d\xi d\tau\\
&\quad - \frac{1}{(2\pi)^{n+1}}\int_{\mathbb{R}^{2n+1}} e^{-\rho\Delta}\Big(\frac{\alpha\rho}{\Delta^2}+\frac{\alpha}{\Delta^3}\Big)e^{i(\sum_{j=1}^{2n}(x_j-y_j)\xi_j+(t-s)\tau)}\tau d\xi d\tau\\
&\quad + \text{terms with weaker singularities} = I + II + t.w.w.s.
\end{aligned}$$

After length calculation, we have

$$I = \frac{C_n}{[2|z-w|^2 + (t-s+2\operatorname{Im}(z\cdot\bar{w}))^2+\rho^2]^n} \quad \text{with} \quad C_n = \frac{2^{n-1}\Gamma(n)}{\pi^{n+1}}$$

and

$$II = \frac{-C_n}{2(n-1)} \frac{1}{[2|z-w|^2 + (t-s+2\,\mathrm{Im}(z\cdot\bar{w}))^2 + \rho^2]^{n-1}}.$$

Therefore, the kernel for the normal component $\mathbf{N}_{n+1}$ can be written as

$$G^\sharp((z,t;\rho),(w,s;\varrho)) = \frac{C_n}{[2|z-w|^2+(t-s+2\mathrm{Im}(z\cdot\bar{w}))^2+(\rho-\varrho)^2]^n} - \frac{C_n}{[2|z-w|^2+(t-s+2\mathrm{Im}(z\cdot\bar{w}))^2+(\rho+\varrho)^2]^n} + t.w.w.s. \tag{2.3}$$

Here "*t.w.w.s*" stands for "terms with weaker singularities". It is easy to see that the "normal component" $\mathbf{N}_{n+1}$ of $\bar{\partial}$-Neumann operator $\mathbf{N}$ gains two in all directions, *i.e.,*

$$\mathbf{N}_{n+1} : L^p_k(\Omega) \to L^p_{k+2}(\Omega)$$

for all $k \in \mathbb{Z}_+$ and $1 < p < \infty$. Now we are left with solving the following subelliptic boundary problem: *Given f on Ω, find a function u on $\bar{\Omega}$ such that*

$$\begin{aligned} \Box^\flat u &= f \quad \text{in } \Omega, \\ \bar{Z}_{n+1} u &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{2.4}$$

where

$$\Box^\flat = -\frac{1}{2}\frac{\partial^2}{\partial\rho^2} - \frac{1}{2}\sum_{j=1}^n \left(Z_j\bar{Z}_j + \bar{Z}_j Z_j\right) - \frac{1}{2}\frac{\partial^2}{\partial t^2} - (n-2)iT.$$

(In general, the operator $\Box^\flat$ should be a matrix. But on the model domain, $\Box^\flat$ can be written as a scalar operator multiply by an identity matrix. Here we omit the index j.)

In order to solve the problem (2.4), we may assume the solution u is given by

$$u(x,t,\rho) = G^\flat(f) + P(u_b) \tag{2.5}$$

where u_b is the "boundary value" of u which we need to determine. Here $G^\flat$ is the Green's function for the Dirichlet problem which is similar to (2.3) and P is a pseudo-differential operator Poisson type. More precisely, we are interested the following coercive boundary value problem

$$\begin{cases} p(x,D)u = 0 & \text{in } \mathbb{R}^{n+1}_+ \\ Q_j(x,D)u = g_j, & \text{on } \mathbb{R}^n, \ \ j = 1,\dots,m \end{cases} \tag{2.6}$$

where $p(x,D)$ is a strongly elliptic differential operator of order $2m$ with coefficients smooth up to boundary, and g_j are given functions on the boundary. By a result of Phong [**31**], operators P_j, $j = 1,\dots,m$ mapping $C^\infty(\mathbb{R}^n)$ to $C^\infty(\overline{\mathbb{R}^{n+1}_+})$ can be constructed such that

$$u = \sum_{j=1}^m P_j(g_j) + S_{-\infty}(u)$$

which satisfies (2.6). Here $S_{-\infty}$ is an infinity smoothing operator. The operators P_j, $j = 1,\dots,m$ play an analogue role to the Poisson kernel in the case of the

Dirichlet problem for the Laplacian. Operators P_j $j = 1, \dots, m$, originally defined on $C^\infty(\mathbb{R}^n)$ can be written as follows:

$$P_j(g_j)(x,\rho) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} e^{ix\cdot\xi} p_j(x,\rho;\xi)\hat{g}_j(\xi)d\xi.$$

Here $p_j(x,\rho;\xi) \in C^\infty(\mathbb{R}^n \times [0,1] \times \mathbb{R}^n)$ is a symbol of Poisson type of order k which satisfies

1. $p_j(x,\rho;\xi)$ has compact support in the (x,ρ) variables;
2. For all multi-indices α, β and integers γ, δ there is a constant $C_{\alpha,\beta,\gamma,\delta}$ so that

$$\left|\rho^\delta\Big(\frac{\partial}{\partial\rho}\Big)^\gamma\Big(\frac{\partial}{\partial x}\Big)^\beta\Big(\frac{\partial}{\partial\xi}\Big)^\alpha p_j(x,\rho;\xi)\right| \le C_{\alpha,\beta,\gamma,\delta}(1+|\xi|)^{k-|\alpha|+\gamma-\delta}.$$

In the case of (2.6), P_j will have order $-k_j$ id $Q_j(x,D)$ has order k_j as a differential operator.

Plugging in the $\bar{\partial}$-Neumann boundary conditions on u in (2.5), *i.e.,*

$$R\bar{Z}_{n+1}(u) = 0, \qquad j = 1,\dots,n$$

where R is the restriction operator to the boundary. Therefore,

$$0 = R\bar{Z}_{n+1}(u) = R\bar{Z}_{n+1}G^\flat(f) + R\bar{Z}_{n+1}P(u_b)$$

$$\textit{i.e.,} \qquad \Box_+(u_b) = R\bar{Z}_{n+1}P(u_b) = -R\bar{Z}_{n+1}G^\flat(f).$$

The operator $\Box_+$ is called the Calderón operator associated to the $\bar{\partial}$-Neumann problem. This is a 1st-order pseudo-differential operator defined on $\partial\Omega$. Hence, in order to solve the $\bar{\partial}$-Neumann problems reduces to invert the operator $\Box_+$.

The principal symbol of the operator $\Box_+$ is

$$\sigma(\Box_+) = \frac{1}{\sqrt{2}}(\tau - \Delta) - \frac{1}{\sqrt{2}}(2-n)\frac{\tau}{\Delta}.$$

where $\Delta = \sqrt{2\sum_{j=1}^n |\sigma(Z_j)|^2 + \tau^2}$. Obviously, $\Box_+$ is a 1st-order pseudo-differential operator which is elliptic when $\tau < 0$ but doubly characteristic on half of the line bundle

$$\Sigma^+ = \{(z,t;\xi,\tau) :\ \tau > \Delta\}$$

on the cotangent bundle $T^*(\partial\Omega)$.

So far, we have dealt exclusively with the $\bar{\partial}$-Neumann problem on the domain Ω^+. However, we may also construct the Calderón operator $\Box_-$ of the $\bar{\partial}$-Neumann problem on

$$\bar{\Omega}^- = \{z \in U :\ \ \rho(z) \le 0\}.$$

Similar calculus gives us the principal symbol of $\Box_-$ is

$$\sigma(\Box_-) = \frac{1}{\sqrt{2}}(\tau + \Delta) + \frac{1}{\sqrt{2}}(2-n)\frac{\tau}{\Delta}.$$

This is a 1st-order pseudo-differential operator characteristic on the half of the line bundle

$$\Sigma^- = \{(z,t;\xi,\tau) :\ \tau < -\Delta\}$$

but elliptic on the characteristics of $\square_+$. Let us try to look the compositions $\square_+ \circ \square_-$ and $\square_- \circ \square_+$.

The important phenomenon is that

$$\square_+ \circ \square_- = -\square_b + \text{zero-order terms}$$

and

$$\square_- \circ \square_+ = -\square_b + \text{zero-order terms}.$$

Here $\square_b$ is the Kohn Laplacian on $(0,1)$-forms defined on the boundary $\partial\Omega$. More precisely,

$$\square_b = \begin{bmatrix} \square'_1 & 0 & \cdots & 0 \\ 0 & \square'_2 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & \square'_n \end{bmatrix} + \mathcal{A}$$

with

$$\square'_\ell = -\frac{1}{2}\sum_{j=1}^{n}\left(Z_j\bar{Z}_j + \bar{Z}_jZ_j\right) + i\Big(\sum_{k=1}^{n}\lambda_k - 2\lambda_\ell\Big)T$$

for $\ell = 1, \dots, n$ and $\mathcal{A} = [\mathcal{A}_{\alpha\beta}]$ of the form

$$\mathcal{A}_{\alpha\beta} = \sum_{k=1}^{n} a^k_{\alpha\beta}Z_k + \sum_{k=1}^{n} b^k_{\alpha\beta}\bar{Z}_k + c_{\alpha\beta}.$$

Now we are ready to write down a parametrix for the $\bar{\partial}$-Neumann problem.

By a result of Folland and Stein [**21**] (see also Beals-Greiner [**1**] and Berenstein-Chang-Tie [**4**]), $\square_b$ has an inverse $\mathbf{K}$ (for $(0,q)$-forms, $1 \le q \le n-1$) such that

$$\square_b \circ \mathbf{K} = \mathbf{I} + \text{smoothing operators}.$$

The kernel for $\mathbf{K}$ has the following form:

$$K(z,t) = \frac{\Gamma(n-1)}{2^{2-2n}\pi^{n+1}} \frac{1}{(|z|^2 - it)^{n-q}(|z|^2 + it)^q}. \tag{2.7}$$

Summarizing the above discussion, for $n \ge 2$ and $\ell = 1, \dots, n$ we have

$$\mathbf{N}_\ell = G^\flat + P(-\mathbf{K}\square_- R\bar{Z}_{n+1}G^\flat) + S_{-\infty}$$

where $S_{-\infty}$ is a smoothing operator. Hence we have the following theorem, see Greiner-Stein [**23**] and Chang [**9**].

Theorem 2.1. *Let $\Omega \subset\subset \mathbb{C}^{n+1}$ be a smoothly bounded, strongly pseudoconvex domain. Then the following operators are bounded on the indicated spaces, for $1 < p < \infty$, $k \ge 0$, and $\alpha > 0$:*

1. $\mathbf{N} : L^p_k(\Omega) \to L^p_{k+1}(\Omega)$;
2. $\mathcal{P}(Z_j, \bar{Z}_j)\mathbf{N} : L^p_k(\Omega) \to L^p_k(\Omega)$. *Here $\mathcal{P}(Z_j, \bar{Z}_j)$ is any quadratic monomial in "horizontal" vector fields;*
3. $\mathbf{N} : \Lambda_\alpha(\Omega) \to \Lambda_{\alpha+1}(\Omega) \cap \Gamma_{\alpha+2}(\Omega)$.

Corollary 2.1. *Let $\Omega \subset\subset \mathbb{C}^n$ be a smoothly bounded, strongly pseudoconvex domain. Then the solving operator $\bar{\partial}^*\mathbf{N}$ of the Kohn solution for the inhomogeneous Cauchy-Riemann equation satisfies the following estimates.*

1. $\bar{\partial}^*\mathbf{N} : L^p_k(\Omega) \to L^p_{k+\frac{1}{2}}(\Omega)$;
2. $X\bar{\partial}^*\mathbf{N} : L^p_k(\Omega) \to L^p_k(\Omega)$ *for any "good" vector fields* X;
3. $\bar{\partial}^*\mathbf{N} : \Lambda_\alpha(\Omega) \to \Lambda_{\alpha+\frac{1}{2}}(\Omega) \cap \Gamma_{\alpha+1}(\Omega)$.

3. $n = 1$ and Ω is pseudoconvex of finite type m

In this case, we just have one holomorphic tangential vector field $Z = X + iY$. Here finite type condition means X and Y satisfying Chow's condition, *i.e.*,

$$\{X, Y, [X, Y], [X, [X, Y]], [Y, [X, Y]], \dots\}$$

spans the tangent bundle of $T^*(\partial\Omega)$. In this case

$$\Box_b = -\frac{1}{2}\big(Z\bar{Z} + \bar{Z}Z\big) - \frac{i}{2}\lambda T = -\bar{Z}Z.$$

However,

$$Zu = f$$

is the Lewy's equation which is not in general locally solvable. However, by a result of Greiner and Stein [**23**], there exists an operator $\widetilde{\mathbf{K}}$ defined on the Heisenberg group whose kernel equals

$$\widetilde{K}(z, t) = \frac{1}{2\pi^2}\frac{1}{|z|^2 - it}\log\Big[\frac{|z|^2 - it}{|z|^2 + it}\Big], \tag{3.1}$$

such that

$$(-\bar{Z}Z)\widetilde{\mathbf{K}} = \widetilde{\mathbf{K}}(-\bar{Z}Z) = \mathbf{I} - \mathbf{S}.$$

Here

$$\mathbf{S} : L^2(\partial\Omega) \to H^2(\partial\Omega)$$

is the Cauchy-Szegö projection (see Stein [**35**]) whose kernel equals

$$S((z, t), (w, s)) = \frac{2^{n-1}\Gamma(n+1)}{\pi^{n+1}}\frac{1}{\big(i(s-t) - 2z\cdot\bar{w}\big)^{n+1}}.$$

Let us consider 0th-order pseudo-differential operators: Γ^+ and Γ^- with symbols in the class $S^0_{1,0}$ such that the principal symbol of Γ^+ equals to 1 on the set

$$\Big\{\Delta < \frac{1}{4}\sigma(T)\Big\}$$

and whose principal symbol equals 0 on the set

$$\Big\{\Delta > \frac{1}{2}\sigma(T)\Big\}.$$

Denote $\Gamma^- = \mathbf{I} - \Gamma^+$. Now we define

$$\mathbf{K}^+ = Q_{\Gamma^-} + \Gamma^+\mathbf{K}\Box_-,$$

where Q_{Γ^-} is the parametrix for $\Box_+$ in the support of Γ^-, *i.e.,*

$$Q_{\Gamma^-}\Box_+ = \mathbf{I} - \Gamma^+ + S_{-\infty}.$$

Then we have

$$\begin{aligned}\mathbf{K}^+\Box_+ &= Q_{\Gamma^-}\Box_+ + \Gamma^+\mathbf{K}\Box_-\Box_+\\ &= \mathbf{I} - \Gamma^+ + \Gamma^+\mathbf{K}(\Box_b + T_0) + S_{-\infty}\\ &= \mathbf{I} - \Gamma^+ + \Gamma^+(\mathbf{I} - \mathbf{S}) + S_{-\infty}\\ &= \mathbf{I} - \Gamma^+\mathbf{S} + S_{-\infty}.\end{aligned}$$

It can be shown that $\Gamma^+\mathbf{S} \in \cap_{k=1}^{\infty} S^k_{1,0}$ by using microlocal analysis where $S^k_{1,0}$ is the class of order k pseudo-differential operators of type $(1,0)$. Therefore, we have for the $\mathbb{C}^2$ case

$$\mathbf{N}_1 = G + P\big((-Q_{\Gamma^-} + \Gamma^+\mathbf{K}\Box_-)R\bar{Z}_2 G\big) + S_{-\infty}. \tag{3.2}$$

For higher step cases, the projection operator is not $\mathbf{S}$ but we still can construct a projection operator $\mathbf{T}$ such that $(-Z\bar{Z})\widetilde{\mathbf{K}} = \widetilde{\mathbf{K}}(-Z\bar{Z})) = \mathbf{I} - \mathbf{T}$ (see Section 7). Then we may obtain a parametrix for the Neumann operator similar to (3.2). Hence we have the following theorem, see Chang, Nagel and Stein [**14**].

Theorem 3.1. *Let $\Omega \subset\subset \mathbb{C}^2$ be a smoothly bounded, pseudoconvex domain of finite type m. Then the following operators are bounded on the indicated spaces, for $1 < p < \infty$, $k \geq 0$, and $\alpha > 0$:*

1. $\mathbf{N}_1 : L^p_k(\Omega) \to L^p_{k+\frac{2}{m}}(\Omega)$;
2. $\mathcal{P}(Z_1, \bar{Z}_1)\mathbf{N}_1 : L^p_k(\Omega) \to L^p_k(\Omega)$. *Here $\mathcal{P}(Z_1, \bar{Z}_1)$ is any quadratic monomial in Z_1 and $\bar{Z}_1$;*
3. $\mathbf{N}_1 : \Lambda_\alpha(\Omega) \to \Lambda_{\alpha+\frac{2}{m}}(\Omega) \cap \Gamma_{\alpha+2}(\Omega)$.

Corollary 3.1. *Let $\Omega \subset\subset \mathbb{C}^2$ be a smoothly bounded, pseudoconvex domain of finite type m. Then the solving operator $\bar{\partial}^*\mathbf{N}$ of the Kohn solution for the inhomogeneous Cauchy-Riemann equation satisfies the following estimates.*

1. $\bar{\partial}^*\mathbf{N} : L^p_k(\Omega) \to L^p_{k+\frac{1}{m}}(\Omega)$;
2. $X\bar{\partial}^*\mathbf{N} : L^p_k(\Omega) \to L^p_k(\Omega)$ *where $X = Z_1$ or $\bar{Z}_1$;*
3. $\bar{\partial}^*\mathbf{N} : \Lambda_\alpha(\Omega) \to \Lambda_{\alpha+\frac{1}{m}}(\Omega) \cap \Gamma_{\alpha+1}(\Omega)$.

4. Ω: decoupled domain of finite type

A domain $\Omega \subset \mathbb{C}^{n+1}$ and its boundary $\mathcal{M}$ are said to be *decoupled* if there are sub-harmonic, nonharmonic polynomials $\mathcal{P}_j$ with $\mathcal{P}_j(0) = 0$ such that

$$\Omega = \Big\{(z_1, \dots, z_n, z_{n+1}) :\ \mathrm{Im}(z_{n+1}) > \sum_{j=1}^{n} \mathcal{P}_j(z_j)\Big\};$$

$$\mathcal{M} = \Big\{(z_1, \dots, z_n, z_{n+1}) :\ \mathrm{Im}(z_{n+1}) = \sum_{j=1}^{n} \mathcal{P}_j(z_j)\Big\}.$$

We call the integer $m_j = 2 + \text{degree}(\Delta\mathcal{P}_j)$ the "degree" of $\mathcal{P}_j$. We identify $\mathcal{M}$ with $\mathbb{C}^n \times \mathbb{R}$ so that the point

$$\Big(z_1, \ldots, z_n, t + i\big(\sum_{j=1}^{n} \mathcal{P}_j(z)\big)\Big) \in \mathcal{M}$$

corresponds to the point $(z_1, \ldots, z_n, t) \in \mathbb{C}^n \times \mathbb{R}$.

We define the "type" of Ω is

$$m = \max\{m_1, \ldots, m_n\}.$$

Remark 4.1. The actual degree of $\mathcal{P}_j$ may be larger, but the addition of a harmonic polynomial to $\mathcal{P}_j$ does not affect our analysis, and can be eliminated by a change of coordinates.

Examples. 1. The Siegel upper half-space:

$$\Omega = \Big\{(z_1, \ldots, z_{n+1}) : \operatorname{Im}(z_{n+1}) > \sum_{j=1}^{n} a_j |z_j|^2\Big\}$$

with $a_j > 0$ for $j = 1, \ldots, n$. This domain is decoupled and strongly pseudoconvex.

2. Decoupled and finite type:

$$\Omega = \Big\{(z_1, \ldots, z_{n+1}) : \operatorname{Im}(z_{n+1}) > \sum_{j=1}^{n} |z_j|^{2m_j}\Big\}.$$

3. The domain

$$\Omega = \Big\{(z_1, \ldots, z_{n+1}) : \operatorname{Im}(z_{n+1}) > \big(\sum_{j=1}^{n} |z_j|^2\big)^{m_j}\Big\}$$

is finite type but not decoupled.

4. The domain

$$\Omega = \big\{(z_1, z_2, z_3) : \operatorname{Im}(z_3) > |z_1|^2 + \exp(-|z_2|^{-2})\big\}$$

is decoupled but not finite type.

Without loss generality, one may concentrate on a model domain

$$\Omega = \big\{(z_1, \ldots, z_{n+1}) : \operatorname{Im}(z_{n+1}) > |z_1|^{2m_1} + \cdots + |z_n|^{2m_n}\big\}$$

with $m_1, \ldots, m_n \in \mathbb{N}$ and $m_1 \le m_2 \le \cdots \le m_n$. Then $\partial\Omega$ can be identified with $\{(z, t) = (z_1, \ldots, z_n, t) \in \mathbb{C}^n \times \mathbb{R}\}$, and

$$\bar{Z}_1 = \frac{\partial}{\partial \bar{z}_1} - i m_1 |z_1|^{2(m_1-1)} z_1 \frac{\partial}{\partial t}, \quad \ldots \quad \bar{Z}_n = \frac{\partial}{\partial \bar{z}_n} - i m_n |z_n|^{2(m_n-1)} z_n \frac{\partial}{\partial t}$$

form a basis of the tangential Cauchy-Riemann vector fields. The eigenvalues $\lambda_1, \ldots, \lambda_n$ of the Levi form at a point $(z_1, \ldots, z_n, t)$ are essentially $|z_1|^{2(m_1-1)}$, $\ldots$, $|z_n|^{2(m_n-1)}$ and are not comparable unless $m_1 = \cdots = m_n$.

We begin by considering separately the component domains

$$\Omega_j = \{(z_j, w_j) \in \mathbb{C}^2 : \ \mathrm{Im}\,|w_j| > |z_j|^{2m_j}\} \cong \{(z_j, t_j, \rho_j) \in \mathbb{C} \times \mathbb{R} \times \mathbb{R}^+\}$$
$$\mathcal{M}_j = \{(z_j, w_j) \in \mathbb{C}^2 : \ \mathrm{Im}\,|w_j| = |z_j|^{2m_j}\} \cong \{(z_j, t_j) \in \mathbb{C} \times \mathbb{R}\},$$

where $\rho_j = \mathrm{Im}\,|w_j| - |z_j|^{2m_1}$. We denote by $\widetilde{\mathcal{M}}$ the Cartesian product $\mathcal{M}_1 \times \cdots \times \mathcal{M}_n$ and we let π be the projection of $\widetilde{\mathcal{M}}$ to $\partial\Omega$ by

$$\pi : (z_1, t_1) \times \cdots \times (z_n, t_n) \ \rightarrow \ (z_1, \ldots, z_n, t_1 + \cdots + t_n).$$

The idea is to deduce the results about regularity of $\Box_b$ on $\partial\Omega$ from corresponding results on $\widetilde{\mathcal{M}}$.

5. Geometry on $\mathcal{M}_j$

Recall that $\bar{Z}_j = X_j + iX_{n+j}$ where $\{X_j, X_{n+j}\}$ are real vector fields on $\mathcal{M}_j$ and satisfy the bracket generating condition. By Chow's theorem [**16**], given $A, B \in \mathcal{M}_j$, there exists a piecewise $\mathcal{C}^1$ curve $\gamma : [0,1] \rightarrow \mathcal{M}_j$ such that $\gamma(0) = A$ and $\gamma(1) = B$, and such that for $s \in [0,1]$:

$$\dot{\gamma}(s) = \alpha_j(s) X_j(\gamma(s)) + \alpha_{n+j}(s) X_{n+j}(\gamma(s)).$$

As we know, for $j = 1, \ldots, n$,

$$X_j = \frac{\partial}{\partial x_j} + m_j x_{n+j}(x_j^2 + x_{j+n}^2)^{m_j - 1}\frac{\partial}{\partial t},$$
$$X_{n+j} = \frac{\partial}{\partial x_{n+j}} - m_j x_j (x_j^2 + x_{j+n}^2)^{m_j - 1}\frac{\partial}{\partial t},$$

Then

$$H = \frac{1}{2}\Big[\big(\xi_j + m_j x_{n+j}(x_j^2 + x_{j+n}^2)^{m_j-1}\theta\big)^2 + \big(\xi_{n+j} - m_j x_j (x_j^2 + x_{j+n}^2)^{m_j-1}\theta\big)^2\Big]$$

is the Hamiltonian function of the sub-Laplacian on the cotangent bundle $T^*\mathcal{M}_j$. A bicharacteristic curve $(x_j(s), x_{n+j}(s), t(s), \xi_{2n+j}(s), \theta(s)) \in T^*\mathcal{M}_j$ is a solution of the Hamilton's system:

$$\dot{t}(s) = H_\theta, \quad \dot{\theta}(s) = -H_t, \quad \dot{x}_j(s) = H_{\xi_j}, \quad \dot{\xi}_j(s) = -H_{x_j}, \quad \dot{\xi}_{n+j}(s) = -H_{x_{n+j}},$$

with boundary conditions,

$$t(0) = t^{(0)}, \quad x_j(0) = x_j^{(0)}, \quad x_{n+j}(0) = x_{n+j}^{(0)},$$
$$t(\tau) = t, \qquad x_j(\tau) = x_j, \qquad x_{n+j}(\tau) = x_{n+j}.$$

The projection $(x_j(s), x_{n+j}(s), t(s))$ of the bicharacteristic curve on $\mathcal{M}_j$ is a geodesic.

By results of Calin, Chang and Greiner [**5,6,8**], there are finitely many geodesic that join the origin to $(z_j, t_j) \Leftrightarrow z_j \neq 0$. In particular, there is only one geodesic connecting the point $(z_j, 0)$ and the origin. These geodesics are parametrized by the solution θ of

$$\frac{|t_j|}{|z_j|^{2m_j}} = \mu_j(\theta) = \frac{2}{2m_j - 1} \frac{\int_0^{(2m_j-1)\theta} \sin^{\frac{2m_j}{2m_j-1}}(v)dv}{\sin^{\frac{2m_j}{2m_j-1}}(\theta)}. \tag{5.1}$$

If $0 \le \theta_1 < \cdots < \theta_N$ are the solutions of (5.1) (see [**2,3,7,8**]), there are exactly N geodesics, and their lengths are given by

$$(d_k^{(j)})^{2m_j} = \nu_j(\theta_k)\big(|t_j| + |z_j|^{2m_j}\big), \quad k = 1, \ldots, N$$

where

$$\nu_j(\theta) = \frac{\Big[\int_0^{(2m_j-1)\theta} \sin^{-\frac{2m_j-2}{2m_j-1}}(v)dv\Big]^{2m_j}}{m_j^{2m_j}(1+\mu_j(\theta)) \sin^{\frac{2m_j}{2m_j-1}}((2m_j-1)\theta)}.$$

If $|z_j| = 0$, then there are uncountably infinitely many geodesics join the origin to a point $(0, t_j)$. Their lengths are $d_1^{(j)}, d_2^{(j)}, \ldots$, where

$$(d_k^{(j)})^{2m_j} = \Big(\frac{k}{2m_j - 1}\Big)^{2m_j-1} \frac{M^{2m_j}}{Q}|t_j|,$$

where the constants M and Q expressed in terms of beta function $\mathcal{B}$:

$$M = \mathcal{B}\Big(\frac{1}{4m_j - 2}, \frac{1}{2}\Big), \qquad Q = 2\mathcal{B}\Big(\frac{4m_j - 1}{4m_j - 2}, \frac{1}{2}\Big).$$

In order to do analysis on $\mathcal{M}$, we combine the ideas in [**8**], [**28**] and [**30**] to quantize the Carnot-Carathéodory distance. Define d_j on $\mathcal{M}_j$ as follows:

$$d_j(p,q) = \inf\big\{\text{length}(\gamma)\colon\ \gamma \text{ horizontal},\ \gamma : [0,1] \to \mathcal{M}_j,\ \ \gamma(0) = p,\ \ \gamma(1) = q\big\}$$

where

$$\text{length}(\gamma) = \int_0^1 \sqrt{|\alpha_j(s)|^2 + |\alpha_{n+j}(s)|}ds.$$

The corresponding non-isotropic ball is

$$B_j(p,\delta) = \big\{q \in \mathcal{M}_j :\ d_j(p,q) < \delta\big\}.$$

Write the commutator

$$[X_j, X_{n+j}] = \lambda_j(p)T + a_j(p)X_j + a_{n+j}(p)X_{n+j}.$$

For $k \ge 2$ set

$$\Lambda_j^k(p) = \sum_{\alpha_1+\cdots+\alpha_\ell \le k-2} \big|X_{\alpha_1} \cdots X_{\alpha_\ell}\lambda_j(p)\big|,$$

where $X_{\alpha_i} = X_j$ or X_{n+j}, $i = 1, \ldots, \ell$. Set

$$\Lambda_j(p,\delta) = \sum_{k=2}^{m_j} \Lambda_j^k(p)\delta^k.$$

Now it is easy to see that there are constants C_1, C_2 depending only on m_j so that for $p \in \mathcal{M}_j$ and $\delta > 0$

$$C_1\delta^2\Lambda_j(p,\delta) \le \big|B_j(p,\delta)\big| \le C_2\delta^2\Lambda_j(p,\delta).$$

Note that for $\delta > 0$, $\delta \mapsto \Lambda_j(p,\delta)$ is a monotone increasing function. Hence there is a unique inverse function $\sigma_j(p,\delta)$ such that for $\delta \ge 0$:

$$\Lambda_j(p,\sigma_j(p,\delta)) = \sigma_j(p,\Lambda_j(p,\delta)) = \delta.$$

We have

$$\sigma_j(p,\delta) \approx \sum_{k=2}^{m_j} \frac{\sqrt[k]{\Lambda_j(p)}}{\delta^{\frac{1}{k}}}.$$

For every $0 < \delta < 1$, every $p \in \mathcal{M}$, we have

$$c\delta^{\frac{1}{2}} \le \sigma_j(p,\delta) \le C\delta^{\frac{1}{m_j}},$$

and for every $0 \le \lambda \le 1$, there are constants C_1', C_2' depending only on m_j so that for $p \in \mathcal{M}_j$ and $\delta > 0$

$$C_1'\delta\cdot\sigma_j^2(p,\delta) \le \big|B_j(p,\delta)\big| \le C_2'\delta\cdot\sigma_j^2(p,\delta).$$

This allows us to define a pseudometric on $\mathcal{M}$ as follows. For $p, q \in \mathcal{M}$,

$$\begin{aligned} d_c(p,q) = \inf\Big\{&\delta > 0 : \exists \text{ piecewise curve } \mathcal{C}^1\ \gamma : [0,1] \to \mathcal{M}, \gamma(0) = p,\ \gamma(1) = q,\\ &\dot\gamma(s) = \sum_{j=1}^{n}\big(a_j(s)X_j + a_{n+j}(s)X_{n+j}\big),\ |a_j(s)|,\\ &|a_{n+j}(s)| \le \sigma_j(p,\delta),\ d_e(p,q) \le \delta\Big\}. \end{aligned}$$

Here $d_e(p,q)$ is the Euclidean distance between p and q.

6. Fundamental solution for operators $\Box_j$

We shall construct the fundamental solution for the Kohn-Laplacian on $\mathcal{M}_j$ (see Beals-Gaveau-Greiner [**2**] and Chang-Greiner [**11**]). It reduces to consider the following operator:

$$\begin{aligned} \Delta_{\lambda,m} &= -\frac{1}{2}(Z\bar Z + \bar Z Z) - \frac{1}{2}\lambda[Z,\bar Z]\\ &= -\frac{\partial^2}{\partial z\partial\bar z} + im|z|^{2m-2}\frac{\partial}{\partial t}\Big(z\frac{\partial}{\partial z} - \bar z\frac{\partial}{\partial\bar z}\Big) - m^2|z|^{4m-2}\frac{\partial^2}{\partial t^2} + im^2\lambda|z|^{2m-2}\frac{\partial}{\partial t} \end{aligned}$$

where

$$Z = \frac{\partial}{\partial z} + im\bar z|z|^{2(m-1)}\frac{\partial}{\partial t}, \quad \bar Z = \frac{\partial}{\partial\bar z} - imz|z|^{2(m-1)}\frac{\partial}{\partial t},$$

with $m \in \mathbb{N}$ and $\lambda \notin \Gamma$. Here Γ is an exceptional set. We need to calculate the fundamental solution $K_{\lambda,m}$ for the operator $\Delta_{\lambda,m}$.

We shall look for a $K_{\lambda,m}$ in the form (see Beals-Gaveau-Greiner [**2,3**] and Chang-Tie [**13**]):

$$K_{\lambda,m} = \int_{-\infty}^{\infty} \frac{E_\lambda(z,w,\tau)V_\lambda(z,w,\tau)}{g_m(z,w,t-s,\tau)} d\tau = -\int_{-\infty}^{\infty} V_\lambda d\big(\log g_m\big),$$

where g_m is the action of the complex Hamiltonian problem which satisfies the Hamilton-Jacobi equation

$$\frac{\partial g_m}{\partial \tau} + \frac{1}{2}\big[(X_j g_m)^2 + (X_{n+j} g_m)^2\big] = 0,$$

and $E = -\frac{\partial g_m}{\partial \tau}$ is the associated energy, the first invariant of motion. The volume element V_λ is the solution of a second-order transport equation

$$\Delta_{\lambda,m}(E_\lambda V_\lambda) + \frac{\partial}{\partial \tau}\Big[T(V_\lambda) + \big(\Delta_{\lambda,m} g_m\big)V_\lambda\Big] = 0,$$

where

$$T = \frac{\partial}{\partial \tau} + (X_j g_m)X_j + (X_{n+j} g_m)X_{n+j}$$

is the differentiation along the bicharacteristics.

Theorem 6.1. *For $-1 < \mathrm{Re}(\lambda) < 1$, the fundamental solution $K_{\lambda,m}$ for the sub-Laplacian $\Delta_{\lambda,m}$ has the following closed form:*

$$K_{\lambda,m}(z,w,t-s) = -\frac{1}{4m\pi^2} \frac{F_{\lambda,m}}{A^{(1-\lambda)/2}\bar{A}^{(1+\lambda)/2}}, \tag{6.1}$$

where

$$A = \frac{1}{2}\left(|z|^{2m} + |w|^{2m} + i(t-s)\right).$$

and

$$\Gamma\Big(\frac{1-\lambda}{2}\Big)\Gamma\Big(\frac{1+\lambda}{2}\Big)F_{\lambda,m}$$
$$= \int_0^1\int_0^1 \Big\{\frac{[s(1-\sigma)]^{-\frac{1+\lambda}{2}}[\sigma(1-s)]^{-\frac{1-\lambda}{2}}}{(1-\mathcal{P}s^{\frac{1}{m}})(1-\bar{\mathcal{P}}\sigma^{\frac{1}{m}})} \frac{1-(s\sigma)^{\frac{1}{m}}|\mathcal{P}|^2}{1-s\sigma|\mathcal{P}|^{2m}}\Big\} ds d\sigma$$

with

$$\mathcal{P} = \begin{cases} \frac{\sqrt[m]{2} z\cdot\bar{w}}{\sqrt[m]{A}}, & \text{if } \; w \neq 0 \\ 0, & \text{if } \; w = 0. \end{cases}$$

Remark 6.2. We can show that $|\mathcal{P}| \le 1$ and $\mathcal{P} = 1 \Leftrightarrow (z,t) = (w,s)$. Hence, $K_{\lambda,m}$ has a unique singularity at (w,s). Moreover, one can show that

$$\big|K_{\lambda,m}(p,q)\big| \le \frac{C d_j(p,q)^2}{|B_j(p, d_j(p,q))|}.$$

Hence $K_{\lambda,m} \in L^1_{\mathrm{loc}}(\mathbb{R}^3)$ and

$$\Delta_{\lambda,m} K_{\lambda,m} = \delta(z,w,t-s).$$

From Theorem 6.1, we obtain the fundamental solution for the operator $\Delta_{0,m}$ as a corollary.

Corollary 6.1. *Assume $(z,t)\neq(w,s)$. Then the fundamental solution $K_{0,m}$ for the sub-Laplacian $\Delta_{0,m}$ has the following closed form:*

$$K_{0,m}=\frac{1}{4m\pi^3|A|}\int_0^1\int_0^1\prod_{\ell=1}^{m-1}\left(1-e^{\frac{2\ell\pi}{m}i}|\mathcal{P}|^2(s\sigma)^{\frac{1}{m}}\right)^{-1}$$
$$\times\frac{ds d\sigma}{\sqrt{s(1-s)\sigma(1-\sigma)}(1-\mathcal{P}s^{\frac{1}{m}})(1-\bar{\mathcal{P}}\sigma^{\frac{1}{m}})}.$$

In particular, when $m=1$, one has

$$K_{0,1}=\frac{1}{4\pi^3|A|}\int_0^1\frac{ds}{\sqrt{s(1-s)}(1-\mathcal{P}s)}\int_0^1\frac{d\sigma}{\sqrt{\sigma(1-\sigma)}(1-\bar{\mathcal{P}}\sigma)}$$
$$=\frac{1}{2\pi}\left[|z-w|^4+\big(t-s+2\operatorname{Im}(z\cdot\bar{w})\big)^2\right]^{-\frac{1}{2}}$$

This coincides with the Folland-Stein formula (2.7) for $q=\frac{n}{2}$.

When $m=2$, one has

$$K_{0,2}(z,w,t-s)=\frac{1}{4\pi d}+\frac{i}{2\pi^2 d}\log[h(\mathcal{P},\bar{\mathcal{P}})],$$

where

$$d=\sqrt{|z^2-w^2|^4+\big(t-s+2\operatorname{Im}(z^2\bar{w}^2)\big)^2}$$

and

$$h(\mathcal{P},\bar{\mathcal{P}})=\frac{|1-\mathcal{P}^2|-i(\mathcal{P}+\bar{\mathcal{P}})}{1+|\mathcal{P}|^2}.$$

This recovers a result of Greiner [**22**].

7. Kernel $K_{1,m}$ and projection operators

If $|\operatorname{Re}(\lambda)|\geq 1\ \Rightarrow$ the integral in formula (6.1) will be divergent. However, we may consider analytic continuation in λ. In fact,

Theorem 7.1. *The function $K_{\lambda,m}$ has a meromorphic extension in λ with simple pole at the exception set Γ:*

$$\Gamma=\left\{\pm\Big(\frac{2k}{m}+1+2\ell\Big),\quad k=0,1,\dots,m-1,\ \ell\in\mathbb{Z}_+\right\}.$$

If $\lambda\notin\Gamma$, $K_{\lambda,m}\in\mathcal{C}^\omega(\mathcal{M}_j\times\mathcal{M}_j\setminus\Gamma)$ and

$$\Delta_{\lambda,m}K_{\lambda,m}=\delta(z,w,t-s).$$

The limiting case $\lambda\to\pm1$ can be evaluated by a residue calculation: for $m>1$,

$$\begin{aligned}K_{1,m}(z,w,t-s)=\frac{1}{4\pi^2 m\bar{A}}\Big\{&\frac{1}{1-\bar{\mathcal{P}}}\log\frac{\bar{A}}{A}+\frac{m}{1-\bar{\mathcal{P}}^m}\log\Big(\frac{1-\bar{\mathcal{P}}}{1-\mathcal{P}}\Big)\\&-\frac{\bar{\mathcal{P}}}{1-\bar{\mathcal{P}}}\sum_{\ell=1}^{m-1}\frac{1-e^{\frac{2\ell\pi}{m}i}}{1-e^{\frac{2\ell\pi}{m}i}\bar{\mathcal{P}}}\log\Big(\frac{1-e^{\frac{2\ell\pi}{m}i}|\mathcal{P}|^2}{1-e^{-\frac{2\ell\pi}{m}i}}\Big)\Big\}.\end{aligned}\tag{7.1}$$

For $m = 1$, we recover the formula (3.1) of Griener and Stein [**23**]. As we mentioned before,

$$\Delta_{1,1} K_{1,1}(z, w, t-s) = I - \mathbf{S}_1.$$

where $\mathbf{S}_1$ is the Cauchy-Szegö projection.

On $\mathcal{M}_j$, the operator

$$\begin{aligned}\Delta_{1,m} &= -\frac{1}{2}\big(Z\bar{Z} + \bar{Z}\,Z\big) - \frac{1}{2}[Z, \bar{Z}] \\ &= -Z\bar{Z}\end{aligned}$$

which is not solvable in general. However, from (7.1), one has

$$\Delta_{1,m} K_{1,m} = I - \mathbf{T}_1$$

where $\mathbf{T}_1$ is a projection operator, see, *e.g.*, Christ [**17**], Fefferman-Kohn [**19**], and Chang-Nagel-Stein [**14**]. In general, $\mathbf{T}_1 \neq \mathbf{S}_1$ (except $m = 1$).

However, one can show $\mathbf{T}_1 = \mathbf{T}_1\mathbf{S}_1$. It follows that

$$\Delta_{1,m} K_{1,m}(I - \mathbf{S}_1) = (I - \mathbf{T}_1)(I - \mathbf{S}_1) = I - \mathbf{S}_1.$$

Using (7.1) again, one has

Theorem 7.2. *Let*

$$\begin{aligned}\widetilde{\mathcal{K}}_j = \ker(K_{1,m}(I - \mathbf{S}_1)) &= K_{1,m} - \ker(K_{1,m}\mathbf{S}_1) \\ &= \frac{-1}{4\pi^2 m\bar{A}}\Big\{\frac{1}{1-\bar{\mathcal{P}}}\log\frac{A}{\bar{A}} + \frac{1}{1-\bar{\mathcal{P}}}\log(1 - |\mathcal{P}|^{2m}) \\ &\quad - m\mathcal{P}\int_0^1 \frac{d\sigma}{(1-\sigma\mathcal{P})(1-\sigma^m|\mathcal{P}|^{2m})}\Big\} \\ &\quad + \frac{1}{8\pi^3 m}\int_{-\infty}^{\infty}\int_0^{\infty} \bar{A}_{(z,t)}^{-1} A_{(w,s)}^{-\frac{1}{m}-1}\log\Big[\frac{A_{(z,t)}}{\bar{A}_{(z,t)}} - \frac{|z\bar{w}'|^{2m}}{|A_{(z,t)}|^2}\Big] \\ &\quad \times \Big\{1 - \big(\frac{z^m\bar{w}^m r^{2m}}{\bar{A}_{(z,t)}A_{(w,s)}}\big)^{\frac{1}{m}}\Big\}^{-2} r\,dr\,ds'.\end{aligned}$$

Then

$$\Delta_{1,m}\widetilde{\mathcal{K}}_j = \delta(z, w, t-s) - \mathbf{S}_1(z, w, t-s).$$

Here

$$\begin{aligned}A_{(z,t)} &= \frac{1}{2}\Big(|z|^{2m} + |w'|^{2m} + i(t-s')\Big), \\ A_{(w,s)} &= \frac{1}{2}\Big(|w|^{2m} + |w'|^{2m} + i(s-s')\Big).\end{aligned}$$

8. Descent to $\mathcal{M}$

The operator

$$\widetilde{\mathcal{K}} = \widetilde{\mathcal{K}}_1 \otimes \widetilde{\mathcal{K}}_2$$

on $\widetilde{\mathcal{M}}$ turns out to be a product-type singular integral on $\mathcal{M}_1 \times \mathcal{M}_2$. Using results on $\widetilde{\mathcal{M}}$, we have

$$\mathcal{K}(z,w,t) = \int_{\Sigma(t)} \widetilde{\mathcal{K}}(z,w,\zeta)d\widetilde{\zeta} = (\widetilde{\mathcal{K}})^\flat(z,w,t),$$

where $d\widetilde{\zeta}$ is the volume element on the surface

$$\Sigma(t) = \big\{\zeta = (\zeta_1, \zeta_2) : \ \zeta_1 + \zeta_2 = t\big\}.$$

It is easy to check that $\mathcal{K}$ is the relative fundamental solutions for $\Box = \Box_1 + \Box_2$ in the sense that

$$\mathcal{K}\Box = \Box\mathcal{K} = I - \mathbf{S}_1 \otimes \mathbf{S}_2 = I - \mathbf{S}.$$

Here $\mathbf{S} = \mathbf{S}_1 \otimes \mathbf{S}_2$ is the orthogonal projection onto the intersection of the null spaces of the operators $\{\Box_1, \Box_2\}$, which is the same as the projection onto the null space of the operator

$$\Box = -(Z_1\bar{Z}_1 + \bar{Z}_2 Z_2) = \Box_1 + \Box_2.$$

Since $\Box$ is translation invariant along the missing direction t, we may take partial Fourier transform of $\Box$ with respect to that variable:

$$\tilde{f}(z,w,\tau) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-it\tau} f(z,w,t)dt.$$

The support of the partial Fourier transform of distribution kernel S_1 of the projection operator $\mathbf{S}_1$ is supported where $\tau \geq 0$. Similarly, the support of the partial Fourier transform of distribution kernel S_2 of the projection operator $\mathbf{S}_2$ is supported where $\tau \leq 0$. Therefore, when $1 \leq q \leq n-1$, we may invert $\Box$. But when $q = 0$ and $q = n$, we need to invert $\Box$ in the orthogonal complements of $\mathbf{S}_1$ or $\mathbf{S}_2$. Summarizing, we have the following theorem.

Theorem 8.1. *Let $\vartheta = \{(j_1, \ldots, j_q) : \ j_1 \leq \cdots \leq j_q, \ 1 \leq j_\ell \leq n, \ 1 \leq \ell \leq q\}$. For each of the $\frac{n!}{q!(n-q)!}$ possible operator $\Box_J$ with $J \in \vartheta_q$, we construct a distribution $\mathcal{K}_J$ on $\mathcal{M} \times \mathcal{M}$ so that if $\mathcal{K}_J$ denotes the linear operator*

$$\big\langle \mathcal{K}_J[\varphi], \psi\big\rangle \ = \ \big\langle \mathcal{K}_J, \varphi \otimes \psi\big\rangle,$$

for φ, $\psi \in \mathcal{C}_0^\infty(\mathcal{M})$. Then

$$\mathcal{K}_J\Box_J = \Box_J\mathcal{K}_J = \begin{cases} I - \mathbf{S}_0, & \text{if } J \in \vartheta_0 \\ I, & \text{if } J \in \vartheta_q,\ 1 \leq q \leq n-1 \\ I - \mathbf{S}_n, & \text{if } J \in \vartheta_n. \end{cases}$$

Let us consider again the model domain in $\mathbb{C}^3$:

$$\mathcal{M} = \big\{(z_1, z_2, z_3) : \ \mathrm{Im}(z_3) = |z_1|^{2m_1} + |z_2|^{2m_2}\big\}.$$

With $\bar{Z}_j = \frac{1}{2}(X_j + iX_{2+j})$, we can construct

$$d_\Sigma \approx |z_1| + |z_2| + |t|^{\frac{1}{2m_2}}$$

the "control metric" defined by the sub-Laplacian $X_1^2 + X_2^2 + X_3^2 + X_4^2$.

However, there is another natural which reflects the "flatness" of the boundary in different complex directions, the "Szegö metric":

$$d_S^{2m_2} \approx |z_1|^{2m_2} + |z_2|^{2m_1} + |t|.$$

Note that if $m_1 < m_2$,

$$d_S(0,p) \approx |z_1| + |z_2|^{\frac{m_1}{m_2}} + |t|^{\frac{1}{2m_2}} \not\approx d_\Sigma(0,p)$$

d_S controls the orthogonal projection on the null-space of the operator $-(Z_1\bar{Z}_1 + Z_2\bar{Z}_2)$. Some mixture of d_Σ and d_S arises in the fundamental solution of the operator

$$\begin{aligned}\Box_b &= -\left(Z_1\bar{Z}_1 + \bar{Z}_2 Z_2\right) = \Box_1 + \Box_2 \\ &= -\frac{1}{2}\sum_{k=1}^{2}\left(Z_k\bar{Z}_k + \bar{Z}_k Z_k\right) - \frac{1}{2}[Z_1, \bar{Z}_1] + \frac{1}{2}[\bar{Z}_2, Z_2].\end{aligned}$$

Theorem 8.2. *For all $J \in \vartheta_q$, the distribution $\mathcal{K}_J$ satisfies the following size estimates: let $\partial_k^{|\alpha_k|}$ be a derivative of order $|\alpha_k|$ made up of the vector fields Z_k and $\bar{Z}_k$ in which each acts in either the variables p_j or q_j. Then for all $\alpha_{\mathbf{k}} = (\alpha_{k_1}, \ldots, \alpha_{k_n})$ there is a constant $C_{\alpha_{\mathbf{k}}}$ such that*

$$\begin{aligned}&\Big|\Big[\prod_{k=1}^{n}\partial_k^{|\alpha_k|}\Big]\mathcal{K}_J(p,q)\Big| \\ &\quad\le C_{\alpha_{\mathbf{k}}}\frac{\big[\prod_{k=1}^{n}\sigma_k(p, d_S(p,q))\big]^2}{\big|B_S(p, d_S(p,q))\big|}\log\Big\{2 + \frac{\big[\prod_{k=1}^{n}\sigma_k(p, d_S(p,q))\big]}{d_\Sigma(p,q)}\Big\} \\ &\qquad\times\prod_{k=1}^{n}\Big[\sigma_k(p, d_S(p,q))^{-1} + d_\Sigma(p,q)^{-1}\Big]^{|\alpha_k|}.\end{aligned}$$

The log *term cannot be removed.*

This phenomena was first discovered by Machedon [**27**]. In fact, Derridj [**18**] has already shown that maximal hypoelliptic estimates are possible only if the eigenvalues of the Levi form degenerate at the same rate. Therefore, analysis on a domain like $\Omega = \{(z_1, z_2, z_3) \in \mathbb{C}^3 : \ \mathrm{Im}(z_3) > |z_1|^4 + |z_2|^4\}$ is easier than analysis on a domain like $\Omega = \{(z_1, z_2, z_3) \in \mathbb{C}^3 : \ \mathrm{Im}(z_3) > |z_1|^2 + |z_2|^4\}$. Readers can also read a survey article by Chang and Fefferman [**10**] for detailed discussion. Moreover, mixed typed homogeneities singular integrals have already seen when we study the Henkin solution for inhomogeneous equation. Sharp estimates for singular integral with kernels like $E_k H_\ell$ which arise from $\bar{\partial}$-Neumann on strongly pseudo-convex domains was obtained by Phong and Stein [**33**], [**34**]. Here E_k is a kernel with Euclidean homogeneity of degree $-k$ and H_ℓ is a kernel with Heisenberg homogeneity of degree $-\ell$. It takes sometime for mathematicians to understand this puzzle when $\partial\Omega$ is a decoupled domain of finite type. However, unlike L^p estimates, it is easier to obtain Hölder estimates for the kernel $\mathcal{K}_J$. We have the following theorems.

Theorem 8.3. *Let $m = \max_{1\le j\le n}\{m_j\}$ be the largest of the degrees of the polynomials $\mathcal{P}_j$. (This is the "type" of the boundary $\mathcal{M}$.) Assume $m > 2$, and suppose that f is a function bounded and supp$(f) \subset B_S(p,1) \subset \mathcal{M}$ where $B_S(p,1)$ is the unit ball induced by the Szegö metric centered at the point p. Then for all $J \in \vartheta_q$ there is a constant C_J such that if $q \in B_S(p,10^{-1}) \subset \mathcal{M}$, the fundamental solution $\mathcal{K}_J$ satisfies*

$$\Big|\mathcal{K}_J(f)(p+q) + \mathcal{K}_J(f)(p-q) - 2\mathcal{K}_J(f)(p)\Big| \le C_J|q|^{\frac{1}{m}}.$$

for $p \in \mathcal{M}$.

In particular, when $m = 1$, *i.e.*, $\mathcal{M}$ is strongly pseudo-convex, one has

Theorem 8.4. *Suppose that f is a bounded function and* supp$(f) \subset B_S(p,1) \subset \mathcal{M}$. *Then there is a constant C_J such that if $q \in B_S(p,10^{-1})$, one has*

$$\Big|\mathcal{K}(f)(p+q) + \mathcal{K}(f)(p-q) - 2\mathcal{K}(f)(p)\Big| \le C\Big[\sum_{j=1}^{n}\sigma_j(p,|q|)^2\Big] \approx C|q|.$$

9. Flag singular integral operators

There is a fundamental issue that arises at this point. Operator like $\widetilde{\mathcal{K}}_J$ is not pseudo-local, because as product-like operators their kernels have singularities on the products of the diagonals of the $\mathcal{M}_j$, and not just on the diagonal of $\widetilde{\mathcal{M}}$. As a result the projection $\mathcal{K}_J = (\widetilde{\mathcal{K}}_J)^\flat$ of such operator on $\mathcal{M}$ is thus in general again not pseudo-local.

We need to obtain the appropriate differential inequalities and cancellation properties satisfied by the kernels of $\mathcal{K}_J$, $J \in \vartheta$ away from diagonal. This leads to a new research direction: flag singular integral operators.

Let

$$\mathbb{R}^N = \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_n},$$

each of which is homogeneous under a given family of possible non-isotropic dilations. We denote the elements $x \in \mathbb{R}^N$ by n-tuples $x = (x_1, \dots, x_n)$ with $x_j \in \mathbb{R}^{m_j}$. On each $\mathbb{R}^{m_j}$ we can choose coordinates $x_j = (x_j^1, \dots, x_j^{m_j})$ so that the dilation by $\delta > 0$ is given by

$$\delta \cdot x_j = \big(\delta^{\lambda_j^1} x_j^1, \dots, \delta^{\lambda_j^{m_j}} x_j^{m_j}\big).$$

We denote by $Q_j = \sum_{k=1}^{m_j} \lambda_j^k$ the homogeneous dimension of $\mathbb{R}^{m_j}$ and $|x_j|$ a smooth homogeneous norm on $\mathbb{R}^{m_j}$. If $\alpha_j = (\alpha_j^1, \dots, \alpha_j^{m_j})$ is a multi-index with m_j components, we denote its weighted length by $|\alpha_j| = \sum_{k=1}^{m_j} \lambda_j^k \alpha_j^k$.

A *k-normalized bump function* on a space $\mathbb{R}^{m_j}$ is a $\mathcal{C}^k$ function supported on the unit ball with $\mathcal{C}^k$-norm bounded by 1.

Definition 9.1. A flag (or filtration) in $\mathbb{R}^N$ is a family of subspaces

$$0 = V_0 \subset V_1 \subset \cdots \subset V_{n-1} \subset V_n = \mathbb{R}^N.$$

For each j let W_j be a complementary subspace V_{j-1} in V_j, *i.e.*,

$$V_j = V_{j-1} \oplus W_j.$$

The family $\{W_j\}$ is called a gradation associated to the filtration $\{V_j\}$.

Definition 9.2. A flag kernel, relative to the flag $\{V_j\}$, is a distribution K on $\mathbb{R}^N$ which coincides with a C^∞ function away from the coordinate subspace $x_n = 0$ and which satisfies:

(1) Differential inequalities: For each

$$\alpha = (\alpha_1, \dots, \alpha_n) \in (\mathbb{Z}_+)^n$$

there is a constant C_α so that

$$\left|\partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n} K(x)\right| \\ \le C_\alpha \Big(|x_1| + \cdots + |x_n|\Big)^{-Q_1 - |\alpha_1|} \times \cdots \times \Big(|x_{n-1}| + |x_n|\Big)^{-Q_{n-1} - |\alpha_{n-1}|} |x_n|^{-Q_n - |\alpha_n|}$$

for $x_n \neq 0$.

(2) Cancellation conditions: These are defined induction on n.

(a) For $n = 1$, given any normalized bump function ϕ and any scaling parameter $R > 0$, the quantity

$$\int K(x)\phi(Rx)dx$$

is bounded uniformly on ϕ and R;

(b) for $n > 1$, given any $j \in \{1, \dots, n\}$, any normalized bump function ϕ on W_j, and any scaling parameter $R > 0$, the distribution

$$K_{\phi,R}(x_1, \dots, \widehat{x}_j, \dots, x_n) = \int K(x)\phi(Rx_j)dx_j$$

is a flag kernel on $\oplus_{k\neq j} W_k$, adapted to the flag

$$0 \subset W_1 \subset \cdots \subset \oplus_{k=1}^{j-1} W_k \subset \oplus_{k\neq j;k=1}^{j+1} W_k \subset \cdots \subset \oplus_{k\neq j;k=1}^{n} W_k.$$

Here are two examples of flag singular integrals.

Example 1. K_1 is the distribution given by integration against the function

$$K_1(x,y) = \frac{1}{x(x+iy)}$$

on the set where $x \neq 0$, and defined as a principal value integral.

Example 2. K_2 is defined in a similar way:

$$K_2(x,y) = \text{P.V.}\frac{1}{xy}\chi\Big(\frac{y}{x}\Big), \quad \text{away from } y \neq 0.$$

Here χ is an even function with compact support.

10. Littlewood-Paley square function associated with flag singular integrals

Let $\psi^1 \in \mathcal{C}_0^\infty(\mathbb{R}^{n+m})$ and $\psi^2 \in \mathcal{C}_0^\infty(\mathbb{R}^m)$ which satisfying

$$\sum_j \Big|\widehat{\psi}^1\Big(\frac{\xi_1}{2^j}, \frac{\xi_2}{2^j}\Big)\Big|^2 = 1 \quad \forall\ (\xi_1,\xi_2) \in \mathbb{R}^{n+m} \setminus \{(0,0)\},$$

$$\sum_k \Big|\widehat{\psi}^2\Big(\frac{\zeta}{2^k}\Big)\Big|^2 = 1 \quad \forall\ \zeta \in \mathbb{R}^m \setminus \{0\}$$

and the moment conditions

$$\int_{\mathbb{R}^{n+m}} x^\alpha y^\beta \psi^1(x,y)dxdy = \int_{\mathbb{R}^m} z^\gamma \psi^2(z)dz = 0$$

for all multiindices α, β, and γ.

Denote

$$\psi^\flat(x,y) = \int_{\mathbb{R}^m} \psi^1(x,y-z)\psi^2(z)dz.$$

For $f \in L^p$, $1 < p < \infty$, the Littlewood-Paley square function $g_F(f)$ is defined by

$$g_F(f)(x,y) = \Big\{\sum_j \sum_k \big|\psi^\flat_{j,k} * f(x,y)\big|^2\Big\}^{\frac{1}{2}}$$

where $\psi^\flat_{j,k}(x,y) = \big(\tau_{2^j,2^k}\psi^\flat\big)(x,y)$. Here

$$\big(\tau_{\delta_1,\delta_2}\psi^\flat\big)(x,y) = \delta_1^{n+m}\delta_2^m \int_{\mathbb{R}^m} \psi^1(\delta_1 x, \delta_1(y-z))\psi^2(\delta_2 z)dz.$$

Theorem 10.1. *There are constants C_1 and C_2 such that for $1 < p < \infty$,*

$$C_1\|f\|_{L^p} \leq \|g_F(f)\|_{L^p} \leq C_2\|f\|_{L^p}.$$

The following theorem was first proved by Nagel, Ricci and Stein [**29**] for L^p, $1 < p < \infty$. Using Littlewood-Paley square functions, the result can be extended to H^p for $0 < p < \infty$.

Theorem 10.2. *Suppose that T is a flag singular integral operator defined on $\mathbb{R}^n \times \mathbb{R}^m$ with the flag kernel*

$$K^\flat(x,y) = \int_{\mathbb{R}^m} \widetilde{K}(x, y-z, z)dz,$$

where the product kernel $\widetilde{K}$ satisfies the conditions in Definition 9.2. Then T is bounded on L^p for $1 < p < \infty$. Moreover, there exists a constant C such that for $f \in L^p$, $1 < p < \infty$,

$$\|T(f)\|_{L^p} \le C\|f\|_{L^p}.$$

11. H^p theory associated with flag singular integrals

In this section, we just give a rough introduction to H^p theory associated with flag singular integral operators (see Han [**24**]). Detailed discussion will appear elsewhere. It is well known that when $0 < p < 1$, H^p is a class of distributions satisfying certain criterion. Hence we first need to introduce suitable test function space on $\mathbb{R}^{n+m} \times \mathbb{R}^m$.

Definition 11.1. A Schwartz test function $\varphi(x,y,z)$ defined on $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ is said to be a test function of product type on $\mathbb{R}^{n+m} \times \mathbb{R}^m$ if

$$\int_{\mathbb{R}^n \times \mathbb{R}^m} \varphi(x,y,z) x^\alpha y^\beta dxdy = \int_{\mathbb{R}^m} \varphi(x,y,z) z^\gamma dz = 0$$

for all multiindices α, β, and γ of nonnegative integers.

We denote $\mathcal{S}_\infty(\mathbb{R}^{n+m} \times \mathbb{R}^m)$ the set of all product Schwartz test functions.

Definition 11.2. A function $\varphi^\flat(x,y)$ defined on $\mathbb{R}^n \times \mathbb{R}^m$ is said to be a text function in $\mathcal{S}_F(\mathbb{R}^n \times \mathbb{R}^m)$ if there exists a function $\varphi \in \mathcal{S}_\infty(\mathbb{R}^{n+m} \times \mathbb{R}^m)$ such that

$$\varphi^\flat(x,y) = \int_{\mathbb{R}^m} \varphi(x, y-z, z)dz. \tag{11.1}$$

If $\varphi^\flat \in \mathcal{S}_F(\mathbb{R}^n \times \mathbb{R}^m)$, then the norm of $\varphi^\flat$ is defined by

$$\|\varphi^\flat\|_{\mathcal{S}_F(\mathbb{R}^n \times \mathbb{R}^m)} = \inf\big\{\|\varphi\|_{\mathcal{S}_\infty(\mathbb{R}^{n+m} \times \mathbb{R}^m)} : \ \forall \text{ representations of } \varphi \text{ in (11.1)}\big\}.$$

We denote by $(\mathcal{S}_F(\mathbb{R}^n \times \mathbb{R}^m))'$ the dual space of $\mathcal{S}_F(\mathbb{R}^n \times \mathbb{R}^m)$.

Since the functions $\psi_{j,k}$ constructed above belong to $\mathcal{S}_F(\mathbb{R}^n \times \mathbb{R}^m)$, so the Littlewood-Paley square function g_F can be defined for all distributions in $(\mathcal{S}_F(\mathbb{R}^n \times \mathbb{R}^m))'$.

Definition 11.3.

$$H^p_F = \big\{f \in (\mathcal{S}_F(\mathbb{R}^n \times \mathbb{R}^m))' : \ g_F(f) \in L^p(\mathbb{R}^{n+m})\big\}.$$

The norm of f is defined by

$$\|f\|_{H^p_F} = \|g_F(f)\|_{L^p}.$$

A question. Does the definition of H^p_F depend on the choice of functions $\psi_{j,k}$? Moreover, one needs to discretize the norm of H^p_F. We have

(1) to prove a Calderón reproducing formula,
(2) to prove a Plancherel-Polya inequality.

However, we are dealing with non-convolution operators here. So we need to develop a local theory for Hardy spaces. Using the $\varphi^\flat_{j,k}$ which was constructed above, one has

$$f(x,y) = \sum_j \sum_k \varphi^\flat_{j,k} * \varphi^\flat_{j,k} * f(x,y).$$

The series converges in the norm of $\mathcal{S}_F$ and in the dual space $(\mathcal{S}_F)'$.

We have the following Plancherel-Polya inequality.

Theorem 11.1. *Suppose* ψ^1, $\phi^1 \in \mathcal{C}^\infty_0(\mathbb{R}^{n+m})$ *and* ψ^2, $\phi^2 \in \mathcal{C}^\infty_0(\mathbb{R}^m)$ *and satisfy the conditions*

$$\sum_j \Big|\widehat{\psi}^1\big(\frac{\xi_1}{2^j}, \frac{\xi_2}{2^j}\big)\Big|^2 = 1 \quad \forall\ (\xi_1,\xi_2) \in \mathbb{R}^{n+m} \setminus \{(0,0)\},$$
$$\sum_k \Big|\widehat{\psi}^2\big(\frac{\zeta}{2^k}\big)\Big|^2 = 1 \quad \forall\ \zeta \in \mathbb{R}^m \setminus \{0\}$$
$$\sum_j \Big|\widehat{\phi}^1\big(\frac{\xi_1}{2^j}, \frac{\xi_2}{2^j}\big)\Big|^2 = 1 \quad \forall\ (\xi_1,\xi_2) \in \mathbb{R}^{n+m} \setminus \{(0,0)\},$$
$$\sum_k \Big|\widehat{\phi}^2\big(\frac{\zeta}{2^k}\big)\Big|^2 = 1 \quad \forall\ \zeta \in \mathbb{R}^m \setminus \{0\}.$$

Define

$$\psi^\flat(x,y) = \int_{\mathbb{R}^m} \psi^1(x, y-z)\psi^2(z)dz,$$
$$\phi^\flat(x,y) = \int_{\mathbb{R}^m} \phi^1(x, y-z)\phi^2(z)dz.$$

Then for $f \in (\mathcal{S}_F)'$ *and* $0 < p < \infty$,

$$\Big\|\Big\{\sum_j \sum_k \sum_J \sum_I \sup_{u\in I, v\in J} \big|\psi^\flat_{j,k} * f(u,v)\big|^2 \chi_I\chi_J\Big\}^{\frac12}\Big\|_{L^p}$$
$$\approx \Big\|\Big\{\sum_j \sum_k \sum_J \sum_I \inf_{u\in I, v\in J} \big|\phi^\flat_{j,k} * f(u,v)\big|^2 \chi_I\chi_J\Big\}^{\frac12}\Big\|_{L^p}$$

where

$$\psi^\flat_{j,k}(x,y) = \big(\tau_{2^j,2^k}\psi^\flat\big)(x,y),$$
$$\phi^\flat_{j,k}(x,y) = \big(\tau_{2^j,2^k}\phi^\flat\big)(x,y),$$

I and J are cubes in $\mathbb{R}^n$ and $\mathbb{R}^m$ with side length 2^{-j-N} and $2^{-k-N}+2^{-j-N}$ for large integer N. Here χ_I and χ_J are the characteristic functions of I and J, respectively.

As a consequence of the Plancherel-Polya inequality, it is easy to see that the Hardy space H^p_F is well defined.

Using the Plancherel-Polya inequality, we can prove the boundedness of flag singular integrals on H^p_F.

Theorem 11.2. *Suppose that T is a flag singular integral with the kernel $K(x,y)$ satisfying the same conditions as in Theorem 10.2. Then T is bounded on h^p_F for $p_m < p \le 1$. Here p_m is an index depending on the type m. Moreover, there exists a constant C such that for $0<p\le 1$,*

$$\|T(f)\|_{h^p_F} \le C\|f\|_{h^p_F}.$$

Theorem 11.3. *There is an operator $\mathbf{K}$ so that, when it is applied to smooth functions with compact support, there is the identity*

$$\mathbf{K}\Box_b = \Box_b\mathbf{K} = (\Box_1+\Box_2)\mathbf{K} = I.$$

Moreover,

1. *The four operators*
$$Z_1\bar{Z}_1\mathbf{K} = \Box_1\mathbf{K},\ \ \bar{Z}_2 Z_2\mathbf{K} = \Box_2\mathbf{K},$$
and
$$\bar{Z}_1\bar{Z}_1\mathbf{K},\ \ Z_2Z_2\mathbf{K}$$
are bounded on $L^p_k(\partial\Omega)$ for $1<p<\infty$, $k\in\mathbb{Z}_+$ and $h^p_F(\partial\Omega)$ for $p_m<p\le 1$. Here p_m is an index depending on the type m of the domain;
2. *Let B_1 and B_2 be bounded functions on $\partial\Omega$ and suppose there are constants C_1, C_2 so that*
$$\lambda_1(z_1)B_1(z,t) \le C_1\lambda_2(z_2);$$
$$\lambda_2(z_2)B_2(z,t) \le C_2\lambda_1(z_1).$$
Then the two operators
$$B_1\bar{Z}_1Z_1\mathbf{K} = B_1\bar{\Box}_1\mathbf{K},\ \ B_2Z_2\bar{Z}_2\mathbf{K} = B_2\bar{\Box}_2\mathbf{K}$$
are bounded on $L^p_k(\partial\Omega)$ for $1<p<\infty$, $k\in\mathbb{Z}_+$ and $h^p_F(\partial\Omega)$ for $p_m<p\le 1$. Here $\lambda_1(z_1) = |z_1|^{2(m_1-1)}$, $\lambda_2(z_2) = |z_2|^{2(m_2-1)}$ are the eigenvalues of the Levi form;
3. *Let B_1 and B_2 be bounded functions on $\partial\Omega$ and suppose there are constants C_1, C_2 so that*
$$B_1(z,t)\le C_1\lambda_2(z_2);\quad B_2(z,t)\le C_2\lambda_1(z_1).$$
Then the two operators
$$B_1Z_1Z_1\mathbf{K},\ \ B_2\bar{Z}_2\bar{Z}_2\mathbf{K}$$
are bounded on $L^p_k(\partial\Omega)$ for $1<p<\infty$, $k\in\mathbb{Z}_+$ and $h^p_F(\partial\Omega)$ for $p_m<p\le 1$;

4. **K** *maps* $L^\infty(\partial\Omega)$ *to the isotropic Lipschitz space* $\Lambda_\alpha(\partial\Omega)$*, where*

$$\alpha = \min\Big\{\frac{1}{m_1}, \frac{1}{m_2}\Big\}.$$

This theorem shows the fundamental solution $\mathbf{K} = (\mathbf{K}_1, \mathbf{K_2})$ of $\Box_b$ will not gain two in all horizontal directions. However, the Kohn solution

$$u = \bar\partial_b^* \bar{\mathbf{K}}(\phi) = -Z_1\bar{\mathbf{K}}_1(\phi_1) - Z_2\bar{\mathbf{K}}_2(\phi_2) = -Z_1\bar{\mathbf{K}}_1(\phi_1) - Z_2\mathbf{K}_1(\phi_2)$$

for the inhomogeneous tangential Cauchy-Riemann equation $\bar\partial_b u = \phi = \phi_1\bar\omega_1 + \phi_2\bar\omega_2$ is well defined in L^p. But, it will not gain one in all horizontal directions.

12. Further results

1. Using the method we mentioned above, construct the "fundamental solution" **N** for the $\bar\partial$-Neumann problem.
2. Obtain possible "sharp" L^p and Lipschitz estimates for the Neumann operator **N**.
3. Obtain the Heikin-Skoda estimate:

$$\big\|\bar\partial^*\mathbf{N}(f)\big\|_{L^1(\Omega)} \le C\Big(\|f\|_{L^1(\Omega)} + \sum_{j=1}^n \Big\|\frac{\mu_j}{\rho} f_j\Big\|_{L^1(\Omega)}\Big).$$

Then use this estimate to characterize zero sets of functions in Nevanlinna class:

$$N^+(\Omega) = \Big\{f \in \mathcal{H}(\Omega) :\ \sup_{\varepsilon>0}\int_{\rho=\varepsilon} \log^+|f(z)|d\sigma(z) < \infty\Big\}.$$

Acknowledgment

Part of this article is based on a series of lectures presented by the author at the Matematisk Institutt, Universitetet i Bergen. The author would like to express his heartfelt gratitude to Professor Alexander Vasiliev and Professor Irina Markina for the invitation and for their help in providing very pleasant working conditions. Thanks are also due to all the colleagues at the Mathematics Department of University of Bergen for the warm hospitality during his visit to Norway. The author would like to thank his teachers, Professor E.M. Stein and Professor P.C. Greiner for their teaching, friendship and encouragement during the past 25 years. The author would also like to thank Professor Y.S. Han for many stimulating conversions on H^p theory for flag singular integral operators.

References

[1] R. Beals and P.C. Greiner, *Calculus on Heisenberg manifolds*, Ann. Math. Studies #**119**, Princeton University Press, Princeton, New Jersey, 1988.

[2] R. Beals, B. Gaveau and P.C. Greiner, On a geometric formula for the fundamental solution of subelliptic Laplacians, *Math. Nachr.*, **181**(1996), 81–163.

[3] R. Beals, B. Gaveau and P.C. Greiner, Hamilton-Jacobi theory and the heat kernel on Heisenberg groups, *J. Math. Pures Appl.*, **79** #7(2000), 633–689.

[4] C. Berenstein, D.C. Chang and J. Tie, *Laguerre Calculus on the Heisenberg Group*, AMS/IP series in Advanced Mathematics, #**22**, American Mathematics Society and International Press, Cambridge, Massachusetts, ISBN 0-8218-2761-8, (2001).

[5] O. Calin, D.C. Chang and P.C. Greiner, On a step $2(k+1)$ subRiemannian manifold, *J. Geom. Anal.*, **12** #1(2004), 1–18.

[6] O. Calin, D.C. Chang and P.C. Greiner, Real and complex Hamiltonian mechanics on some subRiemannian manifolds, *Asian J. Math.*, **18** #1(2004), 137–160.

[7] O. Calin, D.C. Chang, P.C. Greiner and Y. Kannai, On the geometry induced by a Grusin operator, *Proceedings of International Conference on Complex Analysis & Dynamical Systems*, (ed. by L. Karp and L. Zalcman), Contemporary Math., **382**(2005), Amer. Math. Soc., 89–111.

[8] O. Calin, D.C. Chang and P.C. Greiner: *Geometric Analysis on the Heisenberg Group and Its Generalizations*, AMS/IP series in Advanced Mathematics, #**40**, American Mathematics Society and International Press, Cambridge, Massachusetts, ISBN-10: 0-8218-4319-2, (2007).

[9] D.C. Chang, On L^p and Hölder estimates for the $\bar{\partial}$-Neumann problem on strongly pseudoconvex domains, *Math. Ann.*, **282**(1988), 267–297.

[10] D.C. Chang and C. Fefferman, On L^p estimates of the Cauchy-Riemann equation, *"Harmonic Analysis in China"*, (ed. by C.C. Yang), 1–21, Kluwer Academic Publishers, (1995).

[11] D.C. Chang and P.C. Greiner, Analysis and geometry on Heisenberg groups, *"Proceedings of Second International Congress of Chinese Mathematicians"*, New Studies in Advanced Mathematics, International Press, (Ed. by C. Lin and S.T. Yau), 379–405, (2004).

[12] D.C. Chang and P.C. Greiner, Explicit kernel for Kohn Laplacian on a family of pseudoconvex hypersurfaces, reprint, (2007).

[13] D.C. Chang and J. Tie, Estimates for powers of sub-Laplacian on the non-isotropic Heisenberg group, *J. Geom. Anal.*, **10**(2000), 653–678.

[14] D.C. Chang, A. Nagel and E.M. Stein, Estimates for the $\overline{\partial}$-Neumann problem in pseudoconvex domains of finite type in $\mathbf{C^2}$, *Acta Math.*, **169**(1992), 153–227.

[15] S.S. Chern and J.K. Moser, Real hypersurfaces in complex manifolds, *Acta Math.*, **133**(1974), 219–271.

[16] W.L. Chow, Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung, *Math. Ann.*, **117**(1939), 98–105.

[17] M. Christ, Regularity properties of the $\bar{\partial}_b$ equation on weakly pseudoconvex CR manifolds of dimension 3, *J. Amer. Math. Soc.*, **1**(1988), 587–646.

[18] M. Derridj, Regularité pour $\bar{\partial}$ dans quelques domains faiblement pseudoconvexes, *J. Differential Geom.*, **13**(1978), 559–576.

[19] C.L. Fefferman and J.J. Kohn, Hölder estimates on domains of complex dimension two and on three-dimensional CR manifolds, *Adv. Math.*, **69**(1988), 223–303.

[20] G.B. Folland and J.J. Kohn, *The Neumann problem for the Cauchy-Riemann complex*, Ann. of Math. Studies, #**75**, Princeton University Press, Princeton, New Jersey, (1972).

[21] G.B. Folland and E.M. Stein, Estimates for the $\bar{\partial}_b$ complex and analysis on the Heisenberg group, *Comm. Pure Appl. Math.*, **27**(1974), 429–522.

[22] P.C. Greiner, A fundamental solution for a non-ellpitic partial differential operator, *Can. J. Math.*, **31**(1979), 1107–1120.

[23] P.C. Greiner and E.M. Stein, *Estimates for the $\bar{\partial}$-Neumann problem*, Math. Notes, #**19**, Princeton University Press, Princeton, New Jersey, (1977).

[24] Y.S. Han, Fourier analysis associated with certain multi-parameter structures, preprint, (2007).

[25] J.J. Kohn, Harmonic integrals on strongly pseudo-convex manifolds. I. *Ann. of Math.*, **78**(1963), 112–148.

[26] J.J. Kohn, Harmonic integrals on strongly pseudo-convex manifolds. II. *Ann. of Math.*, **79**(1964), 450–472.

[27] M. Machedon, Estimates for the parametrix of the Kohn Laplacian on certain domains, *Invent. Math*, **91**(1988), 339–364.

[28] A. Nagel and E.M. Stein, The $\bar{\partial}_b$-complex on decoupled boundary in $\mathbb{C}^n$, *Ann. of Math.*, **164**(2006), 649–713.

[29] A. Nagel, F. Ricci and E.M. Stein, Singular integrals with flag kernels and analysis on quadratic CR manifolds, *J. Func. Anal.*, **181**(2001), 29–118.

[30] A. Nagel, J.P. Rosay, E.M. Stein and S. Wainger, Estimates for the Bergman and Szegö kernel in $\mathbb{C}^2$, *Ann. of Math.*, **129**(1989), 113–149.

[31] D.H. Phong, On L^p and Hölder estimates for the $\bar{\partial}$ equations on strongly pseudo-convex domains, Ph.D. Dissertation, Princeton University, (1977).

[32] D.H. Phong, On integral representations of the $\bar{\partial}$-Neumann operator, *Proc. Natl. Acad. Sci. USA*, **76**(1979), 1554–1558.

[33] D.H. Phong and E.M. Stein, Hilbert integrals, singular integrals, and Radon transforms. I. *Acta Math.*, **157**(1986), 99–157.

[34] D.H. Phong and E.M. Stein, Phong, Hilbert integrals, singular integrals, and Radon transforms. II. *Invent. Math.*, **86**(1986), 75–113.

[35] E.M. Stein, *Harmonic Analysis – Real Variable Methods, Orthogonality, and Oscillatory Integrals*, Princeton University Press, Princeton, New Jersey, (1993).

Der-Chen Chang
Department of Mathematics
Georgetown University
Washington D.C. 20057, USA

and

National Center for Theoretical Sciences
Mathematics Division
National Tsing Hua University
Hsinchu, Taiwan 30013, ROC
e-mail: `chang@georgetown.edu`

Analysis and Mathematical Physics
Trends in Mathematics, 107–114
© 2009 Birkhäuser Verlag Basel/Switzerland

Numerical Scheme for Laplacian Growth Models Based on the Helmholtz–Kirchhoff Method

A.S. Demidov and J.-P. Lohéac

Abstract. The Helmholtz–Kirchhoff method is an efficient tool for analysing bi-dimensional problems in fluid mechanics. It especially allows to transform a free boundary problem in a fixed boundary problem by introducing a convenient parametrization of the free boundary.

In this paper, it will be shown how it also leads to build numerical schemes. The case of Hele-Shaw flows will be especially studied.

Mathematics Subject Classification (2000). 76D27, 45K05, 65E05.

Keywords. Hele-Shaw flows, free boundary problems, integro-differential equations.

1. Hele-Shaw problems

The core of the Hele-Shaw device is a blob of fluid moving between two glass plates. We are interested in the motion of such a blob when fluid is injected or sucked between the plates. The air which surrounds the blob is another fluid with negligible viscosity. In 1934, L.S. Leibenson gave a simplified mathematical model for this problem when the source of the flow is punctual.

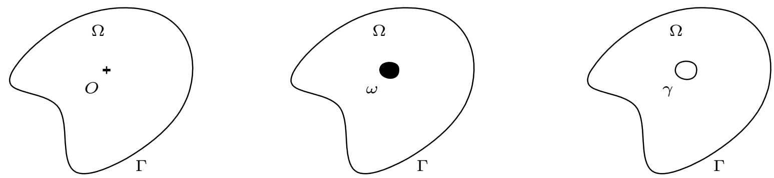
We here consider a class of free boundary problems derived from this so-called Stokes-Leibenson problem.

Let $\Omega_0 \subset \mathbb{R}^2$ be a bounded simply connected domain such that its boundary Γ_0 is smooth enough. This domain will be deformed according to the following law: at time t, we obtain a domain $\Omega_t = \Omega$ of boundary $\Gamma_t = \Gamma$ such that the normal velocity of each point $\mathbf{s} \in \Gamma$ is given by the following kinetic condition,

$$\dot{\mathbf{s}}.\boldsymbol{\nu} = \partial_\nu u \,, \tag{1.1}$$

where u is the solution of some Laplace problem.

Above we denote by $\boldsymbol{\nu}$ the normal unit outwards vector at $\mathbf{s} \in \Gamma$ and by $\partial_\nu u$ the normal derivative of u at this point. We will consider three cases (see Fig. 1).



(1.1) (1.2) (1.3)

FIGURE 1. The three considered cases: (1.1) punctual source, (1.2) distributed internal source, (1.3) distributed boundary source

1.1. Punctual source

Here u is the solution of the following problem,

$$\begin{cases} \Delta u = q\,\delta\,, & \text{in } \Omega\,, \\ u = 0\,, & \text{on } \Gamma\,, \end{cases} \tag{1.2}$$

where δ is the Dirac distribution at some point $O \in \Omega_0$ and q represents the power of the source.

Observe that by an elementary computation, one can obtain the increasing rate of the fluid domain Ω:

$$\int_\Gamma \dot{\mathbf{s}}.\boldsymbol{\nu}\,d\sigma = \int_\Gamma \partial_\nu u\,d\sigma = q\,.$$

1.2. Distributed internal source

In this case, u satisfies

$$\begin{cases} \Delta u = f\,, & \text{in } \Omega\,, \\ u = 0\,, & \text{on } \Gamma\,. \end{cases} \tag{1.3}$$

Here, the support of the right-hand side f is contained in an open set ω such that $\overline{\omega} \subset \Omega_0$.

As well as above, the increasing rate of the fluid domain Ω can be computed

$$\int_\Gamma \dot{\mathbf{s}}.\boldsymbol{\nu}\,d\sigma = \int_\Gamma \partial_\nu u\,d\sigma = \int_\omega f\,d\mathbf{x}\,.$$

1.3. Distributed boundary source

Here, u satisfies

$$\begin{cases} \Delta u = 0\,, & \text{in } \Omega\,, \\ -\partial_\nu u = g\,, & \text{on } \gamma\,, \\ u = 0\,, & \text{on } \Gamma\,, \end{cases} \tag{1.4}$$

and we can write

$$\int_\Gamma \dot{\mathbf{s}}.\boldsymbol{\nu}\,d\sigma = \int_\Gamma \partial_\nu u\,d\sigma = \int_\gamma g\,d\sigma\,.$$

2. Helmholtz–Kirchhoff method

In each case, we consider u as an harmonic function on some subdomain Ω' of Ω and we introduce its harmonically conjugate function v. At every point of Ω', level curves of u and v are orthogonal.

Let us set: $z = x + \imath y$ and: $w = u + \imath v$. the function $z \longmapsto w$ is analytic and univalent from Ω' onto a fixed domain $\Pi \subset \mathbb{C}$.

In Fig. 2, 3, 4, we construct a way composed of level curves of u and v in Ω' and its image in Π.

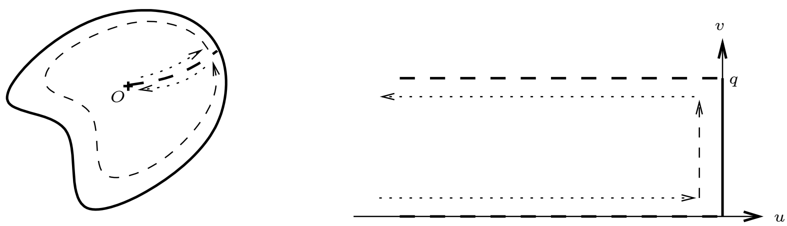


FIGURE 2. The case of a punctual source: $\Omega' = \Omega \setminus \{O\}$ and $\Pi = (-\infty, 0) \times (0, q)$, when $q > 0$.

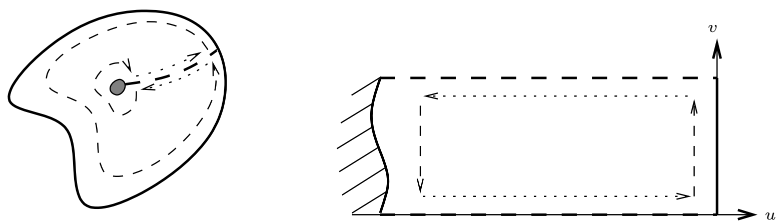


FIGURE 3. The case of an internal distributed source: $\Omega' = \Omega \setminus \overline{\omega}$ and $\Pi \subset (-\infty, 0) \times (0, \int_\omega f\, d\mathbf{x})$, when $\int_\omega f\, d\mathbf{x} > 0$.

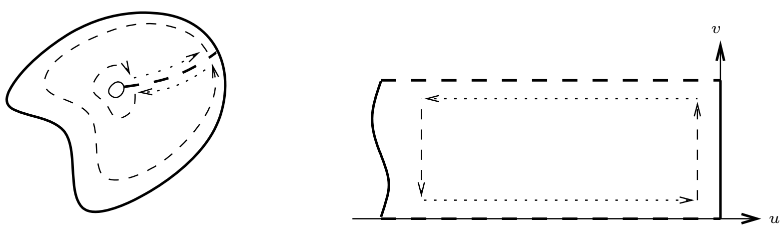


FIGURE 4. The case of a boundary distributed source: $\Omega' = \Omega$ and $\Pi \subset (-\infty, 0) \times (0, \int_\gamma g\, d\sigma)$, when $\int_\gamma g\, d\sigma > 0$.

Now, in order to simplify, we assume that

- there is a constant axis of symmetry,
- the increasing rate of the fluid domain is 2.

Then we can only consider the "upper" part of the fluid domain and we can choose v so that

$$v(t,x,0) = 0\,,\ \text{ if } x > 0\,, \qquad v(t,x,0) = 1\,,\ \text{ if } x < 0\,.$$

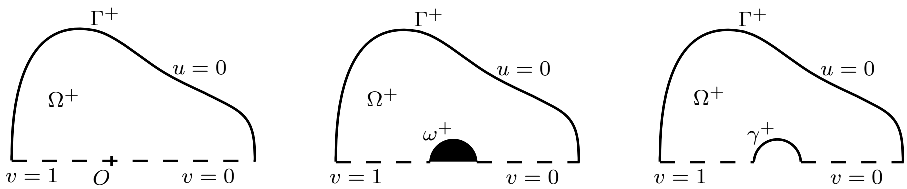


FIGURE 5. The upper part of the domain in the three above cases.

The corresponding domain Π^+ is a simply connected subset of $(-\infty, 0) \times (0, 1)$ and the boundary of Π^+ contains $\{0\} \times (0, 1)$, which corresponds to Γ^+.
Since Γ^+ is a part of a level curve of u, it can be parametrized by the value of v at each point. In Fig. 6, we show an example of computed level curves of u and v (this has been obtained by using a finite elements method).

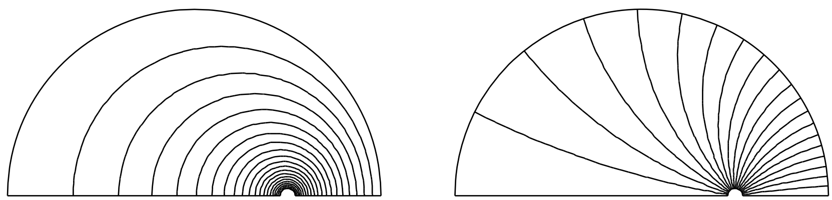


FIGURE 6. Case of an uncentered source in a circular domain: level curves of u (left) and v (right).

Then we can define the **Helmholtz–Kirchhoff** function [1]:

$$A + \imath B = \ln \frac{\partial z}{\partial w}\,,$$

where A and B are harmonically conjugate real functions. We can write

$$(A + \imath B)(t, w) = a_0(t) + \pi w + \sum_{k=1}^{\infty} \beta_k(t)\, \exp(k\pi w)\,,$$

and define

$$a(t,\eta) = A(t,0,\eta)\,, \qquad b(t,\eta) = B(t,0,\eta)\,.$$

This leads to a parametrization of Γ^+,

$$\mathbf{s}(t,\eta) = \mathbf{s}_0(t) + \int_0^\eta \exp(a(t,v))\,\boldsymbol{\tau}(t,v)\,dv\,, \qquad \text{with } 0 \le \eta \le 1\,,$$

where $\boldsymbol{\tau}$ is the tangential unit vector at some point of the moving upper boundary Γ^+ (see Fig. 7). Since a and b are linked by a Hilbert-type transform, the main unknown of our problem becomes the function b which represents the orientation of normal unit vector $\boldsymbol{\nu}$.

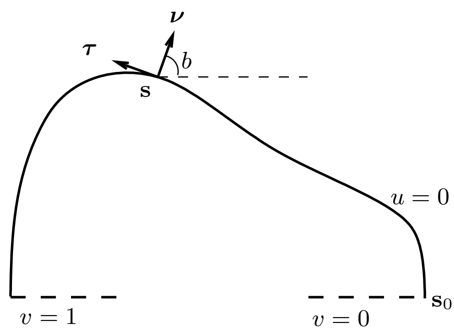


FIGURE 7. Parametrization of Γ^+, unit vectors $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$.

If Γ^+ is regular enough, then we can write, for $t \ge 0$ and $\eta \in [0,1]$,

$$\left(e^a\,\partial_t b\right)(t,\eta) = \left(e^{-a}\,\partial_\eta a\right)(t,\eta) + \partial_\eta b(t,\eta)\int_0^\eta \left(e^a\,\partial_t a - e^{-a}\,\partial_\eta b\right)(t,v)\,dv\,.$$

In other words, we can say that function β such that $\beta(t,\eta) = b(t,\eta) - \pi\eta$, satisfies an integro-differential equation

$$2(t+t_0)\,[\mathbb{K}(\beta)\,\partial_t\beta](t,\eta) = [\mathbb{F}(\beta)](t,\eta)\,, \tag{2.1}$$

where $t_0 > 0$ depends on the measure of the initial domain Ω_0.

This especially leads to prove existence and uniqueness results and qualitative results in some cases of Hele-Shaw flows with a punctual source [2, 3].

3. Numerical model

Our main idea is to restrict above computation to the case of some class of polygonal domains [4].

Hence let us consider the class $\mathcal{P}_m$ ($m \in \mathbb{N}^*$) of simply connected polygonal domains such that

- each of these domains is symmetric with respect to the x-axis and contains the source region,
- its boundary, which we call the "quasi-contour", is a polygonal line with $2m+1$ vertices,

- one vertex belongs to the positive part of the x-axis.

We will compute the behavior of such domains by applying some discrete law inspired by the law of motion of smooth curves in the classical Stokes-Leibenson problem.

For instance, in the case of a punctual source, we obtain the approximated problem:

Let Ω_0^m in $\mathcal{P}_m$ be some "approximation" of Ω_0 and $\Gamma_0^m = \partial\Omega_0^m$.

For $t > 0$, find $\Omega^m = \Omega_t^m$ in $\mathcal{P}_m$ and its boundary $\Gamma^m = \partial\Omega^m$ such that every vertex $\mathbf{p} \in \Gamma^m$ verifies the punctual kinetic law

$$\dot{\mathbf{p}}.\boldsymbol{\nu}(\mathbf{p}) = \int_{\Gamma^m} d_{\mathbf{p}} |\nabla u| \, d\gamma \,, \tag{3.1}$$

with

$$\begin{cases} \Delta u = 2\delta \,, & \text{in } \Omega^m \,, \\ u = 0 \,, & \text{on } \Gamma^m \,. \end{cases} \tag{3.2}$$

As above, we introduce v harmonically conjugate function of u and we use it to get a convenient parametrization of Γ^{m+}.

In Fig. 8, we show an example of computed level curves of u and v for some polygonal domain.

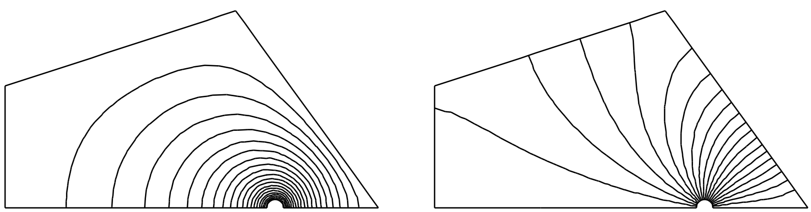


FIGURE 8. Case of an uncentered source in a pentagonal domain ($m = 2$): level curves of u (left) and v (right).

Here, the unknown function b is piecewise constant and can be represented by the finite sequence $\mathbf{N}$ of the orientations of the m normal unit vectors with respect to each edge of Γ^{m+}.

An explicit computation of terms of integro-differential equation (2.1) leads to

$$2(t + t_0) \, [Q(\mathbf{N}) \, \dot{\mathbf{N}}](t) = [P(\mathbf{N})](t) \,, \tag{3.3}$$

where $Q(\mathbf{N})$ is a $m \times m$-matrix and $P(\mathbf{N})$ belongs to $\mathbb{R}^m$.

Then, using initial data, we can perform an explicit Euler scheme for (3.3).

We only present here two examples of computed evolution of quasi-contours: in Fig. 9, the computed evolution of a pentagonal quasi-contour ($m = 2$) and in Fig. 10, the computed evolution of a polygonal quasi-contour ($m = 4$).

In addition, we have got for Hele-Shaw flows with a punctual source the existence of an attractive manifold in the space of quasi-contours. In the case of a

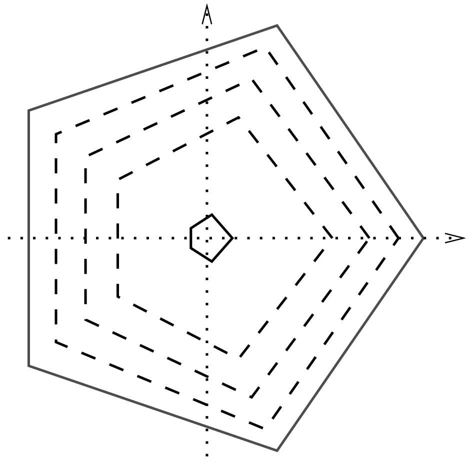


FIGURE 9. An increasing pentagonal quasi-contour: initial shape (center) and shapes at time 1, 2, 3 (dot-lines), 4.

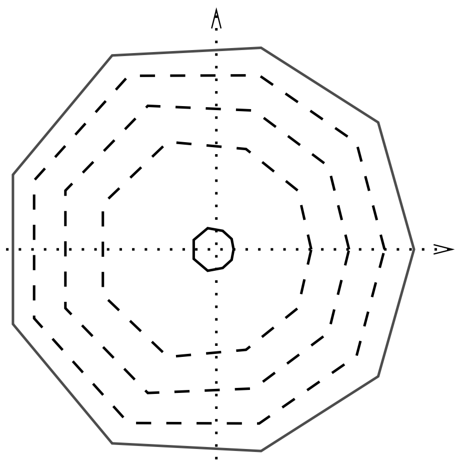


FIGURE 10. An increasing polygonal quasi-contour ($m = 4$): initial shape (center) and shapes at time 1, 2, 3 (dot-lines), 4.

sink, we have to reverse the time-scale. This manifold becomes repulsive and this can explain some fingering phenomenons [4].

For the source case, it has been proved in [2, 3] that the limit of the contour when time tends to ∞ is a circle centered at the source point. Here we obtain in our numerical experiments a similar property: for a fixed m, the quasi-contour tends to a regular polygonal contour centered at the source point when time increases.

In the other hand, the nature of our discrete model does not allow to confirm the property proved in [6] showing that the free boundary becomes instantaneously smooth.

References

[1] A.S. Demidov, *Some Applications of the Helmholtz–Kirchhoff Method.* (*Equilibrium Plasma in Tokamaks, Hele-Shaw Flow, and High-Frequency Asymptotics*). Russian J. Math. Ph. **V. 7, No. 2**, 166–186 (2000).

[2] A.S. Demidov, *Evolution of a perturbation of a circle in a problem for Hele-Shaw flows.* Journ. of Math. Sciences **123, No. 8**, 4381–4403 (2004).

[3] A.S. Demidov, *Evolution of a perturbation of a circle in a problem for Hele-Shaw flows. Part II.* Journ. of Math. Sciences **139, No. 6**, 7064–7078 (2006).

[4] A.S. Demidov, J.-P. Lohéac, *The Stokes–Leibenson Problem for Hele-Shaw Flows.* Patterns and Waves (Eds. A. Abramian, S. Vakulenko, V. Volpert) Saint-Petersburg, 103–124 (2003).

[5] A.S. Demidov, J.-P. Lohéac, *Simulation numérique de phénomènes de digitation dans des écoulements de Hele-Shaw* (in progress).

[6] J.R. King, J.R. Lacey, A.A. Vazquez, *Persistence of corners in free boundaries in Hele-Shaw flow.* Europ. J. Appl. Math. **6** (5), 455–490 (2006).

A.S. Demidov
Moscow State University,
Department of Mechanics and Mathematics,
Leninski gory
Moscow, 119992, Russia
e-mail: `alexandre.demidov@mtu-net.ru`

J.-P. Lohéac
Université de Lyon
École centrale de Lyon
Institut Camille-Jordan
36 avenue Guy-de-Collongue
69134 Écully Cedex, France
e-mail: `Jean-Pierre.Loheac@ec-lyon.fr`

Analysis and Mathematical Physics
Trends in Mathematics, 115–129
© 2009 Birkhäuser Verlag Basel/Switzerland

Gravitational Lensing by Elliptical Galaxies, and the Schwarz Function

C.D. Fassnacht, C.R. Keeton and D. Khavinson

Abstract. We discuss gravitational lensing by elliptical galaxies with some particular mass distributions. Using simple techniques from the theory of quadrature domains and the Schwarz function (cf. [18]) we show that when the mass density is constant on confocal ellipses, the total number of lensed images of a point source cannot exceed 5 (4 bright images and 1 dim image). Also, using the Dive–Nikliborc converse of the celebrated Newton's theorem concerning the potentials of ellipsoids, we show that "Einstein rings" must always be either circles (in the absence of a tidal shear), or ellipses.

1. Basics of gravitational lensing

Imagine n co-planar point-masses (e.g., condensed galaxies, stars, black holes) that lie in one plane, the lens plane. Consider a point light source S (a star, a quasar, etc.) in a plane (a source plane) parallel to the lens plane and perpendicular to the line of sight from the observer, so that the lens plane is between the observer and the light source. Due to deflection of light by masses multiple images $S_1, S_2, \ldots$ of the source may form (cf. Fig. 1). Fig. 2 and Fig. 3 illustrate some further aspects of the lensing phenomenon.

2. Lens equation

In this section we are still assuming that our lens consists of n point masses. Suppose that the light source is located in the position w (a complex number) in the source plane. Then, the lensed image is located at z in the source plane while the masses of the lens L are located at the positions z_j, $j = 1, \ldots, n$ in the lens

The third author gratefully acknowledges partial support from the National Science Foundation under the grant DMS-0701873. The first and third authors are also grateful to Kavli Institute of Theoretical Physics for the partial support of their visit there in 10/2006 under the NSF grant PHY05-51164.

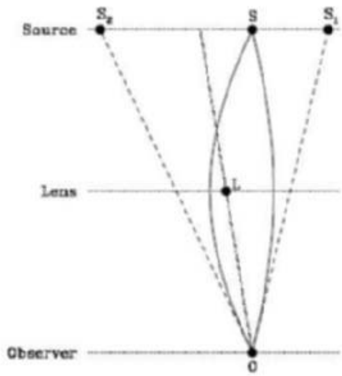


FIGURE 1. The lens L located between source S and observer O produces two images S_1, S_2 of the source S.

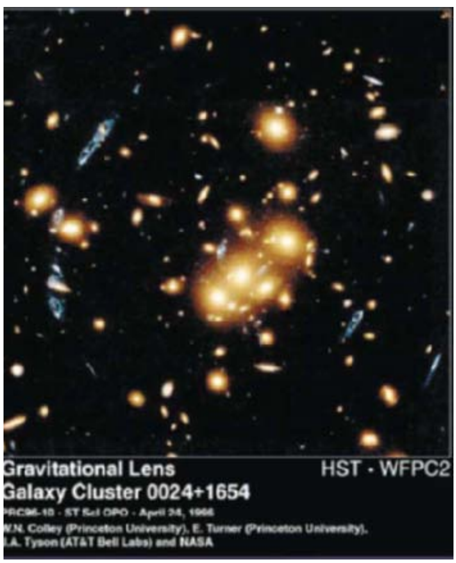


FIGURE 2. Lensing of a galaxy by a cluster of galaxies; the blue spots are all images of a single galaxy located behind the huge cluster of galaxies. (Credit: NASA, W.N. Colley (Princeton), E. Turner (Princeton) and J.A. Tyson (AT&T and Bell Labs).)

plane. The following simple equation, obtained by combining Fermat's Principle of Geometric Optics together with basic equations of General Relativity, connects then the positions of the lensed images, the source and the positions of the masses

FIGURE 3. The bluish bright spots are the lensed images of a quasar (i.e., a quasi-stellar object) behind a bright galaxy in the center. There are 5 images (4 bright + 1 dim), but one cannot really see the dim image in this figure. (Credit: ESA, NASA, K. Sharon (Tel Aviv University) and E. Ofek (Caltech).)

which cause the lensing effect

$$w = z - \sum_{1}^{n} \frac{\sigma_j}{\bar{z} - \bar{z}_j}, \tag{2.1}$$

where σ_j are some physical (real) constants. For more details on the derivation and history of the lensing equation (2.1) we refer the reader to [19], [12], [14], [21]. Sometimes, to include the effect caused by an extra ("tidal") gravitational pull by an object (such as a galaxy) far away from the lens masses, the right-hand side of (2.1) includes an extra linear term $\gamma\bar{z}$, thus becoming

$$w = z - \sum_{1}^{n} \frac{\sigma_j}{\bar{z} - \bar{z}_j} - \gamma\bar{z}, \tag{2.2}$$

where γ is a real constant. The right-hand side of (2.1) or (2.2) is called the lensing map. The number of solutions z of (2.1) (or (2.2)) is precisely the number of images of the source w generated by the lens L. Letting $r(z) = \sum_{1}^{n} \frac{\sigma_j}{z - z_j} + \gamma z + \bar{w}$, the lens equations (2.1) and (2.2) become

$$z - \overline{r(z)} = 0, \tag{2.3}$$

where $r(z)$ is a rational function with poles at z_j, $j = 1, \ldots, n$ and infinity if $\gamma \neq 0$.

3. Historical remarks

The first calculations of the deflection angle by a point mass lens, based on Newton's corpuscular theory of light and the Law of Gravity, go back to H. Cavendish and J. Michel (1784), and P. Laplace (1796) – cf. [20]. J. Soldner (1804) – cf. [21] is usually credited with the first published calculations of the deflection angle and, accordingly, with that of the lensing effect. Since Soldner's calculations were based on Newtonian mechanics they were off by a factor of 2. A. Einstein is usually given credit for calculating the lensing effect in the case of $n = 1$ (one mass lens) around 1933. Yet, some evidence has surfaced recently that he did some of these calculations earlier, around 1912 – cf. [17] and references therein. The recent outburst of activity in the area of lensing is often attributed to dramatic improvements of optics technology that make it possible to check many calculations and predictions by direct visualization.

H. Witt [24] showed by a direct calculation that for $n > 1$ the maximum number of observed images is $\leq n^2 + 1$. Note that this estimate can also be derived from the well-known Bezout theorem in algebraic geometry (cf. [3,8,9,22]). In [11] S. Mao and A.O. Petters and H.J. Witt showed that the maximum possible number of images produced by an n-lens is at least $3n+1$. A.O. Petters in [13], using Morse's theory, obtained a number of estimates for the number of images produced by a non-planar lens. S.H. Rhie [15] conjectured that the upper-bound for the number of lensed images for an n-lens is $5n - 5$. Moreover, she showed in [16] that this bound is attained for every $n > 1$ and, hence, is sharp. Rhie's conjecture was proved in full in [8]. Namely, we have the following result.

Theorem 3.1. *The number of lensed images by an n-mass, $n > 1$, planar lens cannot exceed $5n - 5$ and this bound is sharp* [16]. *Moreover, the number of images is an even number when n is odd and odd where n is even.*

The proof of the above result rests on some simple ideas from complex dynamics (cf. [9, 10]).

4. "Thin" lenses with continuous mass distributions

If we replace point masses by a general, real-valued mass distribution μ, a compactly supported Borel measure in the lens plane, the lens equation with shear (2.2) becomes

$$w = z - \int_{\Omega} \frac{d\mu(\zeta)}{\bar{z} - \bar{\zeta}} - \gamma \bar{z}. \tag{4.1}$$

Here Ω is a bounded domain containing the support of μ. The case of the atomic measure $\mu = \sum_1^n \sigma_j \delta_{z_j}$, $\sigma_i \in \mathbb{R}$ is covered by Theorem 3.1. Also, as noted in [8], if we replace n-point-masses by n non-overlapping radially symmetric masses, the total number of images outside of the region occupied by n-masses is still $5n - 5$

when $\gamma = 0$, and $\leq 5n$ when $\gamma \neq 0$. The reason for that, of course, is that the Cauchy integral

$$\int_{|\zeta - z_j| < R} \frac{d\mu(\zeta)}{z - \zeta}, \quad |z - z_j| > R$$

for any radially symmetric measure $\mu = \mu(|\zeta - z_j|)$ is immediately calculated to be equal $\frac{c}{z - z_j}$, where c is the total mass μ of the disk $\{\zeta : |\zeta - z_j| < R\}$, hence reducing this new situation to the one treated in Theorem 3.1.

Here is another situation that can be treated with help from Theorem 3.1.

Recall that a simply-connected domain Ω is called a *quadrature domain* (of order n) if Ω is obtained from the unit disk $\mathbb{D} := \{z : |z| < 1\}$ via a conformal mapping φ that is a rational function of degree n, $\Omega = \varphi(\mathbb{D})$. Of course, all poles β_j, $j = 1, \ldots, n$ of φ will lie outside $\mathbb{D}$. Then if, say, μ is a uniform mass distribution in Ω, i.e., $\mu = \text{const}\, dx\, dy$, the Cauchy potential term in (4.1) for $z \notin \overline{\Omega}$ becomes

$$\sum_{j=1}^{n} \frac{c_j}{z - z_j}, \quad z_j = \varphi\left(\frac{1}{\overline{\beta_j}}\right), \tag{4.2}$$

where the coefficients c_j are determined by the quadrature formula associated with Ω (cf. [18] for details).

Hence, substituting (4.2) into (4.1) we again obtain that for such thin lens Ω with a uniform density distribution, the number of "bright" images outside Ω cannot exceed $5n - 5$ when no shear is present, or $5n$ otherwise.

In this general context the only previously known (to the best of our knowledge) result is the celebrated Burke's theorem [2]:

Theorem 4.1. *A (finite) number of images produced by a* smooth *mass distribution μ is always odd, provided that $\gamma = 0$ (no shear).*

An elegant complex-analytic proof of Burke's theorem can be found in [19]. The crux of the argument is this. Take $w = 0$ and let n_+, n_- denote, respectively, the number of sense-preserving and sense-reversing zeros of the lens map in (4.1) ($\gamma = 0$).

The argument principle applies to harmonic complex-valued functions in the same way it does to analytic functions. Since the right-hand side of (4.1) behaves like $O(z)$ near ∞, the argument principle then yields that $1 = n_+ - n_-$. Thus, giving us the total number of zeros $N := n_+ + n_- = 2n_- + 1$, an odd number.

5. Ellipsoidal lens

Suppose the lens $\Omega := \left\{\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1, a > b > 0\right\}$ is an ellipse. First assume the mass density to be constant, say 1. Let $c : c^2 = a^2 - b^2$ be the focal distance of Ω. The lens equation (4.1) can be rewritten as

$$\bar{z} - \frac{1}{\pi} \int_{\Omega} \frac{dA(\zeta)}{z - \zeta} - \gamma z = \bar{w}, \tag{5.1}$$

where dA denotes the area measure. Using complex Green's formula (cf., e.g., [19]), we can rewrite (5.1) for $z \in \mathbb{C} \setminus \overline{\Omega}$ as follows:

$$\bar{z} - \frac{1}{2\pi i} \int_{\partial\Omega} \frac{\bar{\zeta}\, d\zeta}{z - \zeta} - \gamma z = \bar{w}. \tag{5.2}$$

As is well known [18], the (analytic) Schwarz function $S(\zeta)$ for the ellipse defined by $S(\zeta) = \bar{\zeta}$ on $\partial\Omega$ can be easily calculated and equals

$$\begin{aligned} S(\zeta) &= \frac{a^2 + b^2}{c^2}\zeta - \frac{2ab}{c^2}\left(\zeta - \sqrt{\zeta^2 - c^2}\right) \\ &= \frac{a^2 + b^2 - 2ab}{c^2}\zeta + \frac{2ab}{c^2}\left(\zeta - \sqrt{\zeta^2 - c^2}\right) \\ &= S_1(\zeta) + S_2(\zeta). \end{aligned} \tag{5.3}$$

Note that S_1 is analytic in $\overline{\Omega}$, while S_2 is analytic outside Ω and $S_2(\infty) = 0$. This is, of course, nothing else but the Plemelj–Sokhotsky decomposition of the Schwarz function $S(\zeta)$ of $\partial\Omega$. From (5.3) and Cauchy's theorem we easily deduce that for $z \in \mathbb{C} \setminus \overline{\Omega}$ the lens equation (5.2) reduces to

$$\bar{z} + \frac{2ab}{c^2}\left(z - \sqrt{z^2 - c^2}\right) - \gamma z = \bar{w}. \tag{5.4}$$

Squaring and simplifying, we arrive from (5.4) at a complex quadratic equation

$$\left[\bar{z} + \left(\frac{2ab}{c^2} z - \gamma\right) z\bar{w}\right]^2 = \frac{2a^2b^2}{c^2}\left(z^2 - c^2\right)$$

which is equivalent to a system of two irreducible real quadratic equations. Bezout's theorem (cf. [8,9], [10], [3]) then implies that (5.1) may only have 4 solutions $z \notin \Omega$. For $z \in \Omega$, using Green's formula and (5.3) we can rewrite the area integral in (5.1)

$$\begin{aligned} -\frac{1}{\pi}\int_{\Omega} \frac{dA(\zeta)}{z - \zeta} &= -\bar{z} + \frac{1}{2\pi i}\int_{\partial\Omega} \frac{\bar{\zeta}\, d\zeta}{\zeta - z} \\ &= -\bar{z} + \frac{1}{2\pi i}\int_{\partial\Omega} \frac{[S_1(\zeta) + S_2(\zeta)]}{\zeta - z} \\ &= -\bar{z} + S_1(z) = -\bar{z} + \frac{a^2 + b^2 - 2ab}{c^2} z \end{aligned} \tag{5.5}$$

We have used here that the Cauchy transform of $S_2 \,|_{\partial\Omega}$ vanishes in Ω since S_2 is analytic in $\overline{\mathbb{C} \setminus \Omega}$ and vanishes at infinity. Substituting (5.5) into (5.1), we arrive at a linear equation

$$\left(\frac{a^2 + b^2 - 2ab}{c^2} - \gamma\right) z = \bar{w} \tag{5.6}$$

for $z \in \Omega$. Equation (5.6), of course, may only have one root in Ω. Thus, we have proved the following

Theorem 5.1. *An elliptic lens* Ω *(say, a galaxy) with a uniform mass density may produce at most four "bright" lensing images of a point light source outside* Ω *and one* ("dim") *image inside* Ω, *i.e., at most* 5 *lensing images altogether.*

This type of result has actually been observed experimentally – cf. Fig. 4, where four bright images are clearly present. It is conceivable that the dim image is also there but we can't see it because it is perhaps too faint compared with the galaxy. Of course, one has to accept Fig. 4 with a grain of salt since we do not expect "real" galaxies to have uniform densities. A model of an elliptical lens, with shear, that produces five images (4 bright + 1 dim) is given in Fig. 5.

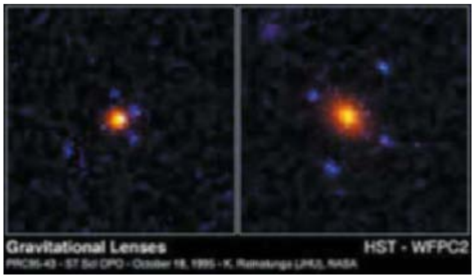


FIGURE 4. Four images of a light source behind the elliptical galaxy. (Credit: NASA, K. Ratnatunga (Johns Hopkins University).)

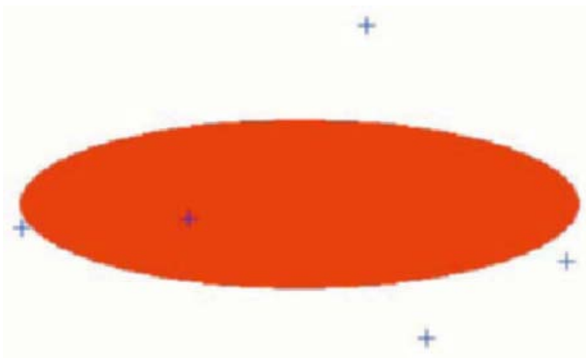


FIGURE 5. A model with five images of a source behind an elliptical lens with axis ratio 0.5 and uniform density 2.

We can extend the previous theorem for a larger class of mass densities. Denote by $q(x,y) := \frac{x^2}{a^2} + \frac{y^2}{b^2} - 1$ the equation of $\Gamma := \partial\Omega$. Let $q_\lambda(x,y) = \frac{x^2}{a^2+\lambda} + \frac{y^2}{b^2+\lambda} - 1$, $-b^2 < \lambda < 0$ stand for the equation of the boundary $\Gamma_\lambda := \partial\Omega_\lambda$ of the ellipse Ω_λ confocal with Ω.

The celebrated MacLaurin theorem (cf. [7]) yields that for any $z \in \mathbb{C} \setminus \overline{\Omega}$

$$\frac{1}{\text{Area}\,(\Omega_\lambda)} \int_{\Omega_\lambda} \frac{dA(\zeta)}{\zeta - z} = \frac{1}{\text{Area}(\Omega)} \int_{\Omega} \frac{dA(\zeta)}{\zeta - z}. \tag{5.7}$$

Thus, if we denote by $u(z,\lambda)$ the Cauchy potential of Ω_λ evaluated at $z \in \mathbb{C} \mid \overline{\Omega}$ we obtain from (5.7)

$$u\,(z,\lambda) = c(\lambda)u_\Omega(z,0), \tag{5.8}$$

where

$$c(\lambda) = \frac{\text{Area}\,(\Omega_\lambda)}{\text{Area}(\Omega)} = \frac{\left(a^2+\lambda\right)^{1/2}\left(b^2+\lambda\right)^{1/2}}{ab}. \tag{5.9}$$

Hence,

$$\frac{\partial u_\lambda(z,\lambda)}{\partial\lambda} = c'(\lambda)u_\Omega(z). \tag{5.10}$$

So, if the mass density $\mu(\lambda)$ in Ω only depends on the elliptic coordinate λ, i.e., is constant on ellipses confocal with Ω inside Ω, its potential outside Ω equals

$$u_{\mu,\Omega}(z) = cu_\Omega(z). \tag{5.11}$$

The constant c is easily calculated from (5.9)–(5.10) and equals

$$c = \int_{-b^2}^{0} \mu(\lambda)c'(\lambda)\,d\lambda. \tag{5.12}$$

It is, of course, natural for physical reasons to assume that $\mu(\lambda) \uparrow \infty$ at the "core" of Ω (i.e., when $\lambda \downarrow -b^2$), the focal segment $[-c,c]$. Yet, from (5.12) since (5.9) yields $c'(\lambda) = O\left(\left(b^2+\lambda\right)^{-1/2}\right)$ near $\lambda_0 = -b^2$, it follows that $\mu(\lambda)$ should not diverge at the core faster than say $O\left(\left(b^2+\lambda\right)^{-1/2+\epsilon}\right)$ for some positive ϵ, so the integral (5.12) converges. Substituting (5.11) into the lens equation (5.1) with constant density replaced by the density $\mu(\lambda)$ and following again the steps in (5.2)–(5.4) we arrive at the following corollary.

Corollary 5.1. *An elliptic lens Ω with mass density that is constant inside Ω on the ellipses confocal with Ω may produce at most four "bright" lensing images of a point light source outside Ω.*

6. Einstein rings

For a one-point mass at z_1 lens with the source at $w = 0$ the lens equation (2.1) without shear becomes

$$z - \frac{c}{\bar{z} - \bar{z}_1} = 0. \tag{6.1}$$

As was already noted by Einstein (cf. [12,19,21] and references cited therein), (6.1) may have two solutions (images) when $z_1 \neq 0$ and a whole circle ("Einstein ring") of solutions when $z_1 = 0$, in other words when the light source, the lens and the observer coalesce – cf. Fig. 6 and Fig. 7.

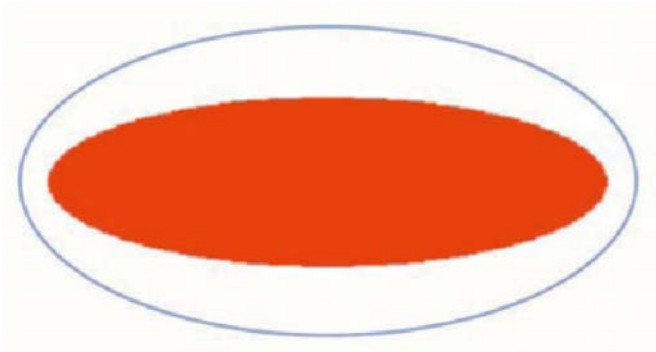


Figure 6. A model of an elliptical Einstein ring surrounding an elliptical lens with axis ratio 0.5 and uniform density 2. The shear in this case must be specially chosen to produce the ring instead of point images. Note that the ring is an ellipse confocal with the lens – cf. Thm. 6.1.

As the following simple theorem shows the "ideal" Einstein rings are limited to ellipses and circles in much more general circumstances.

Theorem 6.1. *Let Ω be any planar ("thin") lens with mass distribution μ_e. If lensing of a point source produces a bounded "image" curve outside of the lens Ω, it must either be a circle when the external shear $\gamma = 0$ or an ellipse.*

Proof. First consider a simpler case when $\gamma = 0$. If the lens produces an image curve Γ outside Ω, the lens equation (4.1) becomes

$$\bar{z} - \bar{w} = \int_\Omega \frac{d\mu(\zeta)}{z-\zeta}, \tag{6.2}$$

for all $z \in \Gamma$. Note that Γ being bounded and also being a level curve of a harmonic function must contain a closed loop surrounding Ω [22]. Without loss of generality, we still denote that loop by Γ. The right-hand side $f(z)$ of (6.2) is a bounded analytic function in the unbounded complement component $\tilde{\Omega}_\infty$ of Γ that vanishes at infinity. Hence $(z-w)f(z)$ is still a bounded and analytic function in $\tilde{\Omega}_\infty$ equal to $|z-w|^2 > 0$ on $\Gamma := \partial\tilde{\Omega}_\infty$. Hence $(z-w)f(z) = \text{const}$ and Γ must be a circle centered at w.

Now suppose $\gamma \neq 0$. Once again we shall still denote by Γ a closed Jordan loop surrounding Ω. Denote by $\tilde{\Omega}$ the interior of Γ, $\tilde{\Omega}_\infty = \mathbb{C} \setminus \operatorname{clos}\left(\tilde{\Omega}\right)$. Also, by translating we can assume that the position of the source w is at the origin.

The equation (4.1) now reads

$$\bar{z} = \int_\Omega \frac{d\mu(\zeta)}{z-\zeta} + \gamma z, \quad z \in \Gamma. \tag{6.3}$$

In other words the right-hand side of (6.3) represents the Schwarz function S of Γ, analytic in $\mathbb{C} \setminus \operatorname{supp}\mu$ with a simple pole at ∞. It is well known that this already

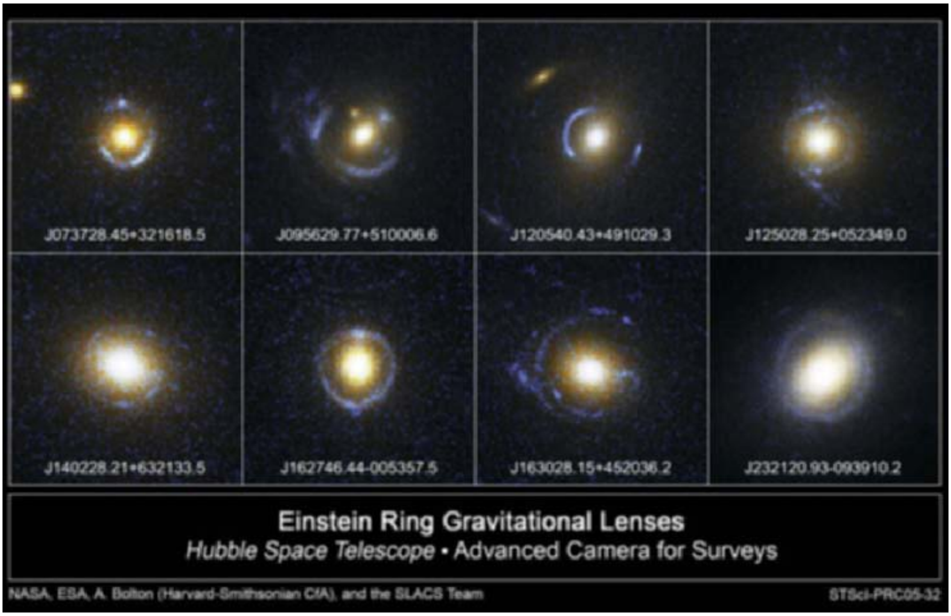


FIGURE 7. Einstein rings. The sources in these observed "realistic" lenses are actually extended, and that is why we see sometimes arcs rather than whole rings. (Credit: ESA, NASA and the SLACS survey team: A. Bolton (Harvard / Smithsonian), S. Burles (MIT), L. Koopmans (Kapteyn), T. Treu (UCSB), and L. Moustakas (JPL/Caltech).)

implies that Γ must be an ellipse (cf. [4,18]) and references therein. For the reader's convenience we supply a simple proof.

Applying Green's formula to (6.3) yields (cf. (5.1)–(5.2)) that for all $z \in \tilde{\Omega}_\infty$

$$\int_\Omega \frac{d\mu(\zeta)}{z-\zeta} = \frac{1}{\pi}\int_{\tilde{\Omega}} \frac{dA(\zeta)}{z-\zeta}, \quad z \in \tilde{\Omega} := \mathbb{C} \setminus \tilde{\Omega}_\infty. \tag{6.4}$$

Let

$$h(z) := \frac{1}{\pi}\int_{\tilde{\Omega}} \frac{dA(\zeta)}{z-\zeta} - \bar{z}, \quad z \in \tilde{\Omega}. \tag{6.5}$$

Then, $h(z)$ is analytic in $\tilde{\Omega}$ (cf. (5.2)) and, in view of (6.3) and (6.4)

$$h\mid_\Gamma = \int_\Omega \frac{d\mu(\zeta)}{z-\zeta}\bigg|_\Gamma - \bar{z}|_\Gamma = \gamma z \mid_\Gamma . \tag{6.6}$$

Thus, $h(z)$ is a linear function and since (6.5) for $z \in \tilde{\Omega}$

$$h(z) := \frac{1}{2}\,\text{grad}\left[\frac{1}{\pi}\int_{\tilde{\Omega}} \log|z-\zeta|\,dA(\zeta) - |z|^2\right] \tag{6.7}$$

we conclude from (6.7) that the potential of $\tilde{\Omega}$

$$u_{\tilde{\Omega}}(z) = \frac{1}{2\pi} \int_{\tilde{\Omega}} \log|z-\zeta|\, dA(\zeta), \quad z \in \tilde{\Omega}$$

equals to a quadratic polynomial inside $\tilde{\Omega}$. The converse of the celebrated theorem of Newton due to P. Dive and N. Nikliborc (cf. [7, Ch. 13–14] and references therein) now yields that $\tilde{\Omega}$ must be an interior of an ellipse, hence $\Gamma := \partial\tilde{\Omega} = \partial\tilde{\Omega}_\infty$ is an ellipse. □

Remark 6.1. One immediately observes that since the converse to Newton's theorem holds in all dimensions the last theorem at once extends to higher dimensions if one replaces the words "image curve" by "image surface".

7. Final remarks

1. The densities considered in §5 are less important from the physical viewpoint than so-called "isothermal density" which is obtained by projecting onto the lens plane the "realistic" three-dimensional density $\sim 1/\rho^2$, where ρ is the (three-dimensional) distance from the origin. This two-dimensional density could be included into the whole class of densities that are constant on all ellipses *homothetic* rather than confocal with the given one. The reason for the term "isothermal" is that when a three-dimensional galaxy has density $\sim 1/\rho^2$ the gas in the galaxy has constant temperature (cf. [5] and the references therein).

Recall that the Cauchy potential of the ellipse

$$\Omega := \left\{ \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1, a > b > 0 \right\}$$

outside of Ω (cf. (5.2)–(5.4)) equals

$$u_0(z) := k\left(z - \sqrt{z^2 - c^2}\right), \quad z \in \mathbb{C} \setminus \overline{\Omega}, \quad c^2 = a^2 - b^2, \tag{7.1}$$

where $k = 2ab/c^2$ is a constant. Replacing the ellipse Ω by a homothetic ellipse $\Omega_t := t\Omega := \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq t^2$, $0 < t < 1$. We obtain using (7.1) for $z \notin \overline{\Omega}$:

$$\begin{aligned} u(z,t) :&= \int_{t\Omega} \frac{dA(\zeta)}{\zeta - z} = t^2 \int_{\Omega} \frac{dA(\zeta)}{t\zeta - z} \\ &= tu\left(\frac{z}{t}; 1\right) = k\left(z - \sqrt{z^2 - c^2 t^2}\right). \end{aligned} \tag{7.2}$$

Thus,

$$\frac{\partial u(z,t)}{\partial t} = k\, \frac{c^2 t}{\sqrt{z^2 - c^2 t^2}}. \tag{7.3}$$

So, if the "isothermal" density $\mu = \frac{1}{t}$ on $\partial\Omega_t$ inside Ω (ignoring constants), we get from (7.3) that the Cauchy potential of such mass distribution outside Ω

equals

$$u_\mu(z) := C_0 \int_0^1 \frac{dt}{\sqrt{z^2 - c^2t^2}}, \tag{7.4}$$

where the constant C_0 depends on Ω only. This is a transcendental function (one of the branches of $\arcsin \frac{c}{z}$), dramatically different from the algebraic potential in (7.1). The lens equation (4.1) now becomes

$$z - C_0 \int_0^1 \frac{dt}{\sqrt{\bar{z}^2 - c^2t^2}} - \gamma\bar{z} = w. \tag{7.5}$$

To the best of our knowledge the precise bound on the maximal possible number of solutions (images) of (7.5) is not known. Up to today, no more than 5 images (4 bright +1 dim) have been observed. However, in [5] there have been constructed explicit models depending on parameters a, b and $0 < \gamma < 1$ having 9 (i.e., 8+1) images. The equation (7.5) essentially differs from all the lens equations considered in this paper since it involves estimating the number of zeros of a transcendental harmonic function with a simple pole at ∞. At this point, we are even reluctant to make a conjecture regarding what this maximal number might be.

Note, that in case of a circle $\Omega = \{x^2 + y^2 < 1\}$ with any radial density $\mu := \varphi(r)$, $r = \sqrt{x^2 + y^2} < 1$, the situation is very simple. The Cauchy potential $u(z)$ outside Ω, as was noted earlier, equals

$$\frac{c}{z}, \quad |z| > 1, \tag{7.6}$$

where c is a constant. Hence, outside Ω the lens equation becomes

$$z - \frac{c}{\bar{z}} - \gamma\bar{z} = w, \tag{7.7}$$

a well-known Chang–Refsdal lens (cf., e.g., [1]) that may have at most 4 solutions except for the degenerate case $\gamma = w = 0$, when the Einstein ring appears. In particular, when $\gamma = 0$, $w \neq 0$, such mass distribution may only produce two bright images outside Ω. For $z : |z| < 1$ inside the lens the potential is still calculated by switching to polar coordinates:

$$\begin{aligned} u(z) &:= \int_0^1 \int_0^{2\pi} \frac{\varphi(r) r dr d\theta}{re^{i\theta} - z} \\ &= \int_{|z|}^1 \varphi(r) dr \int_0^{2\pi} \frac{r d\theta}{re^{i\theta} - z} + \int_0^{|z|} \varphi(r) r dr \int_0^{2\pi} \frac{d\theta}{re^{i\theta} - z} \\ &= \int_{|z|}^1 \varphi(r) dr \int_0^{2\pi} \left(\sum_0^\infty \left(\frac{z}{r}\right) e^{-i(n+1)\theta} \right) d\theta + \frac{1}{z} \int_0^{|z|} \varphi(r) r dr \int_0^{2\pi} \left(\sum_0^\infty \left(\frac{re^{i\theta}}{z} \right)^n d\theta \right. \\ &= \frac{2\pi}{z} \int_0^{|z|} \varphi(r) r dr. \end{aligned} \tag{7.8}$$

In particular, for the "isothermal" density $\varphi(r) \sim \frac{1}{r}$, (7.8) yields for $z : |z| < 1$

$$u(z) = \frac{2\pi}{z}\,|z|,$$

so the lens equation (7.7) becomes

$$\bar{z} - \frac{c}{z}\,|z| - \gamma z = \bar{w}, \tag{7.9}$$

where c is a real constant. Equation (7.9) can have at most two solutions *inside* Ω (only one, if $\gamma = 0$), again, excluding the degenerate case of the Einstein ring. Furthermore, since Burke's theorem allows only an odd number of images, the total maximal number of images for an isothermal sphere cannot exceed 5 (4 bright + 1 dim) as before (or ≤ 3, i.e., $(2+1)$ if $\gamma = 0$). Note, that strictly speaking, Burke's theorem cannot be applied to the isothermal density because of the singularity at the origin. Yet, since the density is radial and smooth everywhere excluding the origin and because it is clear from (7.9) that the origin cannot be a solution, Burke's theorem does apply yielding the above conclusion.

2. The problem of estimating the maximal number of "dim" images inside the lens formed by a uniform mass-distribution inside a quadrature domain Ω (cf. §4) of order n is challenging. In this case the Cauchy potential in (4.1) inside Ω equals to the "analytic" part of the Schwarz function $S(z)$. It is known that $S(z)$ is an algebraic function of degree at most $2n$. Yet, the sharp bounds, similar to those in Theorem 3.1, for the number of zeros of harmonic functions of the form $\bar{z} - a(z)$, where $a(z)$ is an algebraic function, are not known.

3. Another interesting and difficult problem would be to study the maximal number of images by a lens consisting of several elliptical mass distributions. Some rough estimates based on Bezout's theorem can be made by imitating the calculations in §5. Yet, even for 2 uniformly distributed masses these calculations give a rather large possible number of images (≤ 15) while, so far, only 5 images by a two galaxies lens and 6 images by a three galaxies lens have been observed – cf. [6,23].

Added in Proof. Recently, in the 2009 preprint "Transcendental Harmonic Mappings and Gravitational Lensing by Isothermal Galaxies", the third author and E. Lundberg managed to prove that the total number of bright images produced by an isothermal ellipsoidal galaxy without a shear is at most 8. They conjecture however, in correspondence with the models in [5], that the sharp bound for the maximum number of bright images is actually 4 without a shear, and 8 with a shear.

References

[1] J. An and N.W. Evans. The Chang–Refsdal lens revisited. *Mon. Not. R. Astron. Soc.*, 369:317–334, 2006.

[2] W.L. Burke. Multiple gravitational imaging by distributed masses. *Astrophys. J.*, 244:L1, 1981.

[3] Ph. Griffiths and J. Harris. *Principles of Algebraic Geometry*. Pure and Applied Mathematics. Wiley-Interscience, New York, 1978.

[4] B. Gustafsson and H.S. Shapiro. What is a quadrature domain? In *Quadrature domains and their applications*, volume 156 of *Oper. Theory Adv. Appl.*, pages 1–25. Birkhäuser, Basel, 2005.

[5] Ch. Keeton, S. Mao, and H.J. Witt. Gravitational lenses with more than four images, I. classification of caustics. *Astrophys. J.*, pages 697–707, 2000.

[6] Ch. Keeton and J. Winn. The quintuple quasar: Mass models and interpretation. *Astrophys. J.*, 590:39–51, 2003.

[7] D. Khavinson. *Holomorphic partial differential equations and classical potential theory*. Universidad de La Laguna, Departamento de Análisis Matemático, La Laguna, 1996.

[8] D. Khavinson and G. Neumann. On the number of zeros of certain rational harmonic functions. *Proc. Amer. Math. Soc.*, 134(4):1077–1085 (electronic), 2006.

[9] D. Khavinson and G. Neumann. From the fundamental theorem of algebra to astrophysics: a "harmonious" path, Notices Amer. Math. Soc., Vol. 55, Issue 6, 2008, 666–675.

[10] D. Khavinson and G. Świątek. On the number of zeros of certain harmonic polynomials. *Proc. Amer. Math. Soc.*, 131(2):409–414 (electronic), 2003.

[11] S. Mao, A.O. Petters, and H.J. Witt. Properties of point-mass lenses on a regular polygon and the problem of maximum number of images. In T. Piron, editor, *Proc. of the eighth Marcell Grossman Meeting on General Relativity* (*Jerusalem, Israel,* 1977), pages 1494–1496. World Scientific, Singapore, 1998.

[12] R. Narayan and B. Bartelman. Lectures on gravitational lensing. In *Proceedings of the* 1995 *Jerusalem Winter School*, http://cfa-www.harvard.edu/∼narayan/papers/JeruLect.ps, 1995.

[13] A.O. Petters. Morse theory and gravitational microlensing. *J. Math. Phys.*, 33:1915–1931, 1992.

[14] A.O. Petters, H. Levine, and J. Wambsganss. *Singularity Theory and Gravitational Lensing*. Birkhäuser, Boston, MA, 2001.

[15] S.H. Rhie. Can a gravitational quadruple lens produce 17 images? www.arxiv.org/pdf/astro-ph/0103463, 2001.

[16] S.H. Rhie. n-point gravitational lenses with $5n-5$ images, www.arxiv.org/pdf/astro-ph/0305166, 2003.

[17] T. Sauer. Nova Geminorum of 1912 and the origin of the idea of gravitational lensing, lanl.arxiv.org/pdf/0704.0963, 2007.

[18] H.S. Shapiro. *The Schwarz function and its generalization to higher dimensions*, volume 9 of *University of Arkansas Lecture Notes in the Mathematical Sciences.*

[19] N. Straumann. Complex formulation of lensing theory and applications. *Helvetica Phys. Acta*, arXiv:astro-ph/9703103, 70:896–908, 1997.

[20] Ch. Turner. The early history of gravitational lensing, www.nd.edu/∼/turner.pdf, 2006.

[21] J. Wambsganss. Gravitational lensing in astronomy. *Living Rev. Relativity*, www.livingreviews.org/lrr-1998-12, 1:74 pgs., 1998. Last amended: 31 Aug. 2001.

[22] A. Wilmhurst. The valence of harmonic polynomials. *Proc. Amer. Math. Soc.*, 126:2077–2081, 1998.

[23] J. Winn, Ch. Kochanek, Ch. Keeton, and J. Lovell. The quintuple quasar: Radio and optical observations. *Astrophys. J.*, 590:26–38, 2003.

[24] H.J. Witt. Investigations of high amplification events in light curves of gravitationally lensed quasars. *Astron. Astrophys.*, 236:311–322, 1990.

C.D. Fassnacht
Department of Physics
University of California at Davis
Davis, CA 95616, USA
e-mail: `fassnacht@physics.ucdavis.edu`

C.R. Keeton
Department of Physics and Astronomy
Rutgers University
Piscataway, NJ 08854-8019, USA
e-mail: `keeton@physics.rutgers.edu`

D. Khavinson
Department of Mathematics & Statistics
University of South Florida
Tampa, FL 33620-5700, USA
e-mail: `dkhavins@cas.usf.edu`

Analysis and Mathematical Physics
Trends in Mathematics, 131–142
© 2009 Birkhäuser Verlag Basel/Switzerland

Nevanlinna Domains in Problems of Polyanalytic Polynomial Approximation

Konstantin Yu. Fedorovskiy

Abstract. The concept of a Nevanlinna domain, which is the special analytic characteristic of a planar domain, has been naturally appeared in problems of uniform approximation by polyanalytic polynomials. In this paper we study this concept in connection with several allied approximation problems.

Mathematics Subject Classification (2000). Primary 30E10; Secondary 30D55.

Keywords. Nevanlinna domain, polyanalytic function, polyanalytic polynomial, approximation by polyanalytic functions and polyanalytic polynomials.

1. Introduction

In this paper we deal with the notion of a Nevanlinna domain. This notion is the special analytic characteristic of a planar domain that plays a significant role in problems on approximability of functions by polyanalytic polynomials on compact subsets of the complex plane $\mathbb{C}$.

Denote by $\mathbb{D}$ the unit disk $\{z \in \mathbb{C} : \ |z| < 1\}$ and let $\mathbb{T} = \partial\mathbb{D}$ to be the unit circle and $\mathfrak{m}$ to be the normalized Lebesgue measure on $\mathbb{T}$. In what follows we will use the term "almost all" instead of the term "$\mathfrak{m}$-almost all".

For an open set $E \subset \overline{\mathbb{C}}$ let us denote by $\mathrm{H}^\infty(E)$ the class of all bounded holomorphic functions on E. We recall, that for each function $f \in \mathrm{H}^\infty(\mathbb{D})$ and for almost all $\xi \in \mathbb{T}$ there exists the finite angular boundary value $f(\xi)$ of f at ξ from $\mathbb{D}$. The following concept of a Nevanlinna domain was introduced in [1, Definition 2.1]:

Definition 1. One says, that a bounded simply connected domain Ω in $\mathbb{C}$ is called a *Nevanlinna domain* if there exist two functions $u, v \in \mathrm{H}^\infty(\Omega)$ (with $v \not\equiv 0$) such

The author is partially supported by the Russian Foundation for Basic Research (grant no. 07-01-00503) and by the program "Leading Scientific Schools" (grant no. NSh-3877.2008.1).

that the equality

$$\overline{\zeta} = \frac{u(\zeta)}{v(\zeta)} \tag{1.1}$$

holds on $\partial\Omega$ *almost everywhere in the sense of conformal mappings.* It means that the equality of *angular boundary values*

$$\overline{\varphi(\xi)} = \frac{(u \circ \varphi)(\xi)}{(v \circ \varphi)(\xi)},$$

holds for almost all $\xi \in \mathbb{T}$, where φ is some conformal mapping from $\mathbb{D}$ onto Ω.

In fact the definition of a Nevanlinna domain does not depend on the choice of φ and, in view of the Luzin-Privalov boundary uniqueness theorem, the quotient u/v is uniquely defined in Ω (for a Nevanlinna domain Ω). If Ω is a Jordan domain with rectifiable boundary, then the equality (1.1) may be understood directly as the equality of angular boundary values almost everywhere with respect to the Lebesgue measure on $\partial\Omega$. Moreover, the equality (1.1) can be similarly understood on any rectifiable Jordan arc $\gamma \subset \partial\Omega$ such that each point $a \in \gamma$ is not a limit point for the set $\partial\Omega \setminus \gamma$. Notice, that for Jordan domains with rectifiable boundaries the concept of a Nevanlinna domain was introduced in [2] in slightly different terms.

It is easy to find out examples of Nevanlinna (any disk) and not Nevanlinna domains (any domain which is bounded by an ellipse that is not a circle, or by a polygonal line) amongst the domains with piecewise analytic boundaries. However, the construction of Nevanlinna domains with non analytic boundaries as well as with other prescribed analytical and geometrical properties is fairly difficult and delicate problem. This problem was discussed in [3] where, in particular, for each $\alpha \in (0,1)$ the example of Nevanlinna domain with C^1 but not $\mathrm{C}^{1,\alpha}$ boundary was constructed.

The concept of a Nevanlinna domain has been appeared in investigations of problem on uniform approximability of functions by polyanalytic polynomials on compact sets in $\mathbb{C}$. For instance, the criterion of uniform approximability of functions by polyanalytic polynomials on Carathéodory compact sets (see [1, Theorem 2.2]) was formulated in terms of this concept. Let us formulate and briefly discuss these problem and result.

We need to recall, that a function f is called *polyanalytic of order n* (for integer $n > 0$) or, shorter, *n-analytic*, in an open set $U \subset \mathbb{C}$ if it is of the form

$$f(z) = f_0(z) + \overline{z} f_1(z) + \cdots + \overline{z}^{n-1} f_{n-1}(z), \tag{1.2}$$

where $f_0, \ldots, f_{n-1}$ are holomorphic functions in U. Notice that the space $\mathcal{O}_n(U)$ of all n-analytic functions in U consists of all continuous functions f on U such that $\overline{\partial}^n f = 0$ in U in the distributional sense, where $\overline{\partial} = \partial/\partial\overline{z}$ is the Cauchy-Riemann operator. By *n-analytic polynomials* and *n-analytic rational functions* we mean the functions of the form (1.2), where $f_0, \ldots, f_{n-1}$ are polynomials and rational functions in the complex variable z respectively, and we will use the notation $\mathcal{P}(n)$ and $\mathcal{R}(n)$ for these sets of functions. One says, that n-analytic rational function f

has its poles outside some set $E \subset \mathbb{C}$, if all (rational) functions $f_0, \ldots, f_{n-1}$ have their poles outside E.

For a compact set $X \subset \mathbb{C}$ we denote by $\mathrm{C}(X)$ the space of all continuous complex-valued functions on X endowed with the uniform norm and set $\mathrm{A}_n(X) = \mathrm{C}(X) \cap \mathcal{O}_n(X^\circ)$. We also denote by $\mathrm{P}_n(X)$ and $\mathrm{R}_n(X, Y)$, where $Y \supseteq X$ is some compact set, the closures in $\mathrm{C}(X)$ of the subspaces $\{p_{|X} : \ p \in \mathcal{P}(n)\}$ and $\{g_{|X} : g \in \mathcal{R}(n)$, and g has its poles outside $Y\}$ respectively. We are interested in the following approximation problem.

Problem A. *Let $X \subset \mathbb{C}$ be a compact set and $n \geq 2$ be an integer. What conditions on X are necessary and sufficient in order that $\mathrm{A}_n(X) = \mathrm{P}_n(X)$ or $\mathrm{A}_n(X) = \mathrm{R}_n(X, Y)$ for some appropriately chosen compact set $Y \supseteq X$?*

The investigation of this problem was started in 1980th (see [4, 5, 6], where several sufficient approximability conditions were obtained) and until now it remains unsolved in the general case. Some bibliographical notes concerning the matter may be found in [7]. We exclude the case $n = 1$ from the consideration, because in this case the classical theorem by Mergelyan (see [8]) says that $\mathrm{A}_1(X) = \mathrm{P}_1(X)$ if and only if the set $\mathbb{C} \setminus X$ is connected, and therefore, the respective problem is completely solved (in terms of topological properties of X).

Let us recall, that a bounded domain Ω is called a *Carathéodory domain* if $\partial\Omega = \partial\Omega_\infty$, where Ω_∞ is the unbounded connected component of the set $\overline{\mathbb{C}} \setminus \overline{\Omega}$ as well as a compact set X is called a *Carathéodory compact set* if $\partial X = \partial \widehat{X}$, where $\widehat{X}$ denotes the union of X and all bounded connected components of the set $\mathbb{C} \setminus X$. In fact each Carathéodory domain Ω is simply connected and possesses the property $\Omega = (\overline{\Omega})^\circ$. The following result was proved in [1, Theorem 2.2]:

Theorem 1. *Let $n \geq 2$ be an integer.*

1. *If $\Omega \subset \mathbb{C}$ is a Carathéodory domain, then*
$$\mathrm{C}(\partial\Omega) = \mathrm{R}_n(\partial\Omega, \overline{\Omega}) \iff \Omega \textit{ is not a Nevanlinna domain.}$$
2. *If $X \subset \mathbb{C}$ is a Carathéodory compact set, then $\mathrm{A}_n(X) = \mathrm{P}_n(X)$ if and only if $\mathrm{C}(\partial\Omega) = \mathrm{R}_n(\partial\Omega, \overline{\Omega})$ for each bounded connected component Ω of the set $\mathbb{C} \setminus X$ and, therefore, if and only if each bounded connected component of the set $\mathbb{C} \setminus X$ is not a Nevanlinna domain.*

It is worthwhile to mention that the approximability conditions in Theorem 1 are independent on the order of polyanalyticity n. The result of Part 2 of Theorem 1 was obtained in frameworks of the special "reductive" approach, which was elaborated and applied for several different approximation problems in [1, 9, 10, 11]. This approach allow us to conclude (under some suitable assumptions) that one approximability property takes place on a compact set X whenever one has certain similar properties on some appropriately chosen (and more simple) compact subsets of X.

At the same time Problem A for non Carathéodory compact sets is more difficult and remains open. Several interesting and important general results in this problem were obtained in [1, 10, 12]. These results are related with usage of some special refinements of the concept of a Nevanlinna domain, as well as of the mentioned above reductive approach and on studying of special properties of conformal mappings of Carathéodory domains onto the unit disk and new special properties of measures that are orthogonal to rational functions on non Carathéodory compact sets of special type. In [1, Example 4.5] it was shown, that there exists a non Carathéodory compact set Y such that $\mathrm{P}_2(Y) \neq \mathrm{C}(Y)$, but $\mathrm{P}_3(Y) = \mathrm{C}(Y)$. Furthermore, in view of [13, Theorem 1], for each integer $n \geq 1$ there exists a non Carathéodory compact set $X \subset \mathbb{C}$ such that $\mathrm{P}_{2n}(X) = \mathrm{C}(X) \neq \mathrm{P}_n(X)$. Thus the approximability conditions in Problem A are no longer independent on the order of polyanalyticity for non Carathéodory compact sets.

2. Main results and proofs

Let $\mathbb{D}_e = \overline{\mathbb{C}} \setminus \overline{\mathbb{D}}$. For each function $F \in \mathrm{H}^\infty(\mathbb{D}_e)$ and for $\xi \in \mathbb{T}$ we denote by $F(\xi)$ the angular boundary value of F at ξ from $\mathbb{D}_e$, that exists for almost all $\xi \in \mathbb{T}$. Let us recall the notion of a Nevanlinna-type pseudocontinuation of bounded holomorphic functions (see [3, Definition 2]):

Definition 2. Let $f \in \mathrm{H}^\infty(\mathbb{D})$. One says, that a function f admits a *Nevanlinna-type pseudocontinuation*, if there exists two functions $F_1, F_2 \in \mathrm{H}^\infty(\mathbb{D}_e)$ (with $F_2 \not\equiv 0$) such that the equality $f(\xi) = F_1(\xi)/F_2(\xi)$ of angular boundary values holds for almost all points $\xi \in \mathbb{T}$.

This definition is the partial case of the general notion of a pseudocontinuation which was introduced in [14] (see also [15, Definition 2.1.2]). The notions of a Nevanlinna domain and a Nevanlinna-type pseudocontinuation of bounded holomorphic functions are closely related. Indeed, the following characterization of Nevanlinna domains was proved in [1, Proposition 2.1]:

Proposition 1. *Let Ω be a bounded simply connected domain and φ be some conformal mapping from $\mathbb{D}$ onto Ω. Then, Ω is a Nevanlinna domain if and only if φ admits a Nevanlinna-type pseudocontinuation.*

This property was turned out to be rather interesting and useful, and it was used in [3], [10] and [12] in order to obtain several new properties and examples of Nevanlinna domains.

Actually, the notion of a Nevanlinna-type pseudocontinuation of bounded holomorphic functions have also appeared in one other approximation problem that is allied with Problem A, and now we are going on to consider this problem.

Let $p \in [1, \infty]$. In what follows by $\mathrm{L}^p(\mathbb{T})$ one denotes the standard Lebesgue space of functions on $\mathbb{T}$ considering with respect to the measure $\mathfrak{m}$. As usual, the symbol $\mathrm{H}^p(\mathbb{D})$ stands for the Hardy space in the unit disk. We recall that the space $\mathrm{H}^p(\mathbb{D})$ for $p < \infty$ consists of all holomorphic functions f in $\mathbb{D}$ such that

$$\lim_{r \to 1} \int_{\mathbb{T}} |f(r\xi)|^p \, d\mathfrak{m}(\xi) < \infty$$

as well as the space $H^\infty(\mathbb{D})$ consists of all bounded holomorphic in $\mathbb{D}$ functions. For all $p > q > 1$ we have $H^\infty(\mathbb{D}) \subset H^p(\mathbb{D}) \subset H^q(\mathbb{D}) \subset H^1(\mathbb{D})$ and $H^p(\mathbb{D}) \subset N(\mathbb{D})$ for each $p \in [1, \infty]$, where $N(\mathbb{D})$ is the Nevanlinna class in $\mathbb{D}$. Actually, each function $f \in N(\mathbb{D})$ has the form f_1/f_2, where $f_1, f_2 \in H^\infty(\mathbb{D})$.

Let us also denote by $H^p(\mathbb{T})$ the space which consists of all functions $f \in L^p(\mathbb{T})$ such that $\hat{f}(k) := \int_{\mathbb{T}} f(\xi)\,\overline{\xi}^k\, d\mathfrak{m}(\xi) = 0$ for all integer $k < 0$. If $f \in H^p(\mathbb{D})$ for $p \in [1, \infty]$ then, by Fatou's theorem, for almost all $\xi \in \mathbb{T}$ there exists the angular boundary value $f(\xi)$ of f at ξ. These angular boundary values define a function in $H^p(\mathbb{T})$ and the map that takes a function $f \in H^p(\mathbb{D})$ onto its boundary values is an isometric isomorphism of $H^p(\mathbb{D})$ onto $H^p(\mathbb{T})$, as well as for $p = \infty$ this map is also a weak-star homeomorphism. From now on, whenever $f \in H^p(\mathbb{D})$, we will always keep the notation f for the respective boundary function.

For a function $f \in H^\infty(\mathbb{D})$ and for an integer $n \geq 2$, let us define the space

$$M_n^p(\mathbb{T}, f) = H^p(\mathbb{T}) + \overline{f}\, H^p(\mathbb{T}) + \cdots + \overline{f}^{\,n-1} H^p(\mathbb{T}).$$

Theorem 2. *Let $f \in H^\infty(\mathbb{D})$, and let $n \geq 2$ be an integer and $p \in [1, \infty)$.*

1. *The space $M_n^p(\mathbb{T}, f)$ is not dense in $L^p(\mathbb{T})$ if and only if the function f admits a Nevanlinna-type pseudocontinuation.*
2. *The space $M_n^\infty(\mathbb{T}, f)$ is not weak-star dense in $L^\infty(\mathbb{T})$ if and only if the function f admits a Nevanlinna-type pseudocontinuation.*

Notice, that the approximability conditions in Theorem 2 are independent on the order of polyanalyticity n, as it takes place in Theorem 1.

The proof of Theorem 2 is essentially based on the same technical ideas as the proof of the result of [2, Theorem 1]. For the technical reasons we need to reformulate the assertion of Proposition 1 as follows:

Lemma 1. *Let $f \in H^\infty(\mathbb{D})$. Then, f admits a Nevanlinna-type pseudocontinuation if and only if there exists two functions $f_1, f_2 \in H^\infty(\mathbb{D})$ such that $f_2 \not\equiv 0$ and the equality $\overline{f(\xi)} = f_1(\xi)/f_2(\xi)$ holds for almost all points $\xi \in \mathbb{T}$.*

Although the proof of this lemma may be found in [1, Proof of Proposition 2.1] we give it for the reader convenience. Here and in the sequel, for a function ψ we denote by ψ_* the function $\psi_*(z) := \overline{\psi(\overline{z})}$, for all z, where ψ_* is defined.

Proof of Lemma 1. Let f admits a Nevanlinna-type pseudocontinuation and let the functions F_1 and F_2 are taken from Definition 2. For $|z| < 1$ we put $f_1(z) := (F_1)_*(1/z)$ and $f_2(z) := (F_2)_*(1/z)$, thus $f_1, f_2 \in H^\infty(\mathbb{D})$. If, for almost all $\xi \in \mathbb{T}$, the point $z \in \mathbb{D}$ tends non tangentially to ξ then

$$\frac{f_1(z)}{f_2(z)} = \frac{(F_1)_*(1/z)}{(F_2)_*(1/z)} = \overline{\left(\frac{F_1(z')}{F_2(z')}\right)} \to \overline{f(\xi)},$$

because $z' = 1/\overline{z} \in \mathbb{D}_e$ tends non tangentially to ξ.

Conversely, if the functions f_1 and f_2 from $H^\infty(\mathbb{D})$ exist, then for $|z| > 1$ we define $F_1(z) = (f_1)_*(1/z)$ and $F_2(z) = (f_2)_*(1/z)$, so that $F_1, F_2 \in H^\infty(\mathbb{D}_e)$. Thus, for almost all $\xi \in \mathbb{T}$, if $z \in \mathbb{D}_e$ tends to ξ non tangentially, then

$$\frac{F_1(z)}{F_2(z)} = \frac{(f_1)_*(1/z)}{(f_2)_*(1/z)} = \overline{\left(\frac{f_1(z')}{f_2(z')}\right)} \to f(\xi),$$

because $z' = 1/\overline{z} \in \mathbb{D}$ also tends to ξ non tangentially. Therefore, f admits a Nevanlinna-type pseudocontinuation. $\square$

Proof of Theorem 2. We recall that the space of all linear functionals on $\mathrm{L}^p(\mathbb{T})$ can be identified with $\mathrm{L}^q(\mathbb{T})$ via the paring

$$\langle f, g\rangle = \int_{\mathbb{T}} f(\xi)\overline{g(\xi)}\, d\mathfrak{m}(\xi), \tag{2.1}$$

where $f \in \mathrm{L}^p(\mathbb{T})$, $g \in \mathrm{L}^q(\mathbb{T})$ and $q = p/(p-1)$ is the conjugate index for p.

Let us prove the assertion of the part 1. Assume that the function f admits a Nevanlinna-type pseudocontinuation. By Lemma 1 there exist two functions $f_1, f_2 \in \mathrm{H}^\infty(\mathbb{D})$ such that $f_2 \not\equiv 0$ and $\overline{f(\xi)} = f_1(\xi)/f_2(\xi)$ for almost all points $\xi \in \mathbb{T}$. In view of [16, Ch. IX, Sect. 4, Theorem 5], for any functions $h_0, \dots, h_{n-1} \in \mathrm{H}^p(\mathbb{T})$ we have

$$\int_{\mathbb{T}} \Big(\sum_{k=0}^{n-1} h_k(\xi)\overline{f(\xi)}^k\Big) f_2(\xi)^{n-1}\, d\xi = \int_{\mathbb{T}} \Big(\sum_{k=0}^{n-1} h_k(\xi) f_1(\xi)^k f_2(\xi)^{n-k-1}\Big)\, d\xi = 0. \tag{2.2}$$

If we define the function $g(\xi) = \overline{2\pi i\xi f_2(\xi)^{n-1}}$ on $\mathbb{T}$, then $g \in \mathrm{L}^\infty(\mathbb{T})$ and $g \not\equiv 0$ (since $f_2 \not\equiv 0$). Thus the function g defines a linear functional on the space $\mathrm{L}^p(\mathbb{T})$ that annihilates, by formulas (2.1) and (2.2), the space $\mathrm{M}^p_n(\mathbb{T}, f)$. Therefore, $\mathrm{M}^p_n(\mathbb{T}, f)$ is not dense in $\mathrm{L}^p(\mathbb{T})$.

Conversely, let us assume that $\mathrm{M}^p_n(\mathbb{T}, f)$ is not dense in $\mathrm{L}^p(\mathbb{T})$, so that $\mathrm{M}^p_2(\mathbb{T}, f)$ is not dense in $\mathrm{L}^p(\mathbb{T})$. Then, there exists a functional on $\mathrm{L}^p(\mathbb{T})$ that annihilates $\mathrm{M}^p_2(\mathbb{T}, f)$, which means that there exists a function $g \in \mathrm{L}^q(\mathbb{T})$, where $q = p/(p-1)$, such that $g \not\equiv 0$ and the equality

$$\int_{\mathbb{T}} \Big(h_1(\xi) + \overline{f(\xi)} h_2(\xi)\Big)\overline{g(\xi)}\, d\mathfrak{m}(\xi) = 0 \tag{2.3}$$

holds for all functions $h_1, h_2 \in \mathrm{H}^p(\mathbb{T})$. Let $w = \overline{g}$, thus $w \in \mathrm{L}^q(\mathbb{T})$. Taking $h_1 \equiv \xi^\ell$ for integer $\ell > 0$ and $h_2 \equiv 0$ in (2.3) we conclude, that for any integer $k < 0$

$$\hat{w}(k) = \int_{\mathbb{T}} w(\xi)\, \overline{\xi}^k\, d\mathfrak{m}(\xi) = \int_{\mathbb{T}} \xi^{-k}\, w(\xi)\, d\mathfrak{m}(\xi) = 0.$$

It means that $w \in \mathrm{H}^q(\mathbb{T})$. Furthermore, taking in (2.3) $h_1 \equiv 0$ and $h_2 \equiv \xi^\ell$ for integer $\ell > 0$, and arguing analogously we obtain that $w_1 = \overline{f} w \in \mathrm{H}^q(\mathbb{T})$ because of $\hat{w}_1(k) = 0$ for any integer $k < 0$.

Finally, for almost all $\xi \in \mathbb{T}$ one has $\overline{f(\xi)} = w_1(\xi)/w(\xi)$ and since $w, w_1 \in \mathrm{H}^q(\mathbb{T})$ and $w, w_1 \not\equiv 0$, then there exists two functions $f_1, f_2 \in \mathrm{H}^\infty(\mathbb{D})$ such that $f_2 \not\equiv 0$ and the equality

$$\overline{f(\xi)} = \frac{w_1(\xi)}{w(\xi)} = \frac{f_1(\xi)}{f_2(\xi)}$$

holds for almost all $\xi \in \mathbb{T}$. In view of Lemma 1, it gives that the function f admits a Nevanlinna-type pseudocontinuation.

We are going now to prove the assertion of Part 2. The proof of this assertion is very similar to the proof of [2, Theorem 1]. As previously we assume, that the function f admits a Nevanlinna-type pseudocontinuation and take the functions $f_1, f_2 \in \mathrm{H}^\infty(\mathbb{D})$ from Lemma 1. It means that $f_2 \not\equiv 0$ and $\overline{f(\xi)} = f_1(\xi)/f_2(\xi)$ for almost all points $\xi \in \mathbb{T}$. Let us define the measure ν on $\mathbb{T}$ by the formula

$$d\nu(\xi) = 2\pi i \xi f_2(\xi)^{n-1} d\mathfrak{m}(\xi) = f_2(\xi)^{n-1}\, d\xi,$$

so that $\nu \not\equiv 0$ (since $f_2 \not\equiv 0$). Furthermore, for any integer k, $0 \le k < n$ and for any function $h \in \mathrm{H}^\infty(\mathbb{T})$ one has

$$\int_{\mathbb{T}} h(\xi)\overline{f(\xi)}^k\, d\nu(\xi) = \int_{\mathbb{T}} h(\xi)\overline{f(\xi)}^k f_2(\xi)^{n-1}\, d\xi = \int_{\mathbb{T}} h(\xi) f_1(\xi)^k f_2(\xi)^{n-k-1}\, d\xi = 0.$$

Thus $\nu \perp \mathrm{M}_n^\infty(\mathbb{T}, f)$, and therefore $\mathrm{M}_n^\infty(\mathbb{T}, f)$ is not weak-star dense in $\mathrm{L}^\infty(\mathbb{T})$.

Let now $\mathrm{M}_n^\infty(\mathbb{T}, f)$ is not weak-star dense in $\mathrm{L}^\infty(\mathbb{T})$. Then $\mathrm{M}_2^\infty(\mathbb{T}, f)$ is not weak-star dense in $\mathrm{L}^\infty(\mathbb{T})$, and therefore there exists a measure μ on $\mathbb{T}$ orthogonal to $\mathrm{M}_2^\infty(\mathbb{T}, f)$, so that the equality

$$\int_{\mathbb{T}} \big(h_1(\xi) + \overline{f(\xi)} h_2(\xi)\big)\, d\mu(\xi) = 0 \tag{2.4}$$

holds for all functions $h_1, h_2 \in \mathrm{H}^\infty(\mathbb{T})$.

Taking in (2.4) $h_1 \equiv 0$ and $h_2 = \xi^k$ for integer $k \ge 0$ we obtain that the measure $\nu = \overline{f}\mu$ is orthogonal to all polynomials. By F. and M. Riesz's theorem, there exists a function $w_1 \in \mathrm{H}^1(\mathbb{D})$ such that $d\nu = w_1(\xi)\, d\mathfrak{m}$. Similarly, taking in (2.4) $h_1 = \xi^k$ for integer $k \ge 0$ and $h_2 \equiv 0$ we find the function $w_2 \in \mathrm{H}^1(\mathbb{D})$ such that $d\mu = w_2(\xi)\, d\mathfrak{m}(\xi)$. Since $\mu \not\equiv 0$, then $w_1 \not\equiv 0$ and $w_2 \not\equiv 0$.

Thus the equality $\overline{f(\xi)} w_2(\xi) = w_1(\xi)$ of the angular boundary values holds for almost all points $\xi \in \mathbb{T}$, and therefore (since $w_2 \not\equiv 0$) we can rewrite this equality in the form $\overline{f(\xi)} = w_1(\xi)/w_2(\xi)$. Since $w_1, w_2 \in \mathrm{H}^1(\mathbb{D}) \subset \mathrm{N}(\mathbb{D})$, then there exist two functions $f_1, f_2 \in \mathrm{H}^\infty(\mathbb{D})$ such that the quotient w_1/w_2 equals to the quotient f_1/f_2 in $\mathbb{D}$. Then, for almost all $\xi \in \mathbb{T}$ we have $\overline{f(\xi)} = f_1(\xi)/f_2(\xi)$ and, in view of Lemma 1, the function f admits a Nevanlinna-type pseudocontinuation. □

Let us consider one corollary of Theorem 2 that seems to be interesting and useful. For a finite complex-valued Borel measure μ we denote by $\mathrm{L}^p(\mu)$ (where $p \in [1, \infty]$) the standard Lebesgue space of functions with respect to the measure μ. For a Borel measure μ and for an integer $n \ge 2$ we denote by $\mathrm{P}_n^r(\mu)$ the closure of n-analytic polynomials in the space $\mathrm{L}^r(\mu)$ for $r \in [1, \infty)$, as well as by $\mathrm{P}_n^\infty(\mu)$ the weak-star closure of n-analytic polynomials in $\mathrm{L}^\infty(\mu)$.

Let now Ω be a Jordan domain and φ be some conformal mapping from $\mathbb{D}$ onto Ω. In view of Carathéodory extension theorem (see [17, Theorem 2.6]) we may (and shell) assume that φ is extended to a homeomorphism of $\overline{\mathbb{D}}$ onto $\overline{\Omega}$ (let us observe that $\varphi \in \mathrm{A}_1(\overline{\mathbb{D}})$ and $\varphi^{-1} \in \mathrm{A}_1(\overline{\Omega})$ in this case). One has

Corollary 1. *Let Ω and φ be as mentioned above, $r \in [1, \infty]$ and $\omega = \omega(\varphi(0), \cdot, \Omega)$ be the harmonic measure on $\partial\Omega$ evaluated with respect to the point $\varphi(0) \in \Omega$. Then $\mathrm{P}^r_n(\omega) = \mathrm{L}^r(\omega)$ if and only if φ does not admit a Nevanlinna-type pseudocontinuation, or, equivalently, if and only if Ω is not a Nevanlinna domain.*

Let us revert to the assertion of Theorem 2 (Part 1) in the case when $p = 2$. So we are dealing with the Hilbert space $\mathrm{L}^2(\mathbb{T})$ endowed with the scalar product $\langle f, g\rangle = \int_{\mathbb{T}} f(\xi)\overline{g(\xi)}\, d\mathfrak{m}(\xi)$. In what follows, by an operator, one means a bounded linear operator acting on the space $\mathrm{L}^2(\mathbb{T})$ or on some its subspace. For an operator A one denotes by A^* its adjoint operator and by $\ker(A)$ its kernel. For a subset $H \subset \mathrm{L}^2(\mathbb{T})$ one denotes by $H^\perp$ the orthogonal complementary of H in $\mathrm{L}^2(\mathbb{T})$.

For a function $f \in \mathrm{L}^\infty(\mathbb{T})$ let us consider the multiplication operator $M(f)$ on $\mathrm{L}^2(\mathbb{T})$ generated by the function f. It means, that

$$M(f)\, g = fg \in \mathrm{L}^2(\mathbb{T}), \quad \text{for any } g \in \mathrm{L}^2(\mathbb{T}).$$

Furthermore, let Π denotes the Riesz projection operator, i.e., Π is the operator from $\mathrm{L}^2(\mathbb{T})$ onto $\mathrm{H}^2(\mathbb{T})$ acting as follows

$$\text{for a function } g = \sum_{k=-\infty}^{\infty} \lambda_k z^k \in \mathrm{L}^2(\mathbb{T}) \quad \text{one has} \quad \Pi g = \sum_{k=0}^{\infty} \lambda_k z^k.$$

Let us also consider the Toeplitz operator $T(f)$ on $\mathrm{H}^2(\mathbb{T})$ generated by the function f (see [18], in order to find out a review of the theory of Toeplitz operators). It means that the operator $T(f)$ acting in the space $\mathrm{H}^2(\mathbb{T})$ by the formula

$$T(f)\, g = \Pi(M(f)\, g) = \Pi(fg), \quad \text{for a function } g \in \mathrm{H}^2(\mathbb{T}).$$

Henceforth let us assume that $f \in \mathrm{H}^\infty(\mathbb{D})$. Let $\mathrm{H}^2_0(\mathbb{T}) := \{h \in \mathrm{H}^2(\mathbb{T}) :\ h(0) = 0\}$ and $\overline{\mathrm{H}}^2_0(\mathbb{T}) := \{\overline{h} :\ h \in \mathrm{H}^2_0(\mathbb{T})\}$. Then $\mathrm{L}^2(\mathbb{T})$ decomposes into the orthogonal sum of these spaces:

$$\mathrm{L}^2(\mathbb{T}) = \mathrm{H}^2(\mathbb{T}) \oplus \overline{\mathrm{H}}^2_0(\mathbb{T}). \tag{2.5}$$

Accordingly, every operator can be represented as the respective 2×2 operator matrix, and matrix representations of operators $M(f)$ and $M(f)^*$ with respect to this decomposition have the form

$$M(f) = \begin{pmatrix} T(f) & X(f) \\ 0 & Y(f) \end{pmatrix} \quad \text{and} \quad M(f)^* = \begin{pmatrix} T(f)^* & 0 \\ X(f)^* & Y(f)^* \end{pmatrix}. \tag{2.6}$$

Furthermore, we set $\widetilde{T}(f) := Y(f)^*$, $T^k(f) := (T(f))^k$ and $\widetilde{T}^k(f) := (\widetilde{T}(f))^k$, and define, for integer $k \geq 1$, the operators $E_k(f)$ and $\widetilde{E}_k(f)$, related with the operators $T(f)$ and $\widetilde{T}(f)$ by the formulas

$$E_k(f) := T^k(f)^* T^k(f) - T^k(f) T^k(f)^*,$$
$$\widetilde{E}_k(f) := \widetilde{T}^k(f)^* \widetilde{T}^k(f) - \widetilde{T}^k(f) \widetilde{T}^k(f)^*.$$

In the following theorem we calculate explicitly the orthogonal complementary of the space $\mathrm{M}^2_n(\mathbb{T}, f)$ in $\mathrm{L}^2(\mathbb{T})$ and interpret the result of Theorem 2 (Part 1) in terms of special properties of operators $E_k(f)$ and $\widetilde{E}_k(f)$.

Theorem 3. *Let $f \in \mathrm{H}^\infty(\mathbb{D})$.*

1. *For any integer $k \geq 1$*

$$\ker(E_k(f)) = \{h \in \mathrm{H}^2(\mathbb{T}) : \ \overline{f}^k h \in \mathrm{H}^2(\mathbb{T})\}, \tag{2.7}$$

$$\ker(\widetilde{E}_k(f)) = \{b \in \overline{\mathrm{H}}^2_0(\mathbb{T}) : \ f^k b \in \overline{\mathrm{H}}^2_0(\mathbb{T})\}. \tag{2.8}$$

2. *For any integer $n \geq 2$*

$$\mathrm{M}^2_n(\mathbb{T}, f)^\perp = Q_n(f) = \bigcap_{k=1}^{n-1} \ker(\widetilde{E}_k(f)), \quad \text{where} \tag{2.9}$$

$$Q_n(f) := \big\{b \in \overline{\mathrm{H}}^2_0(\mathbb{T}) : \ f^\ell b \in \overline{\mathrm{H}}^2_0(\mathbb{T}), \ \forall\, \ell \in \{1, \dots, n-1\}\big\}.$$

3. *For any integer $k \geq 1$, $\ker(E_k(f)) \neq \varnothing$ if and only if $\ker(\widetilde{E}_k(f)) \neq \varnothing$ and if and only if the function f^k admits a Nevanlinna-type pseudocontinuation.*

Proof. Take an integer $k \geq 1$. Using (2.6) we obtain the following matrix representations for operators $M^k(f) = (M(f))^k$ and $M^k(f)^*$:

$$M^k(f) = \begin{pmatrix} T^k(f) & X_k(f) \\ 0 & \widetilde{T}^k(f)^* \end{pmatrix} \quad \text{and} \quad M^k(f)^* = \begin{pmatrix} T^k(f)^* & 0 \\ X_k(f)^* & \widetilde{T}^k(f) \end{pmatrix}, \tag{2.10}$$

where $X_k(f)$ is the operator from $\overline{\mathrm{H}}^2_0(\mathbb{T})$ to $\mathrm{H}^2(\mathbb{T})$ such that $X_1(f) = X(f)$ and $X_{k+1}(f) = T^k(f)X(f) + X_k(f)\widetilde{T}(f)^*$ for $k > 1$.

Let $N_k(f) := M^k(f)^*M^k(f) - M^k(f)M^k(f)^*$. It follows from (2.10) that

$$N_k(f) = \begin{pmatrix} E_k(f) - X_k(f)X_k(f)^* & T^k(f)^*X_k(f) - X_k(f)\widetilde{T}^k(f) \\ X_k(f)^*T^k(f) - \widetilde{T}^k(f)^*X_k(f)^* & X_k(f)^*X_k(f) - \widetilde{E}_k(f) \end{pmatrix}.$$

Since $M(f)^* = M(\overline{f})$, then for any function $g \in \mathrm{L}^2(\mathbb{T})$ we have

$$N_k(f)\, g = M^k(f)^*M^k(f)\, g - M^k(f)M^k(f)^*\, g = \overline{f}^k f^k g - f^k \overline{f}^k g = 0,$$

which gives $N_k(f) = 0$ and therefore

$$E_k(f) = X_k(f)X_k(f)^* \quad \text{and} \quad \widetilde{E}_k(f) = X_k(f)^*X_k(f),$$

so that

$$\ker(E_k(f)) = \ker(X_k(f)^*) \quad \text{and} \quad \ker(\widetilde{E}_k(f)) = \ker(X_k(f)).$$

In order to prove (2.7) one represents any function $g \in \mathrm{L}^2(\mathbb{T})$ in the "vector" form $g = (h, b)$, where $h = \Pi\, g \in \mathrm{H}^2(\mathbb{T})$ and $b = g - h \in \overline{\mathrm{H}}^2_0(\mathbb{T})$. Taking into account the matrix representation for $M^k(f)^*$ in (2.10), one obtains that

$$M^k(f)^*\, g = \big(T^k(f)^*\, h\,,\, X_k(f)^*\, h + \widetilde{T}^k(f)\, b\big),$$

and finally, that

$$\begin{aligned} \ker(E_k(f)) &= \ker(X_k(f)^*) = \big\{h \in \mathrm{H}^2(\mathbb{T}) : \ X_k(f)^*\, h = 0\big\} \\ &= \big\{h \in \mathrm{H}^2(\mathbb{T}) : \ M^k(f)^*(h, 0) = (T^k(f)^*\, h, 0)\big\} \\ &= \big\{h \in \mathrm{H}^2(\mathbb{T}) : \ M^k(f)^*\, h \in \mathrm{H}^2(\mathbb{T})\big\} = \big\{h \in \mathrm{H}^2(\mathbb{T}) : \ \overline{f}^k h \in \mathrm{H}^2(\mathbb{T})\big\}. \end{aligned}$$

Repeating this computation with minor clear modifications one proves (2.8) and therefore, the assertion of Part 1.

Let us prove the equality (2.9). Since for all functions $g \in \mathrm{L}^2(\mathbb{T})$ and $h_\ell \in \mathrm{H}^2(\mathbb{T})$ for $\ell = 0, 1, \dots, n-1$, one has

$$\Big\langle g, \sum_{\ell=0}^{n-1} \overline{f}^\ell h_\ell \Big\rangle = \sum_{\ell=0}^{n-1} \big\langle g, \overline{f}^\ell h_\ell \big\rangle = \sum_{\ell=0}^{n-1} \big\langle g f^\ell, h_\ell \big\rangle,$$

and since $\langle g f^\ell, h_\ell \rangle = 0$ upon $g f^\ell \in \overline{\mathrm{H}}_0^2(\mathbb{T})$ and $h_\ell \in \mathrm{H}^2(\mathbb{T})$, then each function $g \in Q_n(f)$ is orthogonal to $\mathrm{M}_n^2(\mathbb{T}, f)$. Let us take now some function $g \in \mathrm{L}^2(\mathbb{T})$ such that $g \perp \mathrm{M}_n^2(\mathbb{T}, f)$. Then, for each $\ell \in \{0, 1, \dots, n-1\}$ one has

$$\langle g f^\ell, h \rangle = \langle g, \overline{f}^\ell h \rangle = 0$$

for any function $h \in \mathrm{H}^2(\mathbb{T})$. It means that $g f^\ell \in \overline{\mathrm{H}}_0^2(\mathbb{T})$ for all $\ell \in \{0, 1, \dots, n-1\}$ and thus $g \in Q_n(f)$. The second equality in (2.9) immediately follows from the first one and from the equality (2.8).

We start the proof of Part 3 proving the fact that $\ker(E_k(f)) \neq \varnothing$ if and only if the function f^k admits a Nevanlinna-type pseudocontinuation. Let $\ker(E_k(f)) \neq \varnothing$. It follows from (2.7), then there exists a function $h \in \mathrm{H}^2(\mathbb{T})$ such that $h \not\equiv 0$ and $w := \overline{f}^k h \in \mathrm{H}^2(\mathbb{T})$, thus $\overline{f}^k = w/h$. Since $\mathrm{H}^2(\mathbb{D}) \subset \mathrm{N}(\mathbb{D})$, then there exists two functions $f_1, f_2 \in \mathrm{H}^\infty(\mathbb{D})$ such that $\overline{f(\xi)}^k = w(\xi)/h(\xi) = f_1(\xi)/f_2(\xi)$ for almost all $\xi \in \mathbb{T}$ and in view of Lemma 1 we conclude, that the function f^k admits a Nevanlinna type pseudocontinuation. Conversely, if the function f^k admits a Nevanlinna-type pseudocontinuation and if the functions $f_1, f_2 \in \mathrm{H}^\infty(\mathbb{D})$ (with $f_2 \not\equiv 0$) are taken from Lemma 1 so that $\overline{f(\xi)}^k = f_1(\xi)/f_2(\xi)$, then $f_2 \in \mathrm{H}^2(\mathbb{T})$ and $\overline{f}^k f_2 = f_1 \in \mathrm{H}^2(\mathbb{T})$. Thus, by (2.7), $f_2 \in \ker(E_k(f))$ and, therefore, $\ker(E_k(f)) \neq \varnothing$.

Let us calculate the operators $T(f)$ and $\widetilde{T}(f)$ explicitly. Assuming that

$$h(z) = \sum_{j=0}^{\infty} \lambda_j z^j \in \mathrm{H}^2(\mathbb{T}), \quad b(z) = \sum_{j=1}^{\infty} \beta_j \overline{z}^j \in \overline{\mathrm{H}}_0^2(\mathbb{T}), \quad \text{and that} \quad f(z) = \sum_{j=0}^{\infty} \alpha_j z^j,$$

we obtain by direct computations that

$$T(f)\,h = \sum_{m=0}^{\infty} \Big(\textstyle\sum_{j=0}^{m} \alpha_j \lambda_{m-j} \Big) z^m, \quad \text{and} \quad \widetilde{T}(f)\,b = \displaystyle\sum_{m=1}^{\infty} \Big(\textstyle\sum_{j=0}^{m-1} \overline{\alpha}_j \beta_{m-j} \Big) \overline{z}^m.$$

Therefore, if J is the operator on $\mathrm{L}^2(\mathbb{T})$ acting as $Jf(z) = f(\overline{z})$, then $J^* = J$ and

$$\widetilde{T}(f) = J^*\, T(f_*)\, J,$$

where the function f_* was defined by the formula $f_*(z) = \overline{f(\overline{z})}$. Since $J^2 = 1$ (the identity operator), then

$$\widetilde{T}_k(f) = J^*\, T_k(f)\, J, \quad \text{and hence} \quad \widetilde{E}_k(f) = J^*\, E_k(f_*)\, J.$$

Since the functions f^k and $(f_*)^k = (f^k)_*$ admit or do not admit a Nevanlinna-type pseudocontinuation simultaneously, then the kernels of operators $E_k(f)$ and

$E_k(f_*)$ are empty or non empty simultaneously and it follows from the last formula, that the same is true for kernels of operators $E_k(f)$ and $\widetilde{E}_k(f)$. □

Let us illustrate the results of Theorems 2 and 3 by two simple examples. Let $p = 2$ and let $n \geq 2$ be an integer. We have

Example 1. Let $\varphi_1(z) := z$. It is easy to verify that the function φ_1 admits a Nevanlinna-type pseudocontinuation (indeed, the desired functions in Definition 2 are $F_1 \equiv 1$ and $F_2(z) = 1/z$). By Theorem 2 (Part 1) the space $\mathrm{M}^2_n(\mathbb{T}, \varphi_1)$ is not dense in $\mathrm{L}^2(\mathbb{T})$, and in view of (2.7) and (2.9) one has

$$\begin{aligned} M^2_n(\mathbb{T}, \varphi_1)^\perp &= \left\{ b \in \overline{\mathrm{H}}^2_0(\mathbb{T}) : \ z^k b \in \overline{\mathrm{H}}^2_0(\mathbb{T}), \ k = 1, 2, \dots, n-1 \right\} \\ &= \left\{ b \in \overline{\mathrm{H}}^2_0(\mathbb{T}) : \ b = \sum_{k=1}^{\infty} \beta_k \overline{z}^k, \ \beta_1 = \cdots = \beta_{n-1} = 0 \right\}. \end{aligned}$$

Example 2. Let now $\varphi_2(z) = e^z$. Since $\overline{e^z} = e^{\overline{z}} = e^{1/z}$, and since the function $e^{1/z}$ has an essential singular point at the origin, then (in view of Lemma 1) the function φ_2 does not admit a Nevanlinna-type pseudocontinuation. Thus, using Theorem 2 (Part 1) one concludes that $\mathrm{M}^2_n(\mathbb{T}, \varphi_2)$ is dense in $\mathrm{L}^2(\mathbb{T})$.

Acknowledgment

The author is grateful to the organizers of the Conference in Analysis and Mathematical Physics "New Trends in Complex and Harmonic Analysis" (May 7–12, 2007, Voss, Norway) for the organization of very interesting conference and for their warm hospitality.

References

[1] J.J. Carmona, K.Yu. Fedorovskiy and P.V. Paramonov, *On uniform approximation by polyanalytic polynomials and Dirichlet problem for bianalytic functions*, Sb. Math. **193** (2002), no. 10, 1469–1492.

[2] K.Yu. Fedorovski, *Uniform n-analytic polynomial approximations of functions on rectifiable contours in* $\mathbb{C}$, Math. Notes., 1996, **59** (4), 435–439.

[3] K.Yu. Fedorovskiy, *On some properties and examples of Nevanlinna domains*, Proc. Steklov Inst. Math., 2006, **253**, 186–194.

[4] T. Trent and J.L.-M. Wang, *Uniform approximation by rational modules on nowhere dense sets*, Proc. Amer. Math. Soc., 1981, **81** (1), 62–64.

[5] J.J. Carmona, *Mergelyan approximation theorem for rational modules*, J. Approx. Theory **44** (1985), 113–126.

[6] J.L. Wang, *A localization operator for rational modules*, Rocky Mountain J. of Math., 1989, **19** (4), 999–1002.

[7] K.Yu. Fedorovski, *Approximation and boundary properties of polyanalytic functions*, Proc. Steklov Inst. Math., 2001, **235**, 251–260.

[8] S.N. Mergelyan, *Uniform approximation to functions of a complex variable*, Amer. Math. Soc., Transl., 1954, **101**; translation from Usp. Mat. Nauk, 1952, **7** (2), 31–122.

[9] A. Boivin, P.M. Gauthier and P.V. Paramonov, *Approximation on closed sets by analytic or meromorphic solutions of elliptic equations and applications*, Canadian J. of Math., 2002, **54** (5), 945–969.

[10] A. Boivin, P.M. Gauthier and P.V. Paramonov, *Uniform approximation on closed subsets of* $\mathbb{C}$ *by polyanalytic functions*, Izv. Math., 2004, **68** (3), 447–459.

[11] A.B. Zajtsev, *On uniform approximation of functions by polynomial solutions of second-order elliptic equations on plane compact sets*, Izv. Math., 2004, **68** (6), 1143–1156.

[12] J.J. Carmona and K.Yu. Fedorovskiy, *Conformal maps and uniform approximation by polyanalytic functions*, Oper. Th. Adv. Appl., 2005, **158**, 109–130.

[13] J.J. Carmona and K.Yu. Fedorovskiy, *On the dependence of uniform polyanalytic polynomial approximations on the order of polyanalyticity*, Math. Notes, 2008, **83** (1), 31–36.

[14] H.S. Shapiro, *Generalized analytic continuation*, Symposia on Theor. Phys. and Math., 1968, **8**, 151–163.

[15] R.G. Douglas, H.S. Shapiro and A.L. Shields, *Cyclic vectors and invariant subspaces for the backward shift operator*, Annales de l'Institut Fourier, 1970, **20** (1), 37–76.

[16] G.M. Goluzin, *Geometric theory of functions of a complex variable, 2-nd edition*, "Nauka", Moscow, 1966; English translation: Amer. Math. Soc., Providence, R.I., 1969.

[17] Ch. Pommerenke, *Boundary behaviours of conformal maps*, Springer Verlag, Berlin, 1992.

[18] A. Böttcher, B. Silbermann, *Analysis of Toeplitz Operators*, Springer Verlag, Berlin, 2006.

Konstantin Yu. Fedorovskiy
e-mail: `kfedorovs@yandex.ru`

Analysis and Mathematical Physics
Trends in Mathematics, 143–166
© 2009 Birkhäuser Verlag Basel/Switzerland

Potential Theory in Denjoy Domains

Stephen J. Gardiner and Tomas Sjödin

Abstract. This paper presents an account of Denjoy domains in relation to minimal harmonic functions, the boundary behaviour of the Green function, and to their usefulness as a source of counterexamples in potential theory. The discussion begins with an exposition of key work of Ancona and Benedicks and then moves on to describe several very recent results.

Mathematics Subject Classification (2000). 31B05, 31C35, 31B25, 31C40.

Keywords. Denjoy domain, harmonic function, Martin boundary, Green function, fine potential theory.

1. Introduction

A result of Denjoy [16], dating from 1909, says that a domain of the form $\mathbb{C}\backslash E$, where $E \subset \mathbb{R}$, supports nonconstant bounded analytic functions if and only if E has positive Lebesgue measure. Domains $\Omega \subset \mathbb{C}$ for which $\mathbb{C}\backslash\Omega$ is contained in a line, or, more generally, domains $\Omega \subset \mathbb{R}^n$ $(n \geq 2)$ for which $\mathbb{R}^n\backslash\Omega$ is contained in a hyperplane, have subsequently come to be known as *Denjoy domains.* The purpose of this article is to describe how the special geometry of Denjoy domains has led to some very precise and illuminating potential theoretic results. We will not attempt to provide a comprehensive survey of the subject, but will present some of the key theory and also outline several recent developments.

Denjoy domains sometimes arise naturally in the study of certain domain properties. To take a simple example, let $\mathcal{H}_c(\Omega)$ denote the collection of harmonic functions on a domain Ω that have a finite-valued continuous extension to compactified space $\mathbb{R}^n \cup \{\infty\}$. It is easy to see, by consideration of Poisson integral representations in half-spaces, that $\mathcal{H}_c(\Omega)$ does not separate the points of Ω if Ω is a Denjoy domain. In fact, essentially only Denjoy domains have this non-separation property. More precisely, when $n \geq 3$, domains Ω with this property

This work is part of the programme of the ESF Network "Harmonic and Complex Analysis and Applications" (HCAA), and was also supported by Science Foundation Ireland under Grant 06/RFP/MAT057.

must be of the form $\omega\backslash F$, where ω is a Denjoy domain and F is a relatively closed polar subset of ω; and, when $n = 2$, the characterization is similar, except that ω can be the image of a Denjoy domain under a linear fractional transformation (for this and related results, see [9]).

Denjoy domains also arise naturally in the study of null quadrature domains, that is, domains Ω on which every integrable harmonic function has integral 0. A simple example of a domain with this property is a half-space; for, if h is an integrable harmonic function on $\omega = \mathbb{R}^{n-1} \times (0, \infty)$, then the function

$$t \mapsto \int_{\mathbb{R}} \dots \int_{\mathbb{R}} h(x_1, \dots, x_{n-1}, t)\ dx_1 \dots dx_{n-1} \qquad (t > 0)$$

is of the form $at + b$ for some $a, b \in \mathbb{R}$ (by Theorem 1.5.12 in [8]), so $a = 0 = b$ and hence $\int_\omega h = 0$. It follows that any Denjoy domain also has this property, but it seems that very few other domains have it. In fact, when $n = 2$, Sakai [23] has shown that the only other possibilities are complementary to closed elliptic or parabolic regions, and something similar has been conjectured to be true in higher dimensions (see [22]).

However, while such examples are of some interest, the main potential theoretic motivation for studying Denjoy domains is the desire to gain insight into how the geometry of a domain affects subtle potential theoretic phenomena. One such topic is the Martin boundary (see Chapter 8 of the book [8]), which provides an integral representation for positive harmonic functions on a Greenian domain in the spirit of the Riesz-Herglotz representation in a ball. Since the relationship between the Euclidean and Martin boundaries can be quite complicated, several authors, beginning with Ancona [3] and Benedicks [10], have analysed what can occur in the case of Denjoy domains. Another topic where the case of Denjoy domains has provided insight is the relationship between the geometry of a domain and the boundary behaviour of its Green function (see Carleson and Totik [14] and Carroll and Gardiner [12]). As we will explain later, these two apparently distinct questions are, in fact, intimately related.

Finally, Denjoy domains, and variants of them, can be a fruitful source of counterexamples. In the final section of this article we will illustrate this with reference to recent work of Sjödin [26] concerning integrability and positive harmonic functions, and the solution by Gardiner and Hansen [21] of a long-standing open question about the Riesz decomposition in fine potential theory. We note, in passing, that Denjoy domains have also arisen recently as natural counterexamples in connection with the study of minimal harmonic functions on John domains [2].

Other papers where potential theoretic aspects of Denjoy-type domains have been considered include [1], [4], [5], [6], [7], [13], [15], [20], [24], [25] and [27].

2. Notation and preliminary material

Here we set out our notation and recall some of the potential theoretic facts we need, a convenient reference for which is the book [8].

We will work in Euclidean space $\mathbb{R}^n$ $(n \geq 2)$, using the notation

$$x = (x_1, x_2, \ldots, x_n) = (x', x_n) \text{ and } \widetilde{x} = (x', -x_n), \text{ where } x' \in \mathbb{R}^{n-1},$$

and writing $|x|$ for the Euclidean norm of x. We define

$$H_+ = \{x \in \mathbb{R}^n : x_n > 0\},\ H_- = \{x \in \mathbb{R}^n : x_n < 0\},\ L = \{x \in \mathbb{R}^n : x_n = 0\},$$

$$a_t = (0', t) \quad (t > 0) \quad \text{and} \quad T_t = \{a_s : 0 < s < t\} \quad (0 < t \leq \infty),$$

and write $B(x, r)$ for the open ball of center x and radius r.

To avoid confusion, sequences in $\mathbb{R}^n$ will be denoted $x^{(k)}$ and not x_k (since this is used for the kth coordinate of x). Sequences in $\mathbb{R}$ and sequences of sets will, however, be represented using the familiar subscript notation.

If η is a measure, then

$$\mathcal{L}^p(\eta) := \begin{cases} \{f : |f|^p \text{ is } \eta\text{-integrable}\} & (0 < p < \infty) \\ \{f : |f| \text{ is essentially bounded}\} & (p = \infty) \end{cases}.$$

By $\operatorname{supp}(\eta)$ we will always denote the closed support of a measure η.

For an open set $D \subset \mathbb{R}^n$ we introduce the following classes of real-valued functions on D.

$\mathcal{P}(D)$: the non-negative functions on D,

$\mathcal{L}^p(D)$: the pth power Lebesgue integrable functions on D,

$\mathcal{H}(D)$: the harmonic functions on D.

We also use superpositioning so that, for instance, $\mathcal{HP}(D) = \mathcal{H}(D) \cap \mathcal{P}(D)$. If $D \subset \mathbb{R}^n$ is Greenian, then $G_D(\cdot, \cdot)$ denotes its Green function, and if μ is a measure on D we denote its Green potential by $G_D\mu$.

For a set $A \subset \mathbb{R}^n$ we denote by ∂A its boundary in $\mathbb{R}^n$, and by $\partial^\infty A$ its boundary in $\mathbb{R}^n \cup \{\infty\}$. The harmonic measure on $\partial^\infty D$ for a Greenian open set D with respect to a point $x \in D$ is denoted by λ_x^D.

We recall that, if D is a Greenian domain D, then the Martin kernel normalized at some reference point x_0 is defined on $D \times (D \setminus \{x_0\})$ by

$$M_D(x, y) := \frac{G_D(x, y)}{G_D(x_0, y)}.$$

The Martin compactification $\widehat{D}$ is the smallest compactification of D such that $M_D(x, \cdot)$ can be continuously extended to the boundary $\partial^M D = \widehat{D} \setminus D$ for each x. (We still use $M_D(\cdot, \cdot)$ for the extended function.) For each $h \in \mathcal{HP}(D)$ there is (at least) one representing measure η on $\partial^M D$ for h, that is, for which

$$h(x) = \int M_D(x, y) d\eta(y). \quad (\text{Clearly } \eta(\partial^M D) = h(x_0)).$$

In general, such measures are not unique, but if we require them to be carried by the minimal Martin boundary $\partial_1^M D$, then there is a unique measure, which we denote by ν_h.

We will also need the concept of minimal thinness (see Chapter 9 of [8]). If $y \in \partial_1^M D$ and $A \subset D$, then we recall that A is said to be minimally thin at y (with respect to D) if

$$M_D(\cdot, y) \not\equiv \widehat{R}^A_{M_D(\cdot,y)},$$

where $\widehat{R}^A_{M_D(\cdot,y)}$ denotes the regularized reduction of $M_D(\cdot, y)$ over the set A with respect to positive superharmonic functions on D.

3. Basic theory

3.1. Martin boundary points

We will follow the approach of Ancona [3] to establish a kind of boundary Harnack principle for Denjoy domains, and then use minimal thinness arguments to prove the basic properties of the set of Martin points associated with a Euclidean boundary point for these domains.

Theorem 1. *Suppose that $D \subset \mathbb{R}^n$ is a Greenian domain which contains H_+. There is a constant C_1, depending only on n, such that*

$$G_D(x, y) \leq C_1 G_D(x, a_{|y|}) \quad (x \in T_\infty; y \in D \setminus \{0\}). \tag{1}$$

Proof. (I) *For all $x \in H_+$ and $y \in D \cap H_-$ we have $G_D(x, y) \leq G_D(x, \tilde{y})$.* This holds because the function

$$u(y) := G_D(x, \tilde{y}) - G_D(x, y)$$

is superharmonic and lower bounded in $D \cap H_-$ and has non-negative lower limit quasi-everywhere on $\partial^\infty(D \cap H_-)$. Hence $u \geq 0$ in $D \cap H_-$ by the minimum principle.

We now identify $\mathbb{R}^n$ with $\mathbb{R}^{n-2} \times \mathbb{C}$ by letting

$$(x_1, \ldots, x_{n-2}, x_{n-1}, x_n) = (x_1, \ldots, x_{n-2}, x_{n-1} + ix_n) = (x'', z),$$

and use $\arg z$ to denote the value of the argument in $(-\pi, \pi]$.

(II) *There is a constant d_1, depending only on n, such that every positive harmonic function h on the set*

$$\{(x'', z) : 0 < \arg z < \pi/2\}$$

satisfies

$$h(x'', te^{i\theta}) \leq d_1 h(x'', te^{i\varphi}) \quad (\theta, \varphi \in [\pi/8, 3\pi/8]; t > 0).$$

This is an easy consequence of Harnack's inequalities.

Let s denote the reflection map with respect to the hyperplane

$$\{(w'', z) \in \mathbb{R}^n : \arg z \in \{-7\pi/8, \pi/8\}\}.$$

(III) *For all $x \in T_\infty$ and $y = (y'', z)$ with $\arg z = -\pi/8$ we have*

$$G_D(x, y) \leq d_1 G_D(x, s(y)).$$

This follows by combining (I) and (II).

Now, for fixed $x \in T_\infty$, we apply the minimum principle to the function

$$w(y) := d_1 G_D(x, s(y)) - G_D(x, y)$$

in the set

$$D \cap \{(y'', z) \in \mathbb{R}^n : -\tfrac{\pi}{8} < \arg z < \tfrac{\pi}{8}\}$$

to see from (III) that

$$\sup\left\{G_D(x, (y'', z)) : y'' \in \mathbb{R}^{n-2}, \arg z \in \left[-\tfrac{\pi}{8}, \tfrac{\pi}{8}\right]\right\} \\ \leq d_1 \sup\left\{G_D(x, (y'', z)) : y'' \in \mathbb{R}^{n-2}, \arg z \in \left[\tfrac{\pi}{8}, \tfrac{3\pi}{8}\right]\right\}.$$

By rotating the coordinate system around the x_n-axis and letting $\mathcal{K}$ denote the cone around T_∞ with vertex 0 and half-angle $3\pi/8$ we now obtain:

(IV) *For all $x \in T_\infty$ and $t > 0$ we have*

$$\sup\{G_D(x, y) : |y| = t\} \leq d_1 \sup\{G_D(x, y) : |y| = t, y \in \mathcal{K}\}.$$

Applying Harnack's inequality we get:

(V) *There is a constant d_2, depending only on n, such that, for all $t > 0$, $y \in D \cap \partial B(0, t)$, and $x \in T_\infty$ with $|x| \leq t/2$ or $|x| \geq 2t$,*

$$G_D(x, y) \leq d_2 G_D(x, a_t).$$

To complete the argument let $t = |y|$ and $t/2 < |x| < 2t$. The case $|x| = t$ is trivial, so we assume that $|x| \neq t$. The function $G_D(x, \cdot)$ is harmonic outside $\overline{B}(x, |t - |x|| / 2)$ and is majorized on $\partial B(x, |t - |x|| / 2)$ by $d_3 G_D(x, a_t)$ for some constant d_3 depending only on n, in view of Harnack's inequalities. Letting $C_1 = \max\{d_2, d_3\}$ we now obtain (1). □

Corollary 2. *Suppose $D \subset \mathbb{R}^n$ is a Greenian domain which contains H_+. There is a constant C_2, depending only on n, such that*

$$G_D(x, y) \leq C_2 G_D(x, a_{|y|}) \quad (x \in H_+; y \in D \setminus B(0, 2|x|)). \tag{2}$$

Proof. Let $x = (x', x_n)$ and $a^*_{|y|} = (x', a_{|y-(x',0)|})$. By Theorem 1 we have

$$G_D(x, y) \leq C_1 G_D(x, a^*_{|y|}).$$

But the function $G_D(x, \cdot)$ is positive and harmonic on $H_+ \setminus \overline{B(0, |y|/2)}$ and so, by Harnack's inequalities,

$$G_D(x, a^*_{|y|}) \leq d_1 G_D(x, a_{|y|})$$

for some constant d_1 depending only on n. Thus (2) holds with $C_2 = C_1 d_1$. □

Theorem 3. *Suppose $D \subset \mathbb{R}^n$ is a Greenian domain which contains H_+. There is a constant C_3, depending only on n, such that, for any $\alpha > 0$,*

$$\frac{G_D(x, a_t)}{G_D(x, a_{2\alpha})} \leq C_3 \frac{G_D(y, a_t)}{G_D(y, a_{2\alpha})} \quad (x, y \in H_+ \cap B(0, \alpha); t \geq 10\alpha).$$

Proof. By dilation we may assume, without loss of generality, that $\alpha = 2$. From Corollary 2

$$G_D(x,v) \le C_2 G_D(x,a_4) \quad (x \in H_+ \cap B(0,2); v \in \partial B(0,4) \cap D).$$

Let $R = D \setminus \overline{B(0,4)}$. Now we have

$$G_D(x,v) \le C_2 G_D(x,a_4)\lambda_v^R(\partial B(0,4)) \quad (x \in H_+ \cap B(0,2); v \in R). \tag{3}$$

Fix $\phi \in C^\infty(\mathbb{R}^n)$ with support in $\{3 \le |v| \le 5\}$ such that $0 \le \phi \le 1$, and also $\phi = 1$ on $\partial B(0,4)$. Now

$$\lambda_v^R = \beta_n \Delta G_R(v,\cdot) + \delta_v,$$

where $(\beta_n \max\{1, n-2\})^{-1}$ is the surface area of $\partial B(0,1)$ and $G_R = 0$ outside $R \times R$, so

$$\begin{aligned}\lambda_v^R(\partial B(0,4)) &\le \int \phi(w) d\lambda_v^R(w) \\ &= \int \beta_n G_R(v,w)\Delta\phi(w)dw \quad (v \in D \setminus B(0,5)).\end{aligned}$$

Hence

$$\lambda_v^R(\partial B(0,4)) \le \beta_n ||\Delta\phi||_1 \sup\{G_R(v,w) : |w| \le 5\} \quad (v \in D \setminus B(0,5)). \tag{4}$$

Now suppose that $t > 13$. By Theorem 1 applied to R,

$$\begin{aligned}\sup\{G_R(a_t,w) : |w| \le 5\} &\le \sup\{G_R(a_t,w) : |w - a_4| \le 9\} \\ &\le C_1 G_R(a_t,a_{13}).\end{aligned}$$

In view of (3) and (4) we now have

$$\begin{aligned}\frac{G_D(x,a_t)}{G_D(x,a_4)} &\le C_2 \lambda_{a_t}^R(\partial B(0,4)) \\ &\le \beta_n C_2 ||\Delta\phi||_1 \sup\{G_R(a_t,w) : |w| \le 5\} \\ &\le \beta_n C_1 C_2 ||\Delta\phi||_1 G_R(a_t,a_{13}) \\ &= d_1 G_R(a_t,a_{13}), \quad \text{say,}\end{aligned} \tag{5}$$

for all $x \in H_+ \cap B(0,2)$ and all $t \ge 20$, where d_1 depends only on n and our choice of ϕ. However, by Harnack's inequality, there are positive constants d_2, d_3, depending only on n, such that

$$\frac{G_D(y,w)}{G_D(y,a_4)} \ge d_2 > 0 \,\text{and}\, G_R(a_{13},w) \le d_3 \ (y \in H_+ \cap B(0,2); w \in \partial B(a_{13},2)).$$

By the maximum principle

$$\frac{G_D(y,a_t)}{G_D(y,a_4)} \ge \frac{d_2}{d_3} G_R(a_{13},a_t) \quad (t \ge 20).$$

Hence

$$\frac{G_D(x,a_t)}{G_D(x,a_4)} \le d_1 G_R(a_t,a_{13}) \le \frac{d_1 d_3}{d_2}\frac{G_D(y,a_t)}{G_D(y,a_4)} \quad (t \ge 20),$$

in view of (5), and the result is now established with $C_3 = d_1 d_3 / d_2$. □

Corollary 4. *Suppose $D \subset \mathbb{R}^n$ is a Greenian domain which contains H_+. Then the closure $\overline{T_1}^M$ of T_1 in $\widehat{D}$ intersects $\partial^M D$ in exactly one point y, and $y \in \partial_1^M D$.*

Proof. Let $M_D(z,y)$ denote the Martin kernel of D with reference point a_{10}. By Theorem 3

$$\frac{1}{C_3} M_D(a_{2\alpha}, x) \le M_D(a_{2\alpha}, y) \le C_3 M_D(a_{2\alpha}, x)$$

for all $x, y \in H_+ \cap B(0,\alpha)$ and $\alpha \le 1$. By continuity this inequality also holds for $x, y \in \overline{H_+ \cap B(0,\alpha)}^M$. Now fix $x \in \cap_{\alpha>0} \overline{H_+ \cap B(0,\alpha)}^M$. If $t < 1$, then by Harnack's inequalities,

$$M_D(z,x) \ge d_1 M_D(z,a_t) \frac{M_D(a_{2t},x)}{M_D(a_{2t},a_t)} \qquad (z \in \partial B(a_t, t/2)),$$

where d_1 depends only on n, whence by the minimum principle,

$$M_D(z,x) \ge \frac{d_1}{C_3} M_D(z,a_t) = d_2 M_D(z,a_t), \text{ say} \quad (z \in D \setminus B(a_t, t/2)). \tag{6}$$

If $a_{t_k} \to y \in \overline{T_1}^M \cap \partial^M D$, it follows that

$$M_D(z,x) \ge d_2 M_D(z,y) \qquad (z \in D). \tag{7}$$

Let

$$B_k = B(a_{2^{-k}}, 2^{-k-1}) \text{ and } D_k = \bigcup_{j \ge k} B_j.$$

From (6) we have

$$\begin{aligned} \widehat{R}^{D_k}_{M_D(\cdot,x)}(a_{10}) &\ge \widehat{R}^{B_k}_{M_D(\cdot,x)}(a_{10}) \ge d_2 \widehat{R}^{B_k}_{M_D(\cdot,a_{2^{-k}})}(a_{10}) \\ &\ge d_2 M_D(a_{10}, a_{2^{-k}}) = d_2 \end{aligned}$$

for all k, and so D_1 cannot be minimally thin at all points of $\partial_1^M D$. In particular, $\overline{D_1}^M \cap \partial_1^M D \ne \emptyset$. Further, by (7) and minimality, the sets $\overline{D_1}^M \cap \partial_1^M D$ and $\overline{T_1}^M \cap \partial^M D$ coincide and consist of exactly one point. □

We observe that, in the above proof, we have also shown that

$$\bigcap_{\alpha>0} \overline{H_+ \cap B(0,\alpha)}^M$$

contains exactly one minimal point. (It may also possibly contain non-minimal points.)

By inversion it is easy to see that the preceding results have analogues when a domain satisfies an inner ball condition. In particular, if a Greenian domain $D \subset \mathbb{R}^n$ contains the ball $B(x,r)$ and $y \in \partial D \cap \partial B(x,r)$, then there is exactly one minimal point associated to y that can be reached from the ball $B(x,r)$.

For the remainder of Section 3 we use Ω to denote a Denjoy domain of the form $\mathbb{R}^n \setminus E$, where $E \subset L = \mathbb{R}^{n-1} \times \{0\}$. Further, when $n = 2$, we require that E be non-polar. We define

$$\mathcal{M}_E = \bigcap_{\alpha>0} \overline{(\Omega \cap B(0,\alpha))}^M ;$$

that is, $\mathcal{M}_E$ is the set of all Martin boundary points (not necessarily minimal) associated with 0. Further, we let $\mathcal{P}_E$ denote the positive convex cone generated by the functions $\{M_\Omega(\cdot, y) : y \in \mathcal{M}_E\}$. We will say that $\mathcal{P}_E$ is k-dimensional if the minimum number of functions in $\mathcal{P}_E$ whose positive linear combinations span $\mathcal{P}_E$ is k.

Remark 5. The functions in $\mathcal{P}_E$ are the same as the set of positive harmonic functions which are bounded outside every neighbourhood of the origin and vanish continuously at every regular boundary point apart from 0. (This will become apparent from the proof of the next theorem.)

The following theorem essentially corresponds to Theorems 2 and 3 in Benedicks [10], but has a somewhat different formulation. Chevallier [13] was the first to exploit minimal thinness in this connection.

Theorem 6. *The set $\mathcal{M}_E \cap \partial_1^M \Omega$ consists of either one or two points, and the corresponding functions span $\mathcal{P}_E$. Further:*

(1) *$\mathcal{P}_E$ is one-dimensional if and only if one of the following equivalent conditions holds:*
   a) *all functions in $\mathcal{P}_E$ are symmetric with respect to L;*
   b) *there is no point y in $\mathcal{M}_E \cap \partial_1^M \Omega$ such that $\Omega \cap L$ is minimally thin at y.*

(2) *$\mathcal{P}_E$ is two-dimensional if and only if one of the following equivalent conditions holds:*
   a) *there is a function in $\mathcal{P}_E$ which is not symmetric with respect to L;*
   b) *there is a point y in $\mathcal{M}_E \cap \partial_1^M \Omega$ such that $\Omega \cap L$ is minimally thin at y.*

Proof. By Corollary 4 and the subsequent remark we know that

$$\bigcap_{\alpha>0} \overline{(H_+ \cap B(0,\alpha))}^M \quad \text{and} \quad \bigcap_{\alpha>0} \overline{(H_- \cap B(0,\alpha))}^M$$

each contain exactly one minimal point. We denote these by y_+ and y_-, respectively. (They may or may not be equal.) It is easy to see that

$$\bigcap_{\alpha>0} \overline{(\Omega \cap B(0,\alpha))}^M = \bigcap_{\alpha>0} \overline{(H_+ \cap B(0,\alpha))}^M \cup \bigcap_{\alpha>0} \overline{(H_- \cap B(0,\alpha))}^M .$$

From this we see that $\mathcal{M}_E \cap \partial_1^M \Omega = \{y_+, y_-\}$. We now wish to prove that any function $M_\Omega(\cdot, y)$, where $y \in \mathcal{M}_E$, can be written in the form

$$M_\Omega(\cdot, y) = c_+ M_\Omega(\cdot, y_+) + c_- M_\Omega(\cdot, y_-),$$

for some constants c_+, c_-. To do this it is enough to prove that

$$\widehat{R}^{\Omega\cap B(0,10\alpha)}_{M_\Omega(\cdot,y)} = M_\Omega(\cdot,y) \quad (\alpha > 0), \tag{8}$$

because every point z in $\partial\Omega$ has at most two minimal points associated with it, and, as will become clear below, $\Omega \setminus B(z,r)$ is minimally thin at both of these for any $r > 0$ (similar statements hold for ∞, by inversion). From Corollary 2, Theorem 3 and Harnack's inequality we see that there is a constant d (depending on n and α) such that

$$M_\Omega(z,y) \le d(M_\Omega(a_{|z|},a_t) + M_\Omega(-a_{|z|},-a_t))$$

for all $z \in \Omega \setminus B(0,10\alpha)$, $y \in \Omega \cap B(0,\alpha)$ and $0 < t < \alpha$. By continuity this holds also for $y \in \overline{\Omega \cap B(0,\alpha)}^M$ and $\pm a_t$ replaced by $y_\pm$. Suppose now that $y^{(k)}$ is a sequence in Ω converging to 0 in the Euclidean topology, and to some point $y \in \mathcal{M}_E$. By the above estimates (if we assume $y^{(k)} \in B(0,\alpha)$ for all k) we see that $M_\Omega(\cdot,y^{(k)})$ converges to $M_\Omega(\cdot,y)$ with bounded convergence on $\partial B(0,10\alpha)\cap\Omega$, and so

$$\widehat{R}^{\Omega\cap B(0,10\alpha)}_{M_\Omega(\cdot,y^{(k)})} \to \widehat{R}^{\Omega\cap B(0,10\alpha)}_{M_\Omega(\cdot,y)}.$$

Since

$$\widehat{R}^{\Omega\cap B(0,10\alpha)}_{M_\Omega(\cdot,y^{(k)})} = M_\Omega(\cdot,y^{(k)})$$

for each k, we obtain (8), whence the functions corresponding to $\mathcal{M}_E \cap \partial_1^M\Omega$ span $\mathcal{P}_E$.

We can change our normalization of $M_\Omega(\cdot,\cdot)$ to ensure that

$$M_\Omega(x,y_+) = M_\Omega(\tilde{x},y_-),$$

whence $y_+ \ne y_-$ if and only if $M_\Omega(x,y_+) \ne M_\Omega(\tilde{x},y_+)$ for some $x \in \Omega$.

If $\Omega \cap L$, which is the same as $\Omega \setminus (H_+ \cup H_-)$, is minimally thin at y_+, then so is $\Omega \setminus H_+$ (because either $\Omega \setminus H_+$ or $\Omega \setminus H_-$ must be minimally thin at y_+, by Lemma 9.6.1 in [8], and it cannot be the latter). By symmetry, $\Omega\setminus H_-$ is minimally thin at y_-, so $y_+ \ne y_-$.

On the other hand, if $y_+ \ne y_-$, then $y_+ \notin \overline{H_-}^M$, by the observation following Corollary 4, so $\Omega \setminus H_+$ is certainly minimally thin at y_+, and similarly $\Omega \setminus H_-$ is minimally thin at y_-. □

Again there is nothing special about the boundary point 0 in the above result. The same phenomenon holds for all boundary points, including ∞ (as can be seen by means of an inversion). When constructing certain examples we will actually work with ∞ for reasons of notational convenience.

Some simple examples are as follows. Firstly, if E contains $B(0,\varepsilon) \cap L$ for some $\varepsilon > 0$, then it is clear that $\mathcal{P}_E$ is two-dimensional. In fact, $\mathcal{M}_E$ consists of exactly two points in this case, and so, in particular, it is not connected. On the other hand, if there is a sequence of balls $B(x^{(k)}, r_k)$ in Ω such that $x^{(k)} \in L$, $x^{(k)} \to 0$ and $r_k \ge c|x^{(k)}|$ for some $c \in (0,1)$, then $\mathcal{P}_E$ is one-dimensional. The reason for this is that $a_{|x^{(k)}|}$ can be connected to $-a_{|x^{(k)}|}$ by a Harnack chain whose length does not depend on k, and so the sequences $\left(\pm a_{|x^{(k)}|}\right)$ converge to the same minimal point.

Example 7. We now outline a less trivial example, due to Ancona [3], where $\dim\mathcal{P}_E = 2$. As we will work in the plane it is convenient to use complex notation. Let

$$D = \mathbb{C} \setminus \bigcup_{k\in\mathbb{Z}} [k + 1/4, k + 3/4]\,,$$

and

$$\Omega = \{z \in \mathbb{C} : 1/z \in D\} = \mathbb{C} \setminus \left(\bigcup_{k\in\mathbb{Z}} \left[\frac{1}{k+3/4}, \frac{1}{k+1/4}\right] \cup \{0\} \right).$$

We denote the two (possibly equal) minimal points associated with ∞ on D by w_+ and w_- respectively. We want to prove that $w_+ \neq w_-$. From Theorem 6 we know that $w_+ = w_-$ if and only if the set $D \cap \mathbb{R}$ is not minimally thin at w_+. We also know that $it \to w_+$ as $t \to +\infty$, so it is enough to prove that the unbounded function $M_D(\cdot, w_+)$ is bounded above by some positive constant on $D \cap \mathbb{R}$. To prove this we see from symmetry that

$$G_D(p, it) = G_D(0, p + it) \quad (p \in \mathbb{Z}; t \in (0, \infty)).$$

Hence, by Theorem 1 and Harnack's inequalities,

$$\begin{aligned} G_D(p, it) = G_D(0, p + it) &\le C_1 G_D(0, i\sqrt{p^2 + t^2}) \\ &\le C_1 d_1 G_D(0, it) \quad (p \in \mathbb{Z}; t \ge 2p), \end{aligned}$$

where d_1 is an absolute constant. Thus

$$M_D(p, w_+) = \lim_{t\to\infty} \frac{G_D(p, it)}{G_D(0, it)} \le C_1 d_1 \quad (p \in \mathbb{Z}).$$

Harnack's inequalities can be applied to the circles $\partial B(k, 1)$ to see that there is a constant d_2 such that

$$M_D(x, w_+) \le d_2 \quad (x \in \partial B(k, 1); k \in \mathbb{Z}),$$

and this bound also holds on $B(k, 1) \cap D$ by the maximum principle. In particular, $M_D(\cdot, w_+)$ is bounded on $D \cap \mathbb{R}$, and so $w_+ \neq w_-$.

3.2. A Wiener-type criterion

We have seen that $\mathcal{P}_E$ is either one- or two-dimensional. Benedicks [10] provided an integrated harmonic measure criterion involving "moving cubes" for distinguishing between these two cases. This criterion was recently shown by Carroll and Gardiner [12] to be equivalent to one involving capacity. In this section we will combine ideas from [10] and [12] to give a direct proof of this Wiener-type characterization.

Let $\mathcal{C}(A)$ denote the Newtonian (or logarithmic, if $n = 2$) capacity of a set A. Also, let $\gamma \in (0, 1/3)$ and $D(r) = L \cap \overline{B(0, r)}$ and, for any $k = 0, 1, \dots,$ let $D_k = D(2^{-k})$ and

$$E_k = (E \cap D_k) \cup D(\gamma 2^{-k}) \cup \overline{D_k \backslash D((1-\gamma)2^{-k})}.$$

In the sequel we will use $C(n,\ldots)$ to denote a constant depending at most on $n,\ldots$; its value may change from line to line.

Theorem 8. *For a Denjoy domain $\Omega = \mathbb{R}^n \backslash E$ with $0 \in E$ the following statements are equivalent:*

(a) *$\mathcal{P}_E$ is two-dimensional;*

(b) $\begin{cases} \sum 2^{k(n-2)} \left[\mathcal{C}(D_k) - \mathcal{C}(E_k)\right] < \infty & (n \geq 3) \\ \sum 2^{k} \left[\mathcal{C}(D_k) - \mathcal{C}(E_k)\right] < \infty & (n = 2) \end{cases}$.

In the proof of Theorem 8 we will make use of the following elementary version of the boundary Harnack principle (see Lemma 8.5.1 in [8]).

Lemma 9. *Let $z \in L$ and $0 < \alpha < 1$. If g, h are positive harmonic functions on $B(z,r) \cap H_+$ which continuously vanish on $B(z,r) \cap L$, then g/h has a positive continuous extension to $B(z,r) \cap (H_+ \cup L)$ and*

$$\frac{g(x)}{h(x)} \leq C(n,\alpha)\frac{g(y)}{h(y)} \quad \text{for any } \; x, y \in B(z,\alpha r) \cap H_+.$$

We will give a proof of Theorem 8 when $n \geq 3$ and leave the adjustments required for the plane case to the reader. For any compact set $K \subset \mathbb{R}^n$, we denote by v_K the capacitary potential of K, and by μ_K the associated Riesz measure. We also write

$$A_k = \{x' : 2^{-k} \leq 3\,|x'| \leq 2^{1-k}\} \;\text{ and }\; A_k^* = \{x' : \gamma 2^{-k} \leq |x'| \leq (1-\gamma)2^{-k}\}.$$

Suppose firstly that $\mathcal{P}_E$ is two-dimensional. By Theorem 6 there is a function in $\mathcal{P}_E$ which is not symmetric with respect to L. It follows, by consideration of Poisson integral representations in H_+ and in H_-, and by symmetrization, that there is a function u in $\mathcal{P}_E$ satisfying $u \geq |h|$, where $h(x) = x_n\,|x|^{-n}$. Now $1 - v_{E_k}$ vanishes on $D(\gamma 2^{-k})$ and $D_k \backslash D((1-\gamma)2^{-k})$. It thus follows from Lemma 9 and Harnack's inequalities that

$$\frac{1 - v_{E_k}}{u} \leq \frac{1 - v_{E_k}}{|h|} \leq C(n,\gamma)\frac{1 - v_{E_k}(0', 2^{-k})}{h(0', 2^{-k})} \leq C(n,\gamma)2^{k(1-n)}$$

on $\partial B(0,(1-\gamma/2)2^{-k})\backslash L$ and on $\partial B(0,\gamma 2^{-k-1})\backslash L$. We can therefore apply the maximum principle on the open set

$$B(0,(1-\gamma/2)2^{-k})\backslash[E \cup \overline{B(0,\gamma 2^{-k-1})}]$$

to see that

$$1 - v_{E_k} \leq C(n,\gamma)2^{k(1-n)}u \;\text{ on }\; (A_k^* \times \{0\})\backslash E. \tag{9}$$

Also, $d\mu_{D_0}(x', x_n) = f\left(|x'|\right) dx' d\delta_0$, where δ_0 is the Dirac measure at 0 in $\mathbb{R}$ and $f : [0,1) \to (0,\infty)$ is continuous. (This can be shown using Green's theorem and the fact that the function $x' \mapsto \lim_{t\to 0+}(1 - v_{D_0}(x',t))/t$ is positive and continuous on $\{|x'| < 1\}$, by Lemma 9.) Letting $c_1 = \max_{[0,1-\gamma]} f$, we can thus use dilation to see that

$$d\mu_{D_k} \leq 2^k c_1 dx' d\delta_0 \;\text{ on }\; D((1-\gamma)2^{-k}). \tag{10}$$

Since $v_{D_k} = 1$ on D_k and $E_k \subseteq D_k$, we have

$$\begin{aligned}
&\mathcal{C}(D_k) - \mathcal{C}(E_k) \\
&\quad = \mathcal{C}(D_k) - \int_{E_k} v_{D_k}(y) d\mu_{E_k}(y) = \mathcal{C}(D_k) - \int_{E_k} \int_{D_k} \frac{d\mu_{D_k}(x)}{|x-y|^{n-2}} d\mu_{E_k}(y) \\
&\quad = \int_{D_k} \left(1 - \int_{E_k} \frac{d\mu_{E_k}(y)}{|x-y|^{n-2}} \right) d\mu_{D_k}(x) = \int_{D_k} (1 - v_{E_k})\, d\mu_{D_k} \qquad (11) \\
&\quad \leq 2^k c_1 \int_{A_k^*} (1 - v_{E_k}(x', 0)) dx' \qquad (12) \\
&\quad \leq C(n, \gamma) 2^{k(2-n)} \int_{A_k^*} u(x', 0) dx',
\end{aligned}$$

by (10) and then (9). Condition (b) now follows from the integrability of u on D_0.

The elementary lemma given below will be used in the proof of the converse. Let $B = B(0,1)$ and let σ denote surface area measure on ∂B.

Lemma 10. *Let $q \in (0,1)$ be such that $\sigma(S_1) = \sigma(\partial B)/2$, where $S_1 = \{y \in \partial B : |y_n| \geq q\}$, and let $U = B \backslash F$, where F is a closed subset of L and $0 \notin F$. Then $\lambda_0^U(\partial B) \leq 2\lambda_0^U(S_1)$.*

Proof of the lemma. Let $S_2 = \partial B \backslash S_1$ and $u_i(x) = \lambda_x^B(S_i)$ $(i = 1, 2)$. Clearly $u_1(0) = u_2(0) = \frac{1}{2}$. Further, it follows from the maximum principle and considerations of symmetry that, for any $\varepsilon > 0$, the set $\{u_2 > \frac{1}{2} - \varepsilon\}$ contains $B \cap L$. Hence $u_2 \geq \frac{1}{2} \geq u_1$ on $L \cap B$. Since

$$\lambda_x^U(S_i) = u_i(x) - \int_{B \cap \partial U} u_i d\lambda_x^U \qquad (x \in U)$$

and $B \cap \partial U \subset B \cap L$, we see that

$$\lambda_0^U(S_2) \leq \frac{1}{2} - \int_{B \cap \partial U} u_1 d\lambda_0^U = \lambda_0^U(S_1),$$

and the lemma follows. □

If $x' \in A_k$, we define

$$B_{x'} = B\left((x', 0), \left(\frac{1}{3} - \gamma\right) 2^{-k}\right) \quad \text{and} \quad S_{x'} = \left\{ y \in \partial B_{x'} : y_n > q \left(\frac{1}{3} - \gamma\right) 2^{-k} \right\}.$$

It follows from Lemma 10, symmetry, dilation and the maximum principle that

$$\begin{aligned}
\lambda_{(x',0)}^{B_{x'} \backslash E}(\partial B_{x'}) &\leq 4\lambda_{(x',0)}^{B_{x'} \backslash E}(S_{x'}) \leq 4 \frac{1 - v_{E_k}(x', 0)}{\min_{S_{x'}} (1 - v_{E_k})} \\
&\leq 4 \frac{1 - v_{E_k}(x', 0)}{\min_{S_{x'}} (1 - v_{D_k})} \leq C(n, \gamma) \left(1 - v_{E_k}(x', 0)\right).
\end{aligned}$$

Letting $c_2 = \min_{[0,2/3]} f$, we can now argue as in the first part of the proof to see that

$$\int_{A_k} \lambda_{(x',0)}^{B_{x'}\backslash E}(\partial B_{x'})dx' \le c_2^{-1} 2^{-k} C(n,\gamma) \int_{D_k} (1 - v_{E_k})\, d\mu_{D_k}$$
$$= C(n,\gamma) 2^{-k} \left[\mathcal{C}(D_k) - \mathcal{C}(E_k)\right]. \qquad (13)$$

Now suppose that $\mathcal{P}_E$ is one-dimensional and let $u \in \mathcal{P}_E$. It follows from Theorem 6 that u is symmetric with respect to L and it is not difficult to see that

$$u(0',t) = C(n) \int_{\mathbb{R}^{n-1}} \frac{tu(x',0)}{\left\{|x'|^2 + t^2\right\}^{n/2}} dx' \quad (t > 0). \qquad (14)$$

Since $t^{n-1}u(0',t) \downarrow 0$ as $t \to 0+$, we have

$$t^{n-1}u(0',t) \le 2^{j(1-n)} u(0', 2^{-j}) \quad (0 < t \le 2^{-j}). \qquad (15)$$

It follows easily from Theorem 1 that

$$u(x) \le C(n) u(0, |x|) \qquad (x \in \Omega), \qquad (16)$$

and so, by Harnack's inequalities,

$$u(x',0) = \int u d\lambda_{(x',0)}^{B_{x'}\backslash E} \le C(n) u(0,|x'|) \lambda_{(x',0)}^{B_{x'}\backslash E}(\partial B_{x'}) \quad (x' \ne 0').$$

Combining this with (14), we obtain

$$u(0', 2^{-j}) \le C(n) \int_{\mathbb{R}^{n-1}} \frac{2^{-j} u(0,|x'|)}{\left\{|x'|^2 + 2^{-2j}\right\}^{n/2}} \lambda_{(x',0)}^{B_{x'}\backslash E}(\partial B_{x'}) dx'$$
$$= C(n)\{J_1 + J_2 + J_3\}, \qquad (17)$$

where J_1, J_2 and J_3 are integrals of the preceding integrand over

$$\{|x'| \le 2^{1-j}/3\}, \quad \{2^{1-j}/3 < |x'| \le 2/3\} \text{ and } \{|x'| > 2/3\},$$

respectively. Using (15), and then (13), we see that

$$J_1 \le \int_{\{|x'| \le 2^{1-j}/3\}} \frac{2^{-jn} |x'|^{1-n} u(0', 2^{-j})}{2^{-jn}} \lambda_{(x',0)}^{B_{x'}\backslash E}(\partial B_{x'}) dx'$$
$$\le 3^{n-1} u(0', 2^{-j}) \sum_{k=j}^{\infty} 2^{k(n-1)} \int_{A_k} \lambda_{(x',0)}^{B_{x'}\backslash E}(\partial B_{x'}) dx'$$
$$\le C(n,\gamma) u(0', 2^{-j}) \sum_{k=j}^{\infty} 2^{k(n-2)} \left[\mathcal{C}(D_k) - \mathcal{C}(E_k)\right]. \qquad (18)$$

Since, by the maximum principle and (16),

$$u \le \sup_{\partial B(0,r)} u \le C(n) u(0', r) \text{ on } \Omega \backslash \overline{B(0,r)},$$

we can also use (13) and Harnack's inequalities to see that

$$\begin{aligned} J_2 &\leq C(n) \int_{\{2^{1-j}/3<|x'|\leq 2/3\}} \frac{2^{-j}u(0',2^{1-j}/3)}{|x'|^n} \lambda_{(x',0)}^{B_{x'}\backslash E}(\partial B_{x'})dx' \\ &\leq C(n,\gamma)u(0',2^{-j})2^{-j}\sum_{k=1}^{j-1} 2^{k(n-1)}\left[\mathcal{C}(D_k)-\mathcal{C}(E_k)\right]. \end{aligned} \tag{19}$$

Also,

$$J_3 \leq C(n)u(0',2^{-j})2^{-j}\int_{\{|x'|>2/3\}} |x'|^{-n}\,dx' = C(n)u(0',2^{-j})2^{-j}. \tag{20}$$

Combining (17)–(20) yields

$$\begin{aligned} 1 \leq C(n,\gamma)\Bigg\{ &\sum_{k=j}^{\infty} 2^{k(n-2)}\left[\mathcal{C}(D_k)-\mathcal{C}(E_k)\right] \\ &+ 2^{-j}\sum_{k=1}^{j-1} 2^{k(n-1)}\left[\mathcal{C}(D_k)-\mathcal{C}(E_k)\right] + 2^{-j}\Bigg\} \quad (j\geq 1), \end{aligned}$$

which is incompatible with convergence of $\sum_k 2^{k(n-2)}\left[\mathcal{C}(D_k)-\mathcal{C}(E_k)\right]$. This completes the proof of Theorem 8.

3.3. Examples

We will now briefly illustrate when the criterion in Theorem 8 holds. When $n=2$, we know that $\mathcal{C}(D_k)=2^{-k-1}$ and $\mathcal{C}(E_k)\geq l_1(E_k)/4$, where l_m denotes m-dimensional measure, so from Theorem 8 it follows that $\mathcal{P}_E$ is two-dimensional if $\sum 2^k l_1(D_k\backslash E_k)<\infty$. However, a sharper result is true.

Corollary 11. *Let $n\geq 2$. If $\sum 2^{nk}\left[l_{n-1}(D_k\backslash E_k)\right]^{n/(n-1)}<\infty$, then $\mathcal{P}_E$ is two-dimensional.*

The corollary will follow, by dilation, once we establish the following lemma. (We present the details only for the case where $n\geq 3$.)

Lemma 12. *If $F=D(1)\backslash W$, where W is a relatively open subset of $D(1-\gamma)$, then*

$$\mathcal{C}(D(1))-\mathcal{C}(F)\leq C(n,\gamma)\left[l_{n-1}(W)\right]^{n/(n-1)}.$$

Proof. To see this, we choose ρ such that $l_{n-1}(D(\rho))=2l_{n-1}(W)$. Thus $\rho=C(n)\left(l_{n-1}(W)\right)^{1/(n-1)}$. We may assume, without loss of generality, that $\rho<2^{-3/2}\gamma$. Let

$$U=\{(x',x_n):|x'-z'|<|x_n|<\rho \text{ for some } (z',0)\in W\}\cup W,$$

let $h(x)=\lambda_x^{\omega}(\partial B(0,1))$, where $\omega=(B(0,1)\backslash L)\cup W$, and let $m=\sup_W h$. Elementary estimates of harmonic measure show that there is a constant $c_3\in(0,1)$, depending only on n and γ, such that $\lambda_x^{H_+}(W) \leq c_3$ when $x \in H_+ \cap \partial U$. Let $V = B(0,1) \cap H_+$. Since

$$h(x) = \lambda_x^V(H_+ \cap \partial B(0,1)) + \int_W h\, d\lambda_x^V \qquad (x \in V),$$

$$\int_W h\, d\lambda_x^V \leq \int_W h\, d\lambda_x^{H_+} \leq c_3 m \qquad (x \in H_+ \cap \partial U),$$

and

$$\lambda_x^V(H_+ \cap \partial B(0,1)) \leq C(n,\gamma)x_n \qquad (x \in H_+ \cap B(0, 1-\gamma/2)),$$

we see that

$$h \leq C(n,\gamma)\rho + c_3 m \quad \text{on} \quad H_+ \cap \partial U,$$

whence

$$m = \sup_W h \leq C(n,\gamma)(1-c_3)^{-1}\rho.$$

Finally, we use the fact that $1 - v_F \leq h$ on $B(0,1)\backslash F$ to see (as in (11)) that

$$\begin{aligned}\mathcal{C}(D(1)) - \mathcal{C}(F) &= \int_{D(1)} (1 - v_F)\, d\mu_{D(1)} \leq \int_W h\, d\mu_{D(1)} \\ &\leq C(n,\gamma)\rho\mu_{D(1)}(W) \leq C(n,\gamma)\left[l_{n-1}(W)\right]^{1+1/(n-1)}.\end{aligned}$$

in view of (10) and our choice of ρ. □

Corollary 11 provides a sufficient condition for $\mathcal{P}_E$ to be two-dimensional. A necessary condition is that

$$\int_{\{|x'|\leq 1\}} |x'|^{-n} \operatorname{dist}((x',0),E)dx' < \infty. \tag{21}$$

To see this, we note from Theorem 6 that, if $\mathcal{P}_E$ is two-dimensional, then there is a function u in $\mathcal{P}_E$ satisfying $u(x) \geq x_n |x|^{-n}$ on Ω, whence

$$u(x',0) \geq C(n)u(x', \operatorname{dist}((x',0),E)) \geq C(n)\operatorname{dist}((x',0),E)\,|x'|^{-n}$$

by Harnack's inequalities, and (21) now follows from the local integrability of u on L.

Combining Corollary 11 with the observed necessity of (21), we see that, for

$$E = L\backslash \left(\bigcup_k B(x^{(k)}, r_k)\right), \quad \text{where} \quad x^{(k)} \in L \cap \partial B(0, 2^{-k}) \quad \text{and} \quad r_k < 2^{-k},$$

the cone $\mathcal{P}_E$ is two-dimensional if and only if $\sum 2^{nk} r_k^n < \infty$.

Further illustrations of condition (b) in Theorem 8 may be found in [12]. In particular, it is shown there that, for

$$E = L\backslash \left\{x \in (0,1) \times \mathbb{R}^{n-2} \times \{0\} : \sqrt{x_2^2 + \cdots + x_{n-1}^2} < g(x_1)\right\},$$

where $n \geq 3$ and $g : (0,1) \to (0,\infty)$ is increasing, $\mathcal{P}_E$ is two-dimensional if and only if

$$\int_0^1 t^{-n} \left[g(t)\right]^{n-1} dt < \infty.$$

3.4. Boundary behaviour of the Green function

Up to now we have been discussing the relationship between the Euclidean and Martin boundaries of Denjoy domains. Next we consider the boundary behaviour of their Green functions. Let $x_0 \in \Omega$. We will say that $G_\Omega(x_0,\cdot)$ is *Lipschitz continuous at* 0 if there is a constant $C > 0$ such that $G_\Omega(x_0,x) \leq C\,|x|$ on some neighbourhood of 0, where $G_\Omega(x_0,\cdot)$ is interpreted as 0 on E. This definition is independent of the choice of x_0, in view of Harnack's inequalities. Since $G_{H_+}(x_0,x)/x_n$ has a finite (positive) limit at 0, it is clear that $G_{H_+}(x_0,\cdot)$ is Lipschitz continuous at 0. The next result shows that this remains true of $G_\Omega(x_0,\cdot)$ when Ω is sufficiently like H_+ near 0.

Theorem 13. *For a Denjoy domain $\Omega = \mathbb{R}^n \backslash E$ with $0 \in E$, and for any point $x_0 \in \Omega$, the following statements are equivalent:*

(a) $G_\Omega(x_0,\cdot)$ *is Lipschitz continuous at* 0*;*

(b) $\begin{cases} \sum 2^{k(n-2)} \left[\mathcal{C}(D_k) - \mathcal{C}(E_k)\right] < \infty & (n \geq 3) \\ \sum 2^{k} \left[\mathcal{C}(D_k) - \mathcal{C}(E_k)\right] < \infty & (n = 2) \end{cases}.$

This result was first established for the case $n = 2$ by Carleson and Totik [14], without reference to the work of Benedicks [10]. Below we will present a proof of this result in all dimensions, taken from [12], which explains why the same condition appears in both Theorems 8 and 13.

A few preliminary comments may serve to illuminate this phenomenon. Thinness (in the ordinary sense) of a set A at 0 may be characterized by the existence of a superharmonic function v on a neighbourhood of 0 such that

$$\liminf_{x\to 0, x\in A} v(x) > v(0) = \liminf_{x\to 0} v(x). \tag{22}$$

However, it may equivalently be characterized by the existence of a superharmonic function v on a neighbourhood of 0 such that

$$\liminf_{x\to 0, x\in A} \frac{v(x)}{|x|^{2-n}} > \liminf_{x\to 0} \frac{v(x)}{|x|^{2-n}}. \tag{23}$$

(When $n = 2$, we replace $|x|^{2-n}$ by $\log(1/|x|)$ in (23).) A similar duality occurs in the case of minimal thinness. By analogy with (23), a set $A \subset \Omega$ is minimally thin at $y \in \partial_1^M \Omega$ if and only if there is a positive superharmonic function v on Ω such that

$$\liminf_{x\to y, x\in A} \frac{v(x)}{M_\Omega(x,y)} > \liminf_{x\to y} \frac{v(x)}{M_\Omega(x,y)},$$

but an equivalent characterization is that there is a positive superharmonic function v on Ω such that

$$\liminf_{x\to y, x\in A} \frac{v(x)}{G_\Omega(x_0,x)} > \liminf_{x\to y} \frac{v(x)}{G_\Omega(x_0,x)}$$

(cf. (22)). It is this possibility of characterizing minimal thinness in terms either of minimal harmonic functions or of the Green function that allows us to connect the dimensionality of $\mathcal{P}_E$ with the boundary behaviour of G_Ω. We will now make this connection precise.

Proof. Suppose firstly that condition (b) of Theorem 13 holds, and let $x_0 = (0', 1)$. By Theorems 8 and 6 there is a point $y \in \mathcal{M}_E \cap \partial_1^M \Omega$ such that $\Omega \cap L$ is minimally thin at y. The associated minimal harmonic function u, say, must differ from its Poisson integral in one of the half-spaces, H_+, say. Hence $\Omega \backslash H_+$ is minimally thin at y, and it follows from Theorem 9.5.2 of [8] that

$$\limsup_{x\to y, x\in H_+} \frac{G_{H_+}(x_0,x)}{G_\Omega(x_0,x)} > 0.$$

In view of Remark 5, we know that any sequence of points in H_+ that converges to y must converge to 0 in the Euclidean topology, so

$$\limsup_{x\to 0, x\in H_+} \frac{G_{H_+}(x_0,x)}{G_\Omega(x_0,x)} > 0.$$

Hence, by Theorem 9.3.3(ii) of [8], $G_\Omega(x_0,\cdot)/G_{H_+}(x_0,\cdot)$ has a finite minimal fine limit at 0 relative to H_+, and so a finite nontangential limit there, by Theorem 9.7.4 of [8]. Since $G_\Omega(x_0,x) \le C(n) G_\Omega(x_0,(0',|x|))$, by Theorem 1, we now see that $G_\Omega(x_0,x)/|x|$ is bounded above on $B(0,1/2)\cap\Omega$, and so $G_\Omega(x_0,\cdot)$ is Lipschitz continuous at 0.

Conversely, suppose that condition (a) holds, and let y be the minimal Martin boundary point of Corollary 4 (with $D = \Omega$). Since

$$\limsup_{x\to y, x\in H_+} \frac{G_{H_+}(x_0,x)}{G_\Omega(x_0,x)} \ge C(n) \limsup_{t\to 0+} \frac{t}{G_\Omega(x_0,(0',t))} > 0,$$

by hypothesis, we can apply Theorem 9.5.2 of [8] again to see that $\Omega\backslash H_+$, and hence $L \cap \Omega$, is minimally thin at y. Theorem 6 now shows that $\mathcal{P}_E$ is two-dimensional, and condition (b) follows by Theorem 8. □

4. Applications

4.1. Approximation of positive harmonic functions

Let $D \subset \mathbb{R}^n$ be a Greenian domain. In this section, which extends [26], we shall study the following approximation problem: *when is* $\mathcal{HPL}^p(D)$ *dense* (*with respect to uniform convergence on compact subsets in* D) *in* $\mathcal{HP}(D)$*?* We will deal with the cases $0 < p \le 1$ and $p = \infty$ (so the case $p = \infty$ refers to approximation by bounded positive harmonic functions).

In all cases there are some trivial counterexamples. In the case $p = \infty$ the punctured ball $D = B(0,1) \setminus \{0\}$ is a counterexample. Indeed it is easy to see that $\overline{\mathcal{HPL}^\infty(D)} \neq \mathcal{HP}(D)$ if $\partial D \setminus \operatorname{supp}(\lambda_x^D)$ is non-empty, because the latter set is removable for bounded harmonic functions. Of course, a point such as the origin in the case of the punctured ball corresponds to exactly one Martin boundary point, which is minimal, and the associated minimal harmonic function has the same singularity as the fundamental solution of the Laplacian, and so is integrable. However, we easily obtain counterexamples in the case where $0 < p \leq 1$ if we allow unbounded domains D, for then the half-space $D = H_+$ is a domain for which $\overline{\mathcal{HPL}^p(D)} \neq \mathcal{HP}(D)$ (cf. Section 1).

We now give a general theorem which relates the above approximation problem to "topological" properties of the Martin boundary. Later we will see that Denjoy domains play a role in constructing non-trivial counterexamples to the above approximation properties.

Theorem 14. *Let $D \subset \mathbb{R}^n$ be a Greenian domain.*

(1) *Let $0 < p \leq 1$, and $A_p = \{y \in \partial_1^M D : M_D(\cdot, y) \in \mathcal{L}^p(D)\}$. Then $\overline{\mathcal{HPL}^p(D)} = \mathcal{HP}(D)$ if and only if $\partial_1^M D \subset \overline{A_p}^M$.*

(2) *$\overline{\mathcal{HPL}^\infty(D)} = \mathcal{HP}(D)$ if and only if $\partial_1^M D \subset \operatorname{supp}(\nu_1)$.*

Proof. (1) Firstly we note that, if D_k is an exhaustion of D, then

$$A_p = \partial_1^M D \cap \{y \in \partial^M D : \lim_{k\to\infty} \int_{D_k} (M_D(x,y))^p dx < \infty\},$$

so A_p is a Borel set.

Suppose now that $h \in \mathcal{HPL}^p(D)$, and assume without loss of generality that $h(x_0) = 1$, where x_0 is the reference point of $M_D(\cdot,\cdot)$. Then, by Tonelli's theorem and Jensen's inequality,

$$\begin{aligned} \int_{\partial_1^M D} \int_D (M_D(x,y))^p dx d\nu_h(y) &\leq \int_D \left(\int_{\partial_1^M D} M_D(x,y) d\nu_h(y) \right)^p dx \\ &= \int_D (h(x))^p \, dx < \infty, \end{aligned}$$

and so ν_h is carried by A_p.

Let us now introduce two convex subcones of $\mathcal{HP}(D)$ as follows: $\mathcal{K}_1$ denotes the cone of all (finite) positive linear combinations of functions of the form $M_D(\cdot, y)$ for $y \in A_p$, and $\mathcal{K}_2$ denotes the cone of all those positive harmonic functions h which have a representing measure (not necessarily the one carried by $\partial_1^M D$) supported by $\overline{A_p}^M$.

We note that

$$\mathcal{K}_1 \subset \mathcal{HPL}^p(D) \subset \mathcal{K}_2,$$

and $\mathcal{K}_2$ is closed by weak*-compactness. We will now prove that $\overline{\mathcal{K}_1} = \mathcal{K}_2$ by an application of the Hahn-Banach theorem (about separation of convex cones).

Suppose η is a signed Radon measure with compact support in D such that $\int k d\eta \geq 0$ for all $k \in \mathcal{K}_1$. In particular, by assumption,

$$\int M_D(x,y)d\eta(x) \geq 0 \quad (y \in A_p),$$

and by continuity this also holds for all $y \in \overline{A_p}^M$. Hence, if $h \in \mathcal{K}_2$, so that h has a representing measure ν supported by $\overline{A_p}^M$, then

$$\int h d\eta = \int \int M_D(x,y)d\nu(y)d\eta(x) = \int \int M_D(x,y)d\eta(x)d\nu(y) \geq 0.$$

This proves that $\mathcal{K}_2 \subset \overline{\mathcal{K}_1}$, and hence the first part of the theorem is proved.

(2) This follows by analogous reasoning using the fact that $\mathcal{HPL}^\infty(D)$ is precisely the set of functions representable in the form $\int M_D(\cdot,y)f(y)d\nu_1(y)$, where f is a non-negative bounded Borel measurable function. □

Example 15. Theorem 14 shows that topological information about the minimal Martin boundary points is crucial for these approximation questions. We now claim that the Denjoy domain

$$\Omega = \mathbb{C} \setminus \left(\bigcup_{k\in\mathbb{Z}} \left[\frac{1}{k+3/4}, \frac{1}{k+1/4} \right] \cup \{0\} \right)$$

from Example 7 has the property that the two distinct minimal points y_+ and y_- corresponding to 0 are isolated in $\partial_1^M \Omega$. This is, of course, equivalent to saying that the two distinct minimal points w_+ and w_- corresponding to ∞ for the domain D in the same example are isolated in $\partial_1^M D$. For reasons of notational convenience we will establish this latter assertion. We again choose 0 as our reference point.

To show this, we note from symmetry that every sequence $x^{(k)}$ in $\mathbb{R} \cap D$ converging to ∞ satisfies

$$M_D(\cdot, x^{(k)}) \to \frac{1}{2}\left(M_D(\cdot, w_+) + M_D(\cdot, w_-)\right).$$

It follows by applying Harnack's inequality to the circles $\partial B(k,1)$ that no sequence $\left(x^{(k)}\right)$ with $x^{(k)} \in \partial B(k,1)$ can converge to either w_+ or w_-. From this the claim follows easily using the maximum principle.

We note that Ancona's interest in this example stemmed from the fact that $\partial_1^M \Omega$ is not dense in $\partial^M \Omega$ in this case.

Example 16. We now give non-trivial examples of domains D_p (that is, bounded domains for which $\mathrm{supp}(\lambda_x^{D_p}) = \partial D_p$) where $\overline{\mathcal{HPL}^p(D_p)} \neq \mathcal{HP}(D_p)$.

We begin with the case $p = \infty$. The Denjoy domain Ω of Example 15 has the property that $\overline{\mathcal{HPL}^\infty(\Omega)} \neq \mathcal{HP}(\Omega)$, because the points y_+ and y_- are not in the support of ν_1 (see Theorem 14 (2)). If we define $D_\infty = B(0,1) \cap \Omega$, we get a bounded domain with $\mathrm{supp}(\lambda_x^{D_\infty}) = \partial D_\infty$, yet $\overline{\mathcal{HPL}^\infty(D_\infty)} \neq \mathcal{HP}(D_\infty)$.

However, every positive harmonic function on D_∞ is integrable. To treat the case where $0 < p \le 1$ we fix such a p, choose $k \in \mathbb{N}$ such that $pk \ge 1$, and define $D_p = \{z \in \mathbb{C} : z^{2k} \in D_\infty\}$. Then

$$D_p = B(0,1) \setminus \bigcup_{j=0}^{4k-1} \left\{ \sqrt[2k]{r} \exp\left(i\frac{\pi j}{2k} \right) : r \in \partial\Omega \cap [0,\infty) \right\}.$$

(The boundary of D_p in $B(0,1)$ consists of subsets of $4k$ rays emanating from 0 with angular spacing $\pi/2k$.) From the construction of the Martin boundary it is easy to see that, analogously to the case of D_∞, the origin corresponds to $4k$ distinct points in $\partial_1^M D_p$, and these points are isolated in $\partial_1^M D_p$.

Further, the growth of the corresponding minimal functions is comparable to

$$\operatorname{Im}\left(\frac{1}{z^{2k}}\right) = \frac{1}{r^{2k}} \sin(2\theta k) \text{ on } 0 < \theta < \pi/2k,$$

where $z = re^{i\theta}$. Since $pk \ge 1$, we get

$$\int_0^1 \int_0^{\pi/2k} \left(\frac{1}{r^{2k}} \sin(2\theta k) \right)^p r dr d\theta = \int_0^1 \frac{1}{r^{2pk-1}} dr \int_0^{\pi/2k} \sin^p(2\theta k) d\theta = +\infty.$$

By applying Theorem 14 (1) we now see that $\overline{\mathcal{HPL}^p(D_p)} \neq \mathcal{HP}(D_p)$.

4.2. Minimal harmonic functions associated with an irregular boundary point

As is well known, the minimal harmonic functions on the unit ball B are simply multiples of the Poisson kernel with arbitrary boundary pole; that is, they are multiples of the functions

$$v_z : x \mapsto \frac{1 - |x|^2}{|x - z|^n} \qquad (z \in \partial B).$$

We note that v_z continuously vanishes on $\partial B \backslash \{z\}$ but tends to ∞ along a tangential approach region to z. Now let $U = B \backslash \{x_0\}$, where $x_0 \in B$. The minimal harmonic functions on U comprise all the minimal harmonic functions on B together with multiples of

$$v_{x_0} : x \mapsto G_B(x, x_0) = u_{x_0}(x) - H^B_{u_{x_0}}(x),$$

where

$$u_y(x) = \begin{cases} |x - y|^{2-n} & (n \ge 3) \\ -\log |x - y| & (n = 2) \end{cases}$$

and H_f^V denotes the solution to the Dirichlet problem on V with boundary function f. Clearly v_{x_0} continuously vanishes on $\partial U \backslash \{x_0\}$ and tends to ∞ at x_0. This observation, concerning the irregular boundary point x_0 of U, illustrates the following general fact. Let U be a (Greenian) domain with an irregular boundary point x_0 (so $\mathbb{R}^n \backslash U$ is thin at x_0), and define

$$G_U(x, x_0) = u_{x_0}(x) - H^U_{u_{x_0}}(x) \quad (x \in U).$$

If u is a positive multiple of $G_U(\cdot, x_0)$, then u is a minimal harmonic function on U and $u(x) \to \infty$ as $x \to x_0$ outside a set which is thin at x_0. Brelot [11] observed that the converse to this statement also holds when $n = 2$, but the corresponding question in higher dimensions has remained open until the following recent result [21].

Theorem 17. *Let $n \geq 3$. There is a domain U with irregular boundary point 0, and a minimal harmonic function u on U, such that $u(x) \to \infty$ as $x \to 0$ outside a set which is thin at 0, yet u is* **not** *a multiple of $G_U(\cdot, 0)$.*

We will shortly outline a construction motivated by the theory of Denjoy domains, which was used in the proof of Theorem 17. First we will indicate the significance of this result for fine potential theory.

Recall that, for nonnegative superharmonic functions on a Greenian domain U, the Riesz decomposition says that the following conditions are equivalent:

(i) the only nonnegative harmonic minorant of u is 0;
(ii) the only nonnegative subharmonic minorant of u is 0;
(iii) there is a Borel measure μ on U such that $u = \int G_U(\cdot, y) d\mu(y)$.

The *fine topology* on $\mathbb{R}^n$ is the coarsest topology which renders all superharmonic functions continuous. Fuglede and others have developed, since around 1970, an elegant and powerful theory of finely harmonic and finely superharmonic functions on fine domains (that is, finely connected, finely open sets), an account of which may be found in [17]. The fine topology counterparts of conditions (ii) and (iii) above were shown to be equivalent by Fuglede [18], and so either can be used as the definition of a fine potential. Also, it is obvious that (ii) implies (i). However, it has been a long-standing open question whether the fine topology counterparts of conditions (i) and (ii) are actually equivalent. This question was first raised by Fuglede in 1972 (see p. 105 of [17]), and further emphasized in [19]. We now explain, using an argument from [19], how Theorem 17 leads to a negative answer to this question.

Let U and u be as in Theorem 17. The set $U_0 = U \cup \{0\}$ is then a fine domain. Further, if we define $u(0) = +\infty$, then u certainly satisfies the fine topology analogue of the supermeanvalue property at 0, and so u is "finely superharmonic" on U_0. Any non-negative finely harmonic minorant v of u on U_0 is actually harmonic on the open set Ω (by Theorem 10.16 of [17]), so $v = cu$ for some $c \in [0, 1]$ by the minimality of u on U. Since $v(x_0)$ is finite, we must have $c = 0$ and so $v = 0$. Thus the fine topology version of Property (i) above holds. However, if u were the fine potential of a measure μ on U_0, then $\mu(U) = 0$ by the harmonicity of u on U, and we would be led to the contradictory conclusion that u is a multiple of $G_U(\cdot, 0)$. Thus the fine topology version of Property (iii) (equivalently, (ii)) fails to hold.

We now outline one approach to proving Theorem 17. Let B' denote the unit ball in $\mathbb{R}^{n-1}$, and let $V = B' \times \mathbb{R}$. We are going to exploit the translational invariance of V and the thinness of $\overline{V}$ at infinity when $n \geq 4$. (A more intricate approach is required when $n = 3$: see [21].) Let α denote the square root of the first eigenvalue of $-\sum_{i=1}^{n-1}\partial^2/\partial x_i^2$ on B', and ϕ be the corresponding eigenfunction which satisfies $\phi(0)=1$. The function

$$h:(x',x_n)\mapsto \phi(x')e^{\alpha x_n}$$

is then a minimal harmonic function on V that vanishes on ∂V. We extend h to be a subharmonic function on all of $\mathbb{R}^n$ by defining $h=0$ on $\mathbb{R}^n\backslash V$. Let $\widetilde{V}=\mathbb{R}^n\backslash\overline{V}$. It is enough to establish the following.

Proposition 18. *There is a minimal positive harmonic function u^* on a domain $U^*\supset\mathbb{R}^n\backslash\partial V$ such that u^* continuously vanishes on ∂U^* and satisfies $u^*\geq h$, $u^*=H_{u^*}^{\widetilde{V}}$ (where $u^*(\infty)=0$) and*

$$\left|x'\right|^{n-2}u^*(x',0)\to+\infty \quad \text{as} \quad |x'|\to\infty. \tag{24}$$

Theorem 17 follows from this proposition using the Kelvin transform and inversion in ∂B: the resulting minimal harmonic function u is defined on a domain U which has 0 as an irregular boundary point, so u has a fine limit at 0 which must be infinite, by (24). However, u cannot be a multiple of $G_U(\cdot,0)$ because of its rapid growth along the x_n-axis on approach to 0.

Next, using the symmetries of V, it is not difficult to see that (24) is equivalent to

$$\int_{\partial V}u^*\,d\sigma=\infty, \tag{25}$$

where σ denotes surface area measure on ∂V. This guides our choice of U^*, as follows. We fix $\beta\in(0,1/4)$ and define

$$A_{k,\beta}=\{(x',x_n)\in\partial V:\left|x_n-2^k\right|<\beta e^{-\alpha 2^{k-1}}\}\qquad(k\in\mathbb{N})$$

and $U^*=V\cup\widetilde{V}\cup(\cup_k A_{k,\beta})$. The point here is that, since $h(x',x_n)\approx(1-|x'|)e^{\alpha x_n}$ on V, if there is a harmonic function u^* on U^* such that $u^*\geq h$, then Harnack's inequalities will show that

$$u^*(x',x_n)\geq C(n,\beta)e^{-\alpha 2^{k-1}}e^{\alpha 2^k}=C(n,\beta)e^{\alpha 2^{k-1}}\quad \text{on } A_{k,\beta/2},$$

and so (25) automatically holds.

It therefore remains to show that there is a minimal harmonic function u^* on U^* such that u^* vanishes on ∂U^*, $u^*\geq h$ and $u^*=H_{u^*}^{\widetilde{V}}$. This is analogous to Case (2) of Theorem 6, where there was a minimal harmonic function on Ω which vanishes on E, majorizes $x\mapsto x_n^+\left|x\right|^{-n}$ and equals its Poisson integral in H_-.

The approach taken in [21] to showing this is as follows. We define the cylindrical annular sets

$$W_k=\{(x',x_n):\left||x'|-1\right|<e^{-\alpha 2^{k-1}}\text{ and }\left|x_n-2^k\right|<\beta e^{-\alpha 2^{k-1}}\}$$

and, for any nonnegative continuous function f on $\mathbb{R}^n$, the functions

$$H_jf=\begin{cases} H_f^{\cup_1^j W_k} & \text{on } \cup_1^j W_k \\ f & \text{elsewhere}\end{cases},\qquad H_0f=\begin{cases} h+H_f^V & \text{on } V\\ H_f^{\widetilde{V}} & \text{on } \widetilde{V}\\ f & \text{on } \partial V\end{cases}.$$

(We always interpret $f(\infty)$ as 0.) Next we inductively define a sequence (s_j) of continuous functions on $\mathbb{R}^n$ by

$$s_0 = h, \quad s_{2j-1} = H_j s_{sj-2}, \quad s_{2j} = H_0 s_{2j-1}.$$

It is not difficult to see, using the maximum principle, that each function s_j is subharmonic and the sequence is increasing. Less obvious is the fact that the sequence converges on $\mathbb{R}^n$: this has to be verified using appropriate estimates of harmonic measure. Once this is done it can be checked that the limit function u^* does indeed have all of the desired properties.

References

[1] H. Aikawa, Positive harmonic functions of finite order in a Denjoy type domain. Proc. Amer. Math. Soc. 131 (2003), 3873–3881.

[2] H. Aikawa, K. Hirata and T. Lundh, Martin boundary points of a John domain and unions of convex sets. J. Math. Soc. Japan 58 (2006), 247–274.

[3] A. Ancona, Une propriété de la compactification de Martin d'un domaine euclidien. Ann. Inst. Fourier (Grenoble) 29, 4 (1979), 71–90.

[4] A. Ancona, Régularité d'accès des bouts et frontière de Martin d'un domaine euclidien. J. Math. Pures Appl. (9) 63 (1984), 215–260.

[5] A. Ancona, Sur la frontière de Martin des domaines de Denjoy. Ann. Acad. Sci. Fenn. Ser. A I Math. 15 (1990), 259–271.

[6] A. Ancona and M. Zinsmeister, Fonctions harmoniques positives et compacts de petites dimensions. C. R. Acad. Sci. Paris Sér. I Math. 309 (1989), 305–308.

[7] V.V. Andrievskii, Positive harmonic functions on Denjoy domains in the complex plane. J. Anal. Math. 104 (2008), 83–124.

[8] D.H. Armitage and S.J. Gardiner, Classical Potential Theory. Springer, London, 2001.

[9] D.H. Armitage, S.J. Gardiner and I. Netuka, Separation of points by classes of harmonic functions. Math. Proc. Cambridge Philos. Soc. 113 (1993), 561–571.

[10] M. Benedicks, Positive harmonic functions vanishing on the boundary of certain domains in $\mathbf{R}^n$. Ark. Mat. 18 (1980), 53–72.

[11] M. Brelot, Sur le principe des singularités positives et la topologie de R.S. Martin. Ann. Univ. Grenoble Sect. Sci. Math. Phys. (N.S.), 23 (1948), 113–138.

[12] T. Carroll and S.J. Gardiner, Lipschitz continuity of the Green function in Denjoy domains. Ark. Mat. 46 (2008), 271–283.

[13] N. Chevallier, Frontière de Martin d'un domaine de $\mathbf{R}^n$ dont le bord est inclus dans une hypersurface lipschitzienne. Ark. Mat. 27 (1989), 29–48.

[14] L. Carleson and V. Totik, Hölder continuity of Green's functions. Acta Sci. Math. (Szeged) 70 (2004), 557–608.

[15] M.C. Cranston and T.S. Salisbury, Martin boundaries of sectorial domains. Ark. Mat. 31 (1993), 27–49.

[16] A. Denjoy, Sur les fonctions analytiques uniformes à singularités discontinues. C. R. Acad. Sci. Paris 149 (1909), 258–260.

[17] B. Fuglede, Finely harmonic functions. Lecture Notes in Math. 289, Springer, Berlin, 1972.

[18] B. Fuglede, Représentation intégrale des potentiels fin. C. R. Acad. Sci. Paris Sér. I Math. 300 (1985), 129–132.

[19] B. Fuglede, On the Riesz representation of finely superharmonic functions. Potential theory – surveys and problems, Prague, 1987, pp. 199–201, Lecture Notes in Math. 1344, Springer, Berlin, 1988.

[20] S.J. Gardiner, Minimal harmonic functions on Denjoy domains. Proc. Amer. Math. Soc. 107 (1989), 963–970.

[21] S.J. Gardiner and W. Hansen, The Riesz decomposition of finely superharmonic functions. Adv. Math, 214 (2007), 417–436.

[22] L. Karp and A.S. Margulis, Newtonian potential theory for unbounded sources and applications to free boundary problems. J. Anal. Math. 70 (1996), 1–63.

[23] M. Sakai, Null quadrature domains. J. Analyse Math. 40 (1981), 144–154 (1982).

[24] S. Segawa, Martin boundaries of Denjoy domains. Proc. Amer. Math. Soc. 103 (1988), 177–183.

[25] S. Segawa, Martin boundaries of Denjoy domains and quasiconformal mappings. J. Math. Kyoto Univ. 30 (1990), 297–316.

[26] T. Sjödin, Approximation in the cone of positive harmonic functions. Potential Anal. 27 (2007), 271–280.

[27] M. Sodin, An elementary proof of Benedicks's and Carleson's estimates of harmonic measure of linear sets. Proc. Amer. Math. Soc. 121 (1994), 1079–1085.

Stephen J. Gardiner
School of Mathematical Sciences
University College Dublin
Dublin 4, Ireland
e-mail: `stephen.gardiner@ucd.ie`

Tomas Sjödin
Department of Mathematics
Royal Institute of Technology
SE-100 44 Stockholm, Sweden

Present address
Department of Mathematics
Linköping University
SE-581 83 Linköping, Sweden
e-mail: `tosjo@mai.liu.se`

Analysis and Mathematical Physics
Trends in Mathematics, 167–186
© 2009 Birkhäuser Verlag Basel/Switzerland

Carathéodory Convergence of Immediate Basins of Attraction to a Siegel Disk

Pavel Gumenyuk

Abstract. Let f_n be a sequence of analytic functions in a domain U with a common attracting fixed point z_0. Suppose that f_n converges to f_0 uniformly on each compact subset of U and that z_0 is a Siegel point of f_0. We establish a sufficient condition for the immediate basins of attraction $\mathcal{A}^*(z_0, f_n, U)$ to form a sequence that converges to the Siegel disk of f_0 as to the kernel w. r. t. z_0. The same condition is shown to imply the convergence of the Kœnigs functions associated with f_n to that of f_0. Our method allows us also to obtain a kind of quantitative result for analytic one-parametric families.

Mathematics Subject Classification (2000). Primary 37F45; Secondary 30D05, 37F50.

Keywords. Iteration of analytic functions, Fatou set, Siegel disk, basin of attraction, convergence as to the kernel.

1. Introduction

1.1. Preliminaries

Let U be a domain on the Riemann sphere $\overline{\mathbb{C}}$ and $f : U \to \overline{\mathbb{C}}$ a meromorphic function. Define f^n, the n-fold *iterate* of f, by the following relations: $f^1 : U \to \overline{\mathbb{C}}$, $f^1 := f$, $f^{n+1} : \big(f^n\big)^{-1}(U) \to \overline{\mathbb{C}}$, $f^{n+1} := f \circ f^n$, $n \in \mathbb{N}$. It is convenient to define f^0 as the identity map of U. Denote

$$E(f, U) := \bigcap_{n\in\mathbb{N}} \big(f^n\big)^{-1}(U).$$

The *Fatou set* $\mathcal{F}(f, U)$ of the function f (w. r. t. the domain U) is the set of all interior points z of $E(f, U)$ such that $\{f^n\}_{n\in\mathbb{N}}$ is a normal family in some neigh-

This work is partially supported by the *Research Council of Norway*, the *Russian Foundation for Basic Research* (grant #07-01-00120), and *ESF Networking Programme* "Harmonic and Complex Analysis and its Applications".

bourhood of z. Define the *Julia set* $\mathcal{J}(f, U)$ of f (with respect to the domain U) to be the complement $U \setminus \mathcal{F}(f, U)$ of the Fatou set.

Classically iteration of analytic (meromorphic) functions has been studied for the case of $U \in \{\overline{\mathbb{C}}, \mathbb{C}, \mathbb{C}^* := \mathbb{C} \setminus \{0\}\}$ and $f : U \to U$, see survey papers [1, 2] for the details. As an extension the cases of transcendental meromorphic functions and functions meromorphic in $\overline{\mathbb{C}}$ except for a compact totally disconnected set have been also investigated, see, e.g., [3, 4]. (Note that $f(U) \not\subset U$ for these cases.) In this paper we shall restrict ourselves by the following

Assumption. Suppose that $U, f(U) \subset \mathbb{C}$, i.e., f is an analytic function in a subdomain U of $\mathbb{C}$.

One of the basic problems in iteration theory of analytic functions is to study how the limit behaviour of iterates changes as the function f is perturbed. A large part of papers in this direction are devoted to the continuity property for the dependence of the Fatou and Julia sets on the function to be iterated. We mention the work of A. Douady [5], who investigates the mapping $f \mapsto \mathcal{J}(f, \mathbb{C})$ from the class of polynomials of fixed degree to the set of nonempty plane compacta equipped with the Hausdorff metric $d_H(X, Y) := \max\{\partial(X, Y), \partial(Y, X)\}$, $\partial(X, Y) := \sup_{x \in X} \operatorname{dist}(x, Y)$. We also mention subsequent papers [6]–[11] dealing with other classes of functions. Continuity of Julia sets is closely related to behaviour of connected components of the Fatou set containing periodic points. Now we recall necessary definitions.

Let $z_0 \in U$ be a fixed point of f. The number $\lambda := f'(z_0)$ is called the *multiplier* of z_0. According to the value of λ the fixed point z_0 is said to be *attracting* if $|\lambda| < 1$, *neutral* if $|\lambda| = 1$, and *repelling* if $|\lambda| > 1$. An attracting fixed point is *superattracting* if $\lambda = 0$, or *geometrically attracting* otherwise. Suppose z_0 is a neutral fixed point of f and none of f^n, $n \in \mathbb{N}$, turns into the identity map; then the fixed point z_0 is *parabolic* if $\lambda = e^{2\pi i \alpha}$ for some $\alpha \in \mathbb{Q}$, or *irrationally neutral* otherwise. If an irrationally neutral fixed point belongs to $\mathcal{F}(f, U)$, then it is called a *Siegel point*.

The component of the Fatou set $\mathcal{F}(f, U)$ that contains a fixed point z_0 is called the *immediate basin* of z_0 and denoted by $\mathcal{A}^*(z_0, f, U)$. The immediate basin of a Siegel point is called a *Siegel disk*, and the immediate basin of an attracting fixed point is called an *immediate basin of attraction*. It is a reasonable convention to put by definition $\mathcal{A}^*(z_0, f, U) := \{z_0\}$ for fixed points $z_0 \in \mathcal{J}(f, U)$, in particular for repelling and parabolic ones.

By passing to a suitable iterate of f, the above definitions are naturally extended to periodic points.

1.2. Main results

Consider a sequence $\{f_n : U \to \mathbb{C}\}_{n \in \mathbb{N}}$ of analytic functions with a common attracting fixed point $z_0 \in U$. Suppose that f_n converges to f_0 uniformly on each compact subset of U. It follows easily from arguments of [5] that $\mathcal{A}^*(z_0, f_n, U) \to \mathcal{A}^*(z_0, f_0, U)$ as to the kernel w. r. t. z_0 provided z_0 is an attracting or parabolic fixed point of the limit function f_0. At the same time $\mathcal{A}^*(z_0, f_n, U)$ fails to converge to $\mathcal{A}^*(z_0, f_0, U)$ in general if z_0 is a Siegel point of f_0 (see Example 1 in Section 4). Similarly, the dependence of Julia sets on the function under iteration fails to be continuous at f_0 (with respect to the Hausdorff metric) if f_0 has (generally speaking, periodic) Siegel points. Nevertheless, in the paper [12] devoted to the continuity of Julia sets for one-parametric families of transcendental entire functions H. Kriete established an assertion, which can be stated as follows.

Theorem A. *Suppose $f : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$; $(\lambda, z) \mapsto f_\lambda(z)$ is an analytic family of entire functions $f_\lambda(z) = \lambda z + a_2(\lambda) z^2 + \cdots$ and $\lambda_0 := e^{2\pi i \alpha_0}$, where $\alpha_0 \in \mathbb{R} \setminus \mathbb{Q}$ is a Diophantine number. Let Δ be any Stolz angle at the point λ_0 with respect to the unit disk $\{\lambda : |\lambda| < 1\}$. Then $\mathcal{A}^*(0, f_\lambda, \mathbb{C}) \to \mathcal{A}^*(0, f_{\lambda_0}, \mathbb{C})$ as to the kernel w. r. t. z_0 when $\lambda \to \lambda_0$, $\lambda \in \Delta$.*

Remark 1.1. It was proved by C. Siegel [13] that for a fixed point with multiplier $e^{2\pi i \alpha}$, $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, to be a Siegel point, it is sufficient that α be Diophantine. This condition is not a necessary one even if restricted to the case of quadratic polynomials $f(z) := z^2 + c$, $c \in \mathbb{C}$ (see [14, Th. 6] and [15]). Furthermore, it is easy to construct a nonlinear analytic germ with a Siegel point for any given $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

The Diophantine condition on α_0 is substantially employed in [12], and in view of the above remark it is interesting to find out whether this condition is really essential in Theorem A. Another question to consider is the role of analytic dependence of f_λ on λ. A possible answer is the following statement improving Theorem A.

Theorem 1.2. *Let $f_0 : U \to \mathbb{C}$ be an analytic function with a Siegel point $z_0 \in U$ and $\{f_n : U \to \mathbb{C}\}_{n \in \mathbb{N}}$ a sequence of analytic functions with an attracting fixed point at z_0. Suppose that f_n converges to f_0 uniformly on each compact subset of U and the following conditions hold*

(i) $\left| \arg\left(1 - f_n'(z_0)/f_0'(z_0)\right)\right| < \Theta$ *for some $\Theta < \pi/2$ and all $n \in \mathbb{N}$;*

(ii) *the functions $\left(f_n(z) - f_0(z)\right)/\left(f_n'(z_0) - f_0'(z_0)\right)$, $n \in \mathbb{N}$, are uniformly bounded on each compact subset of U.*

Then $\mathcal{A}^*(z_0, f_n, U)$ converges to $\mathcal{A}^*(z_0, f_0, U)$ as to the kernel w. r. t. z_0.

Condition (i) in this theorem requires that $\lambda_n := f_n'(z_0)$ tends to $\lambda_0 := f_0'(z_0)$ within a Stolz angle, condition (ii) appears instead of analytic dependence of f_λ on λ, and the Diophantine condition on α_0 turns out to be unnecessary. Both conditions (i) and (ii) are essential. We discuss this in Section 4.

Dynamics of iterates in the immediate basin of a fixed point can be described by means of so-called Kœnigs function.

Let z_0 be a fixed point of an analytic function f. The *Kœnigs function* φ associated with the pair (z_0, f) is a solution to the *Schröder functional equation*

$$\varphi\big(f(z)\big) = \lambda \varphi(z), \quad \lambda := f'(z_0), \tag{1.1}$$

analytic in a neighbourhood of z_0 and subject to the normalization $\varphi'(z_0) = 1$.

It is known (see, e.g., [16, pp. 73–76, 116], [17]) that the Kœnigs function exists, is unique, and can be analytically continued all over $\mathcal{A}^*(z_0, f, U)$ provided z_0 is a geometrically attracting or Siegel fixed point. If the Kœnigs function is known, then the iterates can be determined by means of the equality

$$\varphi\big(f^n(z)\big) = \lambda^n \varphi(z), \quad \lambda := f'(z_0).$$

By φ_k, $k \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, denote the Kœnigs function associated with the pair (z_0, f_k). We prove the following

Theorem 1.3. *Under the conditions of Theorem* 1.2*, the sequence φ_n converges to φ_0 uniformly on each compact subset of $\mathcal{A}^*(z_0, f_0, U)$.*

The assertion of Theorem 1.3 should be understood in connection with Theorem 1.2, because the uniform convergence of φ_n on a compact set $K \subset \mathcal{A}^*(z_0, f_0, U)$ requires that K were in the range of definition of φ_n, i.e., in $\mathcal{A}^*(z_0, f_n, U)$, for all $n \in \mathbb{N}$ apart from a finite number.

Assumption. Hereinafter it is convenient to assume without loss of generality that $z_0 = 0$, saving symbol z_0 for other purposes.

For any $a \in \mathbb{C}$ and $A \subset \mathbb{C}$ let us use aA as the short variant of $\{az : z \in A\}$. By $\mathrm{D}(\xi_0, \rho)$ denote the disk $\{\xi : |\xi - \xi_0| < \rho\}$, but reserve the notation $\mathbb{D}$ for the unit disk $\mathrm{D}(0, 1)$.

Remark 1.4. The Kœnigs function φ_0 associated with the Siegel point of f_0 admits another description (see, e.g., [16, p. 116], [17]) as the conformal mapping of the Siegel disk $\mathcal{A}^*(0, f_0, U)$ onto a Euclidean disk $D(0, r)$ that satisfies the condition $\varphi_0(0) = \varphi_0'(0) - 1 = 0$. From this viewpoint it will be convenient to consider the conformal mapping φ, $\varphi(0) = 0$, $\varphi'(0) > 0$, of $\mathcal{A}^*(0, f_0, U)$ onto the unit disk $\mathbb{D}$ instead of the Kœnigs function φ_0. Obviously, $\varphi(z)/\varphi_0(z)$ is constant, and consequently, φ satisfies the Schröder equation (1.1) for $f := f_0$. For shortness, S will stand for $\mathcal{A}^*(0, f_0, U)$. By ψ denote the inverse function to φ and let $S_r := \psi(r\mathbb{D})$, $\mathcal{L}_r := \partial S_r$ for $r \in [0, 1]$. One of the consequences of the fact mentioned above is that f_0 is a conformal automorphism of S and S_r, $r \in (0, 1)$.

During the preparation of this paper another proof of Theorems 1.2 and 1.3 given in [18, p. 3] became known to the author. However, our method allows us also to establish an asymptotic estimate for the rate of covering level-lines of the Siegel disk by basins of attraction for one-parametric analytic families. Let $f : W \times U \to \mathbb{C}$; $(\lambda, z) \mapsto f_\lambda(z)$, where $U \ni 0$ and W are domains in $\mathbb{C}$, be a family of functions and α_0 an irrational number satisfying the following conditions:

(i) $f_\lambda(z)$ depends analytically on both the variable $z \in U$ and the parameter $\lambda \in W$;
(ii) $f_\lambda(0) = 0$ and $f_\lambda'(0) = \lambda$ for all $\lambda \in W$;
(iii) $\lambda_0 := \exp(2\pi i \alpha_0) \in W$ and the function f_{λ_0} has a Siegel point at $z_0 = 0$, with $S := \mathcal{A}^*(0, f_{\lambda_0}, U)$ lying in U along with its boundary ∂S.

Consider the continued faction expansion of α_0 and denote the nth convergent by p_n/q_n. (See, e.g., [19, 20] for a detailed exposition on continued factions.) For $x > 0$ we set

$$n_0(x) := \min\left\{n \in \mathbb{N} : \frac{2q_n q_{n+1}}{q_n + q_{n+1}} \geq x\right\}, \quad \ell(x) := q_{n_0(x)}.$$

Notation φ, ψ, S, and S_r will refer to the limit function f_{λ_0}. Lemma 2.2 with a slight modification can be used to prove the following statement.

Theorem 1.5. *For any Stolz angle Δ at the point λ_0 there exist a constant $C > 0$ and a function $\varepsilon : (0,1) \to (0,+\infty)$ such that for any $r \in (0,1)$ the following statements are true:*

(i) $S_r \subset \mathcal{A}^*(0, f_\lambda, U)$ *for all* $\lambda \in W \cap \Delta$ *satisfying* $|\lambda - \lambda_0| < \varepsilon(r)$;
(ii) $\varepsilon(r) \geq C(1-r)^3/\ell\big((1-r)^{-\gamma}\big)$,

where $\gamma > \gamma_0 := 1 + \max\big\{\beta_\psi(1), \beta_\psi(-1)\big\}$ and β_ψ stands for the integral means spectrum of the function ψ,

$$\beta_\psi(t) := \limsup_{r \to 1-} \frac{\log \int_0^{2\pi} |\psi'(re^{i\theta})|^t \, d\theta}{-\log(1-r)}. \tag{1.2}$$

It is known [21] that $\beta_\psi(1) \leq 0.46$ and $\beta_\psi(-1) \leq 0.403$ for any function ψ bounded and univalent in $\mathbb{D}$. Consequently, $\gamma_0 \leq 1.46$.

Theorem 1.5 has been published in [22]. We sketch its proof and specify the function $\varepsilon(r)$ explicitly in Section 3.

2. Proof of theorems

2.1. Lemmas

Denote $\lambda_k := f_k'(0)$, $k \in \mathbb{N}_0$. Let us fix arbitrary $n_* \in \mathbb{N}$ and consider the linear family

$$f_\lambda[n_*](z) := (1-t)f_0(z) + t f_{n_*}(z), \ t := \frac{\lambda - \lambda_0}{\lambda_{n_*} - \lambda_0}, \quad z \in U, \ \lambda \in \mathbb{C}. \tag{2.1}$$

The number n_* will be not varied throughout the discussion in the present section. So we shall not indicate dependence on n_* until it is necessary. In particular we shall often write f_λ instead of $f_\lambda[n_*]$.

We need the following elementary statement on approximation of integrals by quadrature sums (see, e.g., [23, pp. 55–62]).

Theorem B. *Suppose ϕ is a continuously differentiable function on $[0,1]$. Then for any $N \in \mathbb{N}$ and any set of points $x_0, x_1, \ldots, x_{N-1} \in [0,1]$ the following inequality holds*

$$\left| \int_0^1 \phi(x)\,dx - \frac{1}{N}\sum_{n=0}^{N-1} \phi(x_n) \right| < Q\big(x_0, x_1, \ldots, x_{N-1}\big) \int_0^1 |\phi'(x)|\,dx, \tag{2.2}$$

where $Q\big(x_0, x_1, \ldots, x_{N-1}\big) := \sup_{x\in[0,1]} \big|F(x; x_0, x_1, \ldots, x_{N-1}) - x\big|$ *and*

$$F(x; x_0, x_1, \ldots, x_{N-1}) := \frac{1}{N}\sum_{n=0}^{N-1}\theta(x - x_n),\ \theta(y) := \begin{cases} 1, & \textit{if } y > 0,\\ 0, & \textit{if } y \le 0.\end{cases}$$

Remark 2.1. Consider the sequence $x_n^\beta := \{\alpha_0 n + \beta\}$, where $\{\ \cdot\ \}$ stands for fractional part, α_0 is given by $\lambda_0 = e^{2\pi i\alpha_0}$, and β is an arbitrary real number. Denote

$$Q_{\beta,N} := Q\big(x_0^\beta, x_1^\beta, \ldots, x_{N-1}^\beta\big).$$

Since $\alpha_0 \in \mathbb{R}\backslash\mathbb{Q}$, we have (see, e.g., [23, pp. 102–108]) $Q_{\beta,N} \to 0$ as $N \to +\infty$.

Fix any $r_0 \in (0, 1)$. The following lemma allows us to determine $\varepsilon_* > 0$ such that $S_{r_0} \subset \mathcal{A}^*(0, f_\lambda, U)$ whenever $|\arg(1 - \lambda/\lambda_0)| < \Theta$ and $|\lambda - \lambda_0| < \varepsilon_*$. In order to state this assertion we need to introduce some notation.

Denote

$$k_0(z) := \frac{z}{(1-z)^2},\ z \in \mathbb{D}, \quad k_\gamma(z) := e^{i\gamma}k_0(e^{-i\gamma}z),\ \gamma \in \mathbb{R},$$

$$u(z) := \frac{f_{n_*}(z) - f_0(z)}{\lambda_{n_*} - \lambda_0}, \quad H(\xi) := 1 + \frac{\xi\psi''(\xi)}{\psi'(\xi)},$$

$$J(t) := \frac{\xi u'(\psi(\xi))\psi'(\xi) - u(\psi(\xi))H(\lambda_0\xi)}{\lambda_0\xi\psi'(\lambda_0\xi)},\ \xi := r_0 e^{2\pi i t}.$$

For $\tau \in (0, -\log r_0)$ and $N \in \mathbb{N}$ we put

$$Q_N := \inf_{\beta\in\mathbb{R}} Q_{\beta,N}, \quad a_N := 2\pi Q_N \int_0^1 |J(t)|\, dt,$$

$$\Lambda_N(\tau, \varepsilon) := \frac{\sqrt{1 + 2b^2\cos 2\vartheta + b^4} - 1 + b^2}{2b\cos\vartheta}, \quad \varepsilon > 0,$$

where $\vartheta := \Theta + \arcsin a_N$, $b := \pi\varepsilon N(1 - a_N)/(4\tau)$,

$$\varepsilon_N(\tau) := \frac{1 - k_\pi(r_*)/k_\pi(r^*)}{\sup\limits_{z\in S_{r_*}} |1 - f_{n_*}(z)/f_0(z)|}\,|\lambda_{n_*} - \lambda_0|, \quad r_* := r_0 e^{\tau(1-1/N)},\ r^* := r_0 e^\tau.$$

Lemma 2.2. *Let* $N \in \mathbb{N}$ *and* $\tau \in (0, -\log r_0)$. *If* $a_N < \sin(\pi/2 - \Theta)$, *then* $f_\lambda^N\big(S_{r_0}\big) \subset S_{r_0}$ *for all* λ *such that* $\big|\arg(1 - \lambda/\lambda_0)\big| < \Theta$ *and* $|\lambda - \lambda_0| < \varepsilon_*$, *where* $\varepsilon_* := \varepsilon_N(\tau)\,\Lambda_N\big(\tau, \varepsilon_N(\tau)\big)$.

Remark 2.3. In view of Montel's criterion the inclusion $f_\lambda^N\big(S_{r_0}\big) \subset S_{r_0}$ in Lemma 2.2 implies that $S_{r_0} \subset \mathcal{A}^*(0, f_\lambda, U)$. We will use this simple fact without reference.

Lemma 2.2 in a slightly different form has been proved in [22]. We state its proof here for completeness of the discussion. The scheme of the proof is the following. The main idea is to fix arbitrary $z_0 \in \mathcal{L}_{r_0}$ and consider the function $s_N(\lambda) = s_N(z_0, \lambda) := \varphi(f_\lambda^N(z_0))$. The first step (Lemma 2.4) is to determine a neighbourhood of λ_0 where s_N is well defined, analytic and takes values from a prescribed domain of the form $\{\xi : \rho_1 < |\xi| < \rho_2\}$. The next step (Lemma 2.5) is to calculate the value of $(\partial/\partial\lambda)\log s_N(\lambda)$ at $\lambda = \lambda_0$, which turns out to be equal to

$$A_N(z_0) := \frac{s_N'(\lambda_0)}{s_N(\lambda_0)} = \sum_{k=0}^{N-1} G\big(\lambda_0^k \varphi(z_0)\big),$$

where G is an analytic function in $\mathbb{D}$. The concluding step is to use the equality $\int_0^1 G\big(e^{2\pi i t}\varphi(z_0)\big)\,dt = 1/\lambda_0$ and Theorem B in order to estimate $|A_N(z_0)|$ and $|\arg A_N(z_0)|$. This allows us to employ a consequence of the Schwarz lemma (Proposition 2.6) for proving that $|s_N(\lambda)| \le |\varphi(z_0)|$ for any λ satisfying $|\arg(1 - \lambda/\lambda_0)| < \Theta$ and $|\lambda - \lambda_0| < \varepsilon_*$. Since $z_0 \in \mathcal{L}_{r_0}$ is arbitrary, this means that $f_\lambda^N\big(S_{r_0}\big) \subset S_{r_0}$ for all such values of λ.

Lemma 2.4. *Under the conditions of Lemma* 2.2, $s_N(z,\lambda) := \varphi\big(f_\lambda^N(z)\big)$ *is a well-defined and analytic function for all* $z \in \overline{S_{r_0}}$ *and* $\lambda \in \mathrm{D}(\lambda_0, \varepsilon_N(\tau))$. *Moreover, the following inequality holds*

$$r_0 e^{-\tau} < |s_N(z,\lambda)| < r_0 e^{\tau}, \quad z \in \mathcal{L}_{r_0},\ \lambda \in \mathrm{D}(\lambda_0, \varepsilon_N(\tau)). \tag{2.3}$$

Proof. Let us show that for any $r_1 \in (0,1)$, $r_2 \in (r_1, 1)$ the following inclusion holds

$$B(z_0, r_1, r_2) := \big\{z : |z - z_0| < |z_0|\big(1 - k_\pi(r_1)/k_\pi(r_2)\big)\big\} \subset S_{r_2} \setminus \overline{S_{r_3}}, \tag{2.4}$$

where $z_0 \in \mathcal{L}_{r_1}$ and $r_3 := r_1^2/r_2$. To this end we remark that for any $z_0 \in \mathcal{L}_{r_1}$ the domain $S_{r_2} \setminus \overline{S_{r_3}}$ contains all points z such that

$$\big|\log(z/z_0)\big| < \log\big(k_\pi(r_2)/k_\pi(r_1)\big) \tag{2.5}$$

for some of the branches of log. To make sure this statement is true it is sufficient to employ the following estimate, see, e.g., [24, p. 117, inequal. (18)],

$$\left|\log \frac{z\psi'(z)}{\psi(z)}\right| \le \log \frac{1+|z|}{1-|z|}, \qquad z \in \mathbb{D}, \tag{2.6}$$

Owing to (2.6), for any rectifiable curve $\Gamma \subset \overline{S_{r_2}} \setminus S_{r_3}$ that joins z_0 with $\mathcal{L}_{r_2}$ or $\mathcal{L}_{r_3}$ we have

$$\begin{aligned}\int_\Gamma \left|\frac{dz}{z}\right| &= \int_{\varphi(\Gamma)} \left|\frac{\psi'(\xi)}{\psi(\xi)}\right| |d\xi| \ge \int_{\varphi(\Gamma)} \left|\frac{\psi'(\xi)}{\psi(\xi)}\right| d|\xi| \\ &\ge \min\left\{\int_{r_1}^{r_2} \frac{(1-r)dr}{(1+r)\,r}, \int_{r_3}^{r_1} \frac{(1-r)dr}{(1+r)\,r}\right\} = \log\big(k_\pi(r_2)/k_\pi(r_1)\big).\end{aligned}$$

Using the inequality $|\log(1+\xi)| \le -\log(1-|\xi|)$, $\xi \in \mathbb{D}$, we conclude that for any $z \in B(z_0, r_1, r_2)$,

$$\begin{aligned}\big|\log\big(z/z_0\big)\big| &= \big|\log\big(1 + (z - z_0)/z_0\big)\big| \\ &\le -\log\big(1 - |z - z_0|/|z_0|\big) < \log\big(k_\pi(r_2)/k_\pi(r_1)\big),\end{aligned}$$

i.e., all $z \in B(z_0, r_1, r_2)$ satisfy condition (2.5). Therefore inclusion (2.4) holds.

Let $r \in (0, e^{-\tau/N})$. Set $r' := re^{\tau/N}$ and $r'' := re^{-\tau/N}$. Consider an arbitrary function h subject to the following conditions: h is analytic in S, $h(0) = 0$, and $|h(z) - z| < |z|\big(1 - k_\pi(r)/k_\pi(r')\big)$ for all $z \in \overline{S_r} \setminus \{0\}$.

Set $r_1 := |z_0|$, $r_2 := |z_0|e^{\tau/N}$ for some $z_0 \in \overline{S_r} \setminus \{0\}$. Since $k_\pi(x)/k_\pi(xe^{\tau/N})$ increases with $x \in (0, r]$, the Schwarz lemma can be applied to the function $h(z)-z$ to conclude that $h(z_0) \in B(z_0, r_1, r_2)$ for all $z_0 \in \overline{S_r} \setminus \{0\}$. Therefore (2.4) implies the following inclusions

$$h(\overline{S_r}) \subset S_{r'}, \quad h(\mathcal{L}_r) \subset S_{r'} \backslash \overline{S_{r''}}. \tag{2.7}$$

By considering the function $\big(h(z) - z\big)/z$ with $f_{\lambda_0}(w)$ substituted for z it is easy to check that since the function f_{λ_0} is an automorphism of S_r for any $r \in (0, 1]$ (see Remark 1.4), the above argument can be applied to $h(z) := f_\lambda(f_{\lambda_0}^{-1}(z))$ for all $\lambda \in \mathrm{D}(\lambda_0, \varepsilon_N(\tau))$ and $r \in (0, r_*]$. Thus (2.7) implies that for any $\lambda \in \mathrm{D}(\lambda_0, \varepsilon_N(\tau))$,

$$f_\lambda\left(\overline{S_{r_j}}\right) \subset S_{r_{j+1}}, \qquad j = 0, 1, \ldots, N-1, \tag{2.8}$$

$$f_\lambda\left(\overline{S_{r_j}} \backslash S_{r_{-j}}\right) \subset S_{r_{j+1}} \backslash \overline{S_{r_{-(j+1)}}}, \qquad j = 0, 1, \ldots, N-1, \tag{2.9}$$

where $r_j := r_0 e^{j\tau/N}$, $j = 0, \pm1, \ldots, \pm N$. Applying (2.8) repeatedly, we see that $f_\lambda^N(\overline{S_{r_0}}) \subset S_{r_N}$. Similarly, (2.9) implies that $f_\lambda^N(\mathcal{L}_{r_0}) \subset S_{r_N} \backslash \overline{S_{r_{-N}}}$. The former means that the function $s_N(z, \lambda)$ is well defined and analytic for all $z \in \overline{S_{r_0}}$ and $\lambda \in \mathrm{D}(\lambda_0, \varepsilon_N(\tau))$, while the latter means that inequality (2.3) holds for indicated values of λ. This completes the proof of Lemma 2.4. □

Lemma 2.5. *Under the conditions of Lemma* 2.4*, the following equality holds*

$$A_N(z_0) := \frac{\partial \log s_N(z_0, \lambda)}{\partial \lambda}\bigg|_{\lambda=\lambda_0} = \sum_{k=0}^{N-1} G\big(\lambda_0^k \varphi(z_0)\big), \quad z_0 \in \mathcal{L}_{r_0}, \tag{2.10}$$

where

$$G(\xi) := \frac{u\big(\psi(\xi)\big)}{\lambda_0 \xi \psi'(\lambda_0 \xi)}.$$

Proof. Consider the following function of $n+1$ independent variables

$$g_n(z; \lambda_1, \ldots, \lambda_n) := \begin{cases} \big(f_{\lambda_n} \circ \cdots \circ f_{\lambda_1}\big)(z), & n \in \mathbb{N}, \\ z, & n = 0. \end{cases}$$

Note that

$$A_N(z_0) = \frac{\varphi'\big(f_{\lambda_0}^N(z_0)\big)}{s_N(z_0, \lambda_0)} \cdot \frac{\partial g_N(z_0; \lambda, \ldots, \lambda)}{\partial \lambda}\bigg|_{\lambda=\lambda_0} \qquad \text{and}$$

$$\frac{\partial g_N(z_0; \lambda, \ldots, \lambda)}{\partial \lambda}\bigg|_{\lambda=\lambda_0} = \sum_{k=0}^{N-1} g'_{N,k+1}(z_0; \lambda_0, \ldots, \lambda_0),$$

where $g'_{n,j}$ stands for $(\partial/\partial\lambda_j) g_n$. Using the equality

$$g_N(z; \lambda_1, \ldots, \lambda_n) = g_{N-j}\Big(f_{\lambda_j}\big(g_{j-1}(z; \lambda_1, \ldots, \lambda_{j-1})\big); \lambda_{j+1}, \ldots, \lambda_N\Big),$$

we get

$$g'_{N,k+1}(z_0;\lambda_0,\ldots,\lambda_0) = \big(f_{\lambda_0}^{N-k-1}\big)'\big(f_{\lambda_0}^{k+1}(z_0)\big)\cdot u\big(f_{\lambda_0}^{k}(z_0)\big).$$

Schröder equation (1.1) for $f := f_{\lambda_0}$ allows us to express $f_{\lambda_0}^j$ and $\big(f_{\lambda_0}^j\big)'$ in terms of φ and ψ. Denoting $z_j := f_{\lambda_0}^j(z_0)$, $j\in\mathbb{N}_0$, we obtain

$$\begin{aligned} g'_{N,k+1}(z_0;\lambda_0,\ldots,\lambda_0) &= \lambda_0^{N-k-1}\psi'\big(\lambda_0^{N-k-1}\,\varphi(z_{k+1})\big)\varphi'(z_{k+1})\,u(z_k)\\ &= \lambda_0^{N-k-1}\psi'\big(\lambda_0^{N-k-1}\,\varphi(z_{k+1})\big)\,\frac{u(z_k)}{\psi'\big(\varphi(z_{k+1})\big)}\\ &= \lambda_0^{N-k-1}\psi'\big(\lambda_0^{N}\varphi(z_0)\big)\,\frac{u\big(\psi\big(\lambda_0^k\varphi(z_0)\big)\big)}{\psi'\big(\lambda_0^{k+1}\varphi(z_0)\big)}. \end{aligned}$$

In the same way, we get

$$\frac{\varphi'\big(f_{\lambda_0}^N(z_0)\big)}{s_N(z_0,\lambda_0)} = \frac{1}{\psi'\big(\lambda_0^N\varphi(z_0)\big)\lambda_0^N\varphi(z_0)}.$$

Now one can combine the obtained equalities to deduce (2.10). □

Proposition 2.6. *Let $\tau>0$ and $\Theta\in(0,\pi/2)$. If a function $v(\varsigma)$ is analytic in $\mathbb{D}$ and satisfies the following inequalities*

$$|v(0)|e^{-\tau} < |v(\varsigma)| < |v(0)|e^{\tau},\quad \varsigma\in\mathbb{D}, \tag{2.11}$$

$$\vartheta := \big|\arg\{v'(0)/v(0)\}\big| + \Theta < \pi/2,$$

then the modulus of $t := \pi v'(0)/(4\tau v(0))$ does not exceed 1 and the following inequality holds

$$|v(\varsigma)| \ge |v(0)|,\quad \varsigma\in\Xi(\rho_0), \tag{2.12}$$

where $\Xi(\rho)$ stands for the circular sector $\{\varsigma : |\Im\mathrm{m}\,\varsigma| \le |\varsigma|\sin\Theta \le \rho\sin\Theta\}$ and $\rho_0 := \sqrt{\gamma^2+1}-\gamma$, $\gamma := (1-|t|^2)/(2|t|\cos\vartheta)$.

Proof. Replacing $v(\varsigma)$ with $v(\varsigma)/v(0)$, we can suppose that $v(0)=1$. The multivalued function

$$\phi(\xi) := h\left(\exp\left(\frac{i\pi\log\xi}{2\tau}\right)\right),\quad h(z) := -i\,\frac{z-1}{z+1},$$

maps the annulus $\{\xi : e^{-\tau} < |\xi| < e^{\tau}\}$ conformally onto $\mathbb{D}$ (in the sense of [24, p. 248]) and satisfies the conditions $\phi(1)=0$, $\phi'(1)>0$. Since the composition $f := \phi\circ v$ can be continued analytically along every path in $\mathbb{D}$, it defines an analytic function $f:\mathbb{D}\to\mathbb{D}$, $f(0)=0$. By the Schwarz lemma, $|f'(0)|\le 1$. Since $f'(0)=\phi'(1)v'(0)=\pi v'(0)/(4\tau)=t$, the first part of Proposition 2.6 is proved. To prove the remaining part we note that (2.12) is equivalent to the inequality $\Re\mathrm{e}\, f(\varsigma)\ge 0$. Applying the invariant form of the Schwarz lemma to $f(z)/z$, we obtain

$$\left|\frac{f(\varsigma)-f'(0)\varsigma}{\varsigma-\overline{f'(0)}f(\varsigma)}\right| \le |\varsigma|,\quad \varsigma\in\mathbb{D},$$

It follows that $f(\varsigma)$ lies in the closed disk of radius $R := |\varsigma|^2(1-|t|^2)/(1-|t\varsigma|^2)$ centred at $\sigma_0 := t\varsigma(1-|\varsigma|^2)/(1-|t\varsigma|^2)$. Therefore for the inequality $\Re\mathrm{e}\, f(\varsigma) \geq 0$ to be satisfied, it is sufficient that $\Re\mathrm{e}\,\sigma_0 \geq R$. An easy calculation leads to the following condition

$$\cos(\arg t + \arg \varsigma) \geq \frac{|\varsigma|(1-|t|^2)}{|t|(1-|\varsigma|^2)},$$

which is satisfied for all points of the arc

$$l(\rho) := \{\varsigma : |\Im\mathrm{m}\,\varsigma| \leq |\varsigma| \sin\Theta = \rho \sin\Theta\}, \quad \rho \in (0,1),$$

provided

$$\cos\vartheta \geq \frac{\rho(1-|t|^2)}{|t|(1-\rho^2)}. \tag{2.13}$$

The right-hand of (2.13) increases with $\rho \in (0,1)$ and $\rho := \rho_0$ satisfies (2.13). Therefore inequality (2.12) holds for all $\varsigma \in \bigcup_{\rho\in[0,\rho_0]} l(\rho) = \Xi(\rho_0)$. This completes the proof of Proposition 2.6. □

Proof of Lemma 2.2. Consider the function $s_N(z,\lambda)$ introduced in Lemma 2.4. This lemma states that $s_N(z,\lambda)$ is well defined and analytic for all $z \in \overline{S_{r_0}}$ and $\lambda \in \mathrm{D}(\lambda_0, \varepsilon_N(\tau))$ and satisfies inequality (2.3). According to Remark 1.4, $f_{\lambda_0}(\mathcal{L}_r) = \mathcal{L}_r$ for all $r \in [0,1)$. Consequently $|s_N(z,\lambda_0)| = |\varphi(z)|$, $z \in S$. Therefore for any $z_0 \in \mathcal{L}_{r_0}$ the function $v(\varsigma) := 1/s_N\big(z_0, \lambda_0(1-\varepsilon_N(\tau)\varsigma)\big)$ is analytic in $\mathbb{D}$ and satisfies inequality (2.11).

Let us employ now Proposition 2.6. To this end we compute the logarithmic derivative of $v(\varsigma)$ at $\varsigma = 0$. By Lemma 2.5,

$$\frac{v'(0)}{v(0)} = \lambda_0\varepsilon_N(\tau)A_N(z_0) = \lambda_0\varepsilon_N(\tau)\sum_{k=0}^{N-1} G\big(\lambda_0^k\varphi(z_0)\big).$$

Consider the sum $E_N := \sum_{k=0}^{N-1} G(\lambda_0^k\varphi(z_0))/N$. It can be regarded as an approximate value of the integral $E_* := \int_0^1 G\big(r_0 e^{2\pi i(t+t_0)}\big)\,dt$, where $t_0 \in \mathbb{R}$ is an arbitrary number, which does not affect E_*:

$$E_* = \frac{1}{2\pi i}\int\limits_{|\xi|=r_0} \frac{G(\xi)}{\xi}\,d\xi = \operatorname*{Res}_{\xi=0}\frac{G(\xi)}{\xi} = G(0) = \frac{1}{\lambda_0}.$$

Applying Theorem B to the points $x_n := x_n^\beta$, $\beta := (\arg\varphi(z_0))/(2\pi) - t_0$, and the function $\phi(t) := G\big(r_0 e^{2\pi i(t+t_0)}\big)$, we get the following estimate

$$|E_N - E_*| < Q_{\beta,N}\int_0^1 \big|(d/dt)G\big(r_0 e^{2\pi i(t+t_0)}\big)\big|\,dt.$$

Since $t_0 \in \mathbb{R}$ is arbitrary real, we have

$$|E_N - E_*| \leq Q_N\int_0^1 \big|(d/dt)G\big(r_0 e^{2\pi i(t+t_0)}\big)\big|\,dt. \tag{2.14}$$

The function under the sign $\int_0^1 |\cdot| dt$ is

$$\frac{dG\bigl(r_0 e^{2\pi i(t+t_0)}\bigr)}{dt} = 2\pi i\,\xi G'(\xi) = 2\pi i\,J(t+t_0), \quad \xi := r_0 e^{2\pi i(t+t_0)}.$$

From (2.14) it follows that

$$\left|\frac{1}{N}\cdot\sum_{k=0}^{N-1} G(\lambda_0^k\varphi(z_0)) - \frac{1}{\lambda_0}\right| \le a_N,$$

and hence,

$$\left|\frac{1}{N}\cdot\frac{v'(0)}{v(0)} - \varepsilon_N(\tau)\right| \le a_N\varepsilon_N(\tau). \tag{2.15}$$

Since by condition $0 \le a_N < 1$, inequality (2.15) implies that

$$\left|\frac{v'(0)}{v(0)}\right| \ge N(1-a_N)\varepsilon_N(\tau), \tag{2.16}$$

$$\left|\arg\frac{v'(0)}{v(0)}\right| \le \arcsin a_N. \tag{2.17}$$

Now if we recall that validity of (2.11) has been already verified and take into account (2.16), (2.17), we see that the conditions of Proposition 2.6 are satisfied. Therefore, by elementary reasoning we see that (2.12) holds for all $\varsigma \in \Xi\bigl(\Lambda_N\bigl(\tau, \varepsilon_N(\tau)\bigr)\bigr)$. In terms of s_N this means that

$$|s_N(z_0,\lambda)| \le |s_N(z_0,\lambda_0)| = r_0, \quad \lambda \in \Xi_0, \tag{2.18}$$

where

$$\Xi_0 := \bigl\{\lambda : \bigl|\lambda - \lambda_0\bigr| < \varepsilon_*, \bigl|\arg(1-\lambda/\lambda_0)\bigr| < \Theta\bigr\}.$$

Since $z_0 \in \mathcal{L}_{r_0} = \partial S_{r_0}$ is arbitrary in the above arguments, by the maximum modulus theorem, inequality (2.18) implies that $|\varphi(f_\lambda^N(z))| < r_0$ for all $z \in S_{r_0}$ and $\lambda \in \Xi_0$. Therefore for indicated values of λ we have $f_\lambda^N\bigl(S_{r_0}\bigr) \subset S_{r_0}$. This completes the proof of Lemma 2.2. □

2.2. Proof of Theorem 1.2

Suppose that the sequence $\{f_n\}_{n\in\mathbb{N}}$ satisfies the conditions of Theorem 1.2. Then every subsequence of f_n also meets these conditions. So we have only to prove that $S := \mathcal{A}^*(0, f_0, U)$ is the kernel of the sequence $A_n := \mathcal{A}^*(0, f_0, U)$, that is:

(i) any compact set $K \subset S$ lies in all but a finite number of A_n's;
(ii) S is the largest domain that contains $z = 0$ and satisfies condition (i).

Now we employ Lemma 2.2 in order to prove (i). To this end we should fix any $r_0 \in (0,1)$ such that $S_{r_0} \supset K$, specify appropriate values of N and τ, and trace the dependence on the choice of n_*. As a result we would prove that

$$\varepsilon_*^0 := \inf_{n_*\in\mathbb{N}} \varepsilon_* > 0. \tag{2.19}$$

Since $\lambda_n \to \lambda_0$ as $n \to +\infty$, (2.19) would imply that $K \subset S_{r_0} \subset \mathcal{A}^*(z_0, f_n, U)$ for all $n \in \mathbb{N}$ large enough.

Set $\tau := (1+r_0)/(2r_0)$. In view of condition (ii) of Theorem 1.2,

$$L := \sup_{n_*\in\mathbb{N}} \left(\int_0^1 |J(t)|\, dt \right) < +\infty.$$

Since by Remark 2.1, $Q_N \to 0$ as $N \to +\infty$, there exists $N \in \mathbb{N}$ such that

$$Q_N < \frac{\sin(\pi/4 - \Theta/2)}{2\pi L}.$$

Fix any such value of N. Then $a_N < \sin(\pi/4 - \Theta/2) < \sin(\pi/2 - \Theta)$. Hence Lemma (2.2) is applicable to the specified values of N and τ.

Let us estimate ε_* from below. In view of condition (ii) of Theorem 1.2,

$$\varepsilon_0 := \inf_{n_*\in\mathbb{N}} \varepsilon_N(\tau) > 0.$$

Denote $b := \pi\varepsilon_N(\tau)N(1-a_N)/(4\tau)$, $b_1 := \min\{1, b\}$.

Since $\vartheta = \Theta + \arcsin a_N < \pi/4 + \Theta/2 < \pi/2$, we have

$$\begin{aligned}
\Lambda_N\big(\tau, \varepsilon_N(\tau)\big) &\geq \frac{\sqrt{1 + 2b_1^2\cos 2\vartheta + b_1^4} - 1 + b_1^2}{2b_1\cos\vartheta} \\
&\geq \frac{\sqrt{1 + 2b_1^2\cos 2\vartheta + (b_1^2\cos 2\vartheta)^2} - 1 + b_1^2}{2b_1\cos\vartheta} \\
&> b_1\cos\vartheta > b_1\cos(\pi/4 + \Theta/2) \\
&\geq \cos(\pi/4 + \Theta/2)\min\left\{1, \frac{\pi\varepsilon_0 N\big(1 - \sin(\pi/4 - \Theta/2)\big)}{4\tau}\right\} =: C_0.
\end{aligned}$$

The constant C_0 is positive and does not depend on n_*. From the inequality $\varepsilon_* > \varepsilon_0 C_0$ it follows that (2.19) takes place. This proves assertion (i).

To prove (ii) let us assume the converse. Then there exists a domain $S' \not\subset S$, $0 \in S'$, satisfying (i). Let $z_0 \in S' \setminus S$ and $\Gamma \subset S'$ be a curve that joins points $z = 0$ and z_0. Consider any domain D such that $\Gamma \subset D$ and $K := \overline{D} \subset S'$. By the assumption, $K \subset A_n$ for all n large enough. Now we claim that

$$D \subset E(f_0, U). \tag{2.20}$$

Consider an arbitrary $\zeta_0 \in D$. Suppose that $\zeta_0 \notin E(f_0, U)$. Then there exists $j_0 \in \mathbb{N}$ such that $f_0^{j_0}$ is well defined (and so analytic) in some domain $D_0 \ni \zeta_0$, $D_0 \subset D$, with $f_0^j(\zeta_0) \in U$, $j < j_0$, but $f_0^{j_0}(\zeta_0) \notin U$. Since the sequence f_n converges to f_0 uniformly on each compact subset of U, the sequence $f_n^{j_0}$ converges to $f_0^{j_0}$ uniformly on each compact subset of D_0. According to Hurwitz's theorem, this means that $f_n^{j_0}(D_0) \not\subset U$ for all $n \in \mathbb{N}$ large enough. Consequently, $D \not\subset E(f_n, U)$ for large n. At the same time, $K = \overline{D} \subset A_n \subset E(f_n, U)$ for all n large enough. This contradiction proves (2.20).

The remaining part of the proof depends on the properties of the domain U. Since $U \subset \mathbb{C}$, we have three possibilities:

(Hyp) The domain U is hyperbolic. Then by Montel's criterion, $\mathcal{F}(f_0, U)$ coincides with the interior of $E(f_0, U)$. Since $D \ni 0$ is connected, we conclude that $z_0 \in \Gamma \subset D \subset S$. With this fact contradicting the assumption, the proof of (ii) for the hyperbolic case is completed.

(Euc) The domain U coincides with $\mathbb{C}$. The functions f_n, $n \in \mathbb{N}_0$, are entire functions.

(Cyl) The domain U is the complex plane punctured at one point.

Let us prove (ii) for case (Euc). Since $\Gamma \cap \partial S \neq \emptyset$ and $\partial S \subset \mathcal{J}(f_0, \mathbb{C})$, we have $D \cap \mathcal{J}(f_0, \mathbb{C}) \neq \emptyset$. The classical result proved for entire functions by I.N. Baker [25] asserts that the Julia set coincides with the closure of the set of all repelling periodic points. Therefore, D contains a periodic point of f_0 different from 0. Owing to Hurwitz's theorem, the same is true for f_n provided n is large enough. This leads to a contradiction, because the immediate basin of attraction $\mathcal{A}^*(0, f_n, U)$ contains no periodic points except for the fixed point $z = 0$. Assertion (ii) is now proved for case (Euc).

It remains to consider case (Cyl). Similarly to case (Euc), we need only to show that $D \setminus \{0\}$ contains a periodic point. By means of linear transformations we can assume that $U = \mathbb{C} \setminus \{1\}$. From (2.20) it follows that functions

$$\phi_n(z) := \frac{f_0^n(z) - z}{f_0^n(z) - 1}, \quad n \in \mathbb{N},$$

does not assume values 1 and ∞ in D. Since $D \cap \mathcal{J}(f_0, U) \neq \emptyset$, the family $\{\varphi_n\}_{n \in \mathbb{N}}$ is not normal in D. Hence, due to Montel's criterion, there exists $z_1 \in D$ and $n_0 \in \mathbb{N}$ such that $\phi_{n_0}(z_1) = 0$ and so $z_1 \in D$ is a periodic point of f_0. This completes the proof of (ii) for case (Cyl).

By now (i) and (ii) are shown to be true. Theorem 1.2 is proved. □

2.3. Proof of Theorem 1.3

Fix any $r_0 \in (0, 1)$. As in the proof of Theorem 1.2 one can make use of Lemma 2.2 to show that there exist $n_1, N \in \mathbb{N}$ such that $f_n^N(S_{r_0}) \subset S_{r_0}$ for all $n > n_1$. By Remark 1.4 the function φ_0 maps S conformally onto a Euclidian disk centred at the origin. It is convenient to rescale the dynamic variable, by replacing f_k, $k \in \mathbb{N}_0$, with $r f_k(z/r)$ for some constant $r > 0$, so that $\varphi_0(S) = \mathbb{D}$ (or equivalently $\varphi_0 = \varphi$). Then the functions $g_n(\zeta) := (1/r_0)\big(\varphi_0 \circ f_n^N \circ \varphi_0^{-1}\big)(r_0\zeta)$, $n > n_1$, are defined and analytic in $\mathbb{D}$. Furthermore, $g_n(0) = 0$ and $g_n(\mathbb{D}) \subset \mathbb{D}$ for any $n > n_1$. Let us observe that for any analytic function f with a geometrically attracting or Siegel fixed point z_0 the Kœnigs function φ associated with the pair (z_0, f) is the same as that of the pair (z_0, f^N). Hence it is easy to see that the function $\phi_n(\zeta) := \varphi_n\big(\varphi_0^{-1}(r_0\zeta)\big)/r_0$ is the Kœnigs function associated with $(0, g_n)$. Since $S = \varphi_0^{-1}(\mathbb{D})$ and $r_0 \in (0, 1)$ is arbitrary, it suffices to prove that $\phi_n(\zeta) \to \zeta$ as $n \to +\infty$ uniformly on each compact subset of $\mathbb{D}$.

According to Remark 1.4, the function f_0 is a conformal automorphism of S. Therefore, with f_n converging to f_0 uniformly on each compact subset of $U \supset S$, there exists $n_2 \geq n_1$ such that for all $n > n_2$ functions f_n^N and consequently g_n are univalent in S_{r_0} and in $\mathbb{D}$, respectively. It follows (see, e.g., [26]) that ϕ_n, $n > n_2$, are also univalent in $\mathbb{D}$. The convergence of f_n to f_0 implies also that g_n converges to g_0, $g_0(\zeta) := \lambda_0^N \zeta$, uniformly on each compact subset of $\mathbb{D}$.

We claim that there exists a sequence $\{r_n \in (0,1)\}_{n\in\mathbb{N}}$ converging to 1 such that for all $n > n_2$ the domain $\phi_n(r_n\mathbb{D})$ is contained in some disk $\{\xi : |\xi| < R_n\}$ that lies in $\phi_n(\mathbb{D})$. Owing to the Carathéodory convergence theorem and normality of the family $\{\phi_n : n \in \mathbb{N}, n > n_2\}$, this statement would imply convergence of the sequence ϕ_n to the identity map and hence the proof of Theorem 1.3 would be completed.

By p/q and p'/q' let us denote some successive convergents of the number $\alpha_n := \big(\arg g_n'(0)\big)/(2\pi) = (\arg \lambda_n^N)/(2\pi)$ (regardless of whether α_n is irrational or not). Put $\Omega_n := \phi_n(\mathbb{D})$, $\kappa_n := -\log|g_n'(0)| = -N\log|\lambda_n|$, $a_n := \kappa_n(q-1)$, and $b_n := \pi(1/q + 2/q')$. Consider a point $\zeta_0 \in \mathbb{D}$ and make use of the following inequality (see, e.g., [24, p. 117, inequal. (18)]) from the theory of univalent function

$$\left|\log \frac{\zeta\phi_n'(\zeta)}{\phi_n(\zeta)}\right| \leq \log\frac{1+|\zeta|}{1-|\zeta|}, \quad \zeta \in \mathbb{D},$$

to obtain

$$\int_\Gamma \left|\frac{dw}{w}\right| \geq -\log\big(4k_\pi(|\zeta_0|)\big), \quad k_\pi(z) := \frac{z}{(1+z)^2},\ z \in \mathbb{D}, \tag{2.21}$$

where Γ is any rectifiable curve that joins $\xi_0 := \phi_n(\zeta_0)$ with $\partial\Omega_n$ and lies in Ω_n except for one of the endpoints. The equality in (2.21) can occur only if ϕ_n is a rotation of the Koebe function $k_0(z) := z/(1-z)^2$ and Γ is a segment of a radial half-line. It follows that Ω_n contains the annular sector

$$\Sigma := \big\{\xi_0 e^{x+iy} : |x| \leq a_n,\ |x| \leq b_n,\ x, y \in \mathbb{R}\big\}$$

provided $|\zeta_0| \leq r_n := k_\pi^{-1}\big((1/4)\exp(-\sqrt{a_n^2 + b_n^2})\big)$. Moreover, Ω_n is invariant under the map $\zeta \mapsto \lambda_n^N\zeta$. Indeed,

$$\lambda_n^N\zeta = \lambda_n^N\phi_n\big(\phi_n^{-1}(\zeta)\big) = \phi_n\big(g_n\big(\phi_n^{-1}(\zeta)\big)\big) \in \Omega_n$$

for all $\zeta \in \Omega_n$. Denote

$$\Sigma_0 := \big\{\xi_0 e^{x+iy} : |x| \leq \pi/q, |x| \leq b_n,\ x, y \in \mathbb{R}\big\}, \quad \lambda_* := e^{-\kappa + 2\pi i p/q}.$$

Since p and q are coprime integers, the union of the annular sectors $\lambda_*^j\Sigma_0$, $j = 0, 1, \ldots, q-1$, contains the circle $\xi_0\mathbb{T}$, $\mathbb{T} := \partial\mathbb{D}$. The inequality from the theory of continued fractions $|\alpha_n - p/q| \leq 1/(qq')$ implies that

$$\lambda_*^j\Sigma_0 \subset \big(\lambda_n^N\big)^j\Sigma, \quad j = 0, 1, \ldots, q-1.$$

Therefore, for any $\xi_0 \in \phi_n(r_n\mathbb{D})$ the domain Ω_n contains the circle $\xi_0\mathbb{T}$. It follows that $\phi_n(r_n\mathbb{D})$ is a subset of some disk $\{\xi : |\xi| < R_n\}$ contained in Ω_n.

It remains to choose the successive convergents p/q and p'/q' of α_n in such a way that $r_n \to 1$ as $n \to +\infty$. To this end we fix some successive convergents p/q and p'/q' of $\alpha_* := (\arg \lambda_0^N)/(2\pi)$ and note that p/q and p'/q' are also successive convergents of α_n provided n is large enough, because $\alpha_n \to \alpha_*$ as $n \to +\infty$. Using the fact that $\kappa_n \to 0$ as $n \to +\infty$ and that the denominators of convergents of the irrational number α_* forms unbounded increasing sequence, we see that it is possible to choose p/q for each n in such a way that $\sqrt{a_n^2 + b_n^2} \to 0$ and, consequently, $r_n \to 1$ as $n \to +\infty$. The proof of Theorem 1.3 is now completed. □

3. Proof of Theorem 1.5

In this section we sketch the proof of Theorem 1.5. First of all we note that the proof of Lemma 2.2 does not use the fact that the dependence of $f_\lambda[n_*]$ (see equation (2.1)) on the parameter λ is linear. So Lemma 2.2 can be applied to any analytic family f_λ satisfying conditions (i)–(iii) on page 170, provided some notations are modified to a new (more general) setting. First of all we have to redefine $u(z) := \partial f_\lambda(z)/\partial\lambda|_{\lambda=\lambda_0}$. Then fix any $r \in (0,1)$ and consider the modulus of continuity for the family $h_\lambda := f_\lambda/f_{\lambda_0}$ calculated at $\lambda = \lambda_0$,

$$\omega_r(\delta) := \sup\Big\{\big|1 - f_\lambda(z)/f_{\lambda_0}(z)\big| : z \in S_r, \lambda \in W \cap \mathrm{D}(\lambda_0,\delta)\Big\}, \quad \delta > 0.$$

This quantity, as a function of δ, is defined, continuous, and increasing on the interval $I^* := (0,\delta^*)$, $\delta^* := \mathrm{dist}(\lambda_0, \partial W)$, with $\lim_{\delta\to+0}\omega_r(\delta) = 0$. Therefore there exists an inverse function $\omega_r^{-1} : (0,\epsilon^*) \to (0,+\infty)$, where $\epsilon^* := \lim_{\delta\to\delta^*-0}\omega_r(\delta)$. If $\epsilon^* \neq +\infty$, then we set $\omega_r^{-1}(\epsilon) := \delta^*$ for all $\epsilon \geq \epsilon^*$. Now we can redefine $\varepsilon_N(\tau)$ as

$$\varepsilon_N(\tau) := \omega_{r_*}^{-1}\big(1 - k_\pi(r_*)/k_\pi(r^*)\big), \quad r_* := r_0 e^{\tau(1-1/N)},\ r^* := r_0 e^{\tau}.$$

Finally, define Θ to be equal to the half-angle of Δ. To apply Lemma 2.2 we need the following

Proposition 3.1. *For any $n \in \mathbb{N}$ the following inequality holds*

$$Q_{q_n} < (1/q_n + 1/q_{n+1})/2. \tag{3.1}$$

Proof. Fix $n \in \mathbb{N}$. Due to the inequality $|\alpha_0 - p_n/q_n| < 1/(q_n q_{n+1})$ there exists $\gamma \in (0, 1/q_{n+1})$ such that

$$|\alpha_0 - p_n/q_n| < \gamma/q_n. \tag{3.2}$$

Let $\beta_0 := \big(1/q - (-1)^n\gamma\big)/2$. Taking into account that p_n and q_n are coprime integers one can deduce by means of the inequalities $\gamma < 1/q_{n+1} < 1/q_n$, $(-1)^n(\alpha_0 - p_n/q_n) > 0$, and (3.2) that

$$Q_{\beta_0,q_n} < (1/q_n + 1/q_{n+1})/2. \tag{3.3}$$

This proves the proposition. □

Now let us show how Theorem 1.5 can be proved. Fix $r_0 \in (0,1)$. Define $\varepsilon(r_0)$ in the following way.

According to Proposition 3.1, $0 < a_N < \sin\big(\pi/4 - \theta/2\big)$ for

$$N := \ell\left(2\pi\textstyle\int_0^1 \big|J(t)\big|\,dt\Big/\sin\big(\pi/4-\theta/2\big)\right), \quad \tau := \log\frac{1+2r_0}{3r_0}.$$

Hence, Lemma 2.2 can be used with the specified values of N and τ. Therefore, we can set $\varepsilon(r_0) := \varepsilon_*$, so that statement (i) in Theorem 1.5 becomes true. Let us show that statement (ii) of this theorem is also true, assuming that r_0 is sufficiently close to 1.

Since $\overline{S} \subset U$ there exists $C_1 > 0$ such that

$$\left|1 - \frac{f_\lambda(\psi(\xi))}{f_{\lambda_0}(\psi(\xi))}\right| < C_1|\lambda - \lambda_0|$$

for all $\xi \in \mathbb{D}$ and $\lambda \in \mathrm{D}(\lambda_0, \varepsilon^0)$, where $\varepsilon^0 > 0$ is chosen so that $\overline{\mathrm{D}(\lambda_0,\varepsilon^0)} \subset W$. It follows that $\omega_r^{-1}(s) \geq \min\{\varepsilon^0, s/C_1\}$, for all $s > 0$, $r \in (0,1)$. Elementary calculations show that

$$1 - \frac{k_\pi(r_*)}{k_\pi(r^*)} \geq 1 - \exp\left(-\frac{\tau(1-r^*)}{N(1+r^*)}\right) \geq C_2\frac{(1-r_0)^2}{N}$$

for some constant $C_2 > 0$. Combining these two inequalities we obtain $\varepsilon_N(\tau) \geq C_3(1-r_0)^2/N$, where $C_3 := C_2/C_1$. Now we estimate $\Lambda_N(\tau, \varepsilon_N(\tau))$ in the same way as in the proof of Theorem 1.2 to conclude that $\varepsilon_* \geq C(1-r_0)^3/N$ for some constant $C > 0$. To complete the proof we use the following inequalities (see, e.g., [24, p. 52]):

$$\left|\frac{\xi\psi''(\xi)}{\psi'(\xi)} - \frac{2r^2}{1-r^2}\right| \leq \frac{4r}{1-r^2}, \quad 0 \leq r = |\xi| < 1,$$

$$\left|\frac{\psi'(\xi)}{\psi'(0)}\right| \geq \frac{1-r}{(1+r)^3}, \quad 0 \leq r = |\xi| < 1,$$

which imply that $N \leq \ell\big((1-r_0)^{-\gamma}\big)$ for all $r_0 < 1$ sufficiently close to 1. □

4. Essentiality of conditions in Theorem 1.2

In this section we show that conditions (i) and (ii) in Theorem 1.2 are essential. As for condition (i) this can be regarded as a consequence of lower semi-continuity of the Julia set.

Example 1. Consider the family $f_\lambda(z) := \lambda z + z^2$ in the whole complex plane ($U := \mathbb{C}$). The map $\lambda \mapsto \mathcal{J}(f_\lambda, \mathbb{C})$ is lower semi-continuous [5], i.e.,

$$\mathcal{J}(f_{\lambda_*}, \mathbb{C}) \subset \bigcap_{\varepsilon>0}\bigcup_{\delta>0}\bigcap_{|\lambda-\lambda_*|<\delta} O_\varepsilon\big(\mathcal{J}(f_\lambda,\mathbb{C})\big) \text{ for any } \lambda_* \in \mathbb{C},$$

where $O_\varepsilon(\cdot)$ stands for the ε-neighbourhood of a set. Let $\lambda_0 := e^{2\pi i\alpha_0}$, $\alpha_0 \in \mathbb{R}\backslash\mathbb{Q}$, and $\alpha_n \in \mathbb{Q}$ converge to α_0 as $n \to +\infty$. The point $z_0 := 0$ is a parabolic fixed point of $f_{\lambda_n^0}$, $\lambda_n^0 := \exp(2\pi i\alpha_n)$, and so $0 \in \mathcal{J}(f_{\lambda_n^0}, \mathbb{C})$. Due to lower semi-continuity of $\lambda \mapsto \mathcal{J}(f_\lambda, \mathbb{C})$ at the points λ_n^0, there exists a sequence $\{\mu_n \in (0,1)\}_{n\in\mathbb{N}}$ such that $\mathrm{D}(0,1/n)\cap\mathcal{J}(f_{\lambda_n},\mathbb{C})\neq\emptyset$, $\lambda_n:=\mu_n\lambda_n^0$, $n\in\mathbb{N}$. It follows that $\mathcal{A}^*(0,f_{\lambda_n},\mathbb{C})\to\{0\}$ as to the kernel. Assume that f_{λ_0}, $\lambda_0:=\exp(2\pi i\alpha_0)$, has a Siegel point at $z_0=0$. This is the case if α_0 is a Brjuno number ([14, Th. 6], see also [15]). The sequence $f_n:=f_{\lambda_n}$ satisfies all conditions of Theorem 1.2 except for condition (i), but the conclusion of Theorem 1.2 fails to be true. Therefore condition (i) is an essential one.

It is known [27, p. 44] that condition (ii) can be omitted in Theorem 1.2 provided that the multiplier of the Siegel fixed point $\lambda_0:=f_0'(z_0)$ equals to $\exp(2\pi i\alpha_0)$ for some Brjuno number α_0. However, if no such assumptions concerning α_0 are made, condition (ii) cannot be omitted. This fact is demonstrated by the following

Example 2. Let α_0 be an irrational real number. By q_n denote the denominator of the nth convergent of α_0. Consider the sequence of polynomials

$$f_n(z):=\frac{\lambda_0\left(z+z^{q_n+1}\right)}{1+1/2^{q_n}},\quad \lambda_0:=e^{2\pi i\alpha_0},$$

converging to $f_0(z)=\lambda_0 z$ uniformly on each compact subset of $\mathbb{D}$.

We claim that the sequence of domains $\mathcal{A}^*(0,f_n,\mathbb{D})$ does not converge to $\mathcal{A}^*(0,f_0,\mathbb{D})=\mathbb{D}$ as to the kernel, provided the growth of q_n is sufficiently rapid. Assume the converse. Then for all $n\in\mathbb{N}$ large enough, say for $n>n_0$, the inclusion $D_{48}\subset\mathcal{A}^*(0,f_n,\mathbb{D})$ holds, where $D_j:=j/(j+1)\mathbb{D}$, $j\in\mathbb{N}$. It follows that $f_n^m(D_{48})\subset\mathbb{D}$ for all $n>n_0$, $n\in\mathbb{N}$, and $m\in\mathbb{N}$. Hence the family $\Phi:=\{f_n^m\}_{n>n_0,\,n,m\in\mathbb{N}_0}$ is normal in the disk D_{48}. In particular, there exist constants $C_1>1$, $C_2>0$ such that

$$\left|(f_n^m)'(z)\right|<C_1,\quad z\in D_8,\quad n>n_0,\ n,m\in\mathbb{N}_0, \tag{4.1}$$

$$\left|(f_n^m)''(z)\right|<C_2,\quad z\in D_8,\quad n>n_0,\ n,m\in\mathbb{N}_0. \tag{4.2}$$

Furthermore, by the Schwarz lemma,

$$f_n^m(D_4)\subset D_5,\ f_n^m(D_6)\subset D_7\quad n>n_0,\ n,m\in\mathbb{N}_0, \tag{4.3}$$

Consider functions $g_n:=f_n^{q_n}$, $\tilde g_n:=\tilde f_n^{q_n}$,

$$\tilde f_n(z):=\frac{\exp\left(2\pi i p_n/q_n\right)}{1+1/2^{q_n}}\left(z+z^{q_n+1}\right),\quad z\in\mathbb{D},\quad n>n_0,\quad n\in\mathbb{N},$$

where p_n stands for the numerator of the nth convergent of α_0. Apply the following inequality

$$\left|\tilde f_n(z)-f_n(z)\right|=\left|f_n(z)\right|\cdot\left|\lambda_0-\exp\left(2\pi i p_n/q_n\right)\right|$$
$$\le 4\pi\left|\alpha-\frac{p_n}{q_n}\right|\le\frac{4\pi}{q_nq_{n+1}},\quad z\in\mathbb{D}, \tag{4.4}$$

to prove that

$$\left|\tilde g_n(z)-g_n(z)\right|<\frac{4\pi C_1}{q_{n+1}},\quad z\in D_4, \tag{4.5}$$

for all $n\in\mathbb{N}$ large enough.

Since $q_n \to +\infty$ as $n \to +\infty$, there exists $n_1 \in \mathbb{N}$, $n_1 \ge n_0$, such that

$$\frac{4\pi}{q_n q_{n+1}} < \frac{1}{72} \quad \text{and} \quad \frac{4\pi C_1}{q_{n+1}} < \frac{1}{42}, \qquad n > n_1, \quad n \in \mathbb{N}.$$

We shall show that for all $n > n_1$, $n \in \mathbb{N}$, and $k = 1, 2, \ldots, q_n - 1$ the following implication holds

$$\Big(P(j),\ j = 1, 2, \ldots, k\Big) \Longrightarrow P(k+1), \tag{4.6}$$

where

$$P(j) : \Big[\ \big|\tilde f_n^{j-1}(z)\big| < 1, \quad z \in D_4, \quad \text{and} \quad \big|\tilde f_n^{j}(z) - f_n^{j}(z)\big| < \frac{4j\pi C_1}{q_n q_{n+1}}, \quad z \in D_4.\ \Big] \tag{4.7}$$

Now let $n > n_1$ and $P(j)$ take place for all $j = 1, 2, \ldots, k$. Relations (4.3), (4.4), and (4.7) imply the following inclusions

$$\tilde f_n(D_6) \subset D_8, \qquad \tilde f_n^{j}(D_4) \subset D_6, \quad j = 1, 2, \ldots, k. \tag{4.8}$$

For $j := k$ the latter guarantees that $|\tilde f_n^{k}(z)| < 1$, $z \in D_4$. Fix any $z \in D_4$ and denote $w_j := \tilde f_n^{j}(z)$, $\tilde\xi_j := \tilde f_n(w_j)$, $\xi_j := f_n(w_j)$. According to (4.3) and (4.8), we have $w_j \in D_6$, $\tilde\xi_j, \xi_j \in D_8$, $j = 1, 2, \ldots, k$. Taking this into account, from (4.1) and (4.4), we get the following inequality

$$\begin{aligned}
\Big|\tilde f_n^{k+1}(z) - f_n^{k+1}(z)\Big| &\le \sum_{j=0}^{k} \Big|\big(f_n^{k-j} \circ \tilde f_n^{j+1}\big)(z) - \big(f_n^{k-j+1} \circ \tilde f_n^{j}\big)(z)\Big| \\
&= \sum_{j=0}^{k} \Big|\big(f_n^{k-j} \circ \tilde f_n\big)(w_j) - \big(f_n^{k-j} \circ f_n\big)(w_j)\Big| \\
&= \sum_{j=0}^{k} \Big|f_n^{k-j}(\tilde\xi_j) - f_n^{k-j}(\xi_j)\Big| < \sum_{j=0}^{k} C_1 |\tilde\xi_j - \xi_j| \le \frac{4(k+1)\pi C_1}{q_n q_{n+1}}.
\end{aligned}$$

Therefore, (4.7) holds also for $j := k+1$. This proves implication (4.6).

For $j := 1$ inequality (4.7) follows from (4.4). Hence $P(1)$ is valid. Owing to (4.6), $P(1)$ implies $P(q_n)$. Therefore, inequality (4.5) holds for all $n > n_1$.

The functions $\tilde g_n$ have the fixed point $\tilde z_* := 1/2$. Now we apply (4.5) to show that if

$$q_{n+1} \ge 2^{q_n}, \qquad n \in \mathbb{N}, \tag{4.9}$$

then for any sufficiently large $n \in \mathbb{N}$ the function g_n has also a fixed point $z_* \in D_3 \setminus \{0\}$. Straightforward calculation gives

$$\tilde g_n'(\tilde z_*) = l_n := \left(\frac{1 + (q_n+1)/2^{q_n}}{1 + 1/2^{q_n}}\right)^{q_n} > 1.$$

From (4.2), (4.5), and the Cauchy integral formula it follows that

$$|\tilde g_n''(z)| < C_3 := C_2 + 51200\pi C_1/q_{n+1}, \quad z \in D_3, \quad n > n_1.$$

Now we assume that $n \in \mathbb{N}$ is large enough and apply Rouché's theorem to the functions $\tilde{g}_n(z) - z$ and $g_n(z) - z$ in the disk $B_n := \{z : |z - 1/2| < \rho_n\}$, where $\rho_n := (l_n - 1)/(2C_3)$. Since $B_n \subset D_3$, we have

$$\Re\mathrm{e}\, \frac{d}{dz}\Big(\tilde{g}_n(z) - z\Big) > \frac{l_n - 1}{2}, \qquad z \in B_n.$$

It follows that

$$|\tilde{g}_n(z) - z| \geq \frac{(l_n - 1)\rho_n}{2}, \qquad z \in \partial B_n. \tag{4.10}$$

Inequalities (4.5), (4.9), and (4.10) imply that $|\tilde{g}_n(z) - z| > |\tilde{g}_n(z) - g_n(z)|$ for all $z \in \partial B_n$. Consequently, $g_n(z) - z$ vanishes at some point $z_* \in B_n$. At the same time, the immediate basin $\mathcal{A}^*(0, f_n, \mathbb{D})$ contains no periodic points of f_n except for the fixed point at $z_0 = 0$. Therefore, $D_3 \not\subset \mathcal{A}^*(0, f_n, \mathbb{D})$ for large n. This fact implies that the sequence $\mathcal{A}^*(0, f_n, \mathbb{D})$ does not converge to $\mathbb{D}$ as to the kernel.

It is easy to see that the prescribed sequence f_n satisfies all conditions of Theorem 1.2 with $U := \mathbb{D}$ except for condition (ii), but the conclusion fails to hold. This shows that (ii) is also an essential condition in Theorem 1.2.

References

[1] Bergweiler W., An introduction to complex dynamics, *Textos de Matematica Serie B*, **6**, Universidade de Coimbra, 1995.

[2] Eremenko A., Lyubich M., The dynamics of analytic transformations, *Leningr. Math. J.*, **1** (1990), No. 3, 563–634 (in English); Russian original: *Algebra i analiz*, **1** (1989), No. 3, 1–70.

[3] Bergweiler W., Iteration of meromorphic functions, *Bull. Amer. Math. Soc., NS*, **29** (1993), No. 2, 151–188.

[4] Baker I.N., Domínguez P., Herring M.E., Dynamics of functions meromorphic outside a small set, *Ergod. Th. & Dynam. Sys.*, **21** (2001), No. 3, 647–672.

[5] Douady A., Does a Julia set depend continuously on the polynomial? Proc. Symp. in Appl. Math., **49** (1994) ed. R. Devaney, 91–138.

[6] Kisaka M., Local uniform convergence and convergence of Julia sets, *Nonlinearity*, **8** (1995), No. 2, 237–281.

[7] Kriete H., Repellors and the stability of Julia sets, *Mathematica Gottingensis*, 04/95 (1995), http://www.uni-math.gwdg.de/preprint/mg.1995.04.ps.gz.

[8] Krauskopf B., Kriete H., A note on non-converging Julia sets, *Nonlinearity*, **9** (1996), 601–603.

[9] Kriete H., Continuity of filled-in Julia sets and the closing lemma, *Nonlinearity*, **9** (1996), No. 6, 1599–1608.

[10] Krauskopf B., Kriete H., Hausdorff Convergence of Julia Sets, *Bull. Belg. Math. Soc.*, **6** (1999), No. 1, 69–76.

[11] Wu Sh., Continuity of Julia sets, *Sci. China, Ser. A*, **42** (1999), No. 3, 281–285.

[12] Kriete H., Approximation of indifferent cycles, *Mathematica Gottingensis*, 03/96 (1996), http://www.uni-math.gwdg.de/preprint/mg.1996.03.ps.gz.

[13] Siegel C., Iteration of analytic function, *Ann. of Math.*, **43** (1942), No. 2, 607–612.

[14] Brjuno A., Analytic forms of differential equations, *Trans. Mosc. Math. Soc.*, **25** (1971), 131–288.

[15] Yoccoz J.C., Petits diviseurs en dimension 1, *Astérisque*, **231** (1995) (in French).

[16] Milnor J., *Dynamics in one complex variable. Introductory lectures*, 2nd edition, Vieweg, 2000.

[17] Bargmann D., Conjugations on rotation domains as limit functions of the geometric means of the iterates, *Ann. Acad. Sci. Fenn. Math.*, **23** (1998), No. 2, 507–524.

[18] Buff X., Petersen C., On the size of linearization domains, *Preprint* (2007) http://www.picard.ups-tlse.fr/˜buff/Preprints/Preprints.html.

[19] Bukhshtab A. A., *Theory of numbers*, "Prosvyascheniye", Moscow, 1966 (in Russian).

[20] Coppel W. A., *Number Theory. Part A*, Springer, 2006.

[21] Hedenmalm H., Shimorin S., Weighted Bergman spaces and the integral means spectrum of conformal mappings, *Duke Mathematical Journal* **127** (2005), 341–393.

[22] Gumenuk P.A., Siegel disks and basins of attraction for families of analytic functions, *Izvestiya Saratovskogo Universiteta, NS, Ser. Matem. Mekh. Inform.*, **5** (2005), No. 1, 12–26 (in Russian).

[23] Sobol' I.M., *Multidimensional quadrature formulas and Haar functions*, "Nauka", Moscow, 1969 (in Russian).

[24] Goluzin G.M., *Geometric theory of functions of a complex variable*, 2^{nd} ed., "Nauka", Moscow, 1966 (in Russian); German transl.: Deutscher Verlag, Berlin, 1957; English transl.: Amer. Math. Soc., 1969.

[25] Baker I.N., Repulsive fixpoints of entire functions, *Math. Z.*, **104** (1968), 252–256.

[26] Goryainov V., Kœnigs function and fractional iterates of probability generating functions, *Sb. Math.*, **193** (2002), No. 7, 1009–1025; translation from *Mat. Sb.* **193** (2002), No. 7, 69–86.

[27] Buff X., Disques de Siegel & Ensembles de Julia d'aire strictement positive, *Diplôme D'Habilitation à diriger des Recherches,* Univ. Paul Sabatier, Toulouse 2006 (in French), http://www.picard.ups-tlse.fr/˜buff.

Pavel Gumenyuk
Department of Mathematics, University of Bergen
Johannes Brunsgate 12,
5008 Bergen, Norway

and

Department of Mechanics and Mathematics
Saratov State University,
Astrakhanskaya 83
410012 Saratov, Russia
e-mail: `Pavel.Gumenyuk@math.uib.no`
`Gumenuk@sgu.ru`

Analysis and Mathematical Physics
Trends in Mathematics, 187–192
© 2009 Birkhäuser Verlag Basel/Switzerland

Rings and Lipschitz Continuity of Quasiconformal Mappings

Vladimir Gutlyanskiĭ and Anatoly Golberg

Abstract. Sufficient conditions which guarantee Lipschitz's and Hölder's continuity for quasiconformal mappings in $\mathbb{R}^n$ are established. These conditions are based on a spatial analogue of the planar angular dilatation and can be regarded as Teichmüller-Wittich regularity theorems.

Mathematics Subject Classification (2000). 30C65, 30C62, 26B35.

Keywords. Quasiconformal mappings, Lipschitz and Hölder continuity, module of curve families, module of ring domain.

1. Introduction

A homeomorphism $f : G \to \mathbb{R}^n$ is called **weakly Lipschitz continuous** at $x_0 \in G$ if for every $0 < \alpha < 1$ there is a constant $C > 0$ such that

$$|f(x) - f(x_0)| \leq C|x - x_0|^\alpha \tag{1.1}$$

holds if $|x - x_0| < \delta$, where $\delta > 0$ is a sufficiently small number. If (1.1) holds with $\alpha = 1$, then f is called **Lipschitz continuous** at $x_0 \in G$.

For many questions concerning quasiconformal mappings in space it would be desirable to have criteria for such mappings to be Lipschitz or weakly Lipschitz continuous in a prescribed point. Moreover, it would be desirable to have such criteria written in terms of some integral means of suitable local dilatation coefficients. There are several such results in the plane, see, e.g., [1], [3], [8], [11] and not much seems to be known if $n > 2$, see, e.g., [2], [6], [9], [12].

In this paper we consider the dilatation $D_f(x, x_0)$ of the mapping $f : \Omega \to \mathbb{R}^n$ at the point $x \in \Omega$ with respect to $x_0 \in \mathbb{R}^n$, $x \neq x_0$, which is defined by

$$D_f(x, x_0) = \frac{J_f(x)}{\ell_f^n(x, x_0)}, \tag{1.2}$$

where

$$\ell_f(x,x_0) = \min \frac{|\partial_h f(x)|}{|\langle h, \frac{x-x_0}{|x-x_0|}\rangle|}. \tag{1.3}$$

Here $\partial_h f(x)$ denotes the derivative of f at x in the direction h and the minimum is taken over all unit vectors $h \in \mathbb{R}^n$; $J_f(x)$ is the Jacobian of f at x. For brevity, we will write $D_f(x,0) = D_f(x)$. The dilatation $D_f(x,x_0)$ is a measurable function in Ω and satisfies the inequalities

$$\frac{1}{K_f(x)} \le D_f(x,x_0) \le L_f(x),$$

where

$$K_f(x) = \frac{||f'(x)||^n}{J_f(x)}, \quad L_f(x) = \frac{J_f(x)}{\ell(f'(x))^n}$$

stand for the well-known outer and inner dilatations of f at x, respectively. Here $||f'(x)|| = \sup |f'(x)h|$ and $\ell(f'(x)) = \inf |f'(x)h|$ taken over all unit vectors $h \in \mathbb{R}^n$. For $n = 2$ the dilatation $D_f(z,z_0)$ in the form (1.2) and (1.3) has been defined in [10]. For this case the above double inequality reads as

$$\frac{1-|\mu(z)|}{1+|\mu(z)|} \le \frac{\left|1-\mu(z)\frac{\bar{z}-\bar{z}_0}{z-z_0}\right|^2}{1-|\mu(z)|^2} \le \frac{1+|\mu(z)|}{1-|\mu(z)|}$$

for almost all z, see [4], cf. [10]. Here $\mu(z) = f_{\bar{z}}/f_z$ is the complex dilatation of $f(z)$ at z. In other words, for $n = 2$ the dilatation $D_f(z,z_0)$ coincides with the well-known angular dilatation $D_{\mu,z_0}(z)$ of f at z with respect to z_0, $z \ne z_0$, given by

$$D_{\mu,z_0}(z) = \frac{\left|1-\mu(z)\frac{\bar{z}-\bar{z}_0}{z-z_0}\right|^2}{1-|\mu(z)|^2} = \frac{|\partial_\theta f(z)|^2}{r^2 J_f(z)}.$$

Here $z = z_0 + re^{i\theta}$. Thus, the dilatation $D_f(x,x_0)$ can be viewed as a spatial counterpart of the **angular dilatation** $D_{\mu,z_0}(z)$, which have been used by many authors for the study of quasiconformal mapping in the plane, see, e.g., [4], [5], [7], [8], and the reference therein.

Let $\mathbb{B}^n$ denote the unit ball in $\mathbb{R}^n$. Suppose that $f(x)$ is a quasiconformal mapping in $\mathbb{B}^n$ normalized by $f(0) = 0$. It is known, see [2], [12], that the condition

$$\int_{|x|<R} \frac{K_f(x)-1}{|x|^n} dx < +\infty \tag{1.4}$$

for some $R > 0$ implies for f to have conformal dilatation at the origin, that is, the limit $\lim_{x\to 0} |f(x)|/|x|$ exists and differs from 0 and ∞. This result is referred to as the spatial counterpart of the well-known Teichmüller-Wittich regularity theorem [13], [15].

The sufficient condition for a mapping to be weakly Lipschitz continuous is given in [6]. It has been proved that if $f : \mathbb{B}^n \to \mathbb{R}^n$, $f(0) = 0$, is a quasiconformal mapping, satisfying

$$\lim_{t\to 0}\frac{1}{\Omega_n t^n}\int_{|x|<t} K_f(x)dx = 1, \tag{1.5}$$

then f is weakly Lipschitz continuous at 0, that is Hölder continuous with every exponent less than one.

2. Module of ring domain

Let $\mathcal{E}$ be a family of Jordan arcs or curves in space $\mathbb{R}^n$. A nonnegative and Borel measurable function ρ defined in $\mathbb{R}^n$ is called **admissible** for the family $\mathcal{E}$ if the relation

$$\int\limits_{\gamma} \rho ds \geq 1$$

holds for every locally rectifiable $\gamma \in \mathcal{E}$. The quantity

$$\mathcal{M}(\mathcal{E}) = \inf_{\rho} \int\limits_{\mathbb{R}^n} \rho^n \, dx,$$

where the infimum is taken over all ρ admissible with respect to the family $\mathcal{E}$ is called the **modulus** of the family $\mathcal{E}$ (see, e.g., [14], p. 16).

A ring domain $\mathcal{R} \subset \mathbb{R}^n$ is defined as a finite domain whose complement consists of two components C_0 and C_1. The sets $F_0 = \partial C_0$ and $F_1 = \partial C_1$ are two boundary components of $\mathcal{R}$. For definiteness, let us assume that $\infty \in C_1$.

We say that a curve γ **joins the boundary components in** $\mathcal{R}$ if γ lies in $\mathcal{R}$, except for its endpoints, one of which lies on F_0 and the second on F_1. Denote by $\Gamma_{\mathcal{R}}$ the family of all locally rectifiable curves γ which join the boundary components of $\mathcal{R}$.

The module of a ring domain $\mathcal{R}$ can be represented in the form

$$\operatorname{mod} \mathcal{R} = \left(\frac{\omega_{n-1}}{M(\Gamma_{\mathcal{R}})}\right)^{\frac{1}{n-1}};$$

here ω_{n-1} is the $(n-1)$-dimension Lebesgue measure of the unit sphere S^{n-1} in $\mathbb{R}^n$.

When $\mathcal{R}$ is a spherical annulus $A(x_0; r, R) = \{x \in \mathbb{R}^n : 0 < r < |x - x_0| < R < \infty\}$, its module is given by

$$\operatorname{mod} A(x_0; r, R) = \log \frac{R}{r}.$$

The conformal invariance of the quantity $\operatorname{mod} \mathcal{R}$ in the planar case allows us to restrict ourselves by the circular annulus $\{1 < |z| < e^{\operatorname{mod} \mathcal{R}}\}$. But this is impossible in $\mathbb{R}^n$, with $n \geq 3$.

3. Lipschitz and Hölder continuity of quasiconformal mappings

In this section, we establish the following two theorems.

Theorem 3.1. *Suppose that $f : \mathbb{B}^n \to \mathbb{B}^n$, $f(0) = 0$, is a quasiconformal mapping, such that*

$$\limsup_{r\to 0} \int_{r<|x|<1} \frac{D_f(x) - 1}{|x|^n} dx < +\infty.$$

Then

$$|f(x)| \le C|x|$$

holds in $\mathbb{B}^n$ with some constant C depending only on the value of the above supremum, and therefore, the mapping f is Lipschitz continuous at the origin.

Remark 3.2. The inequality (1.4) implies for f to be Lipschitz continuous at the origin. On the other hand, the mapping

$$f(x) = (x_1 \cos\theta - x_2 \sin\theta, x_2 \cos\theta + x_1 \sin\theta, x_3, \ldots, x_n), \quad f(0) = 0, \quad |x| < 1,$$

where $x = (x_1, \ldots, x_n)$ and $\theta = \log(x_1^2 + x_2^2)$, for which $K_f(x) \equiv (1 + \sqrt{2})^n$ and $D_f(x) \equiv 1$, satisfies the assumption of Theorem 3.1 whereas the integral (1.4) diverges.

Theorem 3.3. *Let $f : \mathbb{B}^n \to \mathbb{B}^n$, $f(0) = 0$, be a quasiconformal mapping, such that*

$$\limsup_{t\to 0} \frac{1}{\Omega_n t^n} \int_{|x|<t} D_f(x) dx \le \frac{1}{M} \tag{3.1}$$

holds for a constant $M > 0$. Then for each $0 < \alpha < M^{1/(n-1)}$, there is a constant C such that

$$|f(x)| \le C|x|^\alpha, \quad x \in \mathbb{B}^n.$$

Remark 3.4. The radial stretching in $\mathbb{R}^n$

$$f(x) = x|x|^{Q-1}, \ \ Q \ge 1$$

has

$$D_f(x) = 1/Q^{n-1}, \quad K_f(x) = Q^{n-1}, \quad L_f(x) = Q.$$

So, the inequality (3.1) holds with $M = Q^{n-1}$. By Theorem 3.3, $|f(x)| \le C|x|^\alpha$, $0 < \alpha < Q$. But the sufficient condition (1.5) is not fulfilled.

The proofs of the above theorems are based on series of lemmas. Let us note only two most important statements.

Lemma 3.5. *Let $f : \Omega \to \mathbb{R}^n$ be a quasiconformal mapping. Suppose that $D_f(x, x_0)$ is locally integrable in the annulus $A = A(x_0; r, R) \subset \Omega$. Then for each nonnegative measurable function $\rho(t)$, $t \in (r, R)$, such that*

$$\int_r^R \rho(t) dt = 1,$$

the following inequality holds

$$\mathcal{M}(f(\Gamma)) \leq \int\limits_A \rho^n(|x - x_0|) D_f(x, x_0)dx,$$

where Γ *stands for the family of curves joining the boundary components of* A *in* A*.*

Lemma 3.6. *Let* $f : \mathbb{R}^n \to \mathbb{R}^n$*,* $f(0) = 0$*,* $n \geq 2$*, be a quasiconformal mapping with the angular dilatation coefficient* $D_f(x)$*. Then for every spherical annulus* $A = A(0; r, R)$ *centered at the origin*

$$\frac{\operatorname{mod} f(A)}{\operatorname{mod} A} \geq \left(\frac{1}{\omega_{n-1} \operatorname{mod} A} \int_A \frac{D_f(x)}{|x|^n} dx \right)^{\frac{1}{1-n}}.$$

If additionally $\operatorname{mod} A - \operatorname{mod} f(A) \geq 0$*, then*

$$\operatorname{mod} A - \operatorname{mod} f(A) \leq \frac{1}{\omega_{n-1}} \int\limits_A \frac{D_f(x) - 1}{|x|^n} dx.$$

References

[1] P.P. Belinskiĭ, *Behavior of a quasi-conformal mapping at an isolated singular point*, L'vov. Gos. Univ. Uč. Zap. Ser. Meh.-Mat. **29** (1954), no. 6, 58–70. (Russian)

[2] Ch. Bishop, V.Ya. Gutlyanskiĭ, O. Martio, M. Vuorinen, *On conformal dilatation in space*, Int. J. Math. Math. Sci. 2003, no. 22, 1397–1420.

[3] M. Brakalova, J.A. Jenkins, *On the local behavior of certain homeomorphisms*, Kodai Math. J. **17** (1994), no. 2, 201–213.

[4] V. Gutlyanskiĭ, O. Martio, T. Sugawa, M. Vuorinen, *On the degenerate Beltrami equation*, Trans. Amer. Math. Soc. **357** (2005), no. 3, 875–900 (electronic).

[5] V. Ya. Gutlyanskiĭ, T. Sugawa, *On Lipschitz Continuity of Quasiconformal Mappings,* Report. Univ. Juväskylä, **83** (2001), 91–108.

[6] V. Gutlyanskiĭ, O. Martio, M. Vuorinen, *On Hölder continuity of quasiconformal maps with VMO dilatation*, Complex Var. Theory Appl. **47** (2002), no. 6, 495–505.

[7] O. Lehto, *On the differentiability of quasiconformal mappings with prescribed complex dilatation*, Ann. Acad. Sci. Fenn. Ser. A I No. **275** (1960), 28 pp.

[8] E. Reich, H. Walczak, *On the behavior of quasiconformal mappings at a point*, Trans. Amer. Math. Soc. **117** (1965), 338–351.

[9] S. Rohde, *Bilipschitz maps and the modulus of rings*, Ann. Acad. Sci. Fenn. Math. **22** (1997), no. 2, 465–474.

[10] V. Ryazanov, U. Srebro, E. Yakubov, *On ring solutions of Beltrami equations*, J. Anal. Math. 96 (2005), 117–150.

[11] A. Schatz, *On the local behavior of homeomorphic solutions of Beltrami's equations*, Duke Math. J. **35** (1968), 289–306.

[12] K. Suominen, *Quasiconformal maps in manifolds*, Ann. Acad. Sci. Fenn. Ser. A I No. **393** (1966), 39 pp.

[13] O. Teichmüller, *Untersuchungen über konforme und quasikonforme Abbildungen*, Deutsche Math. **3** (1938), 621–678.

[14] J. Väisälä, *Lectures on n-dimensional quasiconformal mappings*, Lecture Notes in Mathematics, Vol. 229. Springer-Verlag, Berlin-New York, 1971.

[15] H. Wittich, *Zum Beweis eines Satzes über quasikonforme Abbildungen*, Math. Z. **51** (1948), 278–288.

Vladimir Gutlyanskiĭ
Institute of Applied Mathematics
and Mechanics, NAS of Ukraine
74 Roze Luxemburg
83114 Donetsk, Ukraine
e-mail: `gut@iamm.ac.donetsk.ua`

Anatoly Golberg
Holon Institute of Technology
52 Golomb St.
P.O.Box 305
58102 Holon, Israel
e-mail: `golberga@hit.ac.il`

Analysis and Mathematical Physics
Trends in Mathematics, 193–204
© 2009 Birkhäuser Verlag Basel/Switzerland

A Theoretical Algorithm to get a Schottky Uniformization from a Fuchsian one

Rubén A. Hidalgo

Abstract. Riemann surfaces appear in many different areas of mathematics and physics, as in algebraic geometry, the theory of moduli spaces, topological field theories, cosmology, quantum chaos and integrable systems. A closed Riemann surface may be described in many different forms; for instance, as algebraic curves and by means of different topological classes of uniformizations. The highest uniformization corresponds to Fuchsian groups and the lowest ones to Schottky groups. In this note we discuss a theoretical algorithm which relates a Schottky group from a given Fuchsian group both uniformizing the same closed Riemann surface.

Mathematics Subject Classification (2000). 30F10, 30F40.

Keywords. Riemann surface, Fuchsian groups, Schottky groups.

1. Introduction

Riemann surfaces appear in many different areas of mathematics and physics, as in algebraic geometry, the theory of moduli spaces, topological field theories, cosmology, quantum chaos and integrable systems. For practical use of Riemann surfaces, one needs efficient numerical approaches. In recent years considerable progress has been achieved in the numerical treatment of Riemann surfaces which stimulated further research in the subject and led to new applications. In this note we will be mainly interested on closed Riemann surfaces and its uniformizations. An *uniformization* of a Riemann surface S is provided by a tuple $(\Omega, G, P : \Omega \to S)$, where G is a Kleinian group, Ω is a G-invariant component of its region of discontinuity and $P : \Omega \to S$ is a regular planar covering with G as group of cover transformations. There is a natural partial order on the collection of uniformizations of S; the uniformization $(\Omega_1, G_1, P_1 : \Omega_1 \to S)$ is bigger than the uniformization $(\Omega_2, G_2, P_2 : \Omega_2 \to S)$ if there is a covering map $Q : \Omega_1 \to \Omega_2$ so that $P_1 = P_2 Q$.

This work was completed with the support of project Fondecyt 1070271 and UTFSM 12.08.01.

The highest uniformization is provided by the universal covering (if S is a closed Riemann surface of genus $g \geq 2$, then these are provided by Fuchsian groups G). In Section 2 we recall the definition of Schottky uniformizations of closed Riemann surfaces (these uniformizations correspond to the lowest ones).

A closed Riemann surface may be described by different objects, for instance, by algebraic projective curves, Riemann period matrices and uniformizations. A natural question is, given one of the above representations of a closed Riemann surface, to find explicitly any of the others. Unfortunately, in the general case this has not been possible. In some few cases (closed Riemann surfaces with a large group of conformal automorphisms) one may produce explicitly both a Fuchsian uniformization and an algebraic curve corresponding to the same (conformal equivalent) closed Riemann surface, see for instance the early works [4, 5].

The *general numerical uniformization problem* may be stated as to find numerical algorithms which permits to find, given explicitly one of the representations of a closed Riemann surface, some of the other ones. In this direction, in [9] P. Buser and R. Silhol explain how to obtain an algebraic curve in terms of a Fuchsian uniformization, in [12] Gianni and Seppälä explain how to obtain Riemann period matrices in terms of algebraic curves, in [17, 26] (based in original works of P. Myrberg [25]) Seppälä explains how to obtain Schottky uniformizations in terms of hyperelliptic algebraic curves and in [16] (based in original works of W. Burnside [8]) it is explain how to obtain the algebraic curves in terms of Schottky uniformizations.

At the level of uniformizations, in general if we start with some explicit uniformization for a closed Riemann surface S, then it is not easy to get explicitly all or some of the other uniformizations of it. In this paper we deal within the following *particular numerical uniformization problem*: Given an explicit uniformization of S, to find algorithmically any of the other uniformizations of it. More precisely, if we are given explicitly a Fuchsian uniformization of a closed Riemann surface S, then we provide a theoretical algorithm that permits to obtain an explicit Schottky uniformization of it.

Genus 1

If S is closed Riemann surface of genus $g = 1$ (tori), then everything can be done explicitly. In this case, S is uniformized (universal uniformization) by using a Kleinian group of the form $G_\tau = \langle A(z) = z + 1, B_\tau(z) = z + \tau \rangle \cong \mathbb{Z}^2$, where $\tau \in \mathbb{H}^2$. Any two values $\tau_1, \tau_2 \in \mathbb{H}^2$ provide uniformizations of the same conformal class of tori if and only if there is a Möbius transformation $T(z) = (az+b)/(cz+d)$, where $a, b, c, d \in \mathbb{Z}$ and $ad-bc = 1$, such that $\tau_2 = T(\tau_1)$. Schottky uniformizations of the same S, say given by $\tau \in \mathbb{H}^2$, is provided by the group $K_\lambda = \langle C(z) = \lambda z \rangle$, where $\lambda = \exp(2\pi i \tau)$.

Genus $g \geq 2$

If S is a closed Riemann surface of genus $g \geq 2$, then the uniformization theorem ensures the existence of an uniformization $(\Delta, \Gamma, \pi : \Delta \to S)$, where Δ is the unit disc and Γ is a Fuchsian group isomorphic to the fundamental group of S. Any subgroup of finite index of Γ is again a surface group of higher genus while any subgroup of infinite index must be a free group (known since the early 1900's and proved purely algebraically using Reidemeister-Schreier rewriting by Hoare, Karrass and Solitar in 1971 [18, 19]). Let $(\Omega, G, P : \Omega \to S)$ be an uniformization of S, where Ω is not simply-connected. There is a normal subgroup N so that $\Omega = \Delta/N$ and $G = \Gamma/N$. By the above, N must have infinite index in Δ and, in particular, N is an infinite rank free group. It follows that every finitely generated subgroup of N is in fact a Schottky group. This last fact will be used frequently below. In this case, the numerical uniformization problem ask for some algorithm that permits to get G in terms of Γ.

This paper is organized as follows. In Section 2 we recall some equivalent definitions of Schottky groups and Schottky uniformizations. In Sections 3 and 4 we discuss a theoretical algorithm which relates a Schottky group from a given Fuchsian group both uniformizing the same closed Riemann surface. In Section 5 we discuss two possible numerical implementations. The second numerical implementation relays on the works of D. Crowdy and J.S. Marshall on quadrature domains and the Schottky-Klein prime function [11].

2. Schottky uniformizations

Let C_k, C'_k, $k = 1, \dots, g$, be $2g$ Jordan curves on the Riemann sphere $\widehat{\mathbb{C}}$ such that they are mutually disjoint and bound a $2g$-connected domain, say $\mathcal{D}$. Suppose that for each k there exists a fractional linear transformation A_k so that (i) $A_k(C_k)=C'_k$ and (ii) $A_k(\mathcal{D}) \cap \mathcal{D} = \emptyset$. Let G be the group generated by all these transformations. As consequence of Klein-Maskit's combination theorems, G is a Kleinian group, all its non-trivial elements are loxodromic and a fundamental domain for it is given by $\mathcal{D}$. The group G is called a *Schottky group of genus* g, the set of generators $A_1, \dots, A_g$ is called a *Schottky system of generators* and the loops C_1, $C'_1, \dots, C_g$, C'_g, is called a *fundamental set of loops respect to these generators*. That every set of g generators of G is always a Schottky system of generators is due to V. Chuckrow [10].

In [21] is proved that a Schottky group of rank g is equivalent to a Kleinian group, with non-empty region of discontinuity, which is purely loxodromic and isomorphic to a free group of rank g. As a consequence of Marden's isomorphism theorem [24], a Schottky group of rank g is equivalent to a geometrically finite Kleinian group which is purely loxodromic and isomorphic to a free group of rank g.

If Ω is the region of discontinuity of a Schottky group G, say of rank g, then it is known that Ω is connected and that $S = \Omega/G$ is a closed Riemann surface of genus g. Retrosection theorem (see [6] for a modern proof) asserts that for every closed Riemann surface S of genus g there exists a Schottky group G of genus g with Ω/G holomorphically equivalent to S; we say that S is *uniformized* by the Schottky group G and that $(\Omega, G, P : \Omega \to S)$ is a *Schottky uniformization* of S. As a consequence of the results in [22], the lowest uniformizations of a closed Riemann surface are provided by the Schottky ones.

We will need the following basic property on conformal maps defined on the region of discontinuity of a Schottky group.

Lemma 2.1. *Let G be a Schottky group with region of discontinuity Ω. If $T : \Omega \to \widehat{\mathbb{C}}$ is a one-to-one conformal map, then T is the restriction of a Möbius transformation.*

Proof. The region of discontinuity Ω of a Schottky group G is a domain of class O_{AD} (that is, it admits no holomorphic function with finite Dirichlet norm (see [3, p. 241]). It follows from this (see [3, p. 200]) that any one-to-one conformal map on Ω is necessarily the restriction of a Möbius transformation. □

3. Schottky uniformization from a Fuchsian one

3.1. Starting from a Fuchsian uniformization

Let Γ be a Fuchsian group, acting on the unit disc Δ, so that $\Delta/\Gamma = S$ is a closed Riemann surface of genus $g \geq 2$. The tuple $(\Delta, \Gamma, Q : \Delta \to S)$ is a Fuchsian uniformization of S. A presentation of Γ is given as follows

$$\Gamma = \left\langle A_1, \dots, A_g, B_1, \dots, B_g : \prod_{j=1}^{g} [A_j, B_j] = 1 \right\rangle$$

where $[A, B] = ABA^{-1}B^{-1}$.

3.2. Construction of a Schottky uniformization

Let $N = \langle\langle B_1, \dots, B_g \rangle\rangle$ be the normal envelope of $B_1, \dots, B_g$ inside Γ. Clearly, N is of infinite index in Γ (then a free group of infinite rank) and $G = \Gamma/N$ is a free group of rank g. It is not difficult to see that $\Omega = \Delta/N$ is a planar surface (we only need to note that there are not two elements in N whose axis intersect transversally). Also, G is a group of conformal automorphisms of Ω acting discontinuously. In this way, we have a regular planar covering $P : \Omega \to S$, with G as group of cover transformations. It follows from the results in [23] that we may assume Ω to be a region inside the Riemann sphere $\widehat{\mathbb{C}}$ and G being a geometrically finite Kleinian group containing an invariant component in its region of discontinuity and without parabolic transformations. It follows that G is a Schottky group of rank g and that $(\Omega, G, P : \Omega \to S)$ is a Schottky uniformization of S. As $N \lhd \Gamma$, there is a regular conformal covering map

$$F : \Delta \to \Omega \subset \widehat{\mathbb{C}}$$

whose group of cover transformations is N and so that $Q = PF$. Moreover, for each $A_k \in \Gamma$ there is a Möbius transformation $C_k \in G$ so that $F(A_k(z)) = C_k(F(z))$, for every $z \in \Delta$. The collection $C_1, \dots, C_g$ is a collection of free generators of the Schottky group G.

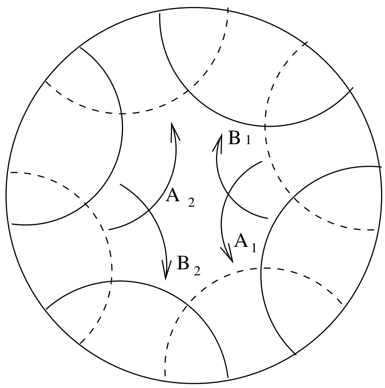


FIGURE 1

3.3. A normalization

Let $\mathcal{D} \subset \Delta$ be a fixed fundamental polygon for N and fix three different values $a, b, c \in \mathcal{D}$. We may assume these three points to be in the interior of a fundamental domain for Γ.

Lemma 3.1. *If we normalize F by requiring that $F(a) = 0$, $F(b) = 1$ and $F(c) = \infty$, then F is unique.*

Proof. This is a consequence of Lemma 2.1. □

In this case, a first step into the numerical uniformization problem is to find an algorithm which permits to generate conformal maps $F_j : \Delta \to F_j(\Delta) = \mathcal{W}_j \subset \widehat{\mathbb{C}}$ converging locally uniformly to F.

4. A convergence process

In this section we maintain all the definitions of the previous section.

4.1. Choosing a sequence of Schottky groups inside N

Let us consider a collection of groups

$$\langle B_1, \ldots, B_g \rangle = G_1 < G_2 < \cdots < N$$

so that

$$\bigcup_{j=1}^{\infty} G_j = N$$

and each G_j is finitely generated. The construction of the groups G_j is always possible (just add one extra element of $N - G_j$ to G_j in order to obtain G_{j+1}). As N is a free group (of infinite rank) and the subgroups of free groups are again free groups, each G_j is a free group of finite rank g_j. As the elements in N are hyperbolic (then loxodromic) and Δ is necessarily contained in the region of discontinuity of G_j, each G_j is in fact a Schottky group of rank g_j.

If we denote by Ω_j the region of discontinuity of G_j, then

$$\Omega_1 \supset \Omega_2 \supset \cdots \supset \Omega_\infty = \Delta \cup \Delta^*,$$

where $\Delta^* = \widehat{\mathbb{C}} - \overline{\Delta}$ and

$$\mathrm{Int}\left(\bigcap_{j\geq 1} \Omega_j\right) = \Omega_\infty.$$

Note that, necessarily, the fixed fundamental domain $\mathcal{D}$ is contained in some fundamental domain for G_j, in particular, $a, b, c \in \Omega_j$, for all j.

4.2. Some conformal maps

If we set $\widehat{G}_j = \langle G_j, \tau(z) = 1/\overline{z}\rangle$, then there are di-analytic covering maps (that is, in local coordinates are either conformal or anti-conformal maps)

$$F_j : \Omega_j \to \Omega_j/\widehat{G}_j = \overline{\mathcal{W}_j} \subset \widehat{\mathbb{C}},$$

where $\overline{\mathcal{W}_j}$ is a planar closed set bounded by $(g_j + 1)$ pairwise disjoint real-analytic simple loops and normalized by the conditions

$$F_j(a) = 0, \quad F_j(b) = 1 \quad \text{and} \quad F_j(c) = \infty.$$

We may assume that $F_j : \Delta \to \mathcal{W}_j$ is conformal, by composing F at the left by the reflection $J(z) = \overline{z}$, if necessary; so $F_j : \Delta^* \to \mathcal{W}_j$ is necessarily anti-conformal.

Unfortunately, our normalization does not makes F_j unique. This in particular seems to be a problem in the process: which choice we need to consider for F_j? Next result asserts that the choice we make for F_j does not matter at the limit process.

Theorem 4.1. *Let us consider a sequence of conformal covering maps*

$$F_j : \Delta \to \mathcal{W}_j \subset \widehat{\mathbb{C}}$$

for the corresponding Schottky groups G_j normalized by the rule that $F_j(a) = 0$, $F_j(b) = 1$ and $F_j(c) = \infty$. Then, such a sequence converges locally uniformly to the covering map F.

Proof. The proof is divided into three parts. In the first one we show that the family is normal. Secondly, we note that every convergent subsequence converges to F and finally we obtain that the complete family converges to F as required.

Normality of the family

Let $R \subset \Delta$ be the union of the orbits under N of the points a, b, c, and set $\Delta_R = \Delta - R$. Montel's theorem asserts that the family $F_j : \Delta_R \to \mathbb{C}$ is a normal family. Unfortunately, this is not enough to ensure this family to be normal on all Δ.

Let us consider any fundamental domain D for N (not necessarily the fixed one $\mathcal{D}$) so that in its interior D^0 are contained points $a_D \in D^0$ in the N-orbit of a, $b_D \in D^0$ in the N-orbit of b and $c_D \in D^0$ in the N-orbit of c. Consider the family of restrictions $F_j : D^0 \to \widehat{\mathbb{C}}$. Clearly, $F_j(a_D) = 0$, $F_j(b_D) = 1$ and $F_j(c_D) = \infty$. It follows from Theorem 2.1 in [20] that $F_j : D^0 \to \widehat{\mathbb{C}}$ is a normal family.

If $D_1, \ldots, D_n$ are fundamental domains for N, so that the interior D_k^0 of D_k always contains a point in the N-orbit of a, b and c, then the previous ensures that $F_j : D_k^0 \to \widehat{\mathbb{C}}$ is a normal family. It follows that $F_j : \cup_{k=1}^n D_k^0 \to \widehat{\mathbb{C}}$ is a normal family.

As Δ is a countable union of interior of fundamental domains as above, we may construct a family of open domains $R_1 \subset R_2 \subset \cdots \subset \Delta$ so that $\cup_{k=1}^{\infty} R_k = \Delta$ and $F_j : R_k \to \widehat{\mathbb{C}}$ is a normal family, for each k.

If we consider any subsequence of $F_j : \Delta \to \widehat{\mathbb{C}}$, there is a subsequence of $F_j : R_1 \to \widehat{\mathbb{C}}$ converging locally uniformly. Now, there is a subsequence of such one whose restriction to R_2 converges locally uniformly. We now consider such a new subsequence and restrict it to R_3 and continue inductively such a process. Now we use the diagonal method to to obtain a subsequence converging locally uniformly on all Δ.

Limit mappings of subsequences

Let us choose any subsequence $F_{j_k} : \Delta \to \widehat{\mathbb{C}}$ that converges locally uniformly, say to $F_\infty : \Delta \to \widehat{\mathbb{C}}$. As $F_{j_k}(a) = 0$, $F_{j_k}(b) = 1$ and $F_{j_k}(c) = \infty$, it follows that $F_\infty(a) = 0$, $F_\infty(b) = 1$ and $F_\infty(c) = \infty$, in particular, F_∞ is non-constant conformal mapping.

Choose any fundamental domain D for N (not necessarily the fixed one $\mathcal{D}$) and let D^0 its interior. As each $G_j < N$, D is contained in some fundamental domain D_j for G_j. Since F_j restricted to D_j is injective, each F_j restricted to D^0 is also injective. We consider the subsequence $F_{j_k} : D^0 \to \widehat{\mathbb{C}}$, which we know that converges locally uniformly to the non-constant conformal mapping $F_\infty : D^0 \to \widehat{\mathbb{C}}$. As the uniform limit of injective conformal mappings is either injective or constant, $F_\infty : D^0 \to \widehat{\mathbb{C}}$ is injective. This proves that F_∞ restricted to D^0 is an homeomorphism onto its image.

As the above holds for any fundamental domain for N and Δ is countable union of such sets, F_∞ is a local homeomorphism on Δ.

Let $\gamma \in N$. Then, there is some j_0 so that $\gamma \in G_j$ for $j \geq j_0$. It follows that for $j \geq j_0$ the equality $F_j(\gamma(z)) = F_j(z)$ holds for each $z \in \Delta$. In particular, $F_\infty(\gamma(z)) = F_\infty(z)$ for every $z \in \Delta$. The above invariance and the fact that the restriction of F_∞ to any of its fundamental domains is an homeomorphism asserts that F_∞ is a covering map for which N is group of cover transformations. Since F_∞ and F acts in the same way at the points a, b and c, then $F = F_\infty$ by Lemma 3.1.

Convergence of the family

Since any subsequence of $F_j : \Delta \to \widehat{\mathbb{C}}$ has a locally uniform convergent subsequence to $F_\infty = F$, the above asserts that the complete sequence converges locally uniformly to F. □

Remark 4.2. As already noted, the normalization $F_j(a) = 0, F_j(b) = 1, F_j(c) = \infty$ does not makes F_j unique. If we consider any two choices for F_j, say $F_{j,k} : \Omega_j \to \overline{\mathcal{W}_{j,k}}$, $k = 1, 2$, then there is a conformal homeomorphism $H_j : \mathcal{W}_{j,1} \to \mathcal{W}_{j,2}$ so that $F_2 = H_j F_1$. We may make a unique choice for each F_j as follows. Let us start with any choice for F_j and let us consider the restriction $F_j : \Delta \to \mathcal{W}_j$ which is a conformal covering map. A classical result [2, 7, 13, 27] asserts that there is a circular domain $\Sigma_j \subset \widehat{\mathbb{C}}$ and a conformal homeomorphism $T_j : \mathcal{W}_j \to \Sigma_j$. As the boundary of $\mathcal{W}_j$ consists of real-analytic simple loops, by Carathéodory's extension theorem, T_j extends as an homeomorphism from $\overline{\mathcal{W}_j}$ onto $\overline{\Sigma_j}$. By composing T_j at the left by a suitable Möbius transformation, we may also assume $T_j(0) = 0$, $T_j(1) = 1$ and $T_j(\infty) = \infty$. Now, we replace $F_j : \Omega_j \to \overline{\mathcal{W}_j}$ by $F_j^c = T_j F_j$, that is, we may now assume $\mathcal{W}_j$ to be a circular domain. We claim that the map F_j^c is unique under the normalization $F_j^c(a) = 0$, $F_j^c(b) = 1$, $F_j^c(c) = \infty$ and the fact that $F_j^c(\Delta)$ is a circular domain. In fact, as previously noted, any other possible map with the same conditions as for F_j^c will be of the form $H_j F_j^c$, where H_j is a conformal homeomorphism between two circular domains and fixing the three points $0, 1, \infty$. By reflection principle, we may extend H to a conformal homeomorphism between two regions U_1 and U_2, where U_1 is the region of discontinuity of the extended Kleinian group K_j generated by the reflections on the boundary circles of $\mathcal{W}_j$ (similarly for U_2). As U_1 is the region of discontinuity of a Schottky group (the orientation-preserving half of K_j) it follows from Lemma 2.1 that H_j is the restriction of a Möbius transformation. As it fixes three different points, $H_j = I$.

4.3. Approximating G

Now, in order to get the generators $C_1, \ldots, C_g$ of G, we proceed as follows. For each $j = 1, \ldots$ and each $k = 1, \ldots, g$, we consider the unique Möbius transformation $C_{j,k}$ so that $C_{j,k}(F_j(a)) = F_j(A_k(a))$, $C_{j,k}(F_j(b)) = F_j(A_k(b))$ and $C_{j,k}(F_j(c)) = F_j(A_k(c))$. At this point we need to recall that a, b, c have been chosen to be in the interior of a fundamental domain for Γ. This asserts that the collection of points

$$F_j(a), F_j(A_k(a)), F_j(b), F_j(A_k(b)), F_j(c), F_j(A_k(c))$$

are pairwise different.

Theorem 4.3. *The Möbius transformations $C_{j,k}$ converge to C_k.*

Proof. This is consequence of the fact that each Möbius transformation is uniquely determined by its action at three different points, that the sequence F_j converges locally uniformly on Δ and that $F(A_k(z)) = C_k(F(z))$ for $z \in \Delta$ and $k = 1, \dots, g$. □

Let $\mathcal{F}_\Gamma$ any fundamental domain for Γ contained in $\mathcal{D}$. As $\mathcal{F}_\Gamma$ is a compact subset of the unit disc Δ and F_j converges locally uniformly to F, we obtain that the compact sets $\mathcal{F}_j = F_j(\mathcal{F}_\Gamma) \subset \widehat{\mathbb{C}}$ converges in Hausdorff topology to a fundamental domain $\mathcal{F}_\infty$ for the Schottky group G.

5. Numerical implementations

5.1. First approach

Let us choose a point $d \in \partial\Delta$ in the boundary of the fixed fundamental domain $\mathcal{D}$ for N, where $\partial\Delta$ denotes the boundary unit circle of Δ. This choice permits to extend each F_j and F continuously to d from inside $\mathcal{D}$. Let T_j be the unique Möbius transformation so that $T_j(0) = 0$, $T_j(1) = 1$ and $T_j(F_j(d)) = \infty$. The map $H_j = T_jF_j : \Delta \to T_j(\mathcal{R}_j)$ is a covering map associated to the group G_j (whose image still a circular domain, one of the circles being a line). In particular, each H_j is analytic inside Δ with a simple pole at d. Let us consider a subsequence F_{j_k} so that $F_{j_k}(d)$ converges, say to d_∞. The Möbius transformations T_{j_k} now converges to the unique Möbius transformation T_∞ so that $T_\infty(0) = 0$, $T_\infty(1) = 1$ and $T_\infty(F_j(d)) = \infty$. The map $H_\infty = T_\infty F : \Delta \to \Omega$ is a covering map associated to the group N. As F_{j_k} converges to F, it follows from the above that H_{j_k} converges locally uniformly to $H_\infty F$. In this case, the Schottky group we obtain is $T_\infty G T_\infty^{-1}$. As H_j is analytic on all Δ, if we take $a = 0$, then this ensures that $H_j(z) = \sum_{k=1}^{\infty} \mu_{jk} z^k$, with radius of convergence 1. If we find numerical approximations of H_j, then we will obtain numerical approximations of both H_∞ and $T_\infty G T_\infty^{-1}$. The idea is to consider a truncation $H_{j,N}(z) = \sum_{k=1}^{N} \mu_{jk,N} z^k$ and some points w_r (well-distributed) on $\mathcal{D}$. The coefficients $\mu_{jk,N}$ can be (over)determined by considering a set of generators of G_j, say $A_1, \dots, A_{g_j} \in G_j$, and asking $H_{j,N}(A_s(w_r)) = H_{j,N}(w_r)$.

5.2. Second approach

The second approach is related to quadrature domains [14] and work of D. Crowdy and J.S. Marshal [11] using the Schottky-Klein prime function.

A bounded domain $D \subset \mathbb{C}$ (we assume its boundary consists of a finite collection of pairwise disjoint simple loops) is said to be a (*classical*) *quadrature domain* [14] if there exists finitely many points $a_1, \dots, a_N \in D$, positive integers $n_1, \dots, n_N$ and complex numbers $c_{kj} \in \mathbb{C}$ ($k = 1, \dots, N$, $j = 0, 1, \dots, n_k - 1$) so that, for every integrable analytic function $h(z)$ in D (extending continuously to the boundary of D) it holds that (the first equality is just Green's formula)

$$\iint_D h(z)dxdy = \frac{1}{2i} \int_{\partial D} h(z)\overline{z}dz = \sum_{k=1}^{N} \sum_{j=0}^{n_k-1} c_{kj} h^{(j)}(a_k).$$

For instance, the unit disc $D=\{z\in\mathbb{C}:|z|<1\}$ is a quadrature domain with $N=1=n_1$, $a_1=0$ and $c_{10}=\pi$. A survey about quadrature domains is, for instance, [15].

A domain D is a quadrature domain if and only if there exists a meromorphic function $h:D\to\widehat{\mathbb{C}}$ so that $h(z)=\overline{z}$ for every $z\in\partial D$ [1]. This last property asserts that the domain D is a quadrature domain if and only if there is a closed Riemann surface S admitting an anticonformal involution $\tau:S\to S$ with maximal number of ovals (components of fixed points) and a meromorphic map $\Psi:S\to\widehat{\mathbb{C}}$ so that, if S_1 and S_2 are the two components of $S-Fix(\tau)$, then $F:S_1\to D$ is a conformal homeomorphism [14].

In our case, for each fixed j, we need to find a regular covering $F_j:\Delta\to F_j(\Delta)=\mathcal{W}_j$, where $\mathcal{W}_j$ is a domain in the Riemann sphere bounded by some finite number of simple loops. It may happens that $\mathcal{W}_j$ is not a quadrature domain. But, as consequence of Theorem 4 in [14], there is a conformal homeomorphism $Q_j:\mathcal{W}_j\to D_j$, where D_j is a quadrature domain.

Now, in [11] it is provided a numerical algorithm to obtain a regular conformal covering $T_j:\Delta\to D_j$ with G_j as group of covering transformations. This numerical implementation is based on the Schottky-Klein prime function of the group G_j [4], in this case, given by

$$\omega(z,\mu)=(z-\mu)\prod_{\gamma\in G_j^*}\frac{(\gamma(z)-\mu)(\gamma(\mu)-z)}{(\gamma(z)-z)(\gamma(\mu)-\mu)}$$

$$z,\mu\in\Omega_j(\text{the region of discontinuity of } G_j)$$

$G_j^*\subset G_j-\{I\}$ and for every $\gamma\in G-\{I\}$ either γ or γ^{-1} belongs to G_j^* (but not both). Let L_j be the Möbius transformation determined by the fact that $L_j(T_j(a))=0$, $L_j(T_j(b))=1$ and $L_j(T_j(c))=\infty$. Then we may consider $F_j=L_j\circ T_j$.

Acknowledgments

The author wishes to express his special thanks to Professors Alexander Vassiliev and Irina Markina for their support to attend the Congress. He would also like to thanks Professor Christian Pommerenke for the discussions relating part of the arguments in this paper during the Congress and to the referee for his/her careful reading and many useful comments and corrections.

References

[1] D. Aharonov and H.S. Shapiro. Domains in which analytic functions satisfy quadrature identities. *J. Analyse Math.* **30** (1976), 39–73.

[2] L. Ahlfors. *Bounded analytic functions.* Duke Math. Journ. **14** (1947), 1–11.

[3] L. Ahlfors and L. Sario. *Riemann Surfaces.* Princeton University Press, Princeton NJ, 1960.

[4] H.F. Baker. *Abel's Theorem and the Allied Theory including the Theory of Theta Functions*. Cambridge University Press, Cambridge, 1897.

[5] H.F. Baker. *Multiply Perodic Functions*. Cambridge University Press, Cambridge, 1907.

[6] L. Bers. Automorphic forms for Schottky groups. Adv. in Math. **16**, 332–361 (1975)

[7] L. Bieberbach. *Über einen Riemannschen Satz aus der Lehre von der konformen Abbildung*. Sitzungsber. Ber. Math. Gesell. **24** (1925), 6–9.

[8] W. Burnside. *On a class of Automorphic Functions*. Proc. London Math. Soc. **23** (1892), 49–88.

[9] P. Buser and R. Silho. *Geodesics, periods and equations of real hyperelliptic curves*. Duke Math. J. **108** (2) (2001), 211–250.

[10] V. Chuckrow. On Schottky groups with applications to Kleinian groups. Annals of Math. **88**, 47–61 (1968)

[11] D.G. Crowdy and J.S. Marshall. Uniformizing the boundaries of multiply connected quadrature domains using Fuchsian groups. *Physica D* **235** (2007), 82–89.

[12] P. Gianni, M. Seppälä et al. *Riemann surfaces, plane algebraic curves and their period matrices*. J. Symbolic Comput. **26** (6) (1998), 789–803.

[13] H. Grunsky. *Über die konforme Abbildung mehrfach zusammenhängender Bereiche auf mehrblättrige Kreise*. Sitzungsber. Preuß. Akad. Wiss. (1937), 1–9.

[14] B. Gustafsson. Quadrature Identities and the Schottky Double. *Acta Applicandae Mathematicae* **1** (1983), 209–240.

[15] B. Gustafsson and H.S. Shapiro. What is a quadrature domain. Ebenfelt, Peter (ed.) et al., Quadrature domains and their applications. The Harold S. Shapiro anniversary volume. Expanded version of talks and papers presented at a conference on the occasion of the 75th birthday of Harold S. Shapiro, Santa Barbara, CA, USA, March 2003. Basel: Birkhäuser. Operator Theory: Advances and Applications 156, 1–25 (2005). ISBN 3-7643-7145-5/hbk

[16] R.A. Hidalgo. *Real surfaces, Riemann matrices and algebraic curves*. Contemp. Math., **311** (2002), 277–299.

[17] R.A. Hidalgo and M. Seppälä. *Numerical Schottky uniformizations: Myrberg's opening process*. Preprint.

[18] A. Hoare, A. Karrass and D. Solitar. *Subgroups of finite index of Fuchsian groups*. Math. Z. **120** (1971), 289–298.

[19] A. Hoare, A. Karrass and D. Solitar. *Subgroups of infinite index in Fuchsian groups*. Math. Z. **125** (1972), 59–69.

[20] O. Lehto. *Univalent Functions and Teichmüller Spaces*. GTM **109**, Springer-Verlag, 1986.

[21] B. Maskit. A characterization of Schottky groups. *J. Analyse Math.* **19** (1967), 227–230.

[22] B. Maskit. A theorem on planar covering surfaces with applications to 3-manifolds. *Ann. of Math.* (2) **81** (1965), 341–355.

[23] B. Maskit. On the classification of Kleinian groups: I. Koebe groups. *Acta Math.* **135** (1975), 249–270.

[24] A. Marden. The geometry of finitely generated Kleinian groups. *Ann. of Math.* **99** (1974), 383462.

[25] P.J. Myrberg. *Über die Numerische Ausführung der Uniformisierung.* Acta Soc. Sci. Fenn. **XLVIII** (7) (1920), 1–53.

[26] M. Seppälä. *Myrberg's numerical uniformization of hyperelliptic curves.* Ann. Acad. Sci. Fenn. Math. **29** (1) (2004), 3–20.

[27] M. Jeong and M. Taniguchi. *Algebraic kernel functions and representation of planar domains.* J. Korean Math. Soc. **40** (2003), 447–460.

Rubén A. Hidalgo
Departamento de Matemática
Universidad Técnica Federico Santa María
Casilla 110-V
Valparaiso, Chile
e-mail: ruben.hidalgo@usm.cl

Analysis and Mathematical Physics
Trends in Mathematics, 205–210
© 2009 Birkhäuser Verlag Basel/Switzerland

Scattering from Sparse Potentials: a Deterministic Approach

Vojkan Jakšić and Philippe Poulin

Abstract. Completeness of the wave operators has been proven for a family of random Schrödinger operators with sparse potentials in the recent paper [17], using a probabilistic approach. As mentioned at Voss, a deterministic result in this direction can also be derived from a Jakšić–Last criterion of completeness [7] and Fredholm's theorem. We present this approach.

Mathematics Subject Classification (2000). Primary 81Q10; Secondary 47A40.

Keywords. Schrödinger operators, sparse potentials, wave operators.

1. Introduction

Since their introduction by Anderson [1], there has been considerable interest in Schrödinger operators with random potentials. These operators represent the energy of a particle affected by a random potential on a lattice. They are of the form $H = \Delta + \lambda V$, where Δ is the centered, discrete Laplacian on $\mathbb{Z}^d$, λ is a real parameter (the so-called *disorder*) and V is a random potential supported on $\mathbb{Z}^d$. In [1], Anderson anticipated the spectral structure of H (i.e., the intervals of localization/delocalization) with respect to the disorder. While the localization aspect of the Anderson conjecture has been mathematically settled in the seminal papers [5, 4, 2, 3], practically nothing is known about the delocalization aspect.

Several research teams have also studied various *sparse models* [6, 9, 10, 11, 12, 13, 14, 15, 17]. In these nonergodic models, spectral properties of H are expected to follow from various geometric constraints on the sites of the potential. These constraints have in common that the minimal distance between two sites becomes arbitrarily large when removing a finite number of them. Examples have been exhibited where all the expected spectral properties are satisfied (almost surely), including completeness [17] of the wave operators on the spectrum of Δ. We present a deterministic extension of this last result.

We consider a discrete Schrödinger operator $H = \Delta + V$ in dimension $d \geqslant 2$, where Δ is the centered Laplacian and V is a bounded potential: for $\varphi \in l^2(\mathbb{Z}^d)$ and $n \in \mathbb{Z}^d$

$$(H\varphi)(n) = \sum_{|m-n|_1=1} \varphi(m) + V(n),$$

where $|n|_1 = \sum_{j=1}^{d} |n^{(j)}|$. We assume that the support of V, which we denote by $\Gamma \subset \mathbb{Z}^d$, satisfies the following sparseness assumption:

(A) *There exists an* $\epsilon > 0$ *such that* $\sum_{m\in\Gamma\setminus\{n\}} |n-m|^{-\frac{1}{2}+\epsilon}$ *is finite for all* $n \in \Gamma$ *and tends to* 0 *when* $|n| \to \infty$ *in* Γ.

This is the case, for instance, if $\Gamma = \{(j^4, 0, \dots, 0) \in \mathbb{Z}^d \ ; \ j \in \mathbb{Z}\}$.

Recall [18] that the *wave operators* on a Borel set $\Theta \subset \mathbb{R}$ are the strong limits $\Omega^{\pm}(H, \Delta) = \lim_{t\to\pm\infty} \mathrm{e}^{\mathrm{i}tH}\mathrm{e}^{-\mathrm{i}t\Delta}\mathbf{1}_{\Theta}(\Delta)$ (if they exist); they are *complete* if their range is $\mathrm{Ran}\mathbf{1}_{\Theta}(H)$ (so Δ and H are unitarily equivalent on Θ). We prove:

Theorem. *Assume* (A). *Then, the wave operators* $\Omega^{\pm}(H, \Delta)$ *exist on* $[-2d, 2d]$. *Moreover, they are complete on* $[-2d, 2d]$ *minus a set of Lebesgue measure zero.*

It is possible to remove the exceptional set in the above by working in the random frame. Then $\{V(n)\}_{n\in\Gamma}$ is a family of independent, identically distributed, absolutely continuous random variables whose common density is compactly supported.[1] Since the essential support of the absolutely continuous spectrum of Δ is $[-2d, 2d]$, and since under Assumption (A) the wave operators exist on this last interval, the Jakšić–Last theorem [8] and the above immediately yield:

Corollary. *In the random frame, Assumption* (A) *implies that the wave operators exist and are complete on* $[-2d, 2d]$, *almost surely.*

This conclusion is stronger than the one we obtained in [17], where only completeness of the wave operators is derived. However, our present assumption is also stronger, since unbounded potentials are discarded.

Here is the outline of the paper. In the sequel $\{n \in \mathbb{Z}^d \ ; \ \inf_{m\in\Gamma} |n-m|_1 \leqslant 1\}$ is denoted by Γ_1, while $\mathbf{1}_0$ and $\mathbf{1}_1$ denote the projections onto $l^2(\Gamma)$ and $l^2(\Gamma_1)$ respectively. Moreover $\delta_m(n)$ denotes the Kronecker delta, where $m, n \in \mathbb{Z}^d$. For $z \in \mathbb{C}_+$ we consider the following restrictions of the free and perturbed resolvents,

$$F_1(z) = \mathbf{1}_1(\Delta - z)^{-1}\mathbf{1}_1, \quad F_0(z) = \mathbf{1}_0(\Delta - z)^{-1}\mathbf{1}_0,$$
$$P_1(z) = \mathbf{1}_1(H - z)^{-1}\mathbf{1}_1, \quad P_0(z) = \mathbf{1}_0(H - z)^{-1}\mathbf{1}_0.$$

Our study is based the following theorem of Jakšić and Last [7]:

Proposition 1. *Let* $U \subset \mathbb{R}$ *be open. Suppose* $\|F_1(e + \mathrm{i}0)\| < \infty$ *and* $\|P_1(e + \mathrm{i}0)\| < \infty$ *for all* $e \in U$. *Then, the wave operators exist and are complete on* U.

[1]Explicitly, the probability space is given by $\mathbb{R}^{\Gamma}$ equipped with its Borel σ-algebra and a probability measure $\prod_{\Gamma} \mu$, where μ is an absolutely continuous, compactly supported measure on $\mathbb{R}$. The variable $V(n)$ is then the projection on the n^{th} coordinate, for $n \in \Gamma$.

For $[a,b] \subset [-2d,2d] \setminus (\{2d, 2d-4, \ldots, -2d+4, -2d\} \cup \{0\})$ and z in the strip $\mathcal{S} := \{e + \mathrm{i}y \ ; \ a < e < b,\ 0 < y < 1\}$, the following *a priori* estimate [16] is available:

Proposition 2. *Let $n = |n|\omega \in \mathbb{Z}^d$. Then, $\lim_{z\to e, z\in\mathbb{C}_+} \langle \delta_0 \,|\, (\Delta - z)^{-1}\delta_n\rangle$ exists and is $O(|n|^{-\frac{1}{2}})$ uniformly in $(e,\omega) \in [a,b]\times \mathrm{S}^{d-1}$. More generally, $\langle \delta_0 \,|\, (\Delta - z)^{-1}\delta_n\rangle = O(|n|^{-\frac{1}{2}} \log |n|)$ uniformly in $(z,\omega) \in \overline{\mathcal{S}} \times \mathrm{S}^{d-1}$.*

Therefore, sparseness will ensure that the matrix elements of $F_1(z)$ are small except about its diagonal. By subtracting a block-diagonal to $F_1(z)$ the remaining part will be compact. We will derive the finiteness of $\|F_1(e + \mathrm{i}0)\|$ first, and then deduce the same for $\|P_1(e + \mathrm{i}0)\|$ by means of Fredholm's theorem:

Proposition 3. *Let $K(z)$ be a function with values in the space of compact operators (endowed with the uniform topology). Suppose $K(z)$ is continuous on $\overline{\mathcal{S}}$ and analytic in $\mathcal{S}$. Then, either $1 - K(z)$ is never invertible on $\overline{\mathcal{S}}$, or it is invertible except on a closed set of Lebesgue measure zero whose intersection with $\mathbb{C}_+$ consists of isolated points.*

2. Proof of the theorem

Let us partition Γ_1 as follows: for all $n \in \Gamma$, we select a neighborhood $\mathcal{B}(n) \subseteq \{m \in \Gamma_1 \ ; \ |m - n|_1 \leqslant 1\}$ containing n in such a way that $\bigcup_{n\in\Gamma} \mathcal{B}(n) = \Gamma_1$ and $\mathcal{B}(n) \cap \mathcal{B}(n') = \emptyset$ if $n \neq n'$. For all $m \in \Gamma_1$ there exists exactly one $n \in \Gamma$ such that $m \in \mathcal{B}(n)$; we then set $\mathcal{B}(m) = \mathcal{B}(n)$.

For $n \in \Gamma_1$, let $S(n) = \sum_{m\in\Gamma_1\setminus\mathcal{B}(n)} \sup_{z\in\mathcal{S}} |\langle \delta_m \,|\, F_1(z)\delta_n\rangle|$. Then,

Lemma 1. *$S(n)$ is finite for all $n \in \Gamma_1$ and tends to 0 when $|n| \to \infty$ in Γ_1.*

Proof. By Proposition 2, $S(n) \leqslant Const \sum_{m\in\Gamma_1\setminus\mathcal{B}(n)} |n - m|^{-\frac{1}{2}+\epsilon}$. Moreover, there exists a $C \geqslant 1$ such that for $\mathcal{B}(m) \neq \mathcal{B}(n)$, $C^{-1}|n - m| \leqslant |n - m'| \leqslant C|n - m|$ for all $m' \in \mathcal{B}(m)$. Since the cardinalities of the $\mathcal{B}(m)$ are bounded, Assumption (A) yields the result. □

For $z \in \mathcal{S}$ let us decompose $F_1(z)$ into two summands: a block-diagonal, $D_1(z) = \sum_{n\in\Gamma_1} \sum_{m\in\mathcal{B}(n)} \langle \delta_n \,|\, F_1(z)\delta_m\rangle \langle \delta_m \,|\, \cdot\rangle \delta_n$, and the other part, $K_1(z) = F_1(z) - D_1(z)$. By Proposition 2, $\langle \delta_n \,|\, F_1(e + \mathrm{i}0)\delta_m\rangle$ exists for $e \in [a,b]$. In particular, letting $F_1(e) := \sum_{m,n\in\Gamma_1} \langle \delta_n \,|\, F_1(e + \mathrm{i}0)\delta_m\rangle \langle \delta_m \,|\, \cdot\rangle \delta_n$, $\lim_{\substack{z\to e\\ z\in\mathbb{C}_+}} F_1(z) = F_1(e)$ weakly. Let us define $D_1(e)$ and $K_1(e)$ in a similar way, so they are weak limits of $D_1(z)$ and $K_1(z)$ respectively.

Lemma 2. *For any $e \in [a,b]$, $\lim_{z\to e, z\in\mathbb{C}_+} D_1(z) = D_1(e)$ uniformly.*

Proof. Let $\{\mathcal{A}_j\}_{j=1}^{L}$ be the list of all subsets of $\{m \in \mathbb{Z}^d \ ; \ |m|_1 \leqslant 1\}$ containing 0. For all $n \in \Gamma$ there exists exactly one j, which we denote by $\mathrm{j}(n)$, such that

$\mathcal{B}(n) - n = \mathcal{A}_j$. Thus, by translational invariance

$$D_1(z) = \sum_{j=1}^{L} \sum_{p,q\in\mathcal{A}_j} \langle \delta_q \,|\, F_1(z)\delta_p\rangle \sum_{\mathrm{j}(n)=j} \langle \delta_{n+p} \,|\, \cdot\rangle \delta_{n+q}.$$

The result follows. □

Lemma 3. *Let $\varepsilon > 0$. There exists a finite-dimensional projection P_ε such that, letting $M_\varepsilon(z){:}{=}P_\varepsilon K_1(z) P_\varepsilon$, $\|K_1(z) - M_\varepsilon(z)\| \leqslant \varepsilon$ for all $z \in \overline{\mathcal{S}}$. Moreover, $M_\varepsilon(z)$ is continuous on $\overline{\mathcal{S}}$ (with respect to the uniform operator topology).*

Proof. By Lemma 1, there exists a finite set $\mathcal{F} \subset \Gamma_1$ such that for all $z \in \overline{\mathcal{S}}$

$$\sup_{n\in\Gamma_1\setminus\mathcal{F}} \sum_{m\in\Gamma_1} |\langle \delta_n \,|\, K_1(z)\delta_m\rangle| + \sup_{n\in\mathcal{F}} \sum_{m\in\Gamma_1\setminus\mathcal{F}} |\langle \delta_n \,|\, K_1(z)\delta_m\rangle| \leqslant \varepsilon. \tag{2.1}$$

Let P_ε be the projection onto the vector space generated by $\{\delta_n\}_{n\in\mathcal{F}}$. Notice that $M_\varepsilon(z)$ is weakly continuous and hence uniformly continuous on $\overline{\mathcal{S}}$. Moreover, $\langle \delta_n \,|\, (K_1(z) - M_\varepsilon(z))\delta_m\rangle = \langle \delta_m \,|\, (K_1(z) - M_\varepsilon(z))\delta_n\rangle$ for all $m, n \in \Gamma_1$, so the equation (2.1) is equivalent to $\|K_1(z) - M_\varepsilon(z)\|_1 = \|K_1(z) - M_\varepsilon(z)\|_\infty \leqslant \varepsilon$. Schur's interpolation theorem then completes the proof. □

Lemma 4. *For any $e \in [a, b]$, $\lim_{z\to e, z\in\mathbb{C}_+} K_1(z) = K_1(e)$ uniformly.*

Proof. Let $\varepsilon > 0$. For $e \in [a, b]$ and $z \in \mathcal{S}$

$$\begin{aligned}\|K_1(z) - K_1(e)\| &\leqslant \|K_1(z) - M_\varepsilon(z)\| + \|K_1(e) - M_\varepsilon(e)\| + \|M_\varepsilon(z) - M_\varepsilon(e)\| \\ &\leqslant \|M_\varepsilon(z) - M_\varepsilon(e)\| + 2\varepsilon.\end{aligned}$$

Since $\lim_{z\to e, z\in\mathbb{C}_+} M_\varepsilon(z) = M_\varepsilon(e)$ uniformly, the proof is complete. □

By Lemmas 2 and 4, $F_1(z)$ has a continuous extension on $\mathbb{C}_+ \cup [a, b]$, so we have reached that $\|F_1(e + \mathrm{i}0)\| < \infty$ for all $e \in [a, b]$. Let us focus on $l^2(\Gamma)$. By the previous work, $F_0(z)$ is continuous on $\overline{\mathcal{S}}$ and analytic in $\mathcal{S}$. Moreover,

Lemma 5. *$F_0(z)$ is invertible in $\mathcal{B}(l^2(\Gamma))$ for all $z \in \mathcal{S}$.*

Proof. Let μ_φ be the spectral measure of a unit vector $\varphi \in l^2(\Gamma)$ with respect to Δ. For a fixed $z = e + \mathrm{i}y \in \mathcal{S}$, $\Im\langle \varphi \,|\, F_0(z)\varphi\rangle = y \int_{-2d}^{2d} ((t - e)^2 + y^2)^{-1}\, \mathrm{d}\mu_\varphi(t)$. This expression is bounded away from zero when φ varies in the unit vectors. Thus, the closure of the numerical range of $F_0(z)$ is included in $\mathbb{C}_+$. The result follows. □

Let $D_0(z) = \mathbf{1}_0 D_1(z)\mathbf{1}_0$ and $K_0(z) = \mathbf{1}_0 K_1(z)\mathbf{1}_0$. By Lemma 3, $K_0(z)$ is a compact operator for any $z \in \overline{\mathcal{S}}$. Moreover, $D_0(z)$ is diagonal; it is indeed a constant (times the identity on $l^2(\Gamma)$) by translational invariance. By Theorem 6.1 in [16], the number $\inf_{z\in\overline{\mathcal{S}}} \Im D_0(z)$, which we denote by I, is positive.

Lemma 6. *$1 + D_0(z)V$ is invertible in $\mathcal{B}(l^2(\Gamma))$ for any $z \in \overline{\mathcal{S}}$.*

Proof. First, $(1+D_0(z)V)^{-1}$ exists, since $I > 0$. If[2] $|V(n)D_0(z)| \leqslant 1/2$, then $|(1+D_0(z)V)^{-1}(n)| \leqslant 2$. Otherwise, $|(1+D_0(z)V)^{-1}(n)| \leqslant 2|D_0(z)|/I$. Hence, $(1+D_0(z)V)^{-1}$ is bounded, as claimed. □

We now transfer our result from the free resolvent to P_0. Since V is bounded, by the argument in Lemma 5, $P_0(z)$ is invertible for each $z \in \mathcal{S}$. Moreover,

Lemma 7. *There exists a closed set of Lebesgue measure zero,* $\mathcal{R} \subset [a,b]$*, such that* P_0 *has a continuous extension* $\overline{\mathcal{S}} \setminus \mathcal{R} \longrightarrow \mathcal{B}(l^2(\Gamma))$.

Proof. Let $z \in \mathcal{S}$. By the resolvent identity, $(1+F_0(z)V)P_0(z) = F_0(z)$. Notice that $1+F_0(z)V$ is invertible, since $P_0(z)$ and $F_0(z)$ are. Thus,

$$P_0(z) = (1+F_0(z)V)^{-1}F_0(z), \tag{2.2}$$

where $z \in \mathcal{S}$. One wonders to which extent $(1+F_0(z)V)^{-1}$ is still invertible when $z \in \partial\mathcal{S}$. Indeed, for any $z \in \overline{\mathcal{S}}$, $1+F_0(z)V = (1-K(z))(1+D_0(z)V)$, where $K(z) := -K_0(z)V(1+D_0(z)V)^{-1}$ is compact. Since for $z \in \mathcal{S}$ both $1+D_0(z)V$ and $1+F_0(z)V$ are invertible, $1-K(z)$ is. By Proposition 3, $1-K(z)$ is thus invertible in $\mathcal{B}(l^2(\Gamma))$ for all $z \in [a,b] \setminus \mathcal{R}$, where $\mathcal{R} \subset [a,b]$ is a closed set of Lebesgue measure zero. Hence, the right side in (2.2) extends continuously up to $\overline{\mathcal{S}} \setminus \mathcal{R}$, as desired. □

Lemma 8. *There exists a closed set of Lebesgue measure zero,* $\mathcal{R} \subset [a,b]$*, such that* P_1 *has a continuous extension* $\overline{\mathcal{S}} \setminus \mathcal{R} \longrightarrow \mathcal{B}(l^2(\Gamma_1))$.

Proof. By the resolvent identity, $F_1(z)\mathbf{1}_0 - P_1(z)\mathbf{1}_0 = F_1(z)VP_0(z)$. Since $F_1(z)$ and $P_0(z)$ extend continuously up to $\overline{\mathcal{S}} \setminus \mathcal{R}$, $P_1(z)\mathbf{1}_0$ also does. By the resolvent identity again, $F_1(z) - P_1(z) = P_1(z)\mathbf{1}_0VF_1(z)$. The result follows. □

In particular, $\|P_1(e+\mathrm{i}0)\| < \infty$ on $[a,b]\setminus\mathcal{R}$. Since the analogous relation has been established for F_1, Proposition 1, the arbitrariness of $[a,b]$, and the absolute continuity of the spectrum of Δ yield the theorem.

Acknowledgements

P. Poulin is grateful to his present supervisors, Kristian Seip and Yurii Lyubarskii, for their kind invitation to the colloquium at Voss. The research of V. Jakšić was partially supported by NSERC.

References

[1] P.W. Anderson, *Absence of diffusion in certain random lattices.* Phys. Rev. **109** (1958), 1492–1505.

[2] M. Aizenman and S. Molchanov, *Localization at large disorder and at extreme energies: an elementary derivation.* Comm. Math. Phys. **157** (1993), 245–278.

[3] M. Aizenman, *Localization at weak disorder: some elementary bounds.* Rev. Math. Phys. **6** (1994), 1163–1182.

[2] $A(n)$ denotes the n^{th} diagonal element of a diagonal operator A.

[4] J. Fröhlich and T. Spencer, *Absence of diffusion in the Anderson tight binding model for large disorder or low energy.* Commun. Math. Phys. **88** (1983), 151–189.

[5] I. Goldsheid, S. Molchanov and L. Pastur, *A pure point spectrum of the stochastic one-dimensional Schrödinger equation.* Funct. Anal. Appl. **11** (1977), 1–10.

[6] D. Hundertmark and W. Kirsch, *Spectral theory of sparse potentials.* In: Stoch. Proc., Phys. and Geom.: new interplays, I. CMS Conf. Proc. **28** (2000), 213–238.

[7] V. Jakšić and Y. Last, *Scattering from subspace potentials for Schrödinger operators on graphs.* Markov Processes and Related Fields **9** (2003), 661–674.

[8] V. Jakšić and Y. Last, *Spectral structure of Anderson type Hamiltonians.* Invent. Math. **141** (2000), 561–577.

[9] W. Kirsch, *Scattering theory for sparse random potentials.* Random Oper. and Stoch. Eqs. **10** (2002), 329–334.

[10] M. Krishna, *Absolutely continuous spectrum for sparse potential.* Proc. Indian Acad. Sci. (math. sci.) **103** (1993), 333–339.

[11] M. Krishna and J. Obermeit, *Localization and mobility edge for sparsely random potentials.* IMSc Preprint (1998), arXiv:math-ph/9805015v2.

[12] S. Molchanov, *Multiscattering on sparse bumps.* In: Advances in Differential Equations and Mathematical Physics. Contemp. Math. **217** (1998), 157–182.

[13] S. Molchanov and B. Vainberg, *Scattering on the system of the sparse bumps: Multidimensional case.* Appl. Anal. **71** (1999), 167–185.

[14] S. Molchanov and B. Vainberg, *Multiscattering by sparse scatterers.* In: Mathematical and Numerical Aspects of Wave Propagation (Santiago de Compostela, 2000). SIAM (2000), 518–522.

[15] S. Molchanov and B. Vainberg, *Spectrum of multidimensional Schrödinger operators with sparse potentials.* Anal. and Comp. Meth. Scattering and Appl. Math. **417** (2000), 231–254.

[16] P. Poulin, *Green's Functions of generalized Laplacians.* In: Probability and Mathematical Physics. CRM Proc. and Lect. Notes **42** (2007), 419–454.

[17] P. Poulin, *Scattering from sparse potentials on graphs.* J. Math. Phys., Anal., Geom. **4** (2008), 151–170.

[18] M. Reed and B. Simon, *Methods of Modern Mathematical Physics,* 3*: Scattering Theory.* Acad. Press, 1978.

Vojkan Jakšić
Department of Mathematics & Statistics
McGill University
805 Sherbrooke Street West
Montréal, QC, H3A 2K6, Canada
e-mail: `jaksic@math.mcgill.ca`

Philippe Poulin
Department of Mathematical Sciences
Norwegian University of Science and Technology
7491 Trondheim, Norway
e-mail: `poulin@math.ntnu.no`

Analysis and Mathematical Physics
Trends in Mathematics, 211–232
© 2009 Birkhäuser Verlag Basel/Switzerland

Application of ATS in a Quantum-optical Model

Anatolii A. Karatsuba and Ekatherina A. Karatsuba

Abstract. The problem of the interaction of a single two-level atom with a single mode of the quantized electromagnetic field in a coherent state in an ideal resonator in the resonance case is considered. The evolution in time of the atomic inversion, represented by the Jaynes-Cummings sum, is studied. On the basis of the application of the theorem on the approximation of a trigonometric sum by a shorter one (ATS), a new efficient method for approximating the Jaynes-Cummings sum is constructed. New asymptotic formulas for the atomic inversion are found, which approximate it on various time intervals, defined by relation between the atom-field coupling constant and the average photon number in the resonator field before the interaction of the field with the atom. The asymptotics that we obtain give the possibility to predetermine the details of the process of the inversion in the Jaynes-Cummings model depending on the field characteristics.

Mathematics Subject Classification (2000). 34A45, 40G99, 41A60, 70K40.

Keywords. Jaynes-Cummings model, atomic inversion, coherent state, trigonometric sums, approximation, theorem on the approximation of a trigonometric sum by a shorter one (ATS), functional equation for the Jacobi Theta-functions.

1. Introduction

We shall consider the behavior of an atom which reacts to one or some modes of the radiation field. Assume the situation where the optical field is almost monochromatic and its frequency is close to one from the frequencies of atomic transitions. Then it is possible not to take into account the transitions between other energy levels and to consider the given atom as the two-level atom. Jaynes and Cummings [4] showed that the equations of motion in the problem on the interaction of a single two-level atom with a single mode of the quantized radiation field are analytically solvable ones. At present, the Jaynes-Cummings model (JCM) [1], [2], [3], [4], [13], [14], [15] is the most simple approximation of the process of interaction of atoms with light, the exact description of which is not possible.

Thus, the JCM and its generalizations (see [1], [4], [12], [14], [15]) can be considered as an approximate answer to the question of what happens with every active atom of a quantum generator. For a long time one referred to the JCM as to purely theoretical construction [15]. However, the creation of one-atom maser and microlaser, and also one-mode resonator of high quality allowed to realize the JCM in practice. The data obtained in experiments confirmed some phenomena theoretically predicted in the JCM. So, by measuring such experimentally observed value as atomic inversion (the difference between the population in the excited state and the ground state of an atom) have been registered the predicted collapses and revivals of the inversion oscillations, which show the corpuscular nature of the radiation field. At present the JCM has a special place in quantum optics also for the reason that it serves for an examination and verification of conjectures concerning of models which are more complicated and close to reality.

We shall consider the dynamics of the JCM where there is only exchange by a single quantum of excitation between fixed atom and field. In this case the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\left|\psi\left(t\right)\right\rangle=\widehat{H}_{\rm int}\left|\psi\left(t\right)\right\rangle \tag{1}$$

for the vector of the state $|\psi\rangle$ of an atom-field system is usually solved (see, for example, [14], [15]) in the so-called "rotating wave approximation", that is simplifying the Hamiltonian of the interaction $\widehat{H}_{\rm int}$ by neglecting the fast oscillating terms. As a result the Hamiltonian takes the form

$$\widehat{H}_{\rm int}=\hbar g\left(\widehat{\sigma}_-\widehat{a}^+e^{-i\Delta t}+\widehat{\sigma}_+\widehat{a}e^{i\Delta t}\right), \tag{2}$$

where g is the atom-field coupling constant, $\widehat{\sigma}_+$, $\widehat{\sigma}_-$ are the Pauli Spin Matrices, $\widehat{a}^+$, $\widehat{a}$ the Bose field creation and annihilation operators, Δ is the difference between the frequency of the atomic transition and the frequency of the resonator field (for exact resonance $\Delta=0$).

Assume that at the initial moment $t=0$ the atom has no connections with the field. Then the solution of the equation (1), (2), which represents the amplitude of the probability $p_{a,m}(t)$ to find the atom in the excited state $|a\rangle$ and m photons in the resonator, or the amplitude of the probability $p_{b,m+1}(t)$ to find the atom in the ground state $|b\rangle$ and $m+1$ photons in the resonator, can be written in the form (see [14], [15])

$$\begin{aligned} p_{a,m}(t)=e^{i\frac{\Delta}{2}t}\Bigg(\left(\cos(\lambda_m t)-i\frac{\Delta}{2\lambda_m}\sin(\lambda_m t)\right)p_{a,m}(0) \\ -i\frac{g\sqrt{m+1}}{\lambda_m}\sin(\lambda_m t)p_{b,m+1}(0)\Bigg), \end{aligned} \tag{3}$$

$$\begin{aligned} p_{b,m+1}(t)=e^{-\,i\frac{\Delta}{2}t}\Bigg(-i\frac{g\sqrt{m+1}}{\lambda_m}\sin(\lambda_m t)p_{a,m}(0) \\ +\left(\cos(\lambda_m t)+i\frac{\Delta}{2\lambda_m}\sin(\lambda_m t)\right)p_{b,m+1}(0)\Bigg), \end{aligned} \tag{4}$$

where λ_m is the parameter which is called the generalized Rabi frequency:

$$\lambda_m = \sqrt{\left(\frac{\Delta}{2}\right)^2 + g^2(m+1)}.$$

Definition 1. *The atomic inversion $W(t)$ is defined as*

$$W(t) = W(t; |a\rangle) - W(t; |b\rangle),$$

where

$$W(t; |a\rangle) = \sum_{m=0}^{\infty} |p_{a,m}(t)|^2, \quad W(t; |b\rangle) = \sum_{m=0}^{\infty} |p_{b,m}(t)|^2,$$

are the probabilities to find atom at the moment t in the excited and in the ground state correspondingly.

Our aim is to study the evolution in time of the atomic inversion $W(t)$. We shall consider the case of exact resonance $\Delta = 0$, that is when

$$\lambda_m = g\sqrt{m+1}. \tag{5}$$

If the atom was in its ground state at $t = 0$, then from (3), (4), (5) we have (see for details [14], [15])

$$W(t) = W_0(t) = -\sum_{m=0}^{+\infty} W_m \cos(2gt\sqrt{m}), \tag{6}$$

and if the atom was in the excited state at $t = 0$, then

$$W(t) = W_1(t) = \sum_{m=0}^{+\infty} W_m \cos(2gt\sqrt{m+1}), \tag{7}$$

where $W_m = W_m(0) = |p_m(0)|^2$ is the probability to find at $t = 0$ in the resonator field m photons.

In quantum optics in the analysis of the JCM the case of the coherent state of the field with the Poissonian photon statistics is more often investigated:

$$W_m = e^{-m_1}\frac{m_1^m}{m!}, \tag{8}$$

where

$$m_1 = \langle m \rangle$$

is the initial average number of photons before the interaction of the field with the atom. Substituting (8) into (6) and (7), we have

$$W(t) = W_j(t) = (-1)^{j+1} e^{-m_1} \sum_{m=0}^{+\infty} \frac{m_1^m}{m!} \cos(2gt\sqrt{m+j}), \quad j = 0, 1. \tag{9}$$

The most direct approach to study $W(t)$ is to approximate the sum (9) by another sum, which is more simple for studying, for example, by the sum of a small amount of summands, in particular, by one summand. Such an approach is applied, for example, in [3], [13] (see also [1]), where the sum (9) is substituted by the infinite sum of integrals or by one integral (without estimating the remainder term), which then are calculated by the saddle point method.

In the present work we shall study the function $W(t)$ from (9), approximating it by a simpler function and by considering certain, interesting, in our opinion, particular cases of this approximation. Some of our formulas coincide with the derived ones in [3], [13]. We note, the greater m_1, the better the formulas obtained. However, for small m_1 these formulas are also of interest.

In what follows, to simplify the calculations for estimating, we shall assume that in Theorems 1, 2

$$m_1 \;\geq\; 100, \quad m_1 \text{ is an integer,} \tag{10}$$

and in Theorems 3, 4

$$m_1 \;\geq\; 2000, \quad m_1 \text{ is an integer.} \tag{11}$$

A general scheme of the method for the approximation of $W(t)$ in the form (9) was described in [11]. It is based on the application of the theorem on the approximation of a trigonometric sum by a shorter one (briefly, ATS) [5]–[8] (on other applications of the ATS to the problems of physics see [9], [10])). Besides, we also use essentially the functional equation for the Jacobi Theta-functions (see [8]).

Let us denote

$$T = 2gt, \quad C(m) = \frac{m_1^m}{m!}. \tag{12}$$

In this connection, from (9)

$$W(t) = W_j(t) = (-1)^{j+1} e^{-m_1} \sum_{m=0}^{+\infty} C(m) \cos\left(T\sqrt{m+j}\right), \quad j = 0, 1. \tag{13}$$

The structure of the paper is the following: Section 2 contains main results, formulated as theorems, Section 3 contains the proofs of auxiliary statements – lemmas, Section 4 contains the proofs of the theorems.

The following notations are used in the paper:

$\theta, \theta_1, \theta_2, \ldots$ – the functions, the module of which doesn't exceed 1, in different formulas they are, generally speaking, different;

for the real x, the function $y = \{x\}$ *fractional part of the number* x, that is $y = \{x\} = x - [x]$, where $[x]$ is *the integral part of* x, that is an integer such that $[x] \leq x < [x] + 1$,

the function $y = ||x|| = \min(\{x\}, 1 - \{x\})$ *distance from* x *to the nearest integer*;

the functions $\rho(x)$ and $\sigma(x)$ are defined by

$$\rho(x) = \frac{1}{2} - \{x\}, \qquad \sigma(x) = \int_0^x \rho(u)du.$$

We note, that for any x

$$0 \leq \{x\} < 1, \quad 0 \leq ||x|| \leq \frac{1}{2}, \quad -\frac{1}{2} < \rho(x) \leq \frac{1}{2}, \quad 0 \leq \sigma(x) \leq \frac{1}{8}. \tag{14}$$

We emphasize that the constants in the obtained asymptotic equalities can be replaced by smaller ones by means of more detailed calculations. However, the role of these constants is a secondary one. We calculate them only to show that the derived asymptotic formulas are effective. Besides, in reality, the accuracy of the formulas, which we obtained, can be much better than this is established in the theorems and lemmas.

2. Statements of main theorems

First of all we formulate the simplest theorem on approximation of the function $W(t)$ by the sum of a finite number of summands for any values of t.

Theorem 1. *Let ν_1 be an arbitrary natural number, such that $1 \le \nu_1 \le m_1^{2/3}$. Then the formula holds:*

$$W(t) = W_j(t) = \frac{(-1)^{j+1}}{\sqrt{2\pi m_1}} \left(1 - \frac{\theta}{12m_1}\right) \left(\sum_{-\nu_1 \le \nu \le \nu_1} \exp\left(-\frac{\nu^2}{2m_1}\right) \cos(T\sqrt{m_1+\nu}) \right.$$
$$\left. + \ 8\theta_{j1} + 4\theta_{j2}\frac{m_1}{\nu_1} \exp\left(-\frac{\nu_1^2}{4m_1}\right)\right); \quad j = 0, 1. \tag{15}$$

Remark 1. *It is natural to assume in* (15) *that* $\nu_1 \ge \sqrt{2m_1 \ln m_1}$.

Then we have the theorems on approximation of $W(t)$ by sums of a small number of summands.

Theorem 2. *Let $T \le m_1$, $\gamma = ||T/(4\pi\sqrt{m_1})||$. Then for $W(t)$ the following asymptotic formula holds:*

$$W(t) = W_j(t) = (-1)^{j+1} \left(1 + \frac{\theta_{j1}}{12m_1}\right) \left(T_1^{-\frac{1}{4}} \exp\left(-\frac{2\pi^2 m_1}{T_1}\gamma^2\right) \cos(D(m_1)) \right.$$
$$\left. + \ 3\theta_{j2} T_1^{-\frac{1}{4}} \exp\left(-\ \frac{\pi^2 m_1}{2T_1}\right) + \frac{5\theta_{j3}}{\sqrt{2\pi m_1}} + \ \frac{\theta_{j4} T}{4\sqrt{2\pi} m_1}\right); \quad j = 0, 1; \tag{16}$$

where

$$T_1 = 1 + \frac{T^2}{16m_1}, \quad D(m_1) = T\sqrt{m_1} + \frac{\pi^2 T\sqrt{m_1}}{2T_1}\gamma^2 - \frac{1}{2}\arctan\left(\frac{T}{4\sqrt{m_1}}\right). \tag{17}$$

Remark 2. *The necessary condition of the relevance of the equation* (16) *is the fulfillment of the inequality*

$$T \le m_1^{\frac{5}{6}}. \tag{18}$$

That is why in Theorem 2 *we assume that* (18) *is satisfied.*

Remark 3. *The inequality*

$$\gamma \le \sqrt{\frac{T_1}{2\pi^2 m_1} \ln \frac{\sqrt{2\pi m_1}}{T_1^{\frac{1}{4}}\left(5 + \frac{T}{4\sqrt{m_1}}\right)}}, \tag{19}$$

is sufficient for the summand with cosine to be the main term of (16)*. Since* $T \le m_1^{5/6}$*, then the right part of* (19) *will be a small number, that is the number* $T/(4\pi\sqrt{m_1})$ *will be close to an integer.*

Remark 4. *Let* $T \le 2\pi\sqrt{m_1}$*, and consequently,* $\gamma = T/(4\pi\sqrt{m_1})$*. Then from* (16), (17) *we have*

$$W(t) = W_j(t) = (-1)^{j+1}\left(1 + \frac{\theta_{j1}}{12m_1}\right)\left(T_1^{-\frac{1}{4}}\exp\left(-\frac{T^2}{8T_1}\right)\cos(D(m_1)) + 3\theta_{j2}\exp\left(-\frac{\pi^2}{7}m_1\right) + \frac{3\theta_{j3}}{\sqrt{m_1}}\right); \quad j = 0, 1; \tag{20}$$

where

$$D(m_1) = T\sqrt{m_1} + \frac{T^3}{32\sqrt{m_1}T_1} - \frac{1}{2}\arctan\left(\frac{T}{4\sqrt{m_1}}\right). \tag{21}$$

Remark 5. *The formula* (20) *will be relevant when*

$$T \le 2\sqrt{\ln m_1}.$$

From here and from (20), (21) *we get*

$$W(t) = W_j(t) = (-1)^{j+1}\exp\left(-\frac{T^2}{8}\right)\cos(T\sqrt{m_1}) + \theta_{j1}\frac{2T^3}{\sqrt{m_1}}; \quad j = 0, 1. \tag{22}$$

Remark 6. *From the previous remarks, it follows that on the interval* $0 < T \le 2\pi\sqrt{m_1}$ (16) *will be asymptotic one only for* $0 < T \le 2\sqrt{\ln m_1}$*. If* $2\sqrt{\ln m_1} < T \le 2\pi\sqrt{m_1}$*, then we can not claim about the inversion* $W(t)$ *anything besides, it is "small" in this interval and is trivially estimated from above by the value* $\sim m_1^{-1/2}$*. At the same time, we note, that in reality,* (22) *can be relevant and also for* $T > 2\sqrt{\ln m_1}$*.*

Remark 7. *Let now* $2\pi\sqrt{m_1} < T \le m_1^{5/6}$*. For any integer k;* $k = 1, 2, 3, \ldots, k_0 = \left[m_1^{1/3}/(4\pi)\right] - 1$*, we shall consider the intervals of the form*

$$\left(4\pi k\sqrt{m_1} - \sqrt{\ln m_1};\ 4\pi k\sqrt{m_1} + \sqrt{\ln m_1}\right),$$

which are inside the interval $(2\pi\sqrt{m_1}, m_1^{5/6})$*. For every k;* $k = 1, 2, 3, \ldots, k_0$ *and every T of the form*

$$T = 4\pi k\sqrt{m_1} + 2x, \qquad |x| \le \sqrt{\frac{1}{2}\ln\sqrt{m_1}},$$

the following asymptotics for $W(t)$ *is valid:*

$$W(t) = W_j(t) = \frac{(-1)^{j+1}}{(1+\pi^2k^2)^{1/4}}\left(\exp\left(-\frac{x^2}{8\pi^2k^2}\right)\right) \times \cos\left(\frac{x^2}{8\pi k} + x - \frac{\pi}{4} + \frac{1}{2}\arctan\frac{1}{\pi k}\right) + \theta_{j1}\frac{k}{2\sqrt{m_1}}; \quad j = 0, 1;$$

where $k = k_0 = \left[m_1^{1/3}/(4\pi)\right] - 1$. *As is obvious from the last asymptotic formula for* $W(t)$, *the maximal values* $|W(t)|$ *decrease with the increasing* k *as values* $1/(2\sqrt{\pi k})$.

Remark 8. *Using the function "fractional part", it is possible to rewrite the asymptotic formula* (16) *in another form. Thus, for* $T \le m_1^{5/6}$ *the following formula is valid*

$$\begin{aligned} W(t) &= W_j(t) \\ &= \frac{(-1)^{j+1}}{T_1^{1/4}} \left(\exp\left(-\frac{2\pi^2 m_1}{T_1}\xi^2 \right) \cos\left(T\sqrt{m_1} + \frac{\pi^2 T}{2T_1}\xi^2 - \frac{1}{2}\arctan\frac{T}{4\sqrt{m_1}} \right) \right. \\ &\quad + \exp\left(-\frac{2\pi^2 m_1}{T_1}(1-\xi)^2 \right) \cos\left(T\sqrt{m_1} + \frac{\pi^2 T}{2T_1}(1-\xi)^2 - \frac{1}{2}\arctan\frac{T}{4\sqrt{m_1}} \right) \\ &\quad \left. + \frac{5\theta_{j1}}{\sqrt{2\pi m_1}} + \frac{\theta_{j2}T}{4\sqrt{2\pi} m_1} \right); \quad j = 0,1; \end{aligned} \tag{23}$$

where $T_1 = 1 + \frac{T^2}{16m_1}$, $\xi = \{T/(4\pi\sqrt{m_1})\}$.

It is clear that if ξ is "close" to 0 or 1, then the main term in (23) will be the first or the second summand respectively. If ξ is "far" from both 0 and 1, then the both terms with cosines are "small" and the inversion is also "small". Therefore, the greatest values of $W(t)$ are situated in small neighborhoods of such t, for which

$$t = t_{\boldsymbol{r}} = \frac{2\pi\sqrt{m_1}}{g}\boldsymbol{r},$$

where $\boldsymbol{r}$ is an integer. We note, that in [3], [13] the value $t_{\boldsymbol{r}}$ is called the "time of revivals".

In the next theorem $W(t)$ is approximated by the sum of a small number (comparatively with the value of $\sqrt{m_1 \ln m_1}$) of summands, for $T \ge m_1^{5/6}$.

Theorem 3. *Let* $\sqrt{m_1} \le T \le \sqrt{m_1^3}$. *Define the number* ν_1 *by the equality*

$$h(\nu_1) = \min_{4X \le \nu \le 8X} h(\nu), \qquad X = \sqrt{m_1 \ln m_1},$$

$$h(\nu) = \left(\left\| \frac{T}{4\pi\sqrt{m_1 - \nu}} \right\| \right)^{-1} + \left(\left\| \frac{T}{4\pi\sqrt{m_1 + \nu}} \right\| \right)^{-1}.$$

Then the following asymptotic formula holds:

$$W(t) = W_j(t) = (-1)^{j+1}\frac{m_1^{\frac{1}{4}}}{\sqrt{T}}\left(1 + 3\theta_{j1}\frac{\nu_1}{m_1}\right)$$

$$\times \sum_{\alpha \leqslant n \leqslant \beta} \exp\left(-\frac{1}{2m_1}\left(\frac{T^2}{16\pi^2 n^2} - m_1 \right)^2 \right) \cos\left(\frac{T^2}{8\pi n} - \frac{\pi}{4} \right) + \frac{\theta_{j2}}{\sqrt{m_1}} + \frac{\theta_{j3}R_0}{2\sqrt{2\pi m_1}};$$

$$j = 0,1;$$

where

$$\alpha = \frac{T}{4\pi\sqrt{m_1+\nu_1}}, \quad \beta = \frac{T}{4\pi\sqrt{m_1-\nu_1}}, \quad R_0 = 1800\left(1+\frac{m_1}{T\sqrt{\ln m_1}}\right)(\ln m_1)^2.$$

It is possible to formulate Theorem 3 in slightly different form, by representing the variable of summation n in the form

$$n = a+\mu, \quad \text{where} \quad a = \left[\frac{T}{4\pi\sqrt{m_1}}\right]; \qquad \mu = 0, \pm 1, \pm 2, \dots.$$

Theorem 4. *In the conditions and notations of Theorem 3 the following asymptotic formula holds:*

$$W(t) = W_j(t) = (-1)^{j+1}\frac{m_1^{frac14}}{\sqrt{T}} \sum_{-M_1 \leqslant \mu \leqslant M_2} \left(1 + 840\theta_{j1}\left(\frac{\pi|\mu-\xi|m_1^{\frac{5}{6}}}{T}\right)^3\right)$$
$$\times \exp\left(-\frac{32\pi^2 m_1^2}{T^2}(\mu-\xi)^2\right)\cos\left(\frac{T^2}{8\pi(a+\mu)}-\frac{\pi}{4}\right) + \theta_{j2}\frac{R_0}{2\sqrt{2\pi m_1}}; \quad (24)$$
$$j = 0, 1;$$

where

$$M_1 = \frac{T\nu_1}{8\pi m_1^{\frac{3}{2}}} - \frac{3}{32\pi}\frac{T\nu_1^2}{m_1^{\frac{5}{2}}}, \quad M_2 = \frac{T\nu_1}{8\pi m_1^{\frac{3}{2}}} + \frac{3}{32\pi}\frac{T\nu_1^2}{m_1^{\frac{5}{2}}},$$
$$a = \left[\frac{T}{4\pi\sqrt{m_1}}\right]; \qquad \xi = \left\{\frac{T}{4\pi\sqrt{m_1}}\right\}.$$

Remark 9. *The application of the ATS made possible to approximate the initial sum* (15) *with* $1+2\nu_1$ *summands* (*for integer* ν_1) *by the sum* (24), *that is the sum of no more than* $1+T\nu_1/(4\pi m_1^{3/2})$ *summands, what is less than* $1+2\nu_1$ *for all* $T \leq \sqrt{m_1^3}$.

Remark 10. *If* $T \leq m_1/(\sqrt{\ln m_1})$ *and the number* $T/(4\pi\sqrt{m_1})$ *is an integer, then* $a = T/(4\pi\sqrt{m_1})$, $\xi = 0$, *and the sum with respect to* μ *in* (24) *consists of one summand with* $\mu = 0$, *and the formula for the inversion obtains the following simple form:*

$$W(t) = W_j(t) = (-1)^{j+1}\frac{m_1^{1/4}}{\sqrt{T}}\cos\left(\frac{T}{2}\sqrt{m_1}\right) + R_{j1}; \ j = 0, 1; \qquad (25)$$

where

$$|R_{j1}| \leq \frac{R_0}{2\sqrt{2\pi m_1}}.$$

Remark 11. *If* $m_1/(\sqrt{\ln m_1}) < T \leq \sqrt{m_1^3}$, *then from* (24)

$$W(t) = W_j(t) = (-1)^{j+1}\frac{m_1^{\frac{1}{4}}}{\sqrt{T}} \sum_{-M_1 \leqslant \mu \leqslant M_2} \exp\left(-\frac{32\pi^2 m_1^2}{T^2}\mu^2\right) \tag{26}$$

$$\times \cos\left(\frac{T^2}{8\pi}\left(\frac{T}{4\pi\sqrt{m_1}}+\mu\right)^{-1} - \frac{\pi}{4}\right) + R_{j2} + \theta_{j2}\frac{R_0}{2\sqrt{2\pi m_1}};\ j = 0,1;$$

where

$$|R_{j2}| \leq 840\pi^3 \frac{m_1^{\frac{1}{4}}}{\sqrt{T}} \sum_{-M_1 \leqslant \mu \leqslant M_2} \left(\frac{|\mu| m_1^{\frac{5}{6}}}{T}\right)^3 \exp\left(-\frac{32\pi^2 m_1^2}{T^2}\mu^2\right)$$

$$\leq 840\pi^3 (\ln m_1)^{3/2} m_1^{-1/4} T^{-1/2}.$$

Remark 12. *For* m_1 *such great that* $M_1 \geq (T\nu_1)/(16\pi m_1^{3/2})$, *that is for* $m_1 \geq \frac{3}{2}\nu_1$, *it is possible to approximate the sum with respect to* μ *in* (26) *by the sum of the same summands with* $|\mu| \leq (T\sqrt{\ln m_1})/(4\pi m_1)$. *For* $T < m_1/(\sqrt{\ln m_1})$ *only one summand with* $\mu = 0$ *remains in* (26), *and this formula will coincide with* (25).

3. Auxiliary statements

The statements presented below are the basis of our proof of main theorems. Some of them have also independent interest.

Lemma 1. *The following equality holds:*

$$W(t) = W_j(t) = (-1)^{j+1}\frac{1}{\sqrt{2\pi m_1}}\left(1 - \frac{\theta}{12 m_1}\right)\widetilde{F}_j(t), \qquad j = 0,1; \tag{27}$$

where

$$\widetilde{F}_j(t) = \sum_{m=0}^{\infty} r(m)\cos(T\sqrt{m+j}),$$

$$r(m) = \begin{cases} \frac{(m+1)\dots(m_1-1)}{m_1^{m_1-m-1}}, & \textit{if} \quad m < m_1 \\ 1, & \textit{if} \quad m = m_1 \\ \frac{m_1^{m-m_1}}{(m_1+1)\dots m}, & \textit{if} \quad m > m_1. \end{cases}$$

Proof. We consider $W(t)$ in the form (13). For any $m \geq 0$ from (12)

$$C(m) \leq C(m_1).$$

Indeed, if $m < m_1$, then

$$\frac{C(m)}{C(m_1)} = \frac{m_1^m}{m!}\frac{m_1!}{m_1^{m_1}} = \frac{m_1(m_1-1)\dots(m_1-m+1)}{m_1^{m_1-m}} < 1. \tag{28}$$

And if $m > m_1$, then

$$\frac{C(m)}{C(m_1)} = \frac{m_1^m}{m!}\frac{m_1!}{m_1^{m_1}} = \frac{m_1^{m-m_1}}{m(m-1)\dots(m_1+1)} < 1. \tag{29}$$

Taking out the factor $C(m_1)$ from the sum in (13), we get

$$W(t) = W_j(t) = (-1)^{j+1}e^{-m_1}C(m_1)\sum_{m=0}^{\infty} r(m)\cos\left(T\sqrt{m+j}\right), \quad j = 0, 1; \tag{30}$$

where

$$r(m) = \frac{m_1!}{m!}m_1^{m-m_1} = \begin{cases} \frac{(m+1)\dots(m_1-1)m_1}{m_1^{m_1-m}}, & \text{if} \quad m < m_1 \\ 1, & \text{if} \quad m = m_1 \\ \frac{m_1^{m-m_1}}{(m_1+1)\dots m}, & \text{if} \quad m > m_1. \end{cases}$$

0n the other hand, according to the Stirling formula

$$C(m_1) = \frac{m_1^{m_1}}{m_1!} = \frac{e^{m_1}}{\sqrt{2\pi m_1}}\left(1 - \frac{\theta}{12m_1}\right).$$

From here and from (30)

$$W(t) = W_j(t) = (-1)^{j+1}\frac{1}{\sqrt{2\pi m_1}}\left(1 - \frac{\theta}{12m_1}\right)\widetilde{F}_j(t), \qquad j = 0, 1. \qquad \square$$

It is easily seen from (28), (29) that the function $r(m_1+\nu) = \frac{C(m_1+\nu)}{C(m_1)}$ is decreasing with increase of $|\nu|$. Let us study the speed of this decreasing.

Lemma 2. *Let ν be an integer. Then the following formulas are valid:*

– *for $0 \le \nu < m_1$,*

$$r(m_1+\nu) = \exp\left(-A(m_1;\nu)\right), \tag{31}$$

where

$$A(m_1;\nu) = \left(m_1 + \nu + \frac{1}{2}\right)\ln\left(1 + \frac{\nu}{m_1}\right) - \nu - J_1(\nu), \tag{32}$$

$$J_1(\nu) = \int_0^{\nu} \frac{\sigma(x)dx}{(m_1+x)^2}; \tag{33}$$

– *for $1 \le \nu < m_1$,*

$$r(m_1 - \nu - 1) = \exp\left(-B(m_1;\nu)\right),$$

where

$$B(m_1;\nu) = \left(m_1 - \nu - \frac{1}{2}\right)\ln\left(1 - \frac{\nu}{m_1}\right) + \nu + J_2(\nu), \tag{34}$$

$$J_2(\nu) = \int_0^{\nu} \frac{\sigma(x)dx}{(m_1-x)^2}. \tag{35}$$

Proof. We consider $r(m_1+\nu)$. If $\nu = 0$, then (31) is valid, since

$$A(m_1; 0) = 0; \quad r(m_1) = 1. \tag{36}$$

Let $\nu \geq 1$. Then

$$r(m_1+\nu) = \frac{m_1^\nu}{(m_1+1)\dots(m_1+\nu-1)(m_1+\nu)} \tag{37}$$

Taking the logarithm of (37), we obtain

$$\ln(r(m_1+\nu)) = \nu \ln m_1 - \sum_{0\,<\,x\,\leqslant\,\nu} \ln(m_1+x). \tag{38}$$

We apply the Euler summation formula (see, for example [8]) to the sum of logarithms from (38). We have

$$\sum_{0<x\leqslant\nu} \ln(m_1+x) = \int_0^\nu \ln(m_1+x)dx + \rho(\nu)\ln(m_1+\nu) - \rho(0)\ln m_1$$
$$-\int_0^\nu \frac{\sigma(x)dx}{(m_1+x)^2} = \nu\ln(m_1+\nu) - \nu + m_1\ln\left(1+\frac{\nu}{m_1}\right) + \frac{1}{2}\ln\left(1+\frac{\nu}{m_1}\right) - J_1(\nu).$$

From here and from (36)–(38) follows the statement of the lemma for $\nu \geq 0$.

Let now $0 < \nu < m_1$. We find successively

$$r(m_1-\nu-1) = \frac{(m_1-\nu)(m_1-\nu+1)\dots(m_1-1)}{m_1^\nu},$$

$$\ln(r(m_1-\nu-1)) = \sum_{0\,<\,x\,\leqslant\,\nu} \ln(m_1-x) - \nu\ln m_1 = \int_0^\nu \ln(m_1-x)dx + \rho(\nu)\ln(m_1-\nu)$$
$$-\rho(0)\ln m_1 - \int_0^\nu \frac{\sigma(x)dx}{(m_1-x)^2} - \nu\ln m_1 = \left(-m_1+\nu+\frac{1}{2}\right)\ln\left(1-\frac{\nu}{m_1}\right) - \nu - J_2(\nu).$$

From here follows the statement of the lemma for $1 \leq \nu < m_1$. □

Lemma 3. *The functions $A(m_1; x)$ and $B(m_1; x)$ of Lemma 2 are monotonically increasing for $0 < x < m_1 - 1$*

Proof. Taking into account (14), we find for $0 \leq x \leq m_1 - 1$

$$\frac{d}{dx}A(m_1;x) = \ln\left(1+\frac{x}{m_1}\right) + \frac{m_1+x+\frac{1}{2}}{m_1+x} - 1 - \frac{\sigma(x)}{(m_1+x)^2} \ > \ 0,$$
$$\frac{d}{dx}B(m_1;x) = -\ln\left(1-\frac{x}{m_1}\right) - \frac{m_1-x-\frac{1}{2}}{m_1-x} + 1 + \frac{\sigma(x)}{(m_1-x)^2} \ > \ 0. \qquad \square$$

Lemma 4. *For any ν_1 such that $2 \leq \nu_1 \leq \frac{m_1}{2}$ the inequalities*

$$A(m_1;\nu_1) \geq \frac{\nu_1^2}{4m_1}, \qquad B(m_1;\nu_1) \geq \frac{\nu_1^2}{4m_1},$$

are satisfied.

Proof. From (32), (34), taking into account (14), we easily find

$$A(m_1;\nu_1) \geq \left(m_1+\nu_1+\frac{1}{2}\right)\left(\frac{\nu_1}{m_1}-\frac{\nu_1^2}{2m_1^2}\right)-\nu_1-\frac{\nu_1}{8m_1^2} \geq \frac{\nu_1^2}{4m_1},$$

$$B(m_1;\nu_1) \geq \left(m_1-\nu_1-\frac{1}{2}\right)\left(-\frac{\nu_1}{m_1}-\frac{\nu_1^2}{2m_1^2}-\frac{7\nu_1^3}{12m_1^3}\right)+\nu_1$$

$$\geq \frac{\nu_1^2}{2m_1}+\frac{\nu_1}{2m_1}-\frac{\nu_1^3}{3m_1^2} \geq \frac{\nu_1^2}{4m_1}. \qquad \square$$

Lemma 5. *Let ν_1 be an arbitrary number, such that $1 < \nu_1 \leq \frac{m_1}{2}$. Then for the function $\widetilde{F}_j(t)$ the following asymptotic formula is valid:*

$$\widetilde{F}_j(t) = F_0(T) + 4\theta\frac{m_1}{\nu_1}\exp\left(-\frac{\nu_1^2}{4m_1}\right); \quad j=0,1; \tag{39}$$

where

$$F_0(T) = \sum_{-\nu_1 \leqslant \nu \leqslant \nu_1} r(m_1+\nu)\cos\left(T\sqrt{m_1+\nu}\right). \tag{40}$$

Proof. Using Lemmas 2–4, we get

$$\sum_{\nu\ >\nu_1} r(m_1+\nu) + \sum_{\nu_1<\nu\leqslant m_1} r(m_1-\nu) \leq 2\int_{\nu_1}^{+\infty} e^{-\frac{x^2}{4m_1}}\,dx \leq \frac{4m_1}{\nu_1}\exp\left(-\frac{\nu_1^2}{4m_1}\right).$$

From here follows the statement of the lemma. $\square$

We remark, that from (39), (40) it follows that the function $\widetilde{F}_j(t)$, $j=0,1$; is well approximated by the sum $F_0(T)$ having the total number of summands of the order $\sqrt{m_1\ln m_1}$ or $\sqrt{m_1}\ln m_1$.

Lemma 6. *For any ν such that $-\nu_1 \leq \nu \leq \nu_1$, where*

$$1 \leq \nu_1^3 \leq \frac{1}{2}m_1^2,$$

the following formulas are valid

$$r(m_1+\nu) = \exp\left(-\frac{\nu^2}{2m_1}\right)\left(1+\ \theta_1\left(\frac{|\nu|^3}{m_1^2}+\frac{|\nu|}{m_1}\right)\right),$$

$$F_0(T) = \sum_{-\nu_1\leqslant\nu\leqslant\nu_1}\exp\left(-\frac{\nu^2}{2m_1}\right)\cos\left(T\sqrt{m_1+\nu}\right)+\ 6\theta_2. \tag{41}$$

Proof. We shall consider (32) and (34). It follows from (33), (35) that it is possible to represent the function $A(m_1;\nu)$ in the form

$$A(m_1;\nu) = \frac{\nu^2}{2m_1} + a(m_1;\nu);$$

where

$$a(m_1;\nu) = \frac{\nu}{2m_1} - \frac{\nu^2\left(\nu+\frac{1}{2}\right)}{2m_1^2} + \left(m_1+\nu+\frac{1}{2}\right)\left(\ln\left(1+\frac{\nu}{m_1}\right)-\frac{\nu}{m_1}+\frac{\nu^2}{2m_1^2}\right) - J_1(\nu). \tag{42}$$

Similarly

$$B(m_1;\nu) = \frac{\nu^2}{2m_1} + b(m_1;\nu);$$

where

$$b(m_1;\nu) = \frac{\nu}{2m_1} + \frac{\nu^2\left(\nu+\frac{1}{2}\right)}{2m_1^2} + \left(m_1-\nu-\frac{1}{2}\right)\left(\ln\left(1-\frac{\nu}{m_1}\right)+\frac{\nu}{m_1}+\frac{\nu^2}{2m_1^2}\right)+J_2(\nu). \tag{43}$$

We estimate easily $|a(m_1;\nu)|$, $|b(m_1;\nu)|$ from above, taking into consideration only that $\nu^3 \le \nu_1^3 \le \frac{1}{2}m_1^3$ ($m_1 \ge 100$). Taking into account (36), we find from (42), (43) by means of simple calculations, that

$$\text{for any } \nu \ge 0: \qquad |a(m_1;\nu)| \le \frac{|\nu|}{2m_1} + \frac{|\nu|^3}{m_1^2},$$

$$\text{for any } \nu \ge 1: \qquad |b(m_1;\nu)| \le \frac{|\nu|}{2m_1} + \frac{|\nu|^3}{m_1^2}.$$

From here and from (32), (34) we have

$$r(m_1+\nu) = \exp\left(-\frac{\nu^2}{2m_1}\right)\left(1+\theta_1\left(\frac{|\nu|^3}{m_1^2}+\frac{|\nu|}{m_1}\right)\right). \tag{44}$$

By substituting (44) into (40), we get

$$F_0(T) = \sum_{-\nu_1 \leqslant \nu \leqslant \nu_1} \exp\left(-\frac{\nu^2}{2m_1}\right)\cos\left(T\sqrt{m_1+\nu}\right) + R_1,$$

where

$$|R_1| \le \sum_{|\nu| \leqslant \nu_1} \exp\left(-\frac{\nu^2}{2m_1}\right)\left(\frac{|\nu|^3}{m_1^2}+\frac{|\nu|}{m_1}\right) \le 2\int_0^{+\infty} e^{-\frac{x^2}{2m_1}}\left(\frac{x^3}{m_1^2}+\frac{x}{m_1}\right)dx = 6.$$

□

One should remark that for many applications, the relation (41) is sufficiently precise. Nevertheless, it is possible to improve the estimate of the remainder term in (41), taking into account the oscillation of the cosine-factor and using the fact that $a(m_1;\nu)$ and $b(m_1;\nu)$ are differentiable as the functions in ν and are the piecewise monotone functions. In the present paper we shall not do it.

We shall approximate the argument of the cosine in (41) by a polynomial of the second degree.

Lemma 7. *Let*

$$\alpha_0 = T\sqrt{m_1}, \quad \alpha_1 = \frac{T}{2\sqrt{m_1}}, \quad \alpha_2 = \frac{T}{8\sqrt{m_1^3}},$$

$$F_1(T) = \sum_{-\nu_1 \leqslant \nu \leqslant \nu_1} \exp\left(-\frac{\nu^2}{2m_1}\right)\cos\left(\alpha_0+\alpha_1\nu-\alpha_2\nu^2\right). \tag{45}$$

Then the following formula holds

$$F_0(T) = F_1(T) + 6\theta_3 + \frac{1}{2}\theta_4 T m_1^{-\frac{1}{2}}. \tag{46}$$

Proof. Using the Taylor formula and the Lagrange finite increments formula, we have

$$\cos\left(T\sqrt{m_1+\nu}\right) = \cos\left(\alpha_0 + \alpha_1\nu - \alpha_2\nu^2\right) + \frac{\theta_1}{8}T|\nu|^3 m_1^{-\frac{5}{2}} \tag{47}$$

Besides,

$$\sum_{-\nu_1 \leqslant \nu \leqslant \nu_1} \exp\left(-\frac{\nu^2}{2m_1}\right)|\nu|^3 \le 2\int_0^\infty x^3 e^{-\frac{x^2}{2m_1}}\,dx = 4m_1^2. \tag{48}$$

From (41), (47), (48) and (45) we get the statement of the lemma. □

We note that, the remainder term in (46) isn't worse than one in (41) for all $T \le 12\sqrt{m_1}$. Furthermore, just as for (41), it is possible to make more precise the estimate of the remainder term in (46) more precise on account of the oscillations of the summands of this sum.

Lemma 8. *Let the function $F_2(T)$ be defined by the equality*

$$F_2(T) = \sum_{\nu=-\infty}^{+\infty} \exp\left(-\frac{\nu^2}{2m_1}\right)\cos\left(\alpha_2\nu^2 - \alpha_1\nu - \alpha_0\right). \tag{49}$$

Then the following formula holds

$$F_1(T) = F_2(T) + 4\theta_5\frac{m_1}{\nu_1}\exp\left(-\frac{\nu_1^2}{2m_1}\right), \tag{50}$$

where $\nu_1^3 \le \frac{1}{2}m_1^2$.

Proof. From (45) and (49) we have

$$|F_1(T) - F_2(T)| \le 2\int_{\nu_1}^{+\infty} e^{-\frac{x^2}{2m_1}}\,dx. \tag{51}$$

From here we find that

$$|F_1(T) - F_2(T)| \le \sqrt{2\pi m_1}, \quad \text{if} \quad \nu_1 \le \sqrt{2m_1},$$

$$|F_1(T) - F_2(T)| \le \frac{2m_1}{\nu_1}\exp\left(-\frac{\nu_1^2}{2m_1}\right), \quad \text{if} \quad \nu_1 > \sqrt{2m_1}.$$

□

For asymptotic evaluation of $F_2(T)$ we need the lemma about the functional equation of Jacobi's Theta-functions.

Lemma 9. *Let $\Re(\tau) > 0$,*

$$\Theta(\tau,\alpha) = \sum_{n=-\infty}^{+\infty} \exp\left(-\pi\tau(n+\alpha)^2\right).$$

Then the following equality is valid

$$\Theta\left(\frac{1}{\tau},\alpha\right)=\sqrt{\tau}\sum_{n=-\infty}^{+\infty}\exp\left(-\pi\tau n^2+2\pi i\alpha n\right). \tag{52}$$

Proof. See proof of this lemma, for example, in [8].

Lemma 10. *For $F_2(T)$ the following formula is valid:*

$$F_2(T)=\frac{\sqrt{2\pi m_1}}{\sqrt[4]{1+4m_1^2\alpha_2^2}}\sum_{n=-\infty}^{+\infty}\exp\left(-\frac{2\pi^2m_1}{1+4m_1^2\alpha_2^2}\left(n-\frac{\alpha_1}{2\pi}\right)^2\right)\cos\left(D(n)\right), \tag{53}$$

where

$$D(n)=\frac{4\pi^2m_1^2\alpha_2}{1+4m_1^2\alpha_2^2}\left(n-\frac{\alpha_1}{2\pi}\right)^2-\frac{\varphi}{2}+\alpha_0, \tag{54}$$

$$\varphi=\arctan(2m_1\alpha_2),\quad \alpha_0=T\sqrt{m_1},\quad \alpha_1=\frac{T}{2\sqrt{m_1}},\quad \alpha_2=\frac{T}{8\sqrt{m_1^3}}.$$

Proof. From (49) we have

$$\begin{aligned}F_2(T)&=\Re\sum_{n=-\infty}^{+\infty}\exp\left(-\frac{n^2}{2m_1}+i\alpha_2n^2-i\alpha_1n-i\alpha_0\right)\\&=\Re\exp(-i\alpha_0)\sum_{n=-\infty}^{+\infty}\exp\left(-\pi\tau n^2+2\pi i\alpha n\right),\end{aligned} \tag{55}$$

where

$$\tau=\frac{1}{2\pi m_1}-i\frac{\alpha_2}{\pi},\qquad \alpha=-\frac{\alpha_1}{2\pi}.$$

Using (52), we find

$$\begin{aligned}&\sum_{n=-\infty}^{+\infty}\exp\left(-\pi\tau n^2+2\pi i\alpha n\right)\\&=\frac{1}{\sqrt{\tau}}\Theta\left(\frac{1}{\tau},\alpha\right)=\frac{\sqrt{2\pi m_1}}{\sqrt[4]{1+4m_1^2\alpha_2^2}}\exp\left(i\frac{\varphi}{2}\right)\\&\quad\times\sum_{n=-\infty}^{+\infty}\exp\left(-\frac{2\pi^2m_1}{1+4m_1^2\alpha_2^2}\left(n-\frac{\alpha_1}{2\pi}\right)^2\right)\exp\left(-i\frac{4\pi^2m_1^2\alpha_2}{1+4m_1^2\alpha_2^2}\left(n-\frac{\alpha_1}{2\pi}\right)^2\right),\end{aligned}$$

where

$$\varphi=\frac{\pi}{2}-\arctan\frac{1}{2m_1\alpha_2}.$$

From here and from (55) we get the statement of the lemma. □

For the proof of Theorem 3 we shall need the following ATS theorem (see [5], [8] and especially [6], [7]):

Theorem (ATS). *Let the real functions $f(x)$, $\varphi(x)$ satisfy on the segment $[a,b]$ the following conditions:*

1) $f^{(4)}(x)$ *and* $\varphi''(x)$ *are continuous;*
2) *there exist the nonnegative numbers* H, U, A, λ, ϕ_0, ϕ_1, ϕ_2, c_1, c_2, c_3 *and* c_4 *such that the inequalities*

$$U \geqslant 1;\ 0 < b-a \leqslant \lambda U;\ |\varphi(x)| \leqslant \phi_0 H;\ |\varphi'(x)| \leqslant \phi_1 H U^{-1};\ |\varphi''(x)| \leqslant \phi_2 H U^{-2};$$
$$c_1 A^{-1} \leqslant f''(x) \leqslant c_2 A^{-1};\ |f^{(3)}(x)| \leqslant c_3 A^{-1} U^{-1};\ |f^{(4)}(x)| \leqslant c_4 A^{-1} U^{-2}$$

are fulfilled. Then, if we define the numbers x_n *by the equation* $f'(x_n) = n$, *we get*

$$\sum_{a<x\leqslant b} \varphi(x)\exp(2\pi i f(x)) = \sum_{\alpha\leqslant n\leqslant\beta} c(n)Z(n) + \theta H\left(K_1\ln(\beta-\alpha+2) + K_2 + K_3(T_a+T_b)\right), \tag{56}$$

where $\alpha = f'(a)$, $\beta = f'(b)$,

$$c(n) = \begin{cases} 1, & \textit{if}\quad \alpha < n < \beta, \\ 0.5, & \textit{if}\quad n = \alpha \quad \textit{or} \quad n = \beta, \end{cases}$$

$$Z(n) = \frac{1+i}{\sqrt{2}}\,\frac{\varphi(x_n)}{\sqrt{f''(x_n)}}\,\exp(2\pi i(f(x_n) - nx_n)),$$

$$T_\mu = \begin{cases} 0, & \textit{if}\quad \|f'(\mu)\| = 0, \\ \min\left(\sqrt{A}, \|f'(\mu)\|^{-1}\right), & \textit{if}\quad \|f'(\mu)\| \neq 0; \end{cases}$$

$$K_1 = \pi^{-1}\left(\phi_0(6.5 + 2c_1^{-1}c_2) + 6.5\lambda\phi_1\right),$$
$$K_2 = (\pi c_1^2)^{-1}\left((\lambda c_2 + 2AU^{-1})K + 2\phi_0 c_2(c_1 + A(b-a)^{-1})\right) + (\phi_0 + \lambda\phi_1)(22.5 + 9c_2A^{-1}),$$
$$K_3 = 2(2+\pi^{-1})\phi_0 + (\pi c_1)^{-1}\left((4 + 2.8\sqrt{c_1} + c_2 + 2c_1^{-1}c_2)\phi_0 + 2\lambda e^{-1}\phi_1\right),$$
$$K = 5\phi_0 c_3 + 2\phi_1 c_1 + \frac{3}{4}\lambda\left(\phi_1 c_3 + \phi_2 c_2 + \frac{\lambda}{3}\phi_2 c_3\right) + \frac{\lambda\phi_0}{2}\max\left(\frac{9}{8}c_4 + \left(\frac{13}{6}\right)^2 c_1^{-1}(c_3 + 0.5kc_4)^2, 2c_2k^{-2}\right),$$
$$k = \min\left(\frac{c_1}{4c_3}, \sqrt{\frac{c_1}{2c_4}}\right), |\theta| \leq 1.$$

4. Proof of main theorems

Using the previous auxiliary statements, it isn't difficult to prove the main theorems.

Proof of Theorem 1 follows from (27), (39), (41).

Proof of Theorem 2. We transform the function $F_2(T)$ which is defined by (53). Let

$$\gamma = \left\|\frac{\alpha_1}{2\pi}\right\| = \left\|\frac{T}{4\pi\sqrt{m_1}}\right\|,$$

and let

$$n_1 = \left[\frac{\alpha_1}{2\pi}\right], \qquad \text{if } \left\{\frac{\alpha_1}{2\pi}\right\} \le \frac{1}{2}: \qquad n_1 = \frac{\alpha_1}{2\pi} - \gamma,$$
$$n_1 = \left[\frac{\alpha_1}{2\pi}\right] + 1, \quad \text{if } \left\{\frac{\alpha_1}{2\pi}\right\} > \frac{1}{2}: \qquad n_1 = \frac{\alpha_1}{2\pi} + \gamma.$$

Then for $n = n_1$,

$$\left(n - \frac{\alpha_1}{2\pi}\right)^2 = \left(n_1 - \frac{\alpha_1}{2\pi}\right)^2 = \gamma^2.$$

Besides, for $T \le m_1$ we have the bound

$$\sum_{\substack{n=-\infty \\ n\neq n_1}}^{+\infty} \exp\left(-\frac{2\pi^2 m_1}{1+4m_1^2\alpha_2^2}\left(n - \frac{\alpha_1}{2\pi}\right)^2\right)$$
$$\le \exp\left(-\frac{\pi^2 m_1}{2\left(1+4m_1^2\alpha_2^2\right)}\right) + 2\sum_{n=1}^{\infty} \exp\left(-\frac{2\pi^2 m_1}{1+4m_1^2\alpha_2^2} n^2\right)$$
$$\le 3\exp\left(-\frac{\pi^2 m_1}{2\left(1+\frac{T^2}{16m_1}\right)}\right).$$

From here and from (53), (54) we find

$$F_2(T) = \frac{\sqrt{2\pi m_1}}{\sqrt[4]{1+\frac{T^2}{16m_1}}} \exp\left(-\frac{2\pi^2 m_1}{1+\frac{T^2}{16m_1}}\right)\cos\left(D(n_1)\right)$$
$$+ 3\theta_4 \frac{\sqrt{2\pi m_1}}{\sqrt[4]{1+\frac{T^2}{16m_1}}} \exp\left(-\frac{\pi^2}{2\left(1+\frac{T^2}{16m_1}\right)}\right), \tag{57}$$

where

$$D(n_1) = \frac{4\pi^2 m_1^2 \alpha_2}{1+4m_1^2\alpha_2^2}\gamma^2 - \frac{\varphi}{2} + \alpha_0.$$

Setting $\nu_1 = 4\sqrt{m_1 \ln m_1}$ in Lemmas 5–8 and using (57), (27), (39), (46), (50) we get the statement of Theorem 2. □

Proof of Theorem 3. We rewrite (41) in the form $F_0(T) = \Re\widetilde{\widetilde{F}}_0(T) + 7\theta_2$, where

$$\widetilde{\widetilde{F}}_0(T) = \sum_{-\nu_1 \leqslant x \leqslant \nu_1} \exp\left(-\frac{x^2}{2m_1}\right)\exp\left(-iT\sqrt{m_1+x}\right).$$

We apply the ATS to the $\widetilde{\widetilde{F}}_0(T)$, setting in it

$$f(x) = -\frac{T}{2\pi}\sqrt{m_1+x}, \quad \varphi(x) = \exp\left(-\frac{x^2}{2m_1}\right), \quad a = -\nu_1, \quad b = \nu_1. \tag{58}$$

We find successively

$$f^{(1)}(x) = -\frac{T}{4\pi\sqrt{m_1+x}}, \quad f^{(2)}(x) = \frac{T}{8\pi(m_1+x)^{3/2}}, \quad \alpha_1 = -\frac{T}{4\pi\sqrt{m_1-\nu_1}},$$

$$\beta_1 = -\frac{T}{4\pi\sqrt{m_1+\nu_1}}, \quad -\frac{T}{4\pi\sqrt{m_1+x_n}} = n, \quad x_n = \left(\frac{T}{4\pi n}\right)^2 - m_1,$$

$$f(x_n) - nx_n = -\frac{T^2}{16\pi^2(-n)} - (-nm_1). \tag{59}$$

(n takes negative values).

We set

$$A = \frac{8\pi m_1^{\frac{3}{2}}}{T}. \tag{60}$$

Replacing $(-n)$ by n we obtain the following formula for $Z(n)$:

$$Z(n) = \exp\left(i\frac{\pi}{4}\right)\frac{\varphi(x_n)}{\sqrt{f''(x_n)}}\exp\left(2\pi i\left(-\frac{T^2}{16\pi^2 n} - nm_1\right)\right),$$

where $\alpha \le n \le \beta$;

$$\alpha = \frac{T}{4\pi\sqrt{m_1+\nu_1}} = -\beta_1, \quad \beta = \frac{T}{4\pi\sqrt{m_1-\nu_1}} = -\alpha_1. \tag{61}$$

In accordance with the choice of ν_1 the values $||\alpha||$ and $||\beta||$ are not equal to zero, and $C(n) = 1$. Since m_1 is an integer, then $\exp(-2\pi i n m_1) = 1$, and the imaginary exponential in $Z(n)$ doesn't depend on m_1. Then we have:

$$\frac{1}{\sqrt{f''(x_n)}} = \frac{T}{2\sqrt{2}\pi n^{\frac{3}{2}}}, \quad \varphi(x_n) = \exp\left(-\frac{1}{2m_1}\left(\frac{T^2}{16\pi^2 n^2} - m_1\right)^2\right).$$

Consequently,

$$Z(n) = \frac{T}{2\sqrt{2}\pi n^{\frac{3}{2}}}\exp\left(-\frac{1}{2m_1}\left(\frac{T^2}{16\pi^2 n^2} - m_1\right)^2\right)\exp\left(2\pi i\left(\frac{-T^2}{16\pi^2 n} + \frac{1}{8}\right)\right),$$

$$F_0(T) = \sum_{\alpha \leqslant n \leqslant \beta}\frac{T}{2\sqrt{2}\pi n^{\frac{3}{2}}}\exp\left(-\frac{1}{2m_1}\left(\frac{T^2}{16\pi^2 n^2} - m_1\right)^2\right)\cos\left(\frac{T^2}{8\pi n} - \frac{\pi}{4}\right)$$
$$+\ 7\theta_2 +\ R_0(T), \tag{62}$$

where $R_0(T)$ is the remainder term of the ATS.

We shall simplify the sum with respect to n in (62), approximating in it the number $n^{-3/2}$ by the number $\left(\frac{T}{4\pi\sqrt{m_1}}\right)^{-3/2}$. We find from (59)

$$n = \frac{T}{4\pi\sqrt{m_1}}\left(1 + \frac{\theta}{2}\frac{\nu_1}{m_1}\right).$$

From here

$$n^{-\frac{3}{2}} = \left(\frac{4\pi\sqrt{m_1}}{T}\right)^{\frac{3}{2}}\left(1 + \theta_1\frac{\nu_1}{m_1}\right).$$

Therefore, it is possible to represent $F_0(T)$ in the form

$$F_0(T) = \frac{2\sqrt{2\pi}m_1^{\frac{3}{4}}}{\sqrt{T}}\left(1+\theta_1\frac{\nu_1}{m_1}\right)\sum_{\alpha\leqslant n\leqslant\beta}\exp\left(-\frac{1}{2m_1}\left(\frac{T^2}{16\pi^2 n^2}-m_1\right)^2\right)$$
$$\times\cos\left(\frac{T^2}{8\pi n}-\frac{\pi}{4}\right)+7\theta_2+R_0(T). \qquad (63)$$

Let us estimate $R_0(T)$. From the theorem condition

$$4\sqrt{m_1\ln m_1}\le\nu_1\le 8\sqrt{m_1\ln m_1}.$$

Using the definition of $h(\nu_1)$, it isn't difficult to prove that

$$h(\nu_1)\le 8\pi\left(\frac{m_1}{T\sqrt{\ln m_1}}+1\right). \qquad (64)$$

From (60), (61) and (58) we find successively the values of the ATS parameters:

$$\phi_0=1,\quad H=1,\quad \phi_1=\frac{\nu_1}{m_1}U,\quad \phi_2=\frac{\nu_1^2}{m_1^2}U^2,\quad C_1=\frac{1}{2},\quad C_2=\frac{3}{2},$$

$$b=\nu_1,\quad a=-\nu_1,\quad \lambda=\frac{2\nu_1}{U},\quad C_3=\frac{3U}{m_1},\quad C_4=\frac{15U^2}{2m_1^2},\quad U\ \geq 1.$$

Setting $U=1$ and taking into account (11), from (56) and (64) we come after simple calculations to the estimate

$$|R_0(T)|\ <\ 1800\left(1+\frac{m_1}{T\sqrt{\ln m_1}}\right)\ln^2 m_1.$$

From here and from (63), (27), (39) the statement of the theorem follows. □

Proof of Theorem 4. For numbers α and β from (61) we find the following asymptotic representations

$$\alpha=\frac{T}{4\pi\sqrt{m_1}}-\frac{T\nu_1}{8\pi m_1^{\frac{3}{2}}}+\frac{3}{32\pi}\frac{T\nu_1^2}{m_1^{\frac{5}{2}}}+\widetilde{R}_1,$$
$$\beta=\frac{T}{4\pi\sqrt{m_1}}+\frac{T\nu_1}{8\pi m_1^{\frac{3}{2}}}+\frac{3}{32\pi}\frac{T\nu_1^2}{m_1^{\frac{5}{2}}}+\widetilde{R}_2,$$

where for the remainder terms $\widetilde{R}_1$, $\widetilde{R}_2$ the relations hold

$$\widetilde{R}_k\le\frac{T\nu_1^3}{m_1^{\frac{7}{2}}}\ll\frac{T\,(\ln m_1)^{\frac{3}{2}}}{m_1^2}\ll\frac{(\ln m_1)^{\frac{3}{2}}}{\sqrt{m_1}}<1;\qquad k=1,2.$$

Let

$$a=\left[\frac{T}{4\pi\sqrt{m_1}}\right];\qquad \xi=\left\{\frac{T}{4\pi\sqrt{m_1}}\right\}.$$

Let us represent the number n from (63) in the form

$$n=a+\mu,$$

where the parameter μ takes the values of integers, such that

$$-M_1 \le \mu \le M_2,$$

$$M_1 = \frac{T\nu_1}{8\pi m_1^{\frac{3}{2}}} - \frac{3}{32\pi}\frac{T\nu_1^2}{m_1^{\frac{5}{2}}}, \qquad M_2 = \frac{T\nu_1}{8\pi m_1^{\frac{3}{2}}} + \frac{3}{32\pi}\frac{T\nu_1^2}{m_1^{\frac{5}{2}}}.$$

Since

$$n = \frac{T}{4\pi\sqrt{m_1}} - \xi + \mu,$$

then, taking (11) into account, we have

$$\begin{aligned} -\frac{1}{2m_1}\left(\frac{T^2}{16\pi^2 n^2} - m_1\right)^2 &= -\frac{m_1}{2}\left(\left(1 - \frac{\xi-\mu}{T}4\pi\sqrt{m_1}\right)^{-2} - 1\right)^2 \\ &= -\frac{32\pi^2 m_1^2(\mu-\xi)^2}{T^2}\left(1 + \theta\frac{5}{2}\frac{(\mu-\xi)4\pi\sqrt{m_1}}{T}\right)^2 \\ &= -\frac{32\pi^2 m_1^2(\mu-\xi)^2}{T^2} + 840\theta_1\frac{\pi^3|\mu-\xi|^3 m_1^{\frac{5}{2}}}{T^3}. \end{aligned}$$

Consequently,

$$\begin{aligned} &\exp\left(-\frac{1}{2m_1}\left(\frac{T^2}{16\pi^2 n^2} - m_1\right)^2\right) \\ &= \left(1 + 840\theta_2\pi^3|\mu-\xi|^3 m_1^{\frac{5}{2}} T^{-3}\right)\left(\exp\left(-\frac{32\pi^2 m_1^2}{T^2}(\mu-\xi)^2\right)\right). \end{aligned} \tag{65}$$

Applying (65), we can represent (63) in the form

$$\begin{aligned} F_0(T) = \frac{2\sqrt{2\pi} m_1^{\frac{3}{4}}}{\sqrt{T}}\left(1 + 2\theta_1\frac{\nu_1}{m_1}\right) &\sum_{-M_1 \leqslant \mu \leqslant M_2}\left(1 + 840\theta_2\left(\frac{\pi|\mu-\xi| m_1^{\frac{5}{6}}}{T}\right)^3\right) \\ \times \exp\left(-\frac{32\pi^2 m_1^2}{T^2}(\mu-\xi)^2\right)&\cos\left(\frac{T^2}{8\pi(a+\mu)} - \frac{\pi}{4}\right) + 7\theta_3 + R_0(T). \end{aligned}$$

The statement of Theorem 4 follows from here and from Theorem 3. □

5. Conclusion

The asymptotic formulas, proved above with the help of the ATS and of other more simple theorems, allow to approximate and to compute with good accuracy the atomic inversion on various time intervals. These formulas reflect the peculiarities of the behavior of the inversion observed in experiments: collapses and revivals of quantum oscillations are repeated, moreover the amplitude of the Rabi oscillations decreases with time, and the duration of revivals increases, the interference of the late revivals with the previous revivals occurs, and we "come out of framework" of the ATS-theorem. As it was shown above, this moment is determined by time of the interaction of the atom with the field and by the parameters: the atom-field coupling constant and the initial average number of photons before the interaction of the field with the atom. The same parameters determine the time of collapses and revivals.

The revivals of the JCM reflect the discrete structure of the photon distribution, which is described by the Jaynes-Cummings sum and which is a pure quantum mechanical phenomenon. The discreteness, expressed in the Jaynes-Cummings sum for the atomic inversion, has caused idea to apply the technics which was constructed in number theory, in particular the ATS, to approximate similar sums.

Acknowledgments

We thank Dr. Diana V. Senchenko for permanent support and stimulation during our work on this paper.

References

[1] S.M. Chumakov, A.B. Klimov, M. Kozierowski. From the Jaynes-Cummings model to collective interactions. Theory of Nonclassical States of Light, 319–370. Edited by V.V. Dodonov and V.I. Man'ko, Taylor and Francis, London–New York (2003).

[2] S.M. Chumakov, M. Kozierowski, J.J. Sanchez-Mondragon. Analytical approach to the photon statistics in the thermal Jaynes-Cummings model with an initially unexcited atom. *Phys. Rev. A* **48:6** (1993), 4594–4597.

[3] M. Fleischhauer, W.P. Schleich. Revivals made simple: Poisson summation formula as a key to the revivals in the Jaynes-Cummings model. *Phys. Rev. A* **47:3** (1993), 4258–4269.

[4] E.T. Jaynes and F.W. Cummings. Comparison of Quantum and Semiclassical Radiation Theory with Application to the Beam Maser. *Proc. IEEE* **51** (1963), 89–109.

[5] A.A. Karatsuba. Approximation of exponential sums by shorter ones. *Proc. Indian Acad. Sci.* (*Math. Sci.*) **97** (1987), 167–178.

[6] A.A. Karatsuba, M.A. Korolev. The theorem on the approximation of a trigonometric sum by a shorter one. *Izv. Ross. Akad. Nauk, Ser. Mat.* **71:2** (2007), 123–150.

[7] A.A. Karatsuba, M.A. Korolev. The theorem on the approximation of a trigonometric sum by a shorter one. *Doklady AN* **412:2** (2007), 159–161.

[8] A.A. Karatsuba, S.M. Voronin. The Riemann Zeta-Function. (W. de Gruyter, Verlag: Berlin, 1992).

[9] E.A. Karatsuba. Approximation of sums of oscillating summands in certain physical problems. *J. Math. Phys.* **45:11** (2004), 4310–4321.

[10] E.A. Karatsuba. Approximation of exponential sums in the problem on the oscillator motion caused by pushes. *Chebyshev's transactions* **6:3(15)** (2005), 205–224.

[11] E.A. Karatsuba. On an approach to the study of the Jaynes-Cummings sum in quantum optics. *J. of Numerical Algorithms* **45:1–4** (2007), 127–137.

[12] L. Mandel and E. Wolf. Optical Coherence and Quantum Optics. “Fizmatlit”, Moscow (2000) (transl. from Cambridge University Press, 1995).

[13] N.B. Narozhny, J.J. Sanchez-Mondragon, and J.H. Eberly . Coherence versus incoherence: Collapse and revival in a simple quantum model. *Phys. Rev. A* **23:1** (1981), 236–247.

[14] M.O. Scully and M.S. Zubairy. Quantum Optics. “Fizmatlit”, Moscow (2003) (transl. from Cambridge University Press, 1997).

[15] W.P. Schleich. Quantum Optics in Phase Space. “Fizmatlit”, Moscow (2005) (transl. from Wiley-VCH, 2001).

Anatolii A. Karatsuba
Steklov Institute of Mathematics of RAS
Gubkina str., 8, Moscow
119991 Russia
fax: +7-499-1350555
e-mail: karatsuba@mi.ras.ru

Ekatherina A. Karatsuba
Dorodnicyn Computing Center of RAS
Vavilova str., 40, Moscow
119333 Russia
fax: +7-499-1356159
e-mail: ekar@ccas.ru

Analysis and Mathematical Physics
Trends in Mathematics, 233–335
© 2009 Birkhäuser Verlag Basel/Switzerland

Geometry of Carnot–Carathéodory Spaces, Differentiability, Coarea and Area Formulas

Maria Karmanova and Sergey Vodop′yanov

Abstract. We compare geometries of two different local Lie groups in a Carnot–Carathéodory space, and obtain quantitative estimates of their difference. This result is extended to Carnot–Carathéodory spaces with $C^{1,\alpha}$-smooth basis vector fields, $\alpha \in [0,1]$, and the dependence of the estimates on α is established. From here we obtain the similar estimates for comparing geometries of a Carnot–Carathéodory space and a local Lie group. These results base on Gromov's Theorem on nilpotentization of vector fields for which we give new and simple proof. All the above imply basic results of the theory: Gromov type Local Approximation Theorems, and for $\alpha > 0$ Rashevskiĭ-Chow Theorem and Ball–Box Theorem, etc. We apply the obtained results for proving hc-differentiability of mappings of Carnot–Carathéodory spaces with continuous horizontal derivatives. The latter is used in proving the coarea formula for smooth contact mappings of Carnot–Carathéodory spaces, and the area formula for Lipschitz (with respect to sub-Riemannian metrics) mappings of Carnot–Carathéodory spaces.

Mathematics Subject Classification (2000). Primary 53C17, 28A75; Secondary 58C35, 93B29.

Keywords. Carnot–Carathéodory space, homogeneous group, nilpotent tangent cone, approximation theorems, differentiability, area and coarea formulas.

The research was partially supported by the Commission of the European Communities (Specific Targeted Project "Geometrical Analysis in Lie groups and Applications", Contract number 028766), the Russian Foundation for Basic Research (Grant 08-01-00531), the State Maintenance Program for Young Russian Scientists and the Leading Scientific Schools of Russian Federation (Grant NSh-5682.2008.1).

Contents

1. Introduction

The geometry of Carnot–Carathéodory spaces naturally arises in the theory of subelliptic equations, contact geometry, optimal control theory, nonholonomic mechanics, neurobiology and other areas (see works by A.A. Agrachev [1], A.A. Agrachev and J.-P. Gauthier [3], A.A. Agrachev and A. Marigo [4], A.A. Agrachev and A.V. Sarychev [6, 7, 8, 9, 10], A.A. Agrachev and Yu.L. Sachkov [5], A. Bellaïche [16], A. Bonfiglioli, E. Lanconelli and F. Uguzzoni [19], S. Buckley, P. Koskela and G. Lu [20], L. Capogna [23, 24], G. Citti, N. Garofalo and E. Lanconelli [33], L. Capogna, D. Danielli and N. Garofalo [25, 26, 27, 28, 29], Ya. Eliashberg [37, 38, 39], G.B. Folland [45, 46], G.B. Folland and E.M. Stein [47], B. Franchi,

R. Serapioni, F. Serra Cassano [56, 57, 58, 59], N. Garofalo [61], N. Garofalo and D.-M. Nhieu [63, 64], R.W. Goodman [66], M. Gromov [70, 71], L. Hörmander [76], F. Jean [77], V. Jurdjevic [84], G.P. Leonardi, S. Rigot [94], W. Liu and H.J. Sussman [96], G. Lu [97], G.A. Margulis and G.D. Mostow [105, 106], G. Metivier [107], J. Mitchell [108], R. Montgomery [109, 110], R. Monti [111, 112], A. Nagel, F. Ricci, E.M. Stein [114, 115], A. Nagel, E.M. Stein and S. Wainger [116], P. Pansu [118, 119, 120, 121], L.P. Rothschild and E.M. Stein [127], R.S. Strichartz [134], A.M. Vershik and V.Ya. Gershkovich [136], S.K. Vodop′yanov [137, 139, 140, 141, 142], S.K. Vodop′yanov and A.V. Greshnov [143], C.J. Xu and C. Zuily [152] for an introduction to this theory and some its applications).

A Carnot–Carathéodory space (and its special case referred below to as a *Carnot manifold*) $\mathbb{M}$ (see, for example, [70, 136]) is a connected Riemannian manifold with a distinguished horizontal subbundle $H\mathbb{M}$ in the tangent bundle $T\mathbb{M}$ that meets some algebraic conditions on the commutators of vector fields $\{X_1, \dots, X_n\}$ constituting a local basis in $H\mathbb{M}$, $n = \dim H_p\mathbb{M}$ for all $p \in \mathbb{M}$.

The distance d_{cc} (the intrinsic Carnot–Carathéodory metric) between points $x, y \in \mathbb{M}$ is defined as the infimum of the lengths of *horizontal* curves joining x and y and is not equivalent to Riemannian distance if $H\mathbb{M}$ is a proper subbundle (a piecewise smooth curve γ is called horizontal if $\dot{\gamma}(t) \in H_{\gamma(t)}\mathbb{M}$). See results on properties of this metric in the paper by A. Nagel, E.M. Stein and S. Wainger [116] and in the monograph by D.Yu. Burago, Yu.D. Burago and S.V. Ivanov [22].

The Carnot–Carathéodory metric is applied in the study of hypoelliptic operators, see C. Fefferman and D.H. Phong [44], L. Hörmander [76], D. Jerison [78], A. Nagel, E.M. Stein and S. Wainger [116], L.P. Rothschild and E.M. Stein [127], A. Sánchez-Calle [128]. Also, this metric and its properties are essentially used in theory of PDE's (see papers by M. Biroli and U. Mosco [17, 18], S.M. Buckley, P. Koskela and G. Lu [20], L. Capogna, D. Danielli and N. Garofalo [25, 26, 27, 28, 29], V.M. Chernikov and S.K. Vodop′yanov [31], D. Danielli, N. Garofalo and D.-M. Nhieu [35], B. Franchi [48], B. Franchi, S. Gallot and R. Wheeden [49], B. Franchi, C.E. Gutiérrez and R.L. Wheeden [50], B. Franchi and E. Lanconelli [51, 52], B. Franchi, G. Lu and R. Wheeden [53, 54], B. Franchi and R. Serapioni [55], R. Garattini [60], N. Garofalo and E. Lanconelli [62], P. Hajłasz and P. Strzelecki [73], J. Jost [79, 80, 81, 82], J. Jost and C.J. Xu [83], S. Marchi [104], K.T. Sturm [135]).

The following results are usually regarded as foundations of the geometry of Carnot–Carathéodory spaces:

1) Rashevskiĭ–Chow Theorem [32, 126] on connection of two points by a horizontal path;
2) Ball–Box Theorem [116] (saying that a ball in Carnot–Carathéodory metric contains a "box" and is a subset of a "box" with controlled "radii");
3) Mitchell's Theorem [108] on convergence of rescaled Carnot–Carathéodory spaces around $g \in \mathbb{M}$ to a nilpotent tangent cone;

4) Gromov's Theorem [70] on convergence of "rescaled" with respect to $g \in \mathbb{M}$ basis vector fields to *nilpotentized* (*at* g) vector fields generating a graded nilpotent Lie algebra (the corresponding connected and simply connected Lie group is called the *nilpotent tangent cone at* g); here $g \in \mathbb{M}$ is an arbitrary point;
5) Gromov Approximation Theorem [70] on local comparison of Carnot–Carathéodory metrics in the initial space and in the nilpotent tangent cone, and its improvements due to A. Bellaïche [16].

The goal of the paper is both to give a new approach to the geometry of Carnot–Carathéodory spaces and to establish some basic results of geometric measure theory on these metric structures including an appropriate differentiability.

New results in the geometry of Carnot–Carathéodory spaces contain essentially new quantitative estimates of closeness of geometries of different tangent cones located one near another. One of peculiarities of the paper is that we solve all problems under minimal assumption on smoothness of the basis vector fields (they are $C^{1,\alpha}$-smooth, $0 \leq \alpha \leq 1$), although all the basic results are new even for C^∞-vector fields. In some parts of this paper, the symbol $C^{1,\alpha}$ means that the derivatives of the basis vector fields are H^α-continuous with respect to some nonnegative symmetric function $\mathfrak{d} : U \times U \to \mathbb{R}$, $U \Subset \mathbb{M}$, such that $\mathfrak{d} \geq C\rho$, $0 < C < \infty$, where C depends only on U, and ρ is the following distance which is comparable with Riemannian one in U: if $y = \exp\Bigl(\sum\limits_{i=1}^{N} y_i X_i\Bigr)(x)$ then

$$\rho(x, y) = \max_{i=1,\dots,N} \{|y_i|\}. \tag{1.0.1}$$

Some additional properties of $\mathfrak{d}$ are described below when it is necessary.

Note that from the very beginning it is unknown whether Rashevskiĭ–Chow Theorem is true for $C^{1,\alpha}$-smooth basis vector fields. Therefore Carnot–Carathéodory distance can not be well defined. We prove that the function $d_\infty : \mathbb{M}\times\mathbb{M} \to \mathbb{R}$, which is defined locally as follows: if $y = \exp\Bigl(\sum\limits_{i=1}^{N} y_i X_i\Bigr)(x)$, then

$$d_\infty(x, y) = \max_{i=1,\dots,N} \{|y_i|^{\frac{1}{\deg X_i}}\},$$

is a quasimetric, and we use it instead of d_{cc}. Note that, in the C^∞-smooth case d_∞ is comparable locally with d_{cc} [116, 70]. One of the main results is the following (see below Theorem 2.4.1 for sharp statement).

Theorem 1.0.1. *Suppose that* $d_\infty(u, u') = C\varepsilon$, $d_\infty(u, v) = \mathcal{C}\varepsilon$ *for some* $C, \mathcal{C} < \infty$,

$$w_\varepsilon = \exp\Bigl(\sum_{i=1}^{N} w_i \varepsilon^{\deg X_i} \widehat{X}_i^u\Bigr)(v) \text{ and } w'_\varepsilon = \exp\Bigl(\sum_{i=1}^{N} w_i \varepsilon^{\deg X_i} \widehat{X}_i^{u'}\Bigr)(v).$$

Then, for $\alpha > 0$, *we have*

$$\max\{d_\infty^u(w_\varepsilon, w'_\varepsilon), d_\infty^{u'}(w_\varepsilon, w'_\varepsilon)\} \leq L\varepsilon\rho(u, u')^{\frac{\alpha}{M}}$$

where L is uniformly bounded in $u, u', v \in U \subset \mathbb{M}$, and in $\{w_i\}_{i=1}^N$ belonging to some compact neighborhood of 0.

In the case of $\alpha = 0$, we have

$$\max\{d_\infty^u(w_\varepsilon, w'_\varepsilon), d_\infty^{u'}(w_\varepsilon, w'_\varepsilon)\} \leq \varepsilon o(1)$$

where $o(1)$ is uniform in $u, u', v \in U \subset \mathbb{M}$, and in $\{w_i\}_{i=1}^N$ belonging to some compact neighborhood of 0 as $\varepsilon \to 0$.

Here we assume that $U \subset \mathbb{M}$ is a compact neighborhood (or compactly embedded neighborhood) small enough, and ρ is a Riemannian metric. The symbol $\widehat{X}_i^u$ ($\widehat{X}_i^{u'}$) denotes a nilpotentized at u (u') vector field (see Gromov's Theorem below) for each $i = 1, \ldots, N$. These vector fields constitute Lie algebra of the nilpotent tangent cone at u (u').

Further, in Theorem 2.3.1 we extend this result to the case of a "chain" consisting of several points.

The content of obtained estimates is very profound: they imply both new properties of Carnot–Carathéodory spaces and the above-mentioned classical ones. Moreover, it allows to obtain these results under minimal smoothness of basis vector fields. We emphasize that many classical results were earlier known only in the cases of sufficiently smooth basis vector fields.

The investigation of sub-Riemannian geometry under minimal smoothness of the basis vector fields is motivated by the recently constructed by G. Citti and A. Sarti, R.K. Hladky and S.D. Pauls, and J. Petitot models of visualization [34, 75, 123]. More exactly, the model of a brain perception of a black-and-white plain image is constructed in these papers. This model makes possible the interpretation on a computer of a human brain's work during the visualization of information. In particular, it is shown how the human brain completes the image a part of which is closed. The geometry of this model is based on a roto-translation group which is a three-dimensional Carnot–Carathéodory space with a tangent cone being a Heisenberg group $\mathbb{H}^1$ at each point. Since by now there are no theorems on regularity of the image created by a human brain, any reduction of smoothness of vector fields is essential for the construction of sharp visualization models.

The main result concerning the geometry of Carnot–Carathéodory spaces is proved in Section 2. The method of proving is new, and it essentially uses Hölder dependence of solutions to ordinary differential equations on parameter (see Theorem 2.1.16). Probably, this dependence is not a new result. For reader's convenience we give its independent proof in Section 5. In Subsection 2.1, all other auxiliary result are formulated.

In Subsection 2.2 we discuss, in particular, the following statements.

A) *Let X_j, $j = 1, \ldots, N$, belong to the class C^1 on a Carnot–Carathéodory space $\mathbb{M}$. On* $\operatorname{Box}(g, \varepsilon r_g)$, *consider the vector fields $\{{}^\varepsilon X_i\} = \{\varepsilon^{\deg X_i} X_i\}$ for all $i = 1, \ldots, N$. Then the uniform convergence*

$$X_i^\varepsilon = \left(\Delta^g_{\varepsilon^{-1}}\right)_* {}^\varepsilon X_i \to \widehat{X}_i^g \quad \text{as } \varepsilon \to 0, \quad i = 1, \ldots, N,$$

(*see the definition of* $\Delta^g_{\varepsilon^{-1}}$ *below in* 2.2.4) *holds at the points of the box* $\mathrm{Box}(g, r_g)$ *and this convergence is uniform in* g *belonging to some compact neighborhood, where the collection* $\{\widehat{X}^g_i\}$, $i = 1, \dots, N$, *of vector fields around* g *constitutes a basis of a graded nilpotent Lie algebra.*

B) *For each compact domain* $U \subset \mathbb{M}$ *small enough, there exists a constant* $Q = Q(U)$, *such that the inequality*

$$d_\infty(u, v) \le Q(d_\infty(u, w) + d_\infty(w, v))$$

holds for every triple of points u, w, $v \in U$, *where* $Q(U)$ *depends on* U.

C) *Given points* $u, v \in \mathbb{M}$, $d_\infty(u, v) = \mathcal{C}\varepsilon$ *for some* $\mathcal{C} < \infty$, *for points*

$$w_\varepsilon = \exp\Bigl(\sum_{i=1}^N w_i \varepsilon^{\deg X_i} X_i\Bigr)(v) \text{ and } w'_\varepsilon = \exp\Bigl(\sum_{i=1}^N w_i \varepsilon^{\deg X_i} \widehat{X}^u_i\Bigr)(v),$$

we have

$$\max\bigl\{d_\infty(w_\varepsilon, w'_\varepsilon), d^u_\infty(w_\varepsilon, w'_\varepsilon)\bigr\} \le \varepsilon o(1)$$

where $o(1)$ *is uniform in* u, v *belonging to a compact neighborhood* $U \subset \mathbb{M}$, *and in* $\{w_i\}_{i=1}^N$ *belonging to some compact neighborhood of* 0 *as* $\varepsilon \to 0$.

Statement **A** is just Gromov's Theorem [70] on the nilpotentization of vector fields. Gromov has formulated it for C^1-smooth fields, however, Example 2.2.15 by Valeriĭ Berestovskiĭ makes evident the fact that arguments of the proof given in [70, pp. 128–133] have to be corrected. In Corollary 2.2.13, in one particular case, we give a new proof of this assertion based on an another idea.

Statement **B** says that d_∞ meets the generalized triangle inequality, i.e., it is a quasimetric. The implication **A** $\Longrightarrow$ **B** is proved in Corollary 2.2.17.

Statement **C** gives an estimate of divergence of integral lines of the given vector fields and the nilpotentized vector fields.

The implication **B** $\Longrightarrow$ **C** is a particular case of Theorem 2.7.1.

In the theory developed in Subsection 2.4, the generalized triangle inequality plays the crucial role.

In Subsection 2.4, we prove one of the basic results of Section 2, namely, Theorem 2.4.1 which compares local geometries of two different local Lie groups. It is essentially based on the main theorem of Subsection 2.3 which compares "global" geometries of different tangent cones (i.e., it looks like Theorem 2.4.1 with $\varepsilon = 1$). Subsection 2.5 is devoted to approximation theorems. In particular, we compare quasimetrics of two tangent cones, and the quasimetric of a tangent cone with the initial one. There we give their proofs and the proofs of some auxiliary properties of the geometry. Further, in Subsection 2.6, we prove Theorem 2.3.1, which is the "continuation" of Theorem 2.4.1. In Subsection 2.7, we compare the geometry of a Carnot–Carathéodory space with the one of a tangent cone. In Subsection 2.8, we give applications of our results to investigation of the sub-Riemannian geometry. From Gromov type theorem on the nilpotentization of vector fields [70], we obtain

Rashevskiĭ–Chow Theorem, Ball–Box Theorem, Mitchell Theorem on Hausdorff dimension of Carnot manifolds and many other corollaries.

Main results of Section 2 are formulated in short communications [144, 145].

Section 3 is devoted to differentiability of mappings in the category of Carnot–Carathéodory spaces.

We recall the classical definition of differentiability for a mapping $\varphi : \mathcal{M} \to \widetilde{\mathcal{M}}$ of two Riemannian manifolds: φ is differentiable at $x \in \mathcal{M}$ if there exists a linear mapping $L : T_x\mathcal{M} \to T_{\varphi(x)}\widetilde{\mathcal{M}}$ of the tangent spaces such that

$$\rho_{\widetilde{\mathcal{M}}}(\varphi(\exp_x h), \exp_{\varphi(x)} Lh) = o(\|h\|_x), \quad h \in T_x\mathcal{M},$$

where $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ and $\exp_{\varphi(x)} : T_{\varphi(x)}\widetilde{\mathcal{M}} \to \widetilde{\mathcal{M}}$ are the exponential mappings, and $\rho_{\widetilde{\mathcal{M}}}$ is the Riemannian metric in $\widetilde{\mathcal{M}}$, $\|h\|_x$ is the length of $h \in T_x\mathcal{M}$.

It is known (see [70, 110]) that the local geometry of a Carnot manifold at a point $g \in \mathbb{M}$ can be modelled as a graded nilpotent Lie group $\mathbb{G}_g\mathbb{M}$. It means that the tangent space $T_g\mathbb{M}$ has an additional structure of a graded nilpotent Lie group. If $\mathbb{M}$ and $\widetilde{\mathbb{M}}$ are two Carnot manifolds and $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$ is a mapping then a suitable concept of differentiability (that is, the *hc*-differentiability) can be obtained from the previous concept in the following way: φ is *hc*-differentiable at $x \in \mathbb{M}$ if there exists a horizontal homomorphism $L : \mathbb{G}_x\mathbb{M} \to \mathbb{G}_{\varphi(x)}\widetilde{\mathbb{M}}$ of the nilpotent tangent cones such that

$$\tilde{d}_{cc}(\varphi(\exp_x h), \exp_{\varphi(x)} Lh) = o(|h|_x), \quad h \in \mathbb{G}_x\mathbb{M},$$

where $\tilde{d}_{cc}$ is the Carnot–Carathéodory distance in $\widetilde{\mathbb{M}}$ and $|\cdot|_x$ is an homogeneous norm in $\mathbb{G}_x\mathbb{M}$.

For us, it is convenient to regard some neighborhood of a point g both as a subspace of the metric space $(\mathbb{M}, d_{cc})$ and as a neighborhood of the unity of the local Lie group $\mathcal{G}^u\mathbb{M}$ with Carnot–Carathéodory metric d^u_{cc} (see Definition 1.2). In the sense explained below (see Definition 2.1.21), $\exp^{-1} : \mathcal{G}^u\mathbb{M} \to \mathbb{G}_u\mathbb{M}$ is an isometrical monomorphism of the Lie structures. Then the last definition of the *hc*-differentiability can be reformulated as follows. Given two Carnot manifolds $(\mathbb{M}, d_{cc})$ and $(\widetilde{\mathbb{M}}, \tilde{d}_{cc})$ and a set $E \subset \mathbb{M}$, a mapping $\varphi : E \to \widetilde{\mathbb{M}}$ is called *hc-differentiable* at a point $u \in E$ (see the paper by S.K. Vodop′yanov and A.V. Greshnov [143], and also [139, 140, 141, 142]) if there exists a horizontal homomorphism $L_u : \left(\mathcal{G}^u\mathbb{M}, d^u_{cc}\right) \to \left(\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}, d^{\varphi(u)}_{cc}\right)$ of the local Lie groups such that

$$d^{\varphi(u)}_{cc}(\varphi(w), L_u(w)) = o(d^u_{cc}(u, w)) \quad \text{as } E \cap \mathcal{G}^u \ni w \to u. \tag{1.0.2}$$

The given definition of the *hc*-differentiability of mappings for Carnot manifolds can be treated as a straightforward generalization of both the classical definition of differentiability and the definition of $\mathcal{P}$-differentiability. Clearly, if the Carnot manifolds are Carnot groups then this definition of *hc*-differentiability is equivalent to the definition of $\mathcal{P}$-differentiability introduced by P. Pansu in [121] for an open set $E \subset \mathbb{G}$. For an arbitrary $E \subset \mathbb{G}$, the last concept was investigated by

S.K. Vodop′yanov [137] and by S.K. Vodop′yanov and A.D. Ukhlov [150] (see also the paper by V. Magnani [98]).

In Section 3, we introduce the notion of *hc*-differentiability, which is adequate to the geometry of Carnot manifold, and study its properties. Moreover, in this section, we prove the *hc*-differentiability of a composition of *hc*-differentiable mappings.

In the same section we prove the *hc*-differentiability of rectifiable curves. In the case of curves, the definition of the *hc*-differentiability is interpreted as follows: a mapping $\mathbb{R} \supset E \ni t \mapsto \gamma(t) \in \widetilde{\mathbb{M}}$ is *hc*-differentiable at a point $s \in E$ in a Carnot manifold $\widetilde{\mathbb{M}}$ if the relation

$$d_c^{\gamma(s)}\big(\gamma(s+\tau), \exp(\tau a)(\gamma(s))\big) = o(\tau) \quad \text{as } \tau \to 0,\ s+\tau \in E, \tag{1.0.3}$$

holds, where $a \in H_{\gamma(s)}\widetilde{\mathbb{M}}$ ((1.0.3) agrees with (1.0.2) when $\mathbb{M} = \mathbb{R}$, see also [105]). On Carnot groups (Carnot manifolds), relation (1.0.3) is equivalent to the $\mathcal{P}$-differentiability (*cc*-differentiability) of curves in the sense of Pansu [121] (Margulis–Mostow [105]). Our proof of differentiability is new even for Carnot groups. We prove step by step the *hc*-differentiability of the absolutely continuous curves, the Lipschitz mappings of subsets of $\mathbb{R}$ into $\mathbb{M}$, and the rectifiable curves. Here we generalize a classical result and obtain the following assertion: *the continuity of horizontal derivatives of a contact mapping defined on an open set implies its pointwise hc-differentiability* (Theorem 3.3.1).

As an important corollary to these assertions, we infer that the nilpotent tangent cone is defined by the horizontal subbundle of the Carnot manifold: *tangent cones found from different collections of basis vector fields with the same span of horizontal vector fields are isomorphic as local Carnot groups* (Corollary 3.3.3). Thus, the correspondence "local basis $\mapsto$ nilpotent tangent cone" is functorial. In the case of C^∞-vector fields, this result was established by A. Agrachev and A. Marigo [4], and G.A. Margulis and G.D. Mostow [106] where a coordinate-free definition of the tangent cone was given.

Main results of Section 3 are formulated in short communications by S.K. Vodopyanov [139, 140] (see some details and more general results on this subject including Rademacher–Stepanov Theorem in [141, 142]).

Section 4 is dedicated to such application of results on *hc*-differentiability as the sub-Riemannian analog of the coarea and area formulas.

It is well known that the coarea formula

$$\int\limits_U \mathcal{J}_k(\varphi, x)\,dx = \int\limits_{\mathbb{R}^k} dz \int\limits_{\varphi^{-1}(z)} d\mathcal{H}^{n-k}(u), \tag{1.0.4}$$

where $\mathcal{J}_k(\varphi, x) = \sqrt{\det(D\varphi(x)D\varphi^*(x))}$, has many applications in analysis on Euclidean spaces. Here we assume that $\varphi \in C^1(U, \mathbb{R}^k)$, $U \subset \mathbb{R}^n$, $n \geq k$. For the first time, it was established by A.S. Kronrod [93] for the case of a function $\varphi : \mathbb{R}^2 \to \mathbb{R}$. Next, it was generalized by H. Federer first for mappings of Riemannian manifolds $\varphi : \mathcal{M}^n \to \mathcal{N}^k$, $n \geq k$, in [41], and then, for mappings of rectifiable sets in Euclidean spaces $\varphi : \mathcal{M}^n \to \mathcal{N}^k$, $n \geq k$, in [42]. Next, in the paper [117], M. Ohtsuka generalized the coarea formula (1.0.4) for mappings $\varphi : \mathbb{R}^n \to \mathbb{R}^m$, $n, m \geq k$, with $\mathcal{H}^k$-σ-finite image $\varphi(\mathbb{R}^n)$. An infinite-dimensional analog of the coarea formula was proved by H. Airault and P. Malliavin in 1988 [11] for the case of Wiener spaces. This result can be found in the monograph by P. Malliavin [102]. See other proofs and applications of the coarea formula in the monographs by L.C. Evans and R.F. Gariepy [40], M. Giaquinta, G. Modica and J. Souček [65], F. Lin and X. Yang [95].

Formula (1.0.4) has applications in the theory of exterior forms, currents, in minimal surfaces problems (see, for example, paper by H. Federer and W.H. Fleming [43]). Also, Stokes formula can be easily obtained by using the coarea formula (see, for instance, lecture notes by S.K. Vodop′yanov [138]). Because of the development of analysis on more general structures, a natural question arise to extend the coarea formula on objects of more general geometry in comparison with Euclidean spaces, especially on metric spaces and structures on sub-Riemannian geometry. In 1999, L. Ambrosio and B. Kirchheim [12] proved the analog of the coarea formula for Lipschitz mappings defined on $\mathcal{H}^n$-rectifiable metric space with values in $\mathbb{R}^k$, $n \geq k$. In 2004, this formula was proved for Lipschitz mappings defined on an $\mathcal{H}^n$-rectifiable metric space with values in an $\mathcal{H}^k$-rectifiable metric space, $n \geq k$, by M. Karmanova [85, 87]. Moreover, necessary and sufficient conditions on the image and the preimage of a Lipschitz mapping defined on an $\mathcal{H}^n$-rectifiable metric space with values in an *arbitrary* metric space for the validity of the coarea formula were found. Independently of this result, the level sets of such mappings were investigated, and the metric analog of Implicit Function Theorem was proved by M. Karmanova [86, 87, 89].

All the above results are connected with rectifiable metric spaces. Note that, their structure is similar to the one of Riemannian manifolds. But there are also *non-rectifiable* metric spaces which geometry is not comparable with the Riemannian one. *Carnot manifolds* are of special interest. Up to now, the problem of the sub-Riemannian coarea formula was one of well-known intrinsic unsolved problems.

A Heisenberg group and a Carnot group are well known particular cases of a Carnot manifold. In 1982, P. Pansu proved the coarea formula for functions defined on a Heisenberg group [118]. Next, in [74], J. Heinonen extended this formula to smooth functions defined on a Carnot group. In [113], R. Monti and F. Serra Cassano proved the analog of the coarea formula for BV-functions defined on a two-step Carnot–Carathéodory space. One more result concerning the analogue of (1.0.4) belongs to V. Magnani. In 2000, he proved a *coarea inequality* for mappings of Carnot groups [100]. The equality was proved only for the case of a mapping defined on a Heisenberg group with values in Euclidean space $\mathbb{R}^k$ [101]. Until now, the question about the validity of the coarea formula even for a model case of a mapping of Carnot groups was open.

The sub-Riemannian analogs of the well-known area formula are proved in recent papers by V. Magnani [98] and S. D. Pauls [122] for Lipschitz mappings of

Carnot groups $\varphi : \mathbb{G} \to \widetilde{\mathbb{G}}$:

$$\int\limits_{\mathbb{G}} \mathcal{J}(\varphi, y)\, d\mathcal{H}^\nu(y) = \int\limits_{\widetilde{\mathbb{G}}} d\mathcal{H}^\nu(x), \tag{1.0.5}$$

However, in their works, these authors do not give the analytic description of sub-Riemannian Jacobian since they construct the Hausdorff measures $\mathbb{H}^\nu$ with respect to Carnot–Carathéodory metrics where ν is the Haudorff dimension of $\mathbb{G}$. The definition is the following:

$$\mathcal{J}(\varphi, y) = \lim_{r\to 0} \frac{\mathcal{H}^\nu(\varphi(B_{cc}(y,r)))}{\mathcal{H}^\nu(B_{cc}(y,r))} = \lim_{r\to 0} \frac{\mathcal{H}^\nu(\widehat{D}\varphi(x)[B_{cc}(y,r)])}{\mathcal{H}^\nu(B_{cc}(y,r))}$$

where $B_{cc}(y,r)$ is the Carnot–Carathéodory ball and $\widehat{D}\varphi(x)$ is the $\mathcal{P}$-differential of φ at y. Obviously, it is impossible to calculate this value.

We prove the sub-Riemannian area formula for Lipschitz (with respect to sub-Riemannian metric) mappings of Carnot–Carath-éodory spaces. The result is new even for mappings of Carnot groups: instead of (1.0.5), for a Lipschitz mapping $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$, we prove that

$$\int\limits_{\mathbb{M}} \sqrt{\det(\widehat{D}\varphi(y)^*\widehat{D}\varphi(y))}\, d\mathcal{H}^\nu(y) = \int\limits_{\widetilde{\mathbb{M}}} f(y)\, d\mathcal{H}^\nu(x) \tag{1.0.6}$$

where $\widehat{D}\varphi(x)$ is the *hc*-differential of φ at y and and Hausdorff measures $\mathcal{H}^\nu$ are constructed with respect to a special (sub-Riemannian) quasimetric d_2 which is locally equivalent to the Carnot–Carathéodory metric d_{cc} and ν is the Haudorff dimension of $\mathbb{M}$. Notice that the sub-Riemannian area formula (1.0.6) looks like in the Riemannian spaces with sub-Riemannian differential instead of Riemannian one.

Main results of Section 4 are formulated in [146, 147, 90].

2. Metric properties of Carnot–Carathéodory spaces

2.1. Preliminary Results

Recall the definition of a Carnot–Carathéodory space.

Definition 2.1.1 (compare with [70, 116]**).** Fix a connected Riemannian C^∞-manifold $\mathbb{M}$ of a dimension N. The manifold $\mathbb{M}$ is called the *Carnot–Carathéodory space* if, in the tangent bundle $T\mathbb{M}$, there exists a tangent subbundle $H\mathbb{M}$ with a finite collection of natural numbers $\dim H_x\mathbb{M} = \dim H_1 < \cdots < \dim H_i < \cdots < \dim H_M = N$, $1 < i < M$, and each point $p \in \mathbb{M}$ possesses a neighborhood $U \subset \mathbb{M}$ with a collection of C^1-smooth vector fields $X_1, \dots, X_N$ on U enjoying the following three properties. For each $v \in U$ we have

(1) $X_1(v), \dots, X_N(v)$ constitutes a basis of $T_v\mathbb{M}$;
(2) $H_i(v) = \operatorname{span}\{X_1(v), \dots, X_{\dim H_i}(v)\}$ is a subspace of $T_v\mathbb{M}$ of a dimension $\dim H_i$, $i = 1, \dots, M$, where $H_1(v) = H_v\mathbb{M}$;

(3) $$[X_i, X_j](v) = \sum_{\deg X_k \leq \deg X_i + \deg X_j} c_{ijk}(v) X_k(v) \qquad (2.1.1)$$

where the *degree* $\deg X_k$ equals $\min\{m \mid X_k \in H_m\}$;

If, additionally, the fourth condition holds then the Carnot–Carathéodory space will be called the *Carnot manifold*:

(4) a quotient mapping $[\cdot,\cdot]_0 : H_1 \times H_j/H_{j-1} \mapsto H_{j+1}/H_j$, induced by Lie brackets by the rule $[X, Y \bmod H_{\deg Y-1}]_0 = [X,Y] \bmod H_{\deg Y}$, is an epimorphism for all $1 \leq j < M$ (here $X \in H_1$, $Y \in H_j$ are C^1-smooth vector fields). Here $H_0 = \{0\}$.

Remark 2.1.2. For any Carnot–Carathéodory space, we prove below, in Theorem 2.1.8, that the coefficients

$$\{c_{ijk}(v)\}_{\deg X_k = \deg X_i + \deg X_j}$$

meet the Jacobi identity (2.1.8) and, thus, they define a nilpotent graded Lie algebra. As soon as

$$[X, fY] = f[X,Y] + (Xf)\cdot Y = f[X,Y] \bmod H_{\deg X + \deg Y - 1}$$

for a C^1-function f, it follows that, at the point $v \in U$, in the standard way, the operator $[\cdot,\cdot]_0$ defines the structure of this algebra on

$$\mathfrak{g}_v = \mathrm{Al}(T_v\mathbb{M}) = \bigoplus_{k=1}^{M} H_k(v)/H_{k-1}(v), \quad H_0(v) = \{0\}.$$

(Here the first equality means that there is an isomorphism between $\mathfrak{g}_v$ and the canonical representation of the algebra $\mathrm{Al}(T_v\mathbb{M})$ independent of the choice of a local basis fields.) Really, for $X_i \in H_{\deg X_i}$ and $X_j \in H_{\deg X_j}$ we define the commutator in $\mathfrak{g}_v$ as

$$\begin{aligned}&[X_i \bmod H_{\deg X_i - 1}, X_j \bmod H_{\deg X_j - 1}]\\ &\quad = \Bigg(\sum_{k\colon \deg X_k = \deg X_i + \deg X_j} c_{ijk}(v) X_k(v)\Bigg) \bmod H_{\deg X_i + \deg X_j - 1}. \qquad (2.1.2)\end{aligned}$$

In view of saying above, we have

$$\begin{aligned}&[X_l \bmod H_{\deg X_l - 1}, [X_i \bmod H_{\deg X_i - 1}, X_j \bmod H_{\deg X_j - 1}]]\\ &= \sum_{k\colon \deg X_k = \deg X_i + \deg X_j} c_{ijk}(v)[X_l \bmod H_{\deg X_l - 1}, X_k(v) \bmod H_{\deg X_i + \deg X_j - 1}]\\ &= \sum_{q\colon \deg X_q = \deg X_l + \deg X_k} \ \sum_{k\colon \deg X_k = \deg X_i + \deg X_j} c_{ijk}(v) c_{lkq}(v) X_q \bmod H_{\deg X_q - 1}.\end{aligned}$$

This equality can be obtained by continuity like in the proof of Theorem 2.1.8. Applying the equalities (2.1.5), we verify that the commutators (2.1.2) meet Jacobi identity.

If we denote $V_k(v) = H_k(v)/H_{k-1}(v)$ then

$$[V_k(v), V_1(v)]_0 \subset V_{k+1}(v) \ \text{ if } \ 1 \leq k < M, \ \text{ and } \ [V_M(v), V_1(v)]_0 = \{0\}. \quad (2.1.3)$$

It is well known [47] that a connected simply connected nilpotent Lie group $\mathbb{G}$ whose Lie algebra $\mathfrak{g}$ is nilpotent and graded, is called the *homogeneous group.*

On a Carnot manifold, instead of (2.1.3) we have a stronger property

$$[V_k(v), V_1(v)]_0 = V_{k+1}(v) \ \text{ if } \ 1 \leq k < M, \ \text{ and } \ [V_M(v), V_1(v)]_0 = \{0\}. \quad (2.1.4)$$

In this case, a connected simply connected nilpotent Lie group $\mathbb{G}$ whose Lie algebra $\mathfrak{g}$ has properties (2.1.4), is called the *stratified homogeneous group* [47] or the *Carnot group* [118].

Notice that all results of Section 2, except of Subsubsections 2.8.1, 2.8.2, and Subsection 3.2, except of Subsubsection 3.2.3, are valid in any Carnot–Carathéodory space. From another side, Condition (4) is necessary only for obtaining results of Subsubsections 2.8.1, 2.8.2 and 3.2.3, Subsection 3.3 and Section 4. The point is that, in the second case, we use the property on connection of any two points of a local Carnot group (see the definition below) by a horizontal (with respect to the local Carnot group) curve that consists of at most L segments of integral lines of horizontal (with respect to the local Carnot group) vector fields. The latter is impossible without Condition (4). By another words, we obtain a part of results only for Carnot manifolds.

Example 2.1.3 (A typical example of a Carnot manifold). For the fixed horizontal subbundle $H\mathbb{M}$, choose C^M-smooth vector fields $X_1, \dots, X_{\dim H_1}$ in U constituting a basis of $H_1 = H\mathbb{M}$ at every $g \in U$ (here $M \in \mathbb{N}$ is a fixed number). Denote by $H_i(g)$ the subset of the tangent space generated by the values of all kinds of commutators of the vector fields $X_1, \dots, X_{\dim H_1}$ up to order $i-1$; we assume here that the commutators of order zero of vector fields are the vector fields themselves. Suppose that $X_1, \dots, X_{\dim H_1}$ meet the Hörmander condition [74] on U; i.e., for each $g \in U$, we have $T_g\mathbb{M} = H_M(g)$ where M is the above-mentioned number. The (*equi*)*regularity* condition [70] is that $\dim H_i(g)$ is independent of the choice of $g \in U$ for every $i \geq 1$; moreover, we assume that $M = \min\{n \in \mathbb{N} \mid T_gU = H_n(g)\}$.

Define by induction vector fields X_i, $i = 1, \dots, N$, constituting a basis at every $g \in U$, as follows: at the first step, to the vector fields $X_1, \dots, X_{\dim H_1}$ that form a basis of H_1, we add vector fields $X_{\dim H_1+1}, \dots, X_{\dim H_2}$ so that the vector fields $X_1, \dots, X_{\dim H_2}$ constitute a basis of H_2; at the $(k-1)$th step, to the vector fields $X_1, \dots, X_{\dim H_{k-1}}$, add vector fields $X_{\dim H_{k-1}+1}, \dots, X_{\dim H_k}$ so that $X_1, \dots, X_{\dim H_k}$ form a basis of H_k. As a result, in $M-1$ steps, we obtain a desired set of vector fields X_i, $i = 1, \dots, N$. The construction of the basis of TU yields $X_k \in C^{M+1-\deg X_k}$. Moreover, we obtain the commutator table of the form 2.1.1 in which $c_{ijk}(g)$ are continuous functions which are identical zero if $\deg X_k > \deg X_i + \deg X_j$.

Example 2.1.4 (A Carnot manifold with C^1-smooth vector fields). Like in the previous example consider the C^M-smooth vector fields $X_1, \dots, X_{\dim H_1} \in H$ in

$U \subset \mathbb{M}$ constituting a basis of $H_1 = H\mathbb{M}$ at every $g \in U$. Suppose that the dimensions of $[H,H](x)$, $[H,[H,H]](x)$,…, $[H,[H,\dots[H,H]\dots]](x)$ do not depend on x (here the last space consists of commutators of $X_1,\dots,X_n$ of the order $M-1$). Choose a basis in $H_2 = \operatorname{span}\{H,[H,H]\}$ by the following way:

$$X_k(v) = \sum_{i,j} a_{ij}^k(v)[X_i,X_j](v) + \sum_l b_l^k(v)X_l,$$

where $a_{ij}^k(v), b_l^k(v) \in C^1$, $i,j,l = 1,\dots,n$, $k = n+1,\dots,\dim H_2$. Similarly, we choose the following basis in $H_{m+1} = \operatorname{span}\{H_m,[H,H_m]\}$, $m = 2,\dots,M-1$:

$$X_k(v) = \sum_{i,j} a_{ij}^k(v)[X_i,X_j](v) + \sum_l b_l^k(v)X_l,$$

where $a_{ij}^k(v), b_l^k(v) \in C^1$, $i = 1,\dots,n$, $j,l = \dim H_{m-1}+1,\dots,\dim H_m$, $k = \dim H_m + 1,\dots,\dim H_{m+1}$.

Remark 2.1.5. Consider a C^2-smooth local diffeomorphism $\eta : U \to \mathbb{R}^N$, $U \subset \mathbb{M}$. Then $\eta_* X_i = D\eta\langle X_i\rangle$ are also C^1-vector fields, $i = 1,\dots,N$. We have the following relations instead of (2.1.1):

$$\eta_*[X_i,X_j](w) = [\eta_*X_i,\eta_*X_j](w) = \sum_{\deg X_k \le \deg X_i + \deg X_j} c_{ijk}(\eta^{-1}(w))\eta_*X_k(w).$$

Denote by $X(w)$ the matrix, the ith column of which consists of the coordinates of $\eta_*X_i(w)$ in the standard basis $\{\partial_j\}_{j=1}^N$. Then the entries of $X(w)$ are C^1-functions. Note that

$$\eta_*[X_i,X_j](w) = X(w)(c_{ij1}(\eta^{-1}(w)),\dots,c_{ijN}(\eta^{-1}(w)))^T.$$

Consequently,

$$(c_{ij1}(\eta^{-1}(w)),\dots,c_{ijN}(\eta^{-1}(w)))^T = (X(w))^{-1}\cdot\eta_*[X_i,X_j](w).$$

From here it follows that all functions $c_{ijk}\circ\eta^{-1}$ are continuous, $i,j,k = 1,\dots,N$. Since η is continuous, then we have that each $c_{ijk} = (c_{ijk}\circ\eta^{-1})\circ\eta$ is also continuous, $i,j,k = 1,\dots,N$.

Similarly, if the vector fields X_i belong to the class $C^{1,\alpha}$, $\alpha \in (0,1]$, $i = 1,\dots,N$, then the functions c_{ijk} belong to the class H^α, $i,j,k = 1,\dots,N$.

Assumption 2.1.6. Throughout the paper, we assume that all the basis vector fields $X_1,\dots,X_N$ are $C^{1,\alpha}$-smooth, and, consequently, their commutators are H^α-continuous, $\alpha \in [0,1]$.

In some parts of this paper, we consider cases when the derivatives of the basis vector fields are H^α-continuous with respect to some nonnegative symmetric function $\mathfrak{d} : U \times U \to \mathbb{R}$, $U \Subset \mathbb{M}$, such that $\mathfrak{d} \ge C\rho$, $0 < C < \infty$, where C depends only on U, and ρ is defined in (1.0.1). Some additional properties of $\mathfrak{d}$ are described below when it is necessary.

Notation 2.1.7. In the paper:

1. The symbol $X \in C^{1,0}$ means that $X \in C^1$, and the symbol $X \in H^0$ means $X \in C$.
2. 0-Hölder continuity means the ordinary continuity. We denote a modulus of continuity of a mapping f by $\omega_f(\cdot)$.
3. We use the distance ρ from (1.0.1) which is equivalent to Riemannian one.

Theorem 2.1.8. *The coefficients*

$$\bar{c}_{ijk} = \begin{cases} c_{ijk}(u) \text{ of } (2.1.1)\,, & \text{if } \deg X_i + \deg X_j = \deg X_k \\ 0, & \text{in other cases} \end{cases}$$

define a graded nilpotent Lie algebra.

Proof. Fix an arbitrary point $u \in \mathbb{M}$ and show that the collection $\{c_{ijk}(u)\}$ with $\deg X_k = \deg X_i + \deg X_j$ defines the structure of a Lie algebra. To do this, we should prove that (see, for example, [125, 19])

1) $\bar{c}_{ijk} = -\bar{c}_{jik}$ for all $i, j, k = 1, \dots, N$;
2) the collection $\{\bar{c}_{ijk}\}$ enjoys Jacobi identity, that is,

$$\sum_k \bar{c}_{ijk}(u)\bar{c}_{kml}(u) + \sum_k \bar{c}_{mik}(u)\bar{c}_{kjl}(u) + \sum_k \bar{c}_{jmk}(u)\bar{c}_{kil}(u) = 0 \qquad (2.1.5)$$

for all $i, j, m, l = 1, \dots, N$.

The property $\bar{c}_{ijk} = -\bar{c}_{jik}$ is evident. Prove that the collection $\{\bar{c}_{ijk}\}$ under consideration enjoys Jacobi identity. Note that, the case of $\deg X_i + \deg X_j + \deg X_m > M$ we have $\bar{c}_{kml}(u) = \bar{c}_{kjl}(u) = \bar{c}_{kil}(u) = 0$ for all $k, l = 1, \dots, N$, thus, such case is trivial.

First step. We may assume without loss of generality that $X_1, \dots, X_N$ are the vector fields on an open set of $\mathbb{R}^N$ (otherwise, consider the local C^2-diffeomorphism η similarly to Remark 2.1.5).

For a vector field $X_i(x) = \sum\limits_{j=1}^{N} \eta_{ij}(x)\partial_j$, consider the mollification $(X_i)_h(x) = \sum\limits_{j=1}^{N} (\eta_{ij} * \psi_h)(x)\partial_j$, $i = 1, \dots, N$, where the function $\psi \in C_0^\infty(B(0,1))$ is such that $\int\limits_{B(0,1)} \psi(x)\, dx = 1$, and $\psi_h(x) = \frac{1}{h^N}\psi\big(\frac{x}{h}\big)$. By the properties of mollification $\eta_{ij} * \psi_h$, $i, j = 1, \dots, N$, we have $(X_i)_h \xrightarrow[h\to 0]{C^1} X_i$ locally in some neighborhood of u. Note that the vector fields $(X_i)_h(v)$, $i = 1, \dots, N$, meet the Jacobi identity, and are a basis of $T_v\mathbb{M}$ for v belonging to some neighborhood of u, if the parameter h is small enough. Consequently, defining the coefficients $\{c^h_{ijk}\}_{i,j,k=1}^N$ by $[(X_i)_h, (X_j)_h] =$

$\sum\limits_{k=1}^{N} c_{ijk}^h (X_k)_h$, we have

$$\sum_k \sum_l c_{ijk}^h c_{kml}^h (X_l)_h + \sum_k \sum_l c_{mik}^h c_{kjl}^h (X_l)_h \\ + \sum_k \sum_l c_{jmk}^h c_{kil}^h (X_l)_h - \sum_l [(X_m)_h c_{ijl}^h](X_l)_h \\ - \sum_l [(X_j)_h c_{mil}^h](X_l)_h - \sum_l [(X_i)_h c_{jml}^h](X_l)_h = 0.$$

Note that, since $(X_i)_h \xrightarrow[h\to 0]{C^1} X_i$ locally, and the vector fields $\{(X_i)_h\}_{i=1}^N$ are linearly independent for all $h \geq 0$ small enough, we have

$$(c_{ij1}^h, \dots, c_{ijN}^h)^T = ((X_1)_h, \dots, (X_N)_h)^{-1}[(X_i)_h, (X_j)_h]$$

and, consequently, $c_{ijk}^h \to c_{ijk}$ as $h \to 0$.

Now, fix $1 \leq l \leq N$. Since the vector fields $\{(X_i)_h\}_{i=1}^N$ are linearly independent for $h > 0$ small enough, we have

$$\sum_k c_{ijk}^h c_{kml}^h + \sum_k c_{mik}^h c_{kjl}^h + \sum_k c_{jmk}^h c_{kil}^h \\ - [(X_m)_h c_{ijl}^h] - [(X_j)_h c_{mil}^h] - [(X_i)_h c_{jml}^h] = 0 \quad (2.1.6)$$

for each fixed l in some neighborhood of u. Fix i, j, m and l such that $\deg X_l = \deg X_i + \deg X_j + \deg X_m$, and consider a test function $\varphi \in C_0^\infty(U)$ on some small compact neighborhood $U \ni u$, $U \Subset \mathbb{M}$. We multiply both sides of (2.1.6) on φ and integrate the result over U. For $h > 0$ small enough, we have

$$0 = \int\limits_U \Big[\sum_k c_{ijk}^h(v) c_{kml}^h(v) + \sum_k c_{mik}^h(v) c_{kjl}^h(v) \\ + \sum_k c_{jmk}^h(v) c_{kil}^h(v)\Big] \cdot \varphi(v)\, dv - \int\limits_U [(X_m)_h c_{ijl}^h](v) \cdot \varphi(v)\, dv \\ - \int\limits_U [(X_j)_h c_{mil}^h](v) \cdot \varphi(v)\, dv - \int\limits_U [(X_i)_h c_{jml}^h](v) \cdot \varphi(v)\, dv.$$

Show that, among the last three integrals, the first one tends to zero as $h \to 0$. Indeed,

$$\int\limits_U [(X_m)_h c_{ijl}^h](v) \cdot \varphi(v)\, dv = -\int\limits_U [(X_m)_h^* \varphi](v) \cdot c_{ijl}^h(v)\, dv,$$

where $(X_i)_h^*$ is an adjoint operator to $(X_i)_h$. The right-hand part integral tends to zero as $h \to 0$ since the value $[(X_m)_h^* \varphi](v)$ is uniformly bounded in U as $h \to 0$, and $(c_{ijl})_h(v) \to 0$ as $h \to 0$ in view of the choice of l. The similar conclusion is true regarding the last two integrals.

Consequently, taking into account the facts that $c^h_{ijk} \to c_{ijk}$ locally, and $c_{ijk} = 0$ for $\deg X_k > \deg X_i + \deg X_j$, and using du Bois–Reymond Lemma for $h \to 0$ we infer

$$\sum_{k:\ \deg X_k \le \deg X_i + \deg X_j} c_{ijk}(v) c_{kml}(v) + \sum_{k:\ \deg X_k \le \deg X_m + \deg X_i} c_{mik}(v) c_{kjl}(v) + \sum_{k:\ \deg X_k \le \deg X_j + \deg X_m} c_{jmk}(v) c_{kil}(v) = 0 \quad (2.1.7)$$

for all $v \in \mathbb{M}$ close enough to u.

Second step. For fixed l, such that $\deg X_l = \deg X_i + \deg X_j + \deg X_m$, investigate the properties of the index k. Consider the first sum. Since $\deg X_l \le \deg X_k + \deg X_m$, we have $\deg X_k \ge \deg X_l - \deg X_m = \deg X_i + \deg X_j$. By (2.1.1), $\deg X_k \le \deg X_i + \deg X_j$, and, consequently, $\deg X_k = \deg X_i + \deg X_j$. The other two cases are considered similarly. Thus, the sum (2.1.7) with $\deg X_l = \deg X_i + \deg X_j + \deg X_m$ and $v = u$ is

$$\sum_{k:\ \deg X_k = \deg X_i + \deg X_j} c_{ijk}(u) c_{kml}(u) + \sum_{k:\ \deg X_k = \deg X_m + \deg X_i} c_{mik}(u) c_{kjl}(u) + \sum_{k:\ \deg X_k = \deg X_j + \deg X_m} c_{jmk}(u) c_{kil}(u) = 0. \quad (2.1.8)$$

The coefficients $\{\bar{c}_{ijk} = c_{ijk}(u)\}_{\deg X_k = \deg X_i + \deg X_j}$ enjoy the Jacobi identity, and, thus, they define the structure of a nilpotent graded Lie algebra. □

We construct the Lie algebra $\mathfrak{g}^u$ from Theorem 2.1.8 as a graded nilpotent Lie algebra of vector fields $\{(\widehat{X}^u_i)'\}_{i=1}^N$ on $\mathbb{R}^N$ such that $(\widehat{X}^u_i)'(0) = e_i$, $i = 1, \dots, N$, and the exponential mapping $(x_1, \dots, x_N) \mapsto \exp\Big(\sum\limits_{i=1}^N x_i (\widehat{X}^u_i)'\Big)(0)$ is identical [124, 19]. By the construction, the relation

$$[(\widehat{X}^u_i)', (\widehat{X}^u_j)'] = \sum_{\deg X_k = \deg X_i + \deg X_j} c_{ijk}(u) (\widehat{X}^u_k)' \quad (2.1.9)$$

holds for the vector fields $\{(\widehat{X}^u_i)'\}_{i=1}^N$ everywhere on $\mathbb{R}^N$.

Notation 2.1.9. We use the following standard notations: for each N-dimensional multi-index $\mu = (\mu_1, \dots, \mu_N)$, its *homogeneous norm* equals $|\mu|_h = \sum\limits_{i=1}^N \mu_i \deg X_i$.

Definition 2.1.10. Consider the ODE

$$\begin{cases} \dot{\gamma}(t) = \sum\limits_{i=1}^N y_i X_i(\gamma(t)) \\ \gamma(0) = x, \ t \in [0, 1], \end{cases}$$

where the vector fields $X_1, \dots, X_N$ are C^1-smooth. Then, for the point $y = \gamma(1)$, we write $y = \exp\Bigl(\sum\limits_{i=1}^{N} y_i X_i\Bigr)(x)$.

Definition 2.1.11. The graded nilpotent Lie (homogeneous) group $\mathbb{G}_u\mathbb{M}$ corresponding to the Lie algebra $\mathfrak{g}^u$, is called the *nilpotent tangent cone* of $\mathbb{M}$ at $u \in \mathbb{M}$. We construct $\mathbb{G}_u\mathbb{M}$ in $\mathbb{R}^N$ as a group algebra [124], that is, the exponential map is identical. By Campbell–Hausdorff formula, the group operation is defined for the basis vector fields $(\widehat{X}_i^u)'$ on $\mathbb{R}^N$, $i = 1, \dots, N$, to be left-invariant [124]: if

$$x = \exp\Bigl(\sum_{i=1}^{N} x_i (\widehat{X}_i^u)'\Bigr),\ y = \exp\Bigl(\sum_{i=1}^{N} y_i (\widehat{X}_i^u)'\Bigr) \text{ then } x \cdot y = z = \exp\Bigl(\sum_{i=1}^{N} z_i (\widehat{X}_i^u)'\Bigr),$$

where

$$\begin{aligned}
&z_i = x_i + y_i, \quad \deg X_i = 1,\\
&z_i = x_i + y_i + \sum_{\substack{|e_l+e_j|_h=2,\\ l<j}} F^i_{e_l,e_j}(u)(x_l y_j - y_l x_j), \quad \deg X_i = 2,\\
&z_i = x_i + y_i + \sum_{\substack{|\mu+\beta|_h=k,\\ \mu>0,\,\beta>0}} F^i_{\mu,\beta}(u) x^\mu \cdot y^\beta \qquad\qquad (2.1.10)\\
&\quad = x_i + y_i + \sum_{\substack{|\mu+e_l+\beta+e_j|_h=k,\\ l<j}} G^i_{\mu,\beta,l,j}(u) x^\mu y^\beta (x_l y_j - y_l x_j), \quad \deg X_i = k,
\end{aligned}$$

and the coefficients $F^i_{\mu,\beta}(u)$ depend on the constants $c_{ijk}(u)$ of (2.1.9).

(Comment: it is just formula (2.1.9) of the paper).

Theorem 2.1.8 implies

Theorem 2.1.12 ([47])**.** *If $\{\frac{\partial}{\partial x_l}\}_{l=1}^N$ is the standard basis in $\mathbb{R}^N$ then the jth coordinate of a vector field $(\widehat{X}_i^u)'(x) = \sum\limits_{j=1}^{N} z_i^j(u,x)\frac{\partial}{\partial x_j}$ can be written as*

$$z_i^j(u,x) = \begin{cases} \delta_{ij} & \text{if} \quad j \le \dim H_{\deg X_i},\\ \sum\limits_{\substack{|\mu+e_i|_h=\deg X_j,\\ \mu>0}} F^j_{\mu,e_i}(u)x^\mu & \text{if} \quad j > \dim H_{\deg X_i}. \end{cases} \qquad (2.1.11)$$

Definition 2.1.13. Suppose that $u \in \mathbb{M}$ and $(v_1, \dots, v_N) \in B_E(0,r)$, where $B_E(0,r)$ is a Euclidean ball in $\mathbb{R}^N$. Define a mapping $\theta_u(v_1, \dots, v_N) : B_E(0,r) \to \mathbb{M}$ as follows:

$$\theta_u(v_1, \dots, v_N) = \exp\Bigl(\sum_{i=1}^{N} v_i X_i\Bigr)(u).$$

It is known, that θ_u is a C^1-diffeomorphism if $0 < r \leq r_u$ for some $r_u > 0$. The collection $\{v_i\}_{i=1}^N$ is called *the normal coordinates* or *the coordinates of the* 1^{st} *kind* (*with respect to* $u \in \mathbb{M}$) of the point $v = \theta_u(v_1, \dots, v_N)$.

Assumption 2.1.14. The compactly embedded neighborhood $\mathcal{U} \Subset \mathbb{M}$ under consideration is such that $\theta_u(B_E(0, r_u)) \supset \mathcal{U}$ for all $u \in \mathcal{U}$.

By means of the exponential map we can push-forward the vector fields $(\widehat{X}_i^u)'$ onto $\mathcal{U}$ for obtaining the vector fields $\widehat{X}_i^u = (\theta_u)_*(\widehat{X}_i^u)'$ where

$$(\theta_u)_*\langle Y\rangle(\theta_u(x)) = D\theta_u(x)\langle Y\rangle,$$

$Y \in T_x\mathbb{R}^N$. Note that $\widehat{X}_i^u(u) = X_i(u)$. Indeed, on the one hand, by the definition, we have $(\theta_u)_*^{-1}\langle X_i\rangle(0) = e_i$. On the other hand, Theorem 2.1.12 implies $(\widehat{X}_i^u)'(0) = e_i$. Thus $\widehat{X}_i^u(u) = X_i(u)$.

Theorem 2.1.15. *The vector fields* $\widehat{X}_i^u$, $i = 1, \dots, N$, *are locally* H^α*-continuous on* u.

The proofs of this theorem and of many other assertions concerning smoothness use often the following lemma (see its proof in Section 5).

Theorem 2.1.16. *Consider the ODE*

$$\begin{cases} \frac{dy}{dt} = f(y, v, u), \\ y(0) = 0 \end{cases} \tag{2.1.12}$$

where $t \in [0, 1]$, $y, v, u \in W \subset \mathbb{R}^N$ *and* $\operatorname{Lip}_y(f) = L < 1$.

1) *If the mapping* $f(y, v, u) = f(y, u) \in C^1(y) \cap H^\alpha(u)$ *then the solution* $y(t, u) \in H^\alpha(u)$ *locally.*
2) *If* $f(y, v, u) \in C^{1,\alpha}(y, u) \cap C^1(v)$ *and* $\frac{\partial f}{\partial v} \in C^{1,\alpha}(y, u)$ *then* $\frac{dy(t,v,u)}{dv} \in H^\alpha(u)$ *locally.*

Notation 2.1.17. If a mapping η is α-Hölder with respect to $\mathfrak{d}$ then we write $\eta \in H_{\mathfrak{d}}^\alpha$.

Remark 2.1.18. The following statements are proved similarly to Theorem 2.1.16.

1) If in (2.1.12) the mapping $f(y, v, u) = f(y, u) \in C^1(y) \cap H_{\mathfrak{d}}^\alpha(u)$, where $\mathfrak{d}$ is a nonnegative symmetric function defined on $\mathcal{U} \times \mathcal{U}$, $\mathcal{U} \Subset \mathbb{M}$, such that $\mathfrak{d} \geq C\rho$, $0 < C < \infty$, where C depends only on $\mathcal{U}$, then the solution $y(t, u) \in H_{\mathfrak{d}}^\alpha(u)$ locally.
2) If $f(y, v, u) \in C^1(y, u) \cap C^1(v)$, its derivatives in y and in u belong to $H_{\mathfrak{d}}^\alpha(y, u)$ locally, $\frac{\partial f}{\partial v} \in C^1(y, u)$, and the derivatives of $\frac{\partial f}{\partial v}$ in y and in u belong to $H_{\mathfrak{d}}^\alpha(y, u)$ locally then $\frac{dy(t,v,u)}{dv} \in H_{\mathfrak{d}}^\alpha(u)$ locally.

Notation 2.1.19. Hereinafter, we denote a nonnegative symmetric function defined on $\mathcal{U} \times \mathcal{U}$, $\mathcal{U} \Subset \mathbb{M}$, possessing properties from item 1 of Remark 2.1.18, by $\mathfrak{d}$.

Proof of Theorem 2.1.15. *First step.* Fix $\mathcal{U} \Subset \mathbb{M}$ small enough. Recall that the vector fields $\{(\widehat{X}_i^u)'\}_{i=1}^N$, $u \in \mathcal{U}$, are smooth on $\theta_u^{-1}(\mathcal{U}) \Subset \mathbb{R}^N$. By (2.1.9), for $v \in \mathcal{U}$ we have the table of commutators

$$[(\widehat{X}_i^u)', (\widehat{X}_j^u)'](v) = \sum_{\deg X_k = \deg X_i + \deg X_j} c_{ijk}(u)(\widehat{X}_k^u)'(v).$$

By means of Assumption 2.1.6 and Definition 2.1.1, the functions $c_{ijk}(u)$ from (2.1.1) are H^α-continuous.

If $X = \sum\limits_{i=1}^N x_i(\widehat{X}_i^u)'$ and $Y = \sum\limits_{i=1}^N y_i(\widehat{X}_i^u)'$ then by Campbell–Hausdorff formula we have $\exp tY \circ \exp tX(g) = \exp Z(t)(g)$ where $Z(t) = tZ_1 + t^2 Z_2 + \dots + t^M Z_M$ and Z_1, $Z_2, \dots$ are some vector fields independent of t. Dynkin formula (see, for instance, [124]) for calculating Z_l, $1 \leq l \leq M$, gives

$$Z_l = \frac{1}{n} \sum_{k=1}^{l} \frac{(-1)^{k-1}}{k} \sum_{(p)(q)} \frac{(\operatorname{ad} Y)^{q_k} (\operatorname{ad} X)^{p_k} \dots (\operatorname{ad} Y)^{q_1} (\operatorname{ad} X)^{p_1 - 1} X}{p_1! q_1! \dots p_k! q_k!}$$
$$= \sum_{(p)(q)} C_{(p)(q)} (\operatorname{ad} Y)^{q_k} (\operatorname{ad} X)^{p_k} \dots (\operatorname{ad} Y)^{q_1} (\operatorname{ad} X)^{p_1 - 1} X,$$

where $C_{(p)(q)} = \text{const}$, $(p) = (p_1, \dots, p_k)$, $(q) = (q_1, \dots, q_k)$. We sum over all natural p_1, q_1, $\dots p_k$, q_k, such that $p_i + q_i > 0$, $p_1 + q_1 + \dots + p_k + q_k = l$, and $(\operatorname{ad} A)B = [A, B]$, $(\operatorname{ad} A)^0 B = B$. According to (2.1.9), each summand can be represented as a sum

$$Z_l(v) = \sum_{j=1}^N d_{j,l}(u, x, y)(\widehat{X}_j^u)'(v),$$

where $d_{j,l}(u, x, y)$ are polynomial functions of $x = (x_1, \dots, x_N)$, $y = (y_1, \dots, y_N)$ coefficients of which are polynomial functions of $\{\bar{c}_{imk}(u)\}_{i,m,k}$ and, consequently, are Hölder in u. More exactly (see, for example, [47]),

$$\sum_{l=2}^M Z_l = \sum_{l=2}^M \sum_{j=1}^N d_{j,l}(u, x, y)(\widehat{X}_j^u)' = \sum_{j=1}^N \left[\sum_{l=2}^M \sum_{\substack{|\mu+\beta|_h = l, \\ \mu > 0, \beta > 0}} F_{\mu,\beta}^j(u) x^\mu \cdot y^\beta \right] (\widehat{X}_j^u)'.$$

Consequently,

$$d_{j,l}(u, x, y) = \sum_{l=2}^M \sum_{\substack{|\mu+\beta|_h = l, \\ \mu > 0, \beta > 0}} F_{\mu,\beta}^j(u) x^\mu \cdot y^\beta.$$

Hence, $F_{\mu,\beta}^j(u)$ are H^α-continuous in u, and $(\widehat{X}_i^u)'$ are also H^α-continuous in u (see (2.1.11)).

Second step. Consider the following Cauchy problem:

$$\begin{cases} \frac{d\Phi(t,u,\xi)}{dt} = \sum\limits_{i=1}^{N} \xi_i X_i(\Phi), \\ \Phi(0,u,\xi) = u, \end{cases} \tag{2.1.13}$$

where $\xi = (\xi_1, \dots, \xi_N)$. Note that $\Phi(t,u,\xi) = \exp\Bigl(\sum\limits_{i=1}^{N} t\xi_i X_i\Bigr)(u)$. We can assume without loss of generality, that $\mathbb{M} = \mathbb{R}^N$. Setting $\Psi = \Phi - u$, rewrite this Cauchy problem the following way:

$$\begin{cases} \frac{d\Psi(t,u,\xi)}{dt} = \sum\limits_{i=1}^{N} \xi_i X_i(\Psi + u), \\ \Psi(0,u,\xi) = 0. \end{cases}$$

If Assumption 2.1.6 holds then the mapping $f(\Psi,\xi,u) = \sum\limits_{i=1}^{N} \xi_i X_i(\Psi+u)$ is $C^{1,\alpha}$-smooth in Ψ and u, and it is C^1-smooth in ξ. Moreover, $\frac{\partial f}{\partial \xi} \in C^{1,\alpha}(\Psi,u)$. Note that, from the definition it follows $\theta_u(\xi) = \Phi(1,u,\xi) = \Psi(1,u,\xi) + u$.

By theorem 2.1.16 (with Ψ instead of y and ξ instead of v) on smooth dependence of ODE solution on parameters (see Section 5 for details), it is easy to see, that the differential $D\theta_u(y)$ is H^α-continuous in u locally.

Since $\widehat{X}_i^u(x) = (\theta_u)_*(\widehat{X}_i^u)'(y)$, $x = \theta_u(y)$, the proposition follows from results of the 1[ST] and 2[ND] steps. □

Remark 2.1.20. If the derivatives of X_i, $i = 1, \dots, N$, are locally Hölder with respect to $\mathfrak{d}$, then $\widehat{X}_i^u$, $i = 1, \dots, N$, are locally Hölder on u with respect to $\mathfrak{d}$.

Definition 2.1.21. To the Lie algebra $\{\widehat{X}_i^u\}_{i=1}^N$ at $u \in \mathbb{M}$, it corresponds a *local homogeneous group* $\mathcal{G}^u\mathbb{M}$. Define it in such a way that the mapping θ_u is a *local group isomorphism* of some neighborhood of the unity of the group $\mathbb{G}_u\mathbb{M}$ and $\mathcal{G}^u\mathbb{M}$.

The canonical Riemannian structure on $\mathcal{G}^u\mathbb{M}$ is defined by scalar product at the unit of $\mathcal{G}^u\mathbb{M}$ coinciding with those in $T_u\mathbb{M}$. The canonical Riemannian structure on the nilpotent tangent cone $\mathbb{G}_u\mathbb{M}$ is defined by such a way that the local group isomorphism θ_u is an isometry.

Assumption 2.1.22. Hereinafter in the paper, we assume that the neighborhood $\mathcal{U}$ under consideration is such that $\mathcal{U} \subset \mathcal{G}^u\mathbb{M}$ for all $u \in \mathcal{U}$. Consider the mapping $\theta_g^u(x_1, \dots, x_N) = \exp\Bigl(\sum\limits_{i=1}^{N} x_i \widehat{X}_i^u\Bigr)(g)$. It is well defined for $(x_1, \dots, x_N) \in B_E(0, r_{g,u})$. We suppose that $\mathcal{U} \subset \theta_g^u(B_E(0, r_{g,u}))$ for all $u, g \in \mathcal{U}$.

Remark 2.1.23. Recall that the vector fields $\widehat{X}_i^u$, $i = 1, \dots, N$, are locally H^α-continuous on $\mathbb{M}$, $\alpha \in [0,1]$. The exponential mapping $\exp\Bigl(\sum\limits_{i=1}^{N} a_i \widehat{X}_i^u\Bigr)(g)$ is not defined correctly for such fields. Therefore, in view of smoothness of $(\theta_u^{-1})_*(\widehat{X}_i^u)$, $i = 1, \dots, N$, we define the point

$$a = \exp\biggl(\sum_{i=1}^{N} a_i \widehat{X}_i^u\biggr)(g)$$

according to Definition 2.1.21: first, we obtain a point

$$a_u = \exp\biggl(\sum_{i=1}^{N} a_i \cdot (\theta_u^{-1})_*\langle\widehat{X}_i^u\rangle\biggr)(\theta_u^{-1}(g)) = \exp\biggl(\sum_{i=1}^{N} a_i \cdot (\widehat{X}_i^u)'\biggr)(\theta_u^{-1}(g)),$$

and then we define $a = \theta_u(a_u)$. Moreover, we similarly define the whole curve corresponding to this exponential mapping. Suppose that

$$\begin{cases} \dot{\gamma_u}(t) = \sum\limits_{i=1}^{N} a_i \cdot (\theta_u^{-1})_*\langle\widehat{X}_i^u\rangle(\gamma_u(t)) \\ \gamma_u(0) = \theta_u^{-1}(g). \end{cases}$$

Then, for the curve $\gamma(t) = \theta_u(\gamma_u(t))$, we have

$$\begin{cases} \dot{\gamma}(t) = \sum\limits_{i=1}^{N} a_i \widehat{X}_i^u(\gamma(t)) \\ \gamma(0) = g. \end{cases}$$

In particular, we have:

1) The exponential mapping $\widehat{\theta}_u(v_1, \dots, v_n) = \exp\Bigl(\sum\limits_{i=1}^{N} v_i \widehat{X}_i^u\Bigr)(u)$ is defined as

$$\theta_u\biggl[\exp\biggl(\sum_{i=1}^{N} v_i (\widehat{X}_i^u)'\biggr)(0)\biggr];$$

and the mapping $\widehat{\theta}_u^w(v_1, \dots, v_n) = \exp\Bigl(\sum\limits_{i=1}^{N} v_i \widehat{X}_i^u\Bigr)(w)$ is defined as

$$\theta_u\biggl[\exp\biggl(\sum_{i=1}^{N} v_i (\widehat{X}_i^u)'\biggr)(\theta_u^{-1}(w))\biggr].$$

2) The inverse mapping $\exp^{-1}$ is also defined by the unique way for vector fields $\{\widehat{X}_i^u\}_{i=1}^N$ since it is defined by the unique way for $\{(\widehat{X}_i^u)'\}_{i=1}^N$.

3) The group operation is defined by the following way: if $x = \exp\Bigl(\sum\limits_{i=1}^{N} x_i \widehat{X}_i^u\Bigr)$, $y = \exp\Bigl(\sum\limits_{i=1}^{N} y_i \widehat{X}_i^u\Bigr)$ then $x \cdot y = \exp\Bigl(\sum\limits_{i=1}^{N} y_i \widehat{X}_i^u\Bigr) \circ \exp\Bigl(\sum\limits_{i=1}^{N} x_i \widehat{X}_i^u\Bigr) = \exp\Bigl(\sum\limits_{i=1}^{N} z_i \widehat{X}_i^u\Bigr)$ where z_i are taken from Definition 2.1.11.

4) Using the normal coordinates $\widehat{\theta}_u^{-1}$, define the action of the *dilation group* δ_ε^u in the local homogeneous group $\mathcal{G}^u\mathbb{M}$: to an element $x = \exp\Big(\sum\limits_{i=1}^{N} x_i \widehat{X}_i^u\Big)(u)$, assign $\delta_\varepsilon^u x = \exp\Big(\sum\limits_{i=1}^{N} x_i \varepsilon^{\deg X_i} \widehat{X}_i^u\Big)(u)$ in the cases where the right-hand side makes sense.

Property 2.1.24. It is easy to see using Property 2.1.23 that the mapping δ_ε^u is differentiable. Moreover, for each vector field $\widehat{X}_i^u$, $i = 1, \dots, N$, we have $(\delta_\varepsilon^u)_* \widehat{X}_i^u(g) = \varepsilon^{\deg X_i} \widehat{X}_i^u(\delta_\varepsilon^u g)$.

This property comes from the "canonical" homogeneous group $T_u\mathbb{M}$ [47].

Lemma 2.1.25 ([139]). *Suppose that* $u \in \mathcal{U}$*. The equality*

$$\sum_{i=1}^{j} \sum_{\substack{|\mu+e_i|_h=\deg X_j,\\ |\mu+e_i|=l,\, \mu>0}} x_i F^j_{\mu,e_i}(u) x^\mu = 0, \quad x = (x_1, \dots, x_N) \in \mathbb{R}^N,$$

holds for all $\deg X_j \geq 2$, $l = 2, \dots, \deg X_j$.

Proof. Consider a vector field $X = \sum\limits_{i=1}^{N} x_i (\widehat{X}_i{}^u)'$. It is known that $\exp rsX \circ \exp rtX(g) = \exp r(s+t)X(g)$. Therefore, by (2.1.10), we have

$$\sum_{\substack{|\mu+\beta|_h=\deg X_j,\\ \mu>0,\, \beta>0}} r^{|\mu+\beta|} F^j_{\mu,\beta}(g) s^{|\mu|} x^\mu \cdot t^{|\beta|} x^\beta = 0 \tag{2.1.14}$$

for all fixed s and t, $\deg X_j \geq 2$. It follows that the coefficients at all powers of r vanish. In particular, if $|\mu+\beta| = l \geq 2$ then

$$\sum_{\substack{|\mu+\beta|_h=\deg X_j,\\ \mu>0,\, \beta>0,\, |\mu+\beta|=l}} F^j_{\mu,\beta}(g) s^{|\mu|} x^\mu \cdot t^{|\beta|} x^\beta = 0$$

for all t and arbitrary fixed x and s. Consequently, if $|\beta| = 1$ then we infer

$$P(s) = \sum_{l=2}^{\deg X_j} s^{l-1} \sum_{i=1}^{j} \sum_{\substack{|\mu+e_i|_h=\deg X_j,\\ |\mu+e_i|=l,\, \mu>0}} x_i F^j_{\mu,e_i}(g) x^\mu \equiv 0,$$

where s is an arbitrarily small parameter. Therefore, all coefficients of the polynomial $P(s)$ at the powers of s vanish. The lemma follows. ☐

Lemma 2.1.26 ([139]). *Let* $u \in \mathcal{U}$ *be an arbitrary point. Then*

$$\exp\Big(\sum_{i=1}^{N} a_i X_i\Big)(u) = \exp\Big(\sum_{i=1}^{N} a_i \widehat{X}_i^u\Big)(u)$$

for all $|a_i| < r_u$, $i = 1, \dots, N$.

Proof. Let $a = (a_1, \ldots, a_N) \in \mathbb{R}^N$. Relations (2.1.10) and (2.1.14) imply the following equalities:

$$\begin{aligned}
&\sum_{i=1}^{N} ta_i(\widehat{X}_i^u)'(ta_1, \ldots, ta_N) \\
&\quad = \sum_{i=1}^{N} ta_i\Big(\frac{\partial}{\partial x_i} + \sum_{j>\dim H_{\deg X_i}} \sum_{\substack{|\mu+e_i|_h=\deg X_j,\\ \mu>0}} F^j_{\mu,e_i}(u)(ta)^\mu \frac{\partial}{\partial x_j}\Big) \\
&\quad = \sum_{i=1}^{N} ta_i \frac{\partial}{\partial x_i} + \sum_{j=2}^{N} \sum_{l=1}^{N+1} t^l \Big(\sum_{i=1}^{j} a_i \sum_{\substack{|\mu+e_i|_h=\deg X_j,\\ |\mu+e_i|=l,\, \mu>0}} F^j_{\mu,e_i}(u)a^\mu\Big)\frac{\partial}{\partial x_j} \\
&\quad = t\sum_{i=1}^{N} a_i \frac{\partial}{\partial x_i}.
\end{aligned}$$

By the definition of the exponential map, we infer

$$\mathbb{R}^N \ni (a_1, \ldots, a_N) = \sum_{i=1}^{N} a_i(\widehat{X}_i^u)'(a_1, \ldots, a_N) = \exp\Big(\sum_{i=1}^{N} a_i(\widehat{X}_i^u)'\Big)$$

since the exponential map is identical. From this, it follows immediately that

$$\begin{aligned}
\theta_u(a_1, \ldots, a_N) &= \theta_u\Big(\sum_{i=1}^{N} a_i(\widehat{X}_i^u)'(a_1, \ldots, a_N)\Big) \\
&= \theta_u\Big(\exp\Big(\sum_{i=1}^{N} a_i(\widehat{X}_i^u)'\Big)\Big) = \exp\Big(\sum_{i=1}^{N} a_i\widehat{X}_i^u\Big)
\end{aligned}$$

according to Remark 2.1.23. □

Definition 2.1.27. Suppose that $\mathbb{M}$ is a Carnot–Carathéodory space, and $u \in \mathbb{M}$. For $a, p \in \mathcal{G}^u\mathbb{M}$, where

$$a = \exp\Big(\sum_{i=1}^{N} a_i\widehat{X}_i^u\Big)(p),$$

we define the value $d_\infty^u(a,p) = \max\limits_i\{|a_i|^{\frac{1}{\deg X_i}}\}$ on $\mathcal{G}^u\mathbb{M}$.

The following properties comes from those on the "canonical" homogeneous group $\mathbb{G}_u\mathbb{M}$ [47].

Property 2.1.28. It is easy to see that $d_\infty^u(x,y)$, $u \in \mathcal{U}$, is a quasimetric on $\mathcal{G}^u\mathbb{M}$ meeting the following properties:

1) $d_\infty^u(x,y) \geq 0$, $d_\infty^u(x,y) = 0$ if and only if $x = y$;
2) $d_\infty^u(u,v) = d_\infty^u(v,u)$;
3) the value $d_\infty^u(x,y)$ is continuous with respect to each of its variables;

4) there exists a constant $Q_\triangle = Q_\triangle(\mathcal{U})$ such that the inequality

$$d_\infty^u(x,y) \le Q_\triangle(d_\infty^u(x,z) + d_\infty^u(z,y))$$

holds for every triple of points $x,y,z \in \mathcal{U} \cap \mathcal{G}^u\mathbb{M}$, i.e., d_∞^u is a quasimetric.

Property 2.1.29. Let

$$w_\varepsilon = \exp\biggl(\sum_{i=1}^N \varepsilon^{\deg X_i} w_i \widehat{X}_i^u\biggr)(v) \quad \text{and} \quad g_\varepsilon = \exp\biggl(\sum_{i=1}^N \varepsilon^{\deg X_i} g_i \widehat{X}_i^u\biggr)(v).$$

Then $d_\infty^u(w_\varepsilon, g_\varepsilon) = \varepsilon d_\infty^u(w_1, g_1)$.

Notation 2.1.30. By $\operatorname{Box}^u(x,r)$ we denote a set $\{y \in \mathbb{M} :\ d_\infty^u(x,y) < r\}$.

Property 2.1.31. We have $\delta_\varepsilon^u(\operatorname{Box}^u(u,r)) = \operatorname{Box}^u(u,\varepsilon r)$.

2.2. Gromov's theorem on the nilpotentization of vector fields and estimate of the diameter of a box

Definition 2.2.1. Suppose that $\mathbb{M}$ is a Carnot–Carathéodory space, and let $\mathcal{U} \subset \mathbb{M}$ be as in Assumption 2.1.14. Given

$$v = \exp\biggl(\sum_{i=1}^N v_i X_i\biggr)(u)$$

$u, v \in \mathcal{U}$, define the value $d_\infty(u,v) = \max\limits_i\{|v_i|^{\frac{1}{\deg X_i}}\}$. By $\operatorname{Box}(x,r)$ we denote a set $\{y \in \mathbb{M} :\ d_\infty(x,y) < r\}$, $r \le r_x$.

Proposition 2.2.2. *The relations*

$$\rho(u,v) \le d_\infty(u,v) \le \rho(u,v)^{\frac{1}{M}}$$

hold for all $u, v \in \mathcal{U} \Subset \mathbb{M}$.

Remark 2.2.3. d_∞ is one of particular cases of $\mathfrak{d}$.

Definition 2.2.4. Using the normal coordinates θ_u^{-1}, define the action of the *dilation group* Δ_ε^u in a neighborhood of a point $u \in \mathbb{M}$: to an element $x = \exp\Bigl(\sum\limits_{i=1}^N x_i X_i\Bigr)(u)$, assign $\Delta_\varepsilon^u x = \exp\Bigl(\sum\limits_{i=1}^N x_i \varepsilon^{\deg X_i} X_i\Bigr)(u)$ in the cases where the right-hand side makes sense.

Property 2.2.5. By Lemma 2.1.26 we have $\Delta_\varepsilon^u x = \delta_\varepsilon^u x$.

Property 2.2.6. By Lemma 2.1.26 we have $\operatorname{Box}^u(u,r) = \operatorname{Box}(u,r)$.

Property 2.2.7. We have $\Delta_\varepsilon^u(\operatorname{Box}(u,r)) = \operatorname{Box}(u,\varepsilon r)$, $r \in (0, r_u]$, $\varepsilon \le 1$.

Property 2.2.8. The value d_∞ has the following properties:

1) $d_\infty(u,v) \geq 0$, $d_\infty(u,v) = 0$ if and only if $u = v$;
2) $d_\infty(u,v) = d_\infty(v,u)$;
3) the value $d_\infty(u,v)$ is continuous with respect to each of its variables;
4) there exists a constant $Q = Q(\mathcal{U})$ such that the inequality

$$d_\infty(u,v) \leq Q(d_\infty(u,w) + d_\infty(w,v))$$

holds for every triple of points $u, w, v \in \mathcal{U}$, i.e., d_∞ is a quasimetric.

Proof. The proof of Properties 1–3 is based on known properties of solutions to ODE's. We prove the generalized triangle inequality at the end of the current subsection (see Corollary 2.2.17). □

Theorem 2.2.9. *Let $X_j \in C^1$ and $M = 2$. Fix $u \in \mathbb{M}$. If $d_\infty(u,w) = C\varepsilon$, then*

$$\widehat{X}_j^u(w) = \sum_{k:\, \deg X_k \leq \deg X_j} [\delta_{kj} + O(\varepsilon)]X_k + \sum_{k:\, \deg X_k > \deg X_j} o(\varepsilon^{\deg X_k - \deg X_j})X_k(w),$$

$j = 1, \dots, N$. All $o(\cdot)$ and $O(\cdot)$ are uniform in u belonging to some compact subset of $\mathcal{U}$.

Proof. First step. Fix $u \in \mathbb{M}$ and put $z_j^k(s) = z_j^k(u,s)$, $k, j = 1, \dots, N$, $s \in \mathbb{R}^N$, from relations (2.1.11). Applying the mapping θ_u^{-1} to each vector field $\widehat{X}_j^u$, $j = 1, \dots, N$, we deduce

$$D\theta_u^{-1}\langle \widehat{X}_j^u \rangle(s) = \sum_{k=1}^N z_j^k(s) \frac{\partial}{\partial x_k},$$

where $\left\{\frac{\partial}{\partial x_k}\right\}_{k=1}^N$ is the collection of the vectors of the standard basis in $\mathbb{R}^N = T_x\mathbb{R}^N$, and by (2.1.11) we have

$$z_j^k(s) = \delta_{kj} + \sum_{|\mu|_h = \deg X_k - \deg X_j > 0} F^k_{\mu, e_j}(u) s^\mu.$$

Note that, here $|s^\mu| = O(\varepsilon^{\deg X_k - \deg X_j})$, since

$$d_\infty(0,s) = d_\infty(\theta_u^{-1}(u), s) = d_\infty^u(u, \theta_u(s)) = O(\varepsilon).$$

Then, for $s = (s_1, \dots, s_N)$, we obtain

$$\begin{aligned}\widehat{X}_j^u(\theta_u(s)) &= \sum_{k=1}^N z_j^k(s) D\theta_u(s)\Big\langle \frac{\partial}{\partial x_k} \Big\rangle \\ &= \sum_{k=1}^N z_j^k(s)\bigg(X_k(\theta_u(s)) + \frac{1}{2}\bigg[X_k, \sum_{l=1}^N s_l X_l\bigg](\theta_u(s))\bigg),\end{aligned}$$

since $D\theta_u(s)\langle\frac{\partial}{\partial x_k}\rangle = X_k(\theta_u(s)) + \frac{1}{2}\Big[X_k, \sum\limits_{l=1}^{N} s_l X_l\Big](\theta_u(s))$. The latter is doubtless true for C^∞-vector fields since in this case Baker-Campbell-Hausdorff formula does all job:

$$\begin{aligned}\theta_u(s+re_k) &= \exp\Big(\sum_{l=1}^{N} s_l X_l + rX_k\Big)(u)\\ &= \exp\Big(\sum_{l=1}^{N} s_l X_l + rX_k\Big)\circ\exp\Big(-\sum_{l=1}^{N} s_l X_l\Big)\circ\exp\Big(\sum_{l=1}^{N} s_l X_l\Big)(u)\\ &= \exp\Big(rX_k + \frac{r}{2}\Big[X_k, \sum_{l=1}^{N} s_l X_l\Big] + o(r)\Big)(\theta_u(s)) \quad \text{as } r\to 0.\end{aligned}$$

It follows directly that, for C^∞-vector fields, we have

$$D\theta_u(s)\Big\langle\frac{\partial}{\partial x_k}\Big\rangle = \frac{\partial}{\partial r}\theta_u(s+re_k)\Big|_{r=0} = X_k(\theta_u(s)) + \frac{1}{2}\Big[X_k, \sum_{l=1}^{N} s_l X_l\Big](\theta_u(s)). \quad (2.2.1)$$

Otherwise, in some local coordinate system around the point u we can approximate the given C^1-vector fields X_l by C^∞-vector fields $X_l^{(q)}$, $q\in\mathbb{N}$, $l=1,\dots,N$, in C^1-topology (see the proof of Theorem 2.1.8 for details). Then, in some neighborhood of 0 small enough, we have correctly defined the mapping $\theta_u^{(q)}(s)$ constructed according to Definition 2.1.13 by means of vector fields $X_l^{(q)}$, $l=1,\dots,N$ (note that the mentioned neighborhood of 0 can be chosen the same for all $q\in\mathbb{N}$). Moreover, $\theta_u^{(q)}(s)$ converges to $\theta_u(s)$ uniformly in some neighborhood of 0 as $q\to\infty$, and $D\theta_u^{(q)}(s)\Big\langle\frac{\partial}{\partial x_k}\Big\rangle = \frac{\partial}{\partial r}\theta_u^{(q)}(s+re_k)\Big|_{r=0}$ converges to $X_k(\theta_u(s)) + \frac{1}{2}\Big[X_k, \sum\limits_{l=1}^{N} s_l X_l\Big](\theta_u(s))$ uniformly in the same neighborhood of 0. By the well-known classical result on the differentiability of a limit mapping, the equality (2.2.1) follows for the given C^1-vector fields X_l, $l=1,\dots,N$.

In view of the properties of the point s, we get $|s_l| = O(\varepsilon^{\deg X_l})$, $l=1,\dots,N$. Moreover, taking into account the table of commutators (2.1.1) from the definition of the Carnot–Carathéodory space, we have

$$\Big[X_k, \sum_{l=1}^{N} s_l X_l\Big](\theta_u(s)) = \sum_{l=1}^{N}\ \sum_{m:\deg X_m\le\deg X_k+\deg X_l} c_{klm}(\theta_u(s))X_m(\theta_u(s)).$$

Consequently,

$$\begin{aligned}\widehat{X}_j^u(\theta_u(s)) &= \sum_{k=1}^{N} z_j^k(s)X_k(\theta_u(s))\\ &\quad+\frac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{N}\ \sum_{m:\,\deg X_m\le\deg X_k+\deg X_l} z_j^k(s)s_l c_{klm}(\theta_u(s))X_m(\theta_u(s))\end{aligned}$$

$$= \sum_{k=1}^{N} \Big[z_j^k(s) + \frac{1}{2} \sum_{m,l\colon \deg X_k \le \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlk}(\theta_u(s)) \Big] X_k(\theta_u(s)).$$

Represent the last sum as

$$\begin{aligned} &\sum_{k\colon \deg X_k < \deg X_j} \Big[z_j^k(s) + \frac{1}{2} \sum_{m,l\colon \deg X_k \le \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlk}(\theta_u(s)) \Big] X_k(\theta_u(s)) \\ + &\sum_{k\colon \deg X_k = \deg X_j} \Big[z_j^k(s) + \frac{1}{2} \sum_{m,l\colon \deg X_k \le \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlk}(\theta_u(s)) \Big] X_k(\theta_u(s)) \\ + &\sum_{k\colon \deg X_k > \deg X_j} \Big[z_j^k(s) + \frac{1}{2} \sum_{m,l\colon \deg X_k \le \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlk}(\theta_u(s)) \Big] X_k(\theta_u(s)). \end{aligned} \tag{2.2.2}$$

Note that, we have $z_j^k(s) = 0$ if $\deg X_k < \deg X_j$. Next, if $\deg X_k < \deg X_j$ and $\deg X_k = \deg X_m + \deg X_l$, we have $\deg X_m < \deg X_j$ and $z_j^m(s) = 0$. Thus, the first sum equals

$$\sum_{k\colon \deg X_k < \deg X_j} \Big[\frac{1}{2} \sum_{m,l\colon \deg X_k < \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlk}(\theta_u(s)) \Big] X_k(\theta_u(s)).$$

Similarly, for the second sum we have $z_j^k(s) = \delta_{kj}$, and if $\deg X_j = \deg X_m + \deg X_l$ then $z_j^m(s) = 0$ since this relation implies $\deg X_m < \deg X_j$. Thus, we obtain

$$\sum_{k\colon \deg X_k = \deg X_j} \Big[\delta_{kj} + \frac{1}{2} \sum_{m,l\colon \deg X_j < \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlj}(\theta_u(s)) \Big] X_j(\theta_u(s)).$$

In the third sum, the functions $z_j^k(s)$ and $z_j^m(s)$ can take any possible values.

Second step. Now, we calculate exact estimates of three sums in (2.2.2).

1) Let $\deg X_k > \deg X_j$ and $\deg X_k = \deg X_m + \deg X_l$. From the above estimate we infer

$$\big| z_j^m(s) s_l \big| = O(\varepsilon^{\deg X_k - \deg X_j}).$$

Next, suppose that $\deg X_k > \deg X_j$ and $\deg X_k < \deg X_m + \deg X_l$. Then all the situations $\deg X_m > \deg X_j$, $\deg X_m = \deg X_j$ and $\deg X_m < \deg X_j$ are possible. Here we have

$$\big| z_j^m(s) s_l \big| = \begin{cases} \varepsilon O(\varepsilon^{\deg X_l}) \le \varepsilon O(\varepsilon^{\deg X_k - \deg X_j}) & \text{if } \deg X_m > \deg X_j, \\ O(\varepsilon^{\deg X_l}) \le \varepsilon O(\varepsilon^{\deg X_k - \deg X_j}) & \text{if } \deg X_m = \deg X_j, \\ 0 & \text{if } \deg X_m < \deg X_j. \end{cases}$$

2) Let now $\deg X_k = \deg X_j$ and $\deg X_k < \deg X_m + \deg X_l$. We again have to consider the situations $\deg X_m > \deg X_j$, $\deg X_m = \deg X_j$ and $\deg X_m <$

$\deg X_j$. It follows

$$\left|z_j^m(s)s_l\right| = \begin{cases} \varepsilon O(\varepsilon^{\deg X_l}) \leq \varepsilon O(1) & \text{if } \deg X_m > \deg X_j, \\ O(\varepsilon^{\deg X_l}) \leq \varepsilon O(1) & \text{if } \deg X_m = \deg X_j, \\ 0 & \text{if } \deg X_m < \deg X_j. \end{cases}$$

3) Finally, let $\deg X_k < \deg X_j$ and $\deg X_k < \deg X_m + \deg X_l$. In three situations $\deg X_m > \deg X_j$, $\deg X_m = \deg X_j$ and $\deg X_m < \deg X_j$, we obtain the same result as in the previous case:

$$\left|z_j^m(s)s_l\right| = \begin{cases} \varepsilon O(\varepsilon^{\deg X_l}) \leq \varepsilon O(1) & \text{if } \deg X_m > \deg X_j, \\ O(\varepsilon^{\deg X_l}) \leq \varepsilon O(1) & \text{if } \deg X_m = \deg X_j, \\ 0 & \text{if } \deg X_m < \deg X_j. \end{cases}$$

Thus, in the first sum of (2.2.2), the coefficients at X_k equal $O(\varepsilon)$, and in the second sum the coefficient at X_j equals $1 + O(\varepsilon)$, and the coefficients at X_k for $k \neq j$ equal $O(\varepsilon)$.

Third step. Consider the last sum (where $\deg X_k > \deg X_j$). Note that,

$$c_{mlk}(\theta_u(s)) = c_{mlk}(u) + o(1). \tag{2.2.3}$$

Then, taking into account the results of the second step, we deduce

$$\begin{aligned} &\sum_{m,l:\deg X_k \leq \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlk}(\theta_u(s)) \\ &= \sum_{m,l:\deg X_k = \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlk}(\theta_u(s)) \\ &\quad + \sum_{m,l:\deg X_k < \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlk}(\theta_u(s)) \\ &= \sum_{m,l:\deg X_k = \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlk}(u) + o(1) \cdot \varepsilon^{\deg X_k - \deg X_j} \\ &\quad + \varepsilon \cdot O(\varepsilon^{\deg X_k - \deg X_j}) \\ &= \sum_{m,l:\deg X_k = \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlk}(u) + o(1) \cdot \varepsilon^{\deg X_k - \deg X_j}. \end{aligned} \tag{2.2.4}$$

Consequently,

$$\begin{aligned} \widehat{X}_j^u(\theta_u(s)) &= \sum_{k:\, \deg X_k \leq \deg X_j} [\delta_{kj} + O(\varepsilon)] X_k \\ &+ \sum_{k:\, \deg X_k > \deg X_j} \Big[z_j^k(s) + \frac{1}{2} \sum_{m,l} z_j^m(s) s_l c_{mlk}(u) + o(\varepsilon^{\deg X_k - \deg X_j}) \Big] X_k(\theta_u(s)), \end{aligned} \tag{2.2.5}$$

where m, l in the last sum are such that $\deg X_k = \deg X_m + \deg X_l$.

Fourth step. It only remains to show that

$$z_j^k(s) + \frac{1}{2} \sum_{m,l:\deg X_k = \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlk}(u) = \delta_{kj}. \qquad (2.2.6)$$

For obtaining this, consider the mapping

$$\widehat{\theta}_u(x) = \exp\Big(\sum_{i=1}^{N} x_i \widehat{X}_i^u\Big)(u) = \theta_u(x),$$

where $x = (x_1, \dots, x_N)$, and apply the arguments of the first step with the following difference: it is known, that the vector fields $\widehat{X}_i^u$, $i = 1, \dots N$, are continuous, but they may not be differentiable, and formally, we cannot consider commutators of such vector fields. Therefore we modify previous arguments. For doing this, we consider the following representation of the identical (see Lemma 2.1.26) mapping:

$$\widetilde{\theta}_0(s) = \exp\bigg(\sum_{i=1}^{N} s_i D\widehat{\theta}_u^{-1}\langle \widehat{X}_i^u \rangle\bigg)(0) = s,$$

and represent $\frac{\partial}{\partial x_k} = D\widetilde{\theta}_0(s)\Big\langle \frac{\partial}{\partial x_k} \Big\rangle$ as before we represented $D\theta_u(s)\Big\langle \frac{\partial}{\partial x_k} \Big\rangle$. It is possible, since the vector fields $D\widehat{\theta}_u^{-1}\langle \widehat{X}_i^u \rangle$, $i = 1, \dots, N$, are smooth. Similarly to the first step, we infer

$$D\widetilde{\theta}_0(s)\Big\langle \frac{\partial}{\partial x_k} \Big\rangle = D\widehat{\theta}_u^{-1}\langle \widehat{X}_k^u \rangle(\widetilde{\theta}_0(s)) + \frac{1}{2}\bigg[D\widehat{\theta}_u^{-1}\langle \widehat{X}_k^u \rangle, \sum_{l=1}^{N} s_l D\widehat{\theta}_u^{-1}\langle \widehat{X}_l^u \rangle\bigg](\widetilde{\theta}_0(s)).$$

Since $\widetilde{\theta}_0(s) = s$, in view of properties of the vector fields $D\widehat{\theta}_u^{-1}\langle \widehat{X}_i^u \rangle$, $i = 1, \dots, N$, we deduce

$$\begin{aligned}\frac{\partial}{\partial x_k} &= D\widehat{\theta}_u^{-1}\langle \widehat{X}_k^u \rangle(s) + \frac{1}{2}\bigg[D\widehat{\theta}_u^{-1}\langle \widehat{X}_k^u \rangle, \sum_{l=1}^{N} s_l D\widehat{\theta}_u^{-1}\langle \widehat{X}_l^u \rangle\bigg](s) \\ &= D\widehat{\theta}_u^{-1}\langle \widehat{X}_k^u \rangle(s) + \frac{1}{2}\sum_{l=1}^{N} s_l \sum_{m:\, \deg X_m = \deg X_k + \deg X_l} c_{klm}(u) D\widehat{\theta}_u^{-1}\langle \widehat{X}_m^u \rangle(s).\end{aligned}$$

It follows

$$D\widehat{\theta}_u(s)\Big\langle \frac{\partial}{\partial x_k} \Big\rangle = \widehat{X}_k^u(\theta_u(s)) + \frac{1}{2}\sum_{l=1}^{N} s_l \sum_{m:\, \deg X_m = \deg X_k + \deg X_l} c_{klm}(u) \widehat{X}_m^u(\theta_u(s)).$$

Applying further the arguments of the first step, we have

$$\widehat{X}_j^u(\theta_u(s)) = \sum_{k=1}^{N}\Big[z_j^k(s) + \frac{1}{2} \sum_{m,\, l:\deg X_k = \deg X_m + \deg X_l} z_j^m(s) s_l c_{mlk}(u)\Big] \widehat{X}_k^u(\theta_u(s)),$$

and thus (2.2.6) is proved.

Taking into account the result (2.2.5) of the third step, we obtain

$$\widehat{X}_j^u(w) = \sum_{k\colon \deg X_k \leq \deg X_j} [\delta_{kj} + O(\varepsilon)]X_k + \sum_{k\colon \deg X_k > \deg X_j} o(\varepsilon^{\deg X_k - \deg X_j})X_k(w),$$

$j = 1, \ldots, N$. The theorem follows. □

Remark 2.2.10.

1) If the vector fields X_i, $i = 1, \ldots, N$, belong to the class $C^{1,\alpha}$, $\alpha \in (0, 1]$, then in (2.2.3) and, consequently, in (2.2.4), we obtain $o(1) = O(\rho(u, \theta_u(s))^\alpha)$. In this case, we have

$$\widehat{X}_j^u(w) = \sum_{k\colon \deg X_k \leq \deg X_j} [\delta_{kj} + O(\varepsilon)]X_k + \sum_{k\colon \deg X_k > \deg X_j} \rho(u, \theta_u(s))^\alpha \cdot O(\varepsilon^{\deg X_k - \deg X_j})X_k(w).$$

2) If the derivatives of the basis vector fields are Hölder with respect to d_∞, we obtain $o(1) = O(d_\infty(u, \theta_u(s))^\alpha) = O(\varepsilon^\alpha)$, and

$$\widehat{X}_j^u(w) = \sum_{k\colon \deg X_k \leq \deg X_j} [\delta_{kj} + O(\varepsilon)]X_k + \sum_{k\colon \deg X_k > \deg X_j} O(\varepsilon^{\deg X_k - \deg X_j + \alpha})X_k(w).$$

3) If the derivatives of the basis vector fields are Hölder with respect to $\mathfrak{d}$, we have

$$\widehat{X}_j^u(w) = \sum_{k\colon \deg X_k \leq \deg X_j} [\delta_{kj} + O(\varepsilon)]X_k + \sum_{k\colon \deg X_k > \deg X_j} \mathfrak{d}(u, \theta_u(s))^\alpha \cdot O(\varepsilon^{\deg X_k - \deg X_j})X_k(w).$$

Corollary 2.2.11. *For $x \in \operatorname{Box}(u, \varepsilon)$, the coefficients $\{a_{j,k}(x)\}_{j,k=1}^N$ from the equality*

$$X_j(x) = \sum_{k=1}^N a_{j,k}(x)\widehat{X}_k^u(x) \tag{2.2.7}$$

enjoy the following property:

$$a_{j,k}(x) = \begin{cases} O(\varepsilon) & \text{if } \deg X_j > \deg X_k, \\ \delta_{kj} + O(\varepsilon) & \text{if } \deg X_j = \deg X_k, \\ o(\varepsilon^{\deg X_k - \deg X_j}) & \text{if } \deg X_j < \deg X_k, \end{cases} \tag{2.2.8}$$

$j = 1, \ldots, N$. All $o(\cdot)$ and $O(\cdot)$ are uniform in u belonging to some compact neighborhood within $\mathcal{U}$.

Proof. According to Theorem 2.2.9, the coefficients $b_{j,k}(x)$ from the relation

$$\widehat{X}_j^u(x) = \sum_{k=1}^N b_{j,k}(x)X_k(x),$$

$j = 1, \dots, N$, have the same properties. Put $A(x) = (a_{j,k}(x))_{j,k=1}^N$ and $B(x) = (b_{j,k}(x))_{j,k=1}^N$. Then $A(x) = B(x)^{-1}$.

We use the well-known formula of calculation of the entries of the inverse matrix to estimate all $a_{j,k}(x)$, $j, k = 1 \dots, N$. We estimate the value $|a_{j,k}(x)| = \frac{|\det B_{j,k}(x)|}{|\det B(x)|}$, where the $(N-1)\times(N-1)$-matrix $B_{j,k}$ is constructed from the matrix $B(x)$ by deleting its jth column and kth line.

It is easy to see that $|\det B(x)| = 1 + O(\varepsilon)$, where $O(\varepsilon)$ is uniform for x and u belonging to some compact neighborhood $\mathcal{U} \subset \mathbb{M}$.

Next, we estimate $|\det B_{j,k}(x)|$. Obviously, $|\det B_{j,j}(x)| = 1 + O(\varepsilon)$ and $|\det B_{j,k}(x)| = O(\varepsilon)$ for $\deg X_j = \deg X_k$ and $j \neq k$, where $O(\varepsilon)$ is uniform for x and u belonging to some compact neighborhood $\mathcal{U} \subset \mathbb{M}$, $j = 1, \dots, N$.

Let now $\deg X_k > \deg X_j$. By construction, the diagonal elements of $B_{j,k}(x)$ with numbers (i, i), $j \leq i < k$, equal $o(\varepsilon^{\deg X_{i+1} - \deg X_i})$, and the elements under these ones equal $1+O(\varepsilon)$. Note that, $\det B_{j,k}(x)$ up to a multiple $(1+O(\varepsilon))$ equals the product of determinants of the following three matrices: the first $P(x) = p_{i,l}(x)$ is a $(j-1) \times (j-1)$-matrix with $p_{i,l}(x) = b_{i,l}(x)$, the second $Q(x) = q_{i,l}(x)$ is a $(k-j) \times (k-j)$-matrix with $q_{i,l}(x) = b_{i+j-1,l+j}(x)$, and the third $R(x) = r_{i,l}(x)$ is an $(N-k) \times (N-k)$-matrix with $r_{i,l}(x) = b_{i+k-1,l+k-1}(x)$.

For the matrices $P(x)$ and $R(x)$, we have

$$|\det P(x)| = 1 + O(\varepsilon) \quad \text{and} \quad |\det R(x)| = 1 + O(\varepsilon).$$

By construction, $q_{i,i}(x) = o(\varepsilon^{\deg X_{i+1} - \deg X_i})$ and $q_{i+1,i}(x) = 1 + O(\varepsilon)$. We have that the product of the diagonal elements of $Q(x)$ equals

$$\prod_{i=j}^{k-1} o(\varepsilon^{\deg X_{i+1} - \deg X_i}) = o(\varepsilon^{\deg X_k - \deg X_j}).$$

It is easy to see that, for all other summands constituting $\det Q(x)$, we have the same estimate.

Similarly, we show that for $\deg X_k < \deg X_j$ we have $|\det B_{j,k}(x)| = O(\varepsilon)$. Here $O(\varepsilon)$ is uniform in x and u belonging to some compact neighborhood $\mathcal{U} \subset \mathbb{M}$. The lemma follows. □

Remark 2.2.12. Similarly to Remark 2.2.10:

1) if $X_i \in C^{1,\alpha}$ then

$$a_{j,k}(x) = \begin{cases} O(\varepsilon) & \text{if } \deg X_j > \deg X_k, \\ \delta_{kj} + O(\varepsilon) & \text{if } \deg X_j = \deg X_k, \\ \rho(u,x)^\alpha \cdot O(\varepsilon^{\deg X_k - \deg X_j}) & \text{if } \deg X_j < \deg X_k, \end{cases}$$

2) if the derivatives of the basis vector fields are Hölder with respect to d_∞ then

$$a_{j,k}(x) = \begin{cases} O(\varepsilon) & \text{if } \deg X_j > \deg X_k, \\ \delta_{kj} + O(\varepsilon) & \text{if } \deg X_j = \deg X_k, \\ O(\varepsilon^{\deg X_k - \deg X_j + \alpha}) & \text{if } \deg X_j < \deg X_k, \end{cases}$$

3) if the derivatives of the basis vector fields are Hölder with respect to $\mathfrak{d}$ then

$$a_{j,k}(x) = \begin{cases} O(\varepsilon) & \text{if } \deg X_j > \deg X_k, \\ \delta_{kj} + O(\varepsilon) & \text{if } \deg X_j = \deg X_k, \\ \mathfrak{d}(u,x)^\alpha \cdot O(\varepsilon^{\deg X_k - \deg X_j}) & \text{if } \deg X_j < \deg X_k, \end{cases}$$

$j = 1, \dots, N$.

Corollary 2.2.11 implies instantly Gromov's Theorem on the nilpotentization of vector fields [70].

Corollary 2.2.13 (Gromov's Theorem [70]**).** *Let* $X_j \in C^1$. *On* $\mathrm{Box}(u, \varepsilon r_u)$, *consider the vector fields* $\{{}^\varepsilon X_i\}=\{\varepsilon^{\deg X_i} X_i\}$, $i = 1, \dots, N$. *Then the uniform convergence*

$$X_i^\varepsilon = \left(\Delta^u_{\varepsilon^{-1}}\right)_* \langle {}^\varepsilon X_i \rangle \to \widehat{X}_i^u \quad \text{as } \varepsilon \to 0, \quad i = 1, \dots, N, \tag{2.2.9}$$

holds at the points of the box $\mathrm{Box}(u, r_u)$ *and this convergence is uniform in u belonging to some compact neighborhood.*

Proof. Really, by (2.2.7), (2.2.8) and in view of Corollary 2.2.11 and Properties 2.1.24 and 2.2.5, we infer

$$X_i^\varepsilon(x) = \left(\Delta^u_{\varepsilon^{-1}}\right)_* \langle {}^\varepsilon X_i \rangle (x) = \varepsilon^{\deg X_i} \sum_{k=1}^N a_{i,k}\left(\Delta^u_\varepsilon(x)\right) \left(\Delta^u_{\varepsilon^{-1}}\right)_* \langle \widehat{X}_k^u \rangle (x)$$

$$= \sum_{k=1}^N \varepsilon^{\deg X_i - \deg X_k} a_{i,k}\left(\Delta^u_\varepsilon(x)\right) \widehat{X}_k^u(x)$$

$$= \sum_{k\colon \deg X_k \le \deg X_i} \varepsilon^{\deg X_i - \deg X_k}(\delta_{ik} + O(\varepsilon))\widehat{X}_k^u(x) + \sum_{k\colon \deg X_k > \deg X_i} o(1)\widehat{X}_k^u(x)$$

as $\varepsilon \to 0$. From here it follows the uniform convergence $X_i^\varepsilon = \left(\Delta^g_{\varepsilon^{-1}}\right)_* {}^\varepsilon X_i \to \widehat{X}_i^g$ as $\varepsilon \to 0$, $i = 1, \dots, N$, at the points of the box $\mathrm{Box}(g, r_g)$ and this convergence is uniform in g belonging to some compact neighborhood. □

Remark 2.2.14. For C^∞-vector fields and arbitrary M, the above corollary is formulated in [107, 127] in another way: $\widehat{X}_i^g$ is a homogeneous part of X_i, $1 = 1, \dots, N$. This statement implies Corollary 2.2.13. It is shown in [68] that, applying arguments based on Campbell–Hausdorff formula, the smoothness of vector fields can be reduced to be $2M + 1$.

Estimates (2.2.8) were written in the proof of [142, Thereom 3.1] as a consequence of the Gromov's Theorem which can be proved by method of [127] under an additional smoothness of vector fields: $X_j \in C^{2M - \deg X_j}$. Corollary 2.2.13 shows that estimates (2.2.8) are not only necessary but also sufficient for the validity of the Gromov's Theorem. In our paper estimates (2.2.8) are obtained under minimal assumption on the smoothness of vector fields: $X_j \in C^1$, $j = 1, \dots, N$. Thus, taking into account the footnote in [142, p. 253]: *If Gromov's Theorem is proved under weaker assumptions then all main results of the present paper and*

[139, 140, 141, 142, 144] *hold under the same assumptions on the smoothness of vector fields*, i.e., all results of the mentioned papers [139, 140, 141, 142, 144] are valid under the minimal assumptions on the smoothness of basis vector fields.

Recall that Gromov [70, p. 130] has formulated the theorem under assumption $X_j \in C^1$. Valeriĭ Berestovskiĭ sent us the following example confirming that arguments of Gromov's proof have to be corrected.

Example 2.2.15. Let $X = \frac{\partial}{\partial x}$, $Y = xy\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + x\frac{\partial}{\partial z}$. Then $Z := [X, Y] = y\frac{\partial}{\partial x} + \frac{\partial}{\partial z}$, $[X, Z] = 0$, $[Y, Z] = \frac{\partial}{\partial x} - y\big(y\frac{\partial}{\partial x} + \frac{\partial}{\partial z}\big) = (1 - y^2)\frac{\partial}{\partial x} - y\frac{\partial}{\partial z}$. One can easily see that X, Y, Z constitutes a global frame of smooth vector fields over the ring of smooth functions in $\mathbb{R}^3$. Also for corresponding one-parameter subgroups $X(x)$, $Y(y)$, $Z(z)$, we have $(X(x) \circ Y(y) \circ Z(z))(0, 0, 0) = (x, y, z)$. Under this $X = \frac{\partial}{\partial x}$ on $\mathbb{R}^3$, $Y = \frac{\partial}{\partial y}$ on $x = 0$, $Z = \frac{\partial}{\partial z}$ on z-line (even on $y = 0$). On the other hand, $\frac{\partial}{\partial y}Z = X \neq [Y, Z]$ (see above) on $x = 0$. This contradicts to Gromov's statement that (A) of [70, p. 131] implies (B) of [70, p. 132] in general case.

Corollary 2.2.16 (Estimate of the diameter of a box). *In a compact neighborhood $\mathcal{U} \subset \mathbb{M}$, for each point $u \in \mathcal{U}$ and each $\varepsilon > 0$ small enough, we have* $\operatorname{diam}(\operatorname{Box}(u, \varepsilon)) \leq L\varepsilon$, *where L depends only on $\mathcal{U}$.*

Proof. Assume the contrary: there exist sequences $\{\varepsilon_k\}_{k\in\mathbb{N}}$, $\{u_k\}_{k\in\mathbb{N}}$, $\{v_k\}_{k\in\mathbb{N}}$ and $\{w_k\}_{k\in\mathbb{N}}$ such that $\varepsilon_k \to 0$ as $k \to \infty$, $d_\infty(u_k, v_k) = \varepsilon_k$ and $d_\infty(u_k, w_k) \leq \varepsilon_k$ but $d_\infty(v_k, w_k) > k\varepsilon_k$. Since $\mathcal{U} \subset \mathbb{M}$ is compact, we may assume without loss of generality that $u_k \to u_0$ as $k \to \infty$. Then $v_k \to u_0$ and $w_k \to u_0$ as $k \to \infty$.

Assume without loss of generality that $\varepsilon^{\deg X_i} D\Delta^{u_k}_{\varepsilon^{-1}} X_i(x) \to \widehat{X}_i^{u_k}(x)$ as $\varepsilon \to 0$ for $x \in \operatorname{Box}(u_0, Kr_0)$, $r_0 \leq 1$, uniformly in u_k, $i = 1, \dots, N$, where $K = \max\{5, 5c^4\}$, c is such that $d^{u_k}_\infty(v, w) \leq c(d^{u_k}_\infty(u, v) + d^{u_k}_\infty(u, w))$ for all $u, v, w \in \operatorname{Box}(u_0, Kr_0)$ and $k \in \mathbb{N}$ (see Corollary 2.2.13). Note that, $c < \infty$ since $c = c(u_k)$ continuously depends on values of $\{F^j_{\mu,\beta}(u_k)\}_{j,\mu,\beta}$, consequently, it depends continuously on u_k. Moreover, the choice of K implies the following:

1) For k big enough, we have that an integral curve of a vector field with constant coefficients connecting $\Delta^{u_k}_{r_0\varepsilon_k^{-1}}(w_k)$ and $\Delta^{u_k}_{r_0\varepsilon_k^{-1}}(v_k)$ in the local homogeneous group $\mathcal{G}^{u_k}\mathbb{M}$ lies in $\operatorname{Box}(u_k, Kr_0)$.
2) We may choose k by the following way: $d_\infty(u_0, u_k) < r_0$ and the Riemannian distance between the integral curves corresponding to the collections $\{\widehat{X}_i^{u_k}\}_{i=1}^N$ and $\{({r_0}^{-1}\varepsilon_k)^{\deg X_i} D\Delta^{u_k}_{r_0\varepsilon_k^{-1}}\langle X_i\rangle\}_{i=1}^N$ (with constant coefficients) that connect points $\Delta^{u_k}_{r_0\varepsilon_k^{-1}}(w_k)$ and $\Delta^{u_k}_{r_0\varepsilon_k^{-1}}(v_k)$, is less than r_0.

Fix $k \in \mathbb{N}$. Then

$$v_k = \exp\Big(\sum_{i=1}^N \xi_i \varepsilon_k^{\deg X_i} X_i\Big)(u_k), \quad w_k = \exp\Big(\sum_{i=1}^N \eta_i \varepsilon_k^{\deg X_i} X_i\Big)(u_k),$$

and $w_k = \exp\Big(\sum\limits_{i=1}^{N} \zeta_i(\varepsilon_k)\varepsilon_k^{\deg X_i} X_i\Big)(v_k)$. Apply the mapping $\Delta^{u_k}_{r_0\varepsilon_k^{-1}}$ to v_k and w_k. We have

$$\Delta^{u_k}_{r_0\varepsilon_k^{-1}}(w_k) = \exp\Bigg(\sum_{i=1}^{N} \zeta_i(\varepsilon)\varepsilon_k^{\deg X_i} D\Delta^{u_k}_{r_0\varepsilon_k^{-1}}\langle X_i\rangle\Bigg)\big(\Delta^{u_k}_{r_0\varepsilon_k^{-1}}(v_k)\big).$$

Note that, $d_\infty\big(u_k, \Delta^{u_k}_{r_0\varepsilon_k^{-1}}(v_k)\big) = r_0$ and $d_\infty\big(u_k, \Delta^{u_k}_{r_0\varepsilon_k^{-1}}(w_k)\big) \leq r_0$. In view of Corollary 2.2.13, the vector fields $({r_0}^{-1}\varepsilon_k)^{\deg X_i} D\Delta^{u_k}_{r_0\varepsilon_k^{-1}}\langle X_i\rangle(x) = \widehat{X}_i^{u_k}(x) + o(1)$, $i = 1, \dots, N$, where $o(1)$ is uniform in x and in u_k. Consequently, since $\dim\operatorname{span}\{\widehat{X}_i^{u_k}(x)\}_{i=1}^N = N$ at each $x \in \operatorname{Box}(u_0, r_0)$, the Riemannian distance between $\Delta^{u_k}_{r_0\varepsilon_k^{-1}}(w_k)$ and $\Delta^{u_k}_{r_0\varepsilon_k^{-1}}(v_k)$ is bounded from above for all $k \in \mathbb{N}$ big enough. Therefore, the coefficients $\zeta_i(\varepsilon_k)$, $i = 1, \dots, N$, are bounded from above for all $k \in \mathbb{N}$ big enough. The assumption $d_\infty(v_k, w_k) > k\varepsilon_k$ contradicts this conclusion.

Thus there exists a constant $L = L(\mathcal{U})$ such that $\operatorname{diam}(\operatorname{Box}(u, \varepsilon)) \leq L\varepsilon$ for $u \in \mathcal{U}$. The statement follows. □

From the previous statement we come immediately to the following

Corollary 2.2.17 (Triangle inequality). *The quasimetric $d_\infty(x, y)$ meets locally the generalized triangle inequality (see Property 2.2.8).*

Corollary 2.2.18 (Decomposition of the basis vector fields). *Fix a point $\theta_u(s) \in \operatorname{Box}(u, O(\varepsilon))$. Remarks 2.2.10 and 2.2.12 imply the following decomposition of $D\theta_u^{-1}\langle X_i\rangle$, $i = 1, \dots, N$:*

$$[D\theta_u^{-1}\langle X_i\rangle(s)]_j = [(\widehat{X}_i^u)'(s)]_j + \sum_{k=1}^{N}(a_{i,k}(\theta_u(s)) - \delta_{ik})[(\widehat{X}_k^u)'(s)]_j.$$

If $d_\infty(u, \theta_u(s)) = O(\varepsilon)$, we have

$$\begin{aligned}[D\theta_u^{-1}\langle X_i\rangle(s)]_j = z_i^j(u, s) &+ \sum_{k\colon \deg X_k \leq \deg X_i} O(\varepsilon) z_k^j(u, s)\\ &+ \sum_{k\colon \deg X_k > \deg X_i} a_{i,k}(\theta_u(s)) z_k^j(u, s).\end{aligned}$$

If $\deg X_j \leq \deg X_i$ then $[D\theta_u^{-1}\langle X_i\rangle(s)]_j = \delta_{ij} + O(\varepsilon)$. For $\deg X_j > \deg X_i$ we have:

1) *If the basis vector fields are C^1-smooth then we deduce $[D\theta_u^{-1}\langle X_i\rangle(s)]_j = z_i^j(u, s) + O(\varepsilon^{\deg X_j - \deg X_i + 1}) + o(1) \cdot \varepsilon^{\deg X_j - \deg X_i}$, and therefore*

$$[D\theta_u^{-1}\langle X_i\rangle(s)]_j = z_i^j(u, s) + o(\varepsilon^{\deg X_j - \deg X_i}).$$

2) *If the derivatives of the basis vector fields are H^α-continuous with respect to $\mathfrak{d}$, then if $\deg X_j > \deg X_i$ we have*

$$[D\theta_u^{-1}\langle X_i\rangle(s)]_j = z_i^j(u, s) + \mathfrak{d}(u, \theta_u(s))^\alpha \cdot O(\varepsilon^{\deg X_j - \deg X_i}).$$

In particular, for $\alpha = 1$ and $\mathfrak{d} = d_\infty$ or $\mathfrak{d} = d_\infty^z$, where $d_\infty(u,z) = O(\varepsilon)$ (see further Local Approximation Theorem 2.5.4), we have

$$[D\theta_u^{-1}\langle X_i\rangle(s)]_j = z_i^j(u,s) + O(\varepsilon^{\deg X_j - \deg X_i + 1}).$$

Remark 2.2.19. Next we assume that Gromov's Theorem 2.2.13 holds for arbitrary M. In particular, it follows that all corollaries of this subsection are valid under the same assumption.

2.3. Comparison of geometries of tangent cones

The goal of Subsections 2.3, 2.4 and 2.6 is to compare the geometries of two local homogeneous groups. The main result of Section 2 is the following

Theorem 2.3.1. *Let $u, u' \in \mathcal{U}$ be such that $d_\infty(u,u') = C\varepsilon$. For a fixed $Q \in \mathbb{N}$, consider points w_0, $d_\infty(u,w_0) = \mathcal{C}\varepsilon$, and*

$$w_j^\varepsilon = \exp\biggl(\sum_{i=1}^N w_{i,j}\varepsilon^{\deg X_i}\widehat{X}_i^u\biggr)(w_{j-1}^\varepsilon), \quad w_j^{\varepsilon\prime} = \exp\biggl(\sum_{i=1}^N w_{i,j}\varepsilon^{\deg X_i}\widehat{X}_i^{u'}\biggr)(w_{j-1}^{\varepsilon\prime}),$$

$w_0^{\varepsilon\prime} = w_0^\varepsilon = w_0$, $j = 1,\dots,Q$. (Here $Q \in \mathbb{N}$ is such that all these points belong to the neighborhood $\mathcal{U} \subset \mathbb{M}$, for all $\varepsilon > 0$.) Then, for $\alpha > 0$,

$$\max\{d_\infty^u(w_Q^\varepsilon, w_Q^{\varepsilon\prime}), d_\infty^{u'}(w_Q^\varepsilon, w_Q^{\varepsilon\prime})\} = \varepsilon\cdot[\Theta(C,\mathcal{C},Q,\{F_{\mu,\beta}^j|_{\mathcal{U}}\}_{j,\mu,\beta})]\rho(u,u')^{\frac{\alpha}{M}}. \tag{2.3.1}$$

In the case of $\alpha = 0$, we have

$$\max\{d_\infty^u(w_Q^\varepsilon, w_Q^{\varepsilon\prime}), d_\infty^{u'}(w_Q^\varepsilon, w_Q^{\varepsilon\prime})\} = \varepsilon\cdot[\Theta(C,\mathcal{C},Q,\{F_{\mu,\beta}^j|_{\mathcal{U}}\}_{j,\mu,\beta})][\omega(\rho(u,u'))]^{\frac{1}{M}}$$

where $\omega \to 0$ is a modulus of continuity. (Here Θ is a bounded measurable function: $|\Theta| \le C_0 < \infty$, it is uniform in $u, u', w_0 \in \mathcal{U}$ and $\{w_{i,j}\}$, $i = 1,\dots,N$, $j = 1,\dots,Q$, belonging to some compact neighborhood of 0, and it depends on Q and on $\{F_{\mu,\beta}^j|_{\mathcal{U}}\}_{j,\mu,\beta}$.)

Notation 2.3.2. Throughout the paper, by the symbol Θ, we denote some bounded function absolute values of which do not exceed some $0 \le C < \infty$, where C depends only on the whole neighborhood where Θ is defined (i.e., it does not depend on *points* of this neighborhood).

Remark 2.3.3. If the derivatives of X_i, $i = 1,\dots,N$, are locally Hölder with respect to $\mathfrak{d}$, then we have $\mathfrak{d}(u,u')^{\frac{\alpha}{M}}$ instead of $\rho(u,u')^{\frac{\alpha}{M}}$ in (2.3.1).

In the current subsection we prove the "base" of the main result, i.e., we obtain it for $Q = 1$ and $\varepsilon = 1$. The proof for the general case is written in Subsection 2.6.

Fix points $u, u' \in \mathcal{U}$, where $\mathcal{U}$ is such that Assumption 2.1.14 holds. Recall that the collections of vector fields $\{\widehat{X}_i^u\}_{i=1}^N$ and $\{\widehat{X}_i^{u'}\}_{i=1}^N$ are frames in $\mathcal{G}^u\mathbb{M}$ and in $\mathcal{G}^{u'}\mathbb{M}$ respectively.

Definition 2.3.4. By $\widehat{X}^p(q)$, we denote the matrix, such that its ith column consists of the coordinates of the vector $\widehat{X}_i^p(q)$, $i = 1, \dots, N$, $p \in \mathbb{M}$, $q \in \mathcal{G}^p\mathbb{M}$, in the frame $\{X_j\}_{j=1}^N$.

Lemma 2.3.5. *Suppose that Assumption* 2.1.6 *holds. Let* $\Xi(u, u', q)$, $u, u', q \in \mathcal{U}$, *be the matrix such that*

$$\widehat{X}^{u'}(q) = \widehat{X}^u(q)\Xi(u, u', q). \tag{2.3.2}$$

Then the entries of $\Xi(u, u', q)$ *are* (*locally*) H^α*-continuous in* u *and* u'.

Proof. This statement is a direct consequence of Theorem 2.1.15. Indeed, the latter asserts that the vector fields $\{\widehat{X}_i^u\}_{i=1}^N$ are locally H^α-continuous in u. Since we prove a local property, and $\mathbb{M}$ is a Riemannian manifold, then, instead of $\mathbb{M}$, we may consider without loss of generality some neighborhood $\mathcal{U} \subset \mathbb{R}^N$ containing u and u'. Then it is easy to see that the entries of the matrices $\widehat{X}^u$ and $\widehat{X}^{u'}$ are (locally) H^α-continuous on $\mathcal{U} \times \mathcal{U}$. Since both matrices are non-degenerate in $\mathcal{U} \subset \mathbb{M}$, we have that $\Xi(u, u', q) = \widehat{X}^u(q)^{-1}\widehat{X}^{u'}(q)$ is also non-degenerate, and its entries $\Xi_{ij}(u, u', q)$ belong locally to $H^\alpha(\mathcal{U} \times \mathcal{U})$, $i, j = 1, \dots, N$. Moreover, these Hölder constants are the same for all $q \in \mathcal{U}$. □

Remark 2.3.6. If the derivatives of X_i, $i = 1, \dots, N$, are locally Hölder with respect to $\mathfrak{d}$, then the entries of Ξ are also locally Hölder with respect to $\mathfrak{d}$ (see Remark 2.1.20).

Remark 2.3.7. Suppose that Assumption 2.1.14 holds. Since $\Xi(u, u', q)$ equals the unit matrix if $u = u'$ then $\Xi_{ij} = \delta_{ij} + \Theta\rho(u, u')^\alpha$, where $\Theta = \Theta(u, u', q)$ is a bounded measurable function: $|\Theta| \le C_0 < \infty$, and the constant $C_0 \ge 0$ depends only on the neighborhood $\mathcal{U} \subset \mathbb{M}$.

Indeed, it follows immediately from the α-Hölder continuity of all vector fields: we have $|\Xi_{ij}(u, u', q) - \delta_{ij}| \le C_0(\rho(u, u')^\alpha)$, where

$$C_0 = \sup_{u,u',q \in \mathcal{U}} \frac{|\Xi_{ij}(u, u', q) - \delta_{ij}|}{\rho(u, u')^\alpha} < \infty$$

depends only on the neighborhood $\mathcal{U} \subset \mathbb{M}$.

Remark 2.3.8. If the derivatives of X_i, $i = 1, \dots, N$, are locally Hölder with respect to $\mathfrak{d}$, then $\Xi_{ij}(u, u', q) = \delta_{ij} + \Theta\mathfrak{d}(u, u')^\alpha$.

Theorem 2.3.9. *Suppose that* $u, v \in \mathcal{U}$. *Consider points*

$$w = \exp\Big(\sum_{i=1}^N w_i\widehat{X}_i^u\Big)(v), \quad \textit{and} \quad w' = \exp\Big(\sum_{i=1}^N w_i\widehat{X}_i^{u'}\Big)(v).$$

Then, for $\alpha > 0$, *we have*

$$\max\{d_\infty^u(w, w'), d_\infty^{u'}(w, w')\} = \Theta[\rho(u, u')^\alpha\rho(v, w)]^{\frac{1}{M}}, \tag{2.3.3}$$

where $u, u', v \in \mathcal{U}$, $\{w_i\}_{i=1}^N \in U(0) \subset \mathbb{R}^N$.

In the case of $\alpha = 0$,

$$\max\{d_\infty^u(w,w'), d_\infty^{u'}(w,w')\} = \Theta[\omega_\Xi(\rho(u,u'))\rho(v,w)]^{\frac{1}{M}}.$$

Here the symbol ω_Ξ *denotes the modulus of continuity of* Ξ *on the variables* u *and* u'.

Remark 2.3.10. Here (see Notation 2.3.2), the value

$$\sup|\Theta(u,u',v,\{w_i\}_{i=1}^N)| < \infty$$

depends only on $\mathcal{U} \subset \mathbb{M}$ and $U(0) \subset \mathbb{R}^N$.

Proof of Theorem 2.3.9. *First step.* Fix $q \in \mathcal{U} \Subset \mathbb{M}$. Notice that both collections of vectors $\{\widehat{X}_i^u(q)\}_{i=1}^N$ and $\{\widehat{X}_i^{u'}(q)\}_{i=1}^N$, $u,u' \in \mathcal{U}$, are frames of $T_q\mathbb{M}$. Consequently, there exists the transition $(N\times N)$-matrix $\Xi(u,u',q) = (\Xi(u,u',q))_{i,k}$ such that

$$\widehat{X}_i^{u'}(q) = \sum_{k=1}^N (\Xi(u,u',q))_{k,i}\widehat{X}_k^u(q). \tag{2.3.4}$$

Remark 2.3.7 implies that

$$\Xi(u,u',q)_{i,j} = \begin{cases} 1+\Theta_{i,j}\rho(u,u')^\alpha & \text{if } i=j \\ \Theta_{i,j}\rho(u,u')^\alpha & \text{if } i\neq j, \end{cases} \tag{2.3.5}$$

where the values $\Theta_{i,j}$ are bounded uniformly in all $u,u',q \in \mathcal{U}$. Thus $\widehat{X}_i^{u'}(q) = \widehat{X}_i^u(q) + \widehat{X}^u(q)[\Xi(u,u',q) - I]$, where $|[\Xi(u,u',q)-I]|_{k,j} = \Theta_{k,j}\rho(u,u')^\alpha$ for all $k,j = 1,\dots,N$.

Second step. Consider the integral line $\gamma(t)$ of the vector field $\sum\limits_{i=1}^N w_i\widehat{X}_i^{u'}$ starting at v with the endpoint w'. Rewrite the tangent vector to $\gamma(t)$ in the frame $\{\widehat{X}_i^u\}_{i=1}^N$ as $\dot\gamma(t) = \sum\limits_{i=1}^N w_i^u(\gamma(t))\widehat{X}_i^u(\gamma(t))$. From (2.3.4) it follows that

$$w_i^u(q) = \sum_{k=1}^N w_k(\Xi(u,u',q))_{i,k}.$$

From (2.3.5) we can estimate the coefficient w_i^u at $\widehat{X}_i^u$:

$$w_i^u = w_i + \sum_{k=1}^N [w_k\Theta_{i,k}\rho(u,u')^\alpha], \quad i=1,\dots,N. \tag{2.3.6}$$

Third step. Next, we estimate the Riemannian distance between w and w'. By $\kappa(t)$ denote the integral line of the vector field $\sum\limits_{i=1}^N w_i\widehat{X}_i^u$ connecting v and w, i.e., a line such that $\kappa(0) = v$ and

$$\dot\kappa(t) = \sum_{i=1}^N w_i\widehat{X}_i^u(\kappa(t)).$$

By means of the mapping θ_u^{-1} we transport $\kappa(t)$ and $\gamma(t)$ to $\mathbb{R}^N$. Let $\kappa_u(t) = \theta_u^{-1}(\kappa(t))$ and $\gamma_u(t) = \theta_u^{-1}(\gamma(t))$. Then

$$\dot{\kappa}_u(t) = (\theta_u^{-1})_*\langle\dot{\kappa}(t)\rangle = \sum_{i=1}^N w_i(\widehat{X}_i^u)'(\kappa_u(t))$$

and similarly

$$\dot{\gamma}_u(t) = \sum_{i=1}^N w_i(\theta_u^{-1})_*\langle\widehat{X}_i^{u'}\rangle = \sum_{i=1}^N w_i^u(t)(\widehat{X}_i^u)'(\gamma_u(t))$$

since $(\theta_u^{-1})_*\langle\widehat{X}_i^{u'}\rangle(\theta_u^{-1}(q)) = \sum\limits_{k=1}^N(\Xi(u,u',q))_{k,i}(\widehat{X}_i^u)'(\theta_u^{-1}(q))$ (see (2.3.4)). Using formula (2.1.11), rewrite the tangent vectors in Cartesian coordinates:

$$\dot{\kappa_u}(t) = \sum_{i=1}^N w_i\sum_{j=1}^N z_i^j(u,\kappa_u(t))\frac{\partial}{\partial x_j} = \sum_{j=1}^N W_j(u,\kappa_u(t))\frac{\partial}{\partial x_j}$$

where

$$W_j(u,\kappa_u(t)) = \sum_{i=1}^N w_i z_i^j(u,\kappa_u(t)) = w_j + \sum_{i=1}^{j-1} w_i z_i^j(u,\kappa_u(t))$$

since $z_i^j = 0$ if $j < i$. Similarly

$$\dot{\gamma_u}(t) = \sum_{j=1}^N W_j^u(u,\gamma_u(t))\frac{\partial}{\partial x_j}$$

where

$$W_j(u,\gamma_u(t)) = w_j^u(t) + \sum_{i=1}^{j-1} w_i^u(t)z_i^j(u,\gamma_u(t)).$$

Now we estimate the length of the curve $\lambda_u(t) = \gamma_u(t) - \kappa_u(t) + \theta_u^{-1}(w)$ with endpoints $\theta_u^{-1}(w)$ and $\theta_u^{-1}(w')$. The tangent vector to $\lambda_u(t)$ equals

$$\begin{aligned}\dot{\lambda_u}(t) = \dot{\gamma_u}(t) - \dot{\kappa_u}(t) &= \sum_{j=1}^N[W_j^u(u,\gamma_u(t)) - W_j(u,\kappa_u(t))]\frac{\partial}{\partial x_j}\\ &= \sum_{j=1}^N\Big[(w_j^u(t) - w_j) + \sum_{i<j} w_i(z_i^j(u,\gamma_u(t)) - z_i^j(u,\kappa_u(t)))\Big]\frac{\partial}{\partial x_j}\\ &\quad + \sum_{j=1}^N\Big[\sum_{i:\, i<j}^N (w_i^u(t) - w_i)z_i^j(u,\gamma_u(t))\Big]\frac{\partial}{\partial x_j}.\end{aligned} \tag{2.3.7}$$

Notice that for the last sum we have

$$\Big|\sum_{j=1}^N\Big[\sum_{i:\, i<j}^N (w_i^u(t) - w_i)z_i^j(u,\gamma_u(t))\Big]\frac{\partial}{\partial x_j}\Big| = \Theta\rho(u,u')^\alpha\rho(v,w)$$

since $w_i^u(t) = w_i + \Theta\rho(u,u')^\alpha\rho(v,w)$ by (2.3.6). By properties of z_i^j,

$$z_i^j(u,\gamma_u(t)) - z_i^j(u,\kappa_u(t)) = \Theta\Big[\sum_{|\mu|=1} |\gamma_u^\mu(t) - \kappa_u^\mu(t)|\Big].$$

Notice that

$$|\gamma_u(t) - \kappa_u(t)| \le \int_0^t |\dot{\gamma_u}(\tau) - \dot{\kappa_u}(\tau)|\,d\tau.$$

Consequently

$$\max_t |\gamma_u(t) - \kappa_u(t)| \le \max_t |\dot{\gamma_u}(t) - \dot{\kappa_u}(t)| = \max_t |\dot{\lambda_u}(t)|.$$

Applying these estimates to (2.3.7) and taking into account (2.3.6), we obtain

$$\max_t |\dot{\lambda_u}(t)| = \Theta\rho(u,u')^\alpha\rho(v,w) + \Theta\rho(v,w)\max_t |\dot{\lambda_u}(t)|.$$

From here it follows

$$\max_t |\dot{\lambda_u}(t)| = \frac{\Theta\rho(u,u')^\alpha\rho(v,w)}{1-\Theta\rho(v,w)} \le 2\Theta\rho(u,u')^\alpha\rho(v,w)$$

if $\Theta\rho(v,w) \le \frac{1}{2}$. Further, we denote the function 2Θ also by the symbol Θ. Thus,

$$\rho(\theta_u^{-1}(w),\theta_u^{-1}(w')) \le \int_0^1 |\dot{\lambda_u}(t)|\,dt \le \max_t |\dot{\lambda_u}(t)| = \Theta\rho(u,u')^\alpha\rho(v,w),$$

and $\rho(w,w') \le \Theta\rho(u,u')^\alpha\rho(v,w)$.

Fourth step. By the inequality $d_\infty^u(p,q) \le C\rho(p,q)^{\frac{1}{M}}$, we obtain the estimate of $d_\infty^u(w,w')$:

$$d_\infty^u(w,w') = \Theta[\rho(u,u')^\alpha\rho(v,w)]^{\frac{1}{M}}$$

in a compact neighborhood $\mathcal{U}$. The same estimate is true for $d_\infty^{u'}(w,w')$. The theorem follows. $\square$

Remark 2.3.11. If the derivatives of X_i, $i = 1,\dots,N$, are locally Hölder with respect to $\mathfrak{d}$, then we have $\mathfrak{d}(u,u')^\alpha$ instead of $\rho(u,u')^\alpha$ in (2.3.3) (the proof is similar, see Remarks 2.1.20 and 2.3.8).

2.4. Comparison of local geometries of tangent cones

Consider points

$$w_\varepsilon = \exp\Big(\sum_{i=1}^N w_i\varepsilon^{\deg X_i}\widehat{X}_i^u\Big)(v) \text{ and } w'_\varepsilon = \exp\Big(\sum_{i=1}^N w_i\varepsilon^{\deg X_i}\widehat{X}_i^{u'}\Big)(v).$$

Theorem 2.4.1. *Assume that $u,u',v \in \mathcal{U} \Subset \mathbb{M}$. Suppose that $d_\infty(u,u') = C\varepsilon$ and $d_\infty(u,v) = \mathcal{C}\varepsilon$ for some $C,\mathcal{C} < \infty$. Then, for $\alpha > 0$, we have*

$$\max\{d_\infty^u(w_\varepsilon,w'_\varepsilon), d_\infty^{u'}(w_\varepsilon,w'_\varepsilon)\} = \varepsilon[\Theta(C,\mathcal{C})]\rho(u,u')^{\frac{\alpha}{M}}. \tag{2.4.1}$$

In the case of $\alpha = 0$*, we have*

$$\max\{d^u_\infty(w_\varepsilon, w'_\varepsilon), d^{u'}_\infty(w_\varepsilon, w'_\varepsilon)\} = \varepsilon[\Theta(C,\mathcal{C})]\max\{\omega_\Xi(\rho(u,u')), \omega_{\Delta^u_{\varepsilon^{-1},v}\circ\Delta^{u'}_{\varepsilon,v}}(\rho(u,u'))\}^{\frac{1}{M}},$$

where $\Delta^u_{\varepsilon^{-1},v}$ *is defined below in* (2.4.5) *and* (2.4.6)*.* (*Here* Θ *is uniform in* $u, u', v \in \mathcal{U} \subset \mathbb{M}$*, and in* $\{w_i\}_{i=1}^N$ *belonging to some compact neighborhood of* 0 (*see Notation* 2.3.2)*, and all the values* $\omega_{\Delta^u_{\varepsilon^{-1},v}\circ\Delta^{u'}_{\varepsilon,v}}(\cdot)$ *are uniform in* $u, u', v \in \mathcal{U}$ *and* $\varepsilon > 0$)*.*

Remark 2.4.2. If the derivatives of X_i, $i = 1, \dots, N$, are locally α-Hölder with respect to $\mathfrak{d}$ (instead of ρ), then we have $\mathfrak{d}(u,u')^{\frac{\alpha}{M}}$ instead of $\rho(u,u')^{\frac{\alpha}{M}}$ in (2.4.1) (the proof is similar, see also Remark 2.3.11).

Proof of Theorem 2.4.1. *First step.* We put $w = w_1$ and $w' = w'_1$. In the frame $\{\widehat{X}^u_i\}_{i=1}^N$, we have

$$w' = \exp\Big(\sum_{i=1}^N w'_i \widehat{X}^u_i\Big)(v).$$

Consider the point

$$\omega_\varepsilon = \exp\Big(\sum_{i=1}^N w'_i \varepsilon^{\deg X_i}\widehat{X}^u_i\Big)(v).$$

Note that $\omega_1 = w'$. In view of the generalized triangle inequality, $d^u_\infty(w_\varepsilon, w'_\varepsilon) \le c(d^u_\infty(w_\varepsilon, \omega_\varepsilon) + d^u_\infty(\omega_\varepsilon, w'_\varepsilon))$. By the above estimate

$$d^u_\infty(\omega_\varepsilon, w_\varepsilon) = \varepsilon d^u_\infty(w, w') = \varepsilon\Theta(\rho(u,u')^\alpha \rho(v,w))^{\frac{1}{M}}. \tag{2.4.2}$$

Note that, if $\alpha = 0$, then we obtain here $\varepsilon\Theta(\omega_\Xi(\rho(u,u'))\rho(v,w))^{\frac{1}{M}}$ instead of $\varepsilon\Theta(\rho(u,u')^\alpha\rho(v,w))^{\frac{1}{M}}$.

Now we estimate the distance $d^u_\infty(\omega_\varepsilon, w'_\varepsilon)$. Represent w'_ε in the frame $\widehat{X}^u_i$, $i = 1, \dots, N$:

$$w'_\varepsilon = \exp\Big(\sum_{i=1}^N \alpha_i(\varepsilon)\varepsilon^{\deg X_i}\widehat{X}^u_i\Big)(v), \tag{2.4.3}$$

and consider the point

$$\omega' = \exp\Big(\sum_{i=1}^N \alpha_i(\varepsilon)\widehat{X}^u_i\Big)(v).$$

Here the coefficients $\alpha_i(\varepsilon)$, $i = 1, \dots, N$, depend on u, u', $\{w_i\}_{i=1}^N$ and $\varepsilon > 0$. Note that, in view of definition, we have $d^u_\infty(\omega_\varepsilon, w'_\varepsilon) = \varepsilon d^u_\infty(w', \omega')$.

Second step. Next, we show that the coefficients $\alpha_i(\varepsilon)$, $i = 1, \dots, N$, are uniformly bounded for all $\varepsilon > 0$ uniformly on u, u', v and $\{w_i\}_{i=1}^N$. By another words, we prove that there exists $S < \infty$ such that $d^u_\infty(v, w'_\varepsilon) \le S\varepsilon$ for all $\varepsilon > 0$ small enough and all u and $\{w_i\}_{i=1}^N$. Indeed, by the generalized triangle inequality for Carnot groups, we have

$$d^u_\infty(v, w'_\varepsilon) \le c(d^u_\infty(u, v) + d^u_\infty(u, w'_\varepsilon)).$$

Next, $d^u_\infty(u,w'_\varepsilon) = d_\infty(u,w'_\varepsilon)$. Since $d_\infty(u,v) = \mathcal{C}\varepsilon$, it is enough to show that $d_\infty(u,w'_\varepsilon) \leq K\varepsilon$. We obtain it via estimating the value $d_\infty(u',w'_\varepsilon)$ and taking into account the fact that $d_\infty(u,u') = C\varepsilon$. Since $d_\infty(u',w'_\varepsilon) = d^{u'}_\infty(u',w'_\varepsilon)$, then in view of the generalized triangle inequality for homogeneous groups, we have

$$d^{u'}_\infty(u',w'_\varepsilon) \leq c(d^{u'}_\infty(u',v) + d^{u'}_\infty(v,w'_\varepsilon)). \tag{2.4.4}$$

The conditions $d_\infty(u,u') = C\varepsilon$, $d_\infty(u,v) = \mathcal{C}\varepsilon$ and Corollary 2.2.16 imply

$$d^{u'}_\infty(u',v) = d_\infty(u',v) \leq L\max\{C,\mathcal{C}\}\varepsilon.$$

Applying (2.4.4) and Corollary 2.2.16 to points u, u' and w'_ε, we infer

$$d_\infty(u,w'_\varepsilon) \leq K\varepsilon.$$

From here and from the fact that $d_\infty(u,v) = \mathcal{C}\varepsilon$, we have

$$d^u_\infty(v,w'_\varepsilon) \leq S\varepsilon$$

for all $\varepsilon > 0$ small enough, u, u', $v \in \mathcal{U}$ and $\{w_i\}_{i=1}^N$ belonging to some compact neighborhoods.

From here, we have that all $\alpha_i(\varepsilon)$, $i = 1,\dots,N$, are bounded uniformly in $\varepsilon > 0$.

Third step. Note that $d^u_\infty(\omega_\varepsilon,w'_\varepsilon) = \varepsilon d^u_\infty(w',\omega')$. Consider the mapping

$$\Delta^u_{\varepsilon,v}(x) = \exp\Big(\sum_{i=1}^N x_i\varepsilon^{\deg X_i}\widehat{X}^u_i\Big)(v). \tag{2.4.5}$$

More exactly, if we fix $u,v \in \mathcal{U}$, then

$$\mathcal{U} \ni x \mapsto \{x_1,\dots,x_N\} \text{ by such a way that } x = \exp\Big(\sum_{i=1}^N x_i\widehat{X}^u_i\Big)(v)$$
$$\stackrel{\Delta^u_{\varepsilon,v}}{\longmapsto} \exp\Big(\sum_{i=1}^N x_i\varepsilon^{\deg X_i}\widehat{X}^u_i\Big)(v). \tag{2.4.6}$$

Show that the coordinate functions are H^α-continuous in $u \in \mathcal{U}$ uniformly on $\varepsilon > 0$.

1. The case of $\alpha > 0$**.** Indeed, the mapping

$$\theta_{v,u}(x_1,\dots,x_N) = \exp\Big(\sum_{i=1}^N x_i\widehat{X}^u_i\Big)(v),$$

where $(x_1,\dots,x_N) \in \mathrm{Box}(0,T\varepsilon)$, is H^α-continuous in $u \in \mathcal{U}$ as a solution to an equation with H^α-continuous in u right-hand part (see Section 5), and its Hölder constant does not depend on v belonging to some compact neighborhood. This mapping is also quasi-isometric on $(x_1\varepsilon^{-\deg X_1},\dots,x_N\varepsilon^{-\deg X_N})\in \mathrm{Box}(0,T)$ with respect to the metric ρ: there exist $0<K_1\le K_2<\infty$, such that

$$K_1|(y_1,\dots,y_N)-(z_1,\dots,z_N)|\le \rho(\theta_{v,u}(y),\theta_{v,u}(z))\\ \le K_2|(y_1,\dots,y_N)-(z_1,\dots,z_N)|,$$

where $y=(y_1,\dots,y_N)$, $z=(z_1,\dots,z_N)$, $y,z\in \mathrm{Box}(0,T)$. Consider now the inverse mapping, which assigns to a given point $x\in\mathbb{M}$, $d^u_\infty(v,x)\le T\varepsilon$, the "coordinates" $x_1(u,x)\varepsilon^{-\deg X_1},\dots,x_N(u,x)\varepsilon^{-\deg X_N}$ such that

$$x=\exp\Bigl(\sum_{i=1}^N x_i(u,x)\widehat{X}^u_i\Bigr)(v).$$

Note that the quasi-isometric coefficients of the mapping $\theta_{v,u}$ are independent of $(x_1,\dots,x_N)$, u and v belonging to some compact set (here we suppose that $d^u_\infty(v,x)\le T\varepsilon$). Show that the functions $x_1(u,x)\varepsilon^{-\deg X_1}$, …, $x_N(u,x)\varepsilon^{-\deg X_N}$ are H^α-continuous in $u\in\mathcal{U}$ for a fixed $x\in\mathbb{M}$, and their Hölder constants are bounded locally uniformly in x, v and in $\varepsilon>0$. (Here, to guarantee the uniform boundedness of $x_1(u,x)\varepsilon^{-\deg X_1},\dots,x_N(u,x)\varepsilon^{-\deg X_N}$, we assume that

1) both values $d_\infty(u,v)$ and $d^u_\infty(v,x)$ are comparable to ε;
2) the point u can be changed only by such a point u', that the distance $d_\infty(u,u')$ is also comparable to ε (see second step).)

The latter statement follows from the fact, that $\theta_{u,v}(x_1,\dots,x_N)$ is locally Hölder in u, and its Hölder constant is independent of v belonging to some compact neighborhood, and of $(x_1,\dots,x_N)$ belonging to some compact neighborhood $U(0)$ of zero. Since we prove a local property of a mapping then we may assume that u, u', x and v meet our above condition on d_∞-distances and they belong to some compact neighborhood $\mathcal{U}$ such that the mapping $\theta_{u,v}$ is bi-Lipschitz on $(x_1\varepsilon^{-\deg X_1},\dots,x_N\varepsilon^{-\deg X_N})$ if $u\in\mathcal{U}$: there exist constants $0<Q_1\le Q_2<\infty$ such that

$$Q_1|(x^1_1\varepsilon^{-\deg X_1},\dots,x^1_N\varepsilon^{-\deg X_N})-(x^2_1\varepsilon^{-\deg X_1},\dots,x^2_N\varepsilon^{-\deg X_N})|\\ \le \rho\bigl(\theta_{u,v}(x^{1,\varepsilon}),\theta_{u,v}(x^{2,\varepsilon})\bigr)\\ \le Q_2|(x^1_1\varepsilon^{-\deg X_1},\dots,x^1_N\varepsilon^{-\deg X_N})-(x^2_1\varepsilon^{-\deg X_1},\dots,x^2_N\varepsilon^{-\deg X_N})|$$

for $x^{1,\varepsilon}=(x^1_1\varepsilon^{-\deg X_1},\dots,x^1_N\varepsilon^{-\deg X_N})$ and $x^{2,\varepsilon}=(x^2_1\varepsilon^{-\deg X_1},\dots,x^2_N\varepsilon^{-\deg X_N})$. Moreover, its bi-Lipschitz coefficients are independent of points u, v, and

$$(x_1\varepsilon^{-\deg X_1},\dots,x_N\varepsilon^{-\deg X_N})$$

belonging to some compact neighborhoods. Indeed, consider the mapping

$$\theta_v(u,x_1,\dots,x_N)=\theta_{u,v}(x_1,\dots,x_N)$$

and suppose that for any $L>0$ there exist $\varepsilon>0$, points $v,x\in\mathcal{U}$, a level set $\theta_v^{-1}(x)$, and points $(u,x_1(u)\varepsilon^{-\deg X_1},\dots,x_N(u)\varepsilon^{-\deg X_N})$ and $(u',x_1(u')\varepsilon^{-\deg X_1},$

$\ldots, x_N(u')\varepsilon^{-\deg X_N})$ on it such that

$$\Big|(x_1(u)\varepsilon^{-\deg X_1}, \ldots, x_N(u)\varepsilon^{-\deg X_N}) \\ - (x_1(u')\varepsilon^{-\deg X_1}, \ldots, x_N(u')\varepsilon^{-\deg X_N})\Big| \geq L\rho(u,u')^\alpha \quad (2.4.7)$$

for some u and u'. The assumption (2.4.7) leads to the following contradiction:

$$\begin{aligned} 0 &= \rho\Big(\theta_v(u, x_1(u)\varepsilon^{-\deg X_1}, \ldots, x_N(u)\varepsilon^{-\deg X_N}),\\ &\qquad \theta_v(u', x_1(u')\varepsilon^{-\deg X_1}, \ldots, x_N(u')\varepsilon^{-\deg X_N})\Big)\\ &\geq \rho\Big(\theta_v(u, x_1(u)\varepsilon^{-\deg X_1}, \ldots, x_N(u)\varepsilon^{-\deg X_N}),\\ &\qquad \theta_v(u, x_1(u')\varepsilon^{-\deg X_1}, \ldots, x_N(u')\varepsilon^{-\deg X_N})\Big)\\ &\quad - \rho\Big(\theta_v(u, x_1(u')\varepsilon^{-\deg X_1}, \ldots, x_N(u')\varepsilon^{-\deg X_N}),\\ &\qquad \theta_v(u', x_1(u')\varepsilon^{-\deg X_1}, \ldots, x_N(u')\varepsilon^{-\deg X_N})\Big)\\ &\geq C_x\Big|(x_1(u)\varepsilon^{-\deg X_1}, \ldots, x_N(u)\varepsilon^{-\deg X_N})\\ &\quad - (x_1(u')\varepsilon^{-\deg X_1}, \ldots, x_N(u')\varepsilon^{-\deg X_N})\Big|\\ &\quad - C_u\rho(u,u')^\alpha \geq (LC_x - C_u)\rho(u,u')^\alpha > 0 \qquad \text{if } L > C_u/C_x. \end{aligned} \quad (2.4.8)$$

Note that $\omega' = \Delta^u_{\varepsilon^{-1},v}(\Delta^{u'}_{\varepsilon,v}(w'))$, and $w' = \Delta^{u'}_{\varepsilon^{-1},v}(\Delta^{u'}_{\varepsilon,v}(w'))$. Here, for the point $w'_\varepsilon = \Delta^{u'}_{\varepsilon,v}(w')$, we have $x_i(u, w'_\varepsilon) = \alpha_i(\varepsilon)\cdot\varepsilon^{\deg X_i}$ on the one hand, and we have $x_i(u', w'_\varepsilon) = w_i\cdot\varepsilon^{\deg X_i}$ on the other hand, $i = 1, \ldots, N$. Since the points u, u', v and w'_ε meet our assumption on points (see above two assumptions marked by "•"), we have that the Hölder constants of $x_i(u,x)\varepsilon^{-\deg X_i}$ are bounded uniformly in $\{w_j\}_{j=1}^N$ belonging to some neighborhood of zero. Hence, $\rho(\omega', w') = \Theta\rho(u,u')^\alpha$, and

$$d^u_\infty(\omega', w') = \Theta\rho(u,u')^{\frac{\alpha}{M}}. \quad (2.4.9)$$

2. The case of $\alpha = 0$ is proved similarly to the previous case. We prove that the functions $x_1(u,x)\varepsilon^{-\deg X_1}, \ldots, x_N(u,x)\varepsilon^{-\deg X_N}$ are uniformly continuous in $u \in \mathcal{U}$ for a fixed $x \in \mathbb{M}$, and this continuity is uniform in x, v and $\varepsilon > 0$. The points under consideration meet the above condition:

1) both values $d_\infty(u,v)$ and $d^u_\infty(v,x)$ are comparable to ε;
2) the point u can be changed only by such a point u', that the distance $d_\infty(u,u')$ is also comparable to ε (see second step).

To prove our result, we assume the contrary that there exists $\sigma > 0$ such that for any $\delta > 0$ there exist $\varepsilon > 0$, points $v, x \in \mathcal{U}$, a level set $\theta_v^{-1}(x)$, and points $(u, x_1(u)\varepsilon^{\deg X_1}, \ldots, x_N(u)\varepsilon^{\deg X_N})$ and $(u', x_1(u')\varepsilon^{\deg X_1}, \ldots, x_N(u')\varepsilon^{\deg X_N})$ on it such that $\rho(u,u') < \delta$, and in the right-hand part of (2.4.7) instead of $L\rho(u,u')^\alpha$, we obtain σ.

Repeating further the scheme of the proof almost verbatim and replacing $(LC_x - C_u)\rho(u,u')^\alpha$ by $\sigma C_x - \omega_{\theta_v}(\rho(u,u'))$ in the right-hand part of (2.4.8), we obtain the similar contradiction and deduce

$$\rho(\omega', w') = \omega_{\Delta^u_{\varepsilon^{-1},v}\circ\Delta^{u'}_{\varepsilon,v}}(\rho(u,u')). \tag{2.4.10}$$

We may assert without loss of generality, that $\omega_{\Delta^u_{\varepsilon^{-1},v}\circ\Delta^{u'}_{\varepsilon,v}}$ does not depend on x and v (see (2.4.7) and (2.4.8)).

Fourth step. Taking (2.4.2), (2.4.9) and (2.4.10) into account we obtain

$$d^u_\infty(\omega_\varepsilon, w'_\varepsilon) = \varepsilon[\Theta(C,\mathcal{C})]\rho(u,u')^{\frac{\alpha}{M}} \quad \text{and} \quad d^u_\infty(w_\varepsilon, w'_\varepsilon) = \varepsilon[\Theta(C,\mathcal{C})]\rho(u,u')^{\frac{\alpha}{M}}$$

for $\alpha > 0$. Similarly, we obtain the theorem for $\alpha = 0$:

$$d^u_\infty(w_\varepsilon, w'_\varepsilon) = \varepsilon[\Theta(C,\mathcal{C})]\max\{\omega_\Xi(\rho(u,u')), \omega_{\Delta^u_{\varepsilon^{-1},v}\circ\Delta^{u'}_{\varepsilon,v}}(\rho(u,u'))\}^{\frac{1}{M}}.$$

The theorem follows. □

Corollary 2.4.3. 1. *Note that $d_\infty(u,u') = C\varepsilon$ implies $\rho(u,u') \le C\varepsilon$. Then, for $\alpha > 0$, we have*

$$d^u_\infty(w_\varepsilon, w'_\varepsilon) = O(\varepsilon^{1+\frac{\alpha}{M}}) \text{ as } \varepsilon \to 0$$

where O is uniform in $u, u', v \in \mathcal{U} \subset \mathbb{M}$, and in $\{w_i\}_{i=1}^N$ belonging to some compact neighborhood of 0, and depends on C and $\mathcal{C}$.

2. *If $\alpha = 0$ then*

$$d^u_\infty(w_\varepsilon, w'_\varepsilon) = o(\varepsilon) \text{ as } \varepsilon \to 0$$

where o is uniform in $u, u', v \in \mathcal{U} \subset \mathbb{M}$, and in $\{w_i\}_{i=1}^N$ belonging to some compact neighborhood of 0, and depends on C and $\mathcal{C}$.

Remark 2.4.4. The estimate $O(\varepsilon^{1+\frac{\alpha}{M}})$ is also true for the case of vector fields X_i, $i = 1,\dots,N$, which are Hölder with respect to such $\mathfrak{d}$ that $d_\infty(u,u') = C\varepsilon$ implies $\mathfrak{d}(u,u') = K\varepsilon$, where K is bounded for $u, u' \in \mathcal{U}$.

A particular case is $\mathfrak{d} = d^z_\infty$, where $d_\infty(z,u) \le Q\varepsilon$ (see Local Approximation Theorem 2.5.4, case $\alpha = 0$, below).

2.5. The approximation theorems

In this subsection, we prove two Approximation Theorems. Their proofs use the following geometric property.

Proposition 2.5.1. *For a neighborhood $\mathcal{U} \Subset \mathbb{M}$, there exist positive constants $C > 0$ and $r_0 > 0$ depending on $\mathcal{U}$, M, N and values $\{F^j_{\mu,\beta}|_\mathcal{U}\}_{j,\mu,\beta}$ such that for any points u and v from a neighborhood $\mathcal{U}$ the following inclusion is valid:*

$$\bigcup_{x\in \mathrm{Box}^u(v,r)} \mathrm{Box}^u(x,\xi) \subseteq \mathrm{Box}^u(v, r + C\xi), \quad 0 < \xi,\ r \le r_0.$$

Proof. Suppose $x = \exp\Big(\sum\limits_{i=1}^{N} x_i \widehat{X}_i^u\Big)(v)$, $d_\infty^u(v,x) \le r$, and $z = \exp\Big(\sum\limits_{i=1}^{N} z_i \widehat{X}_i^u\Big)(x)$, $d_\infty^u(x,z) \le \xi$. We estimate the distance $d_\infty^u(v,z)$ applying formulas (2.1.10) to points x and z. Denote by $\{\zeta_i\}_{i=1}^N$ the coordinates of z with respect to v: $z = \exp\Big(\sum\limits_{i=1}^{N} \zeta_i \widehat{X}_i^u\Big)(v)$.

Case of $\deg X_i = 1$. Then $|\zeta_i| \le |x_i| + |z_i| \le (r+\xi)^{\deg X_i}$.

Case of $\deg X_i = 2$. Then

$$\begin{aligned} |\zeta_i| &\le |x_i| + |z_i| + \sum_{\substack{|e_l+e_j|_h=2,\\ l<j}} |F^i_{e_l,e_j}(u)||x_l z_j - z_l x_j| \\ &\le r^2 + \xi^2 + c_i(u) r\xi \le r^2 + 2r\frac{c_i(u)}{2}\xi + \Big(\frac{c_i(u)}{2}\xi\Big)^2 \\ &= \Big(r + \frac{c_i(u)}{2}\xi\Big)^{\deg X_i} = (r + C_i(u)\xi)^{\deg X_i}. \end{aligned}$$

Here we assume without loss of generality that $C_i(u) \ge 1$.

Case of $\deg X_i = k > 2$. We obtain similarly to the previous case

$$\begin{aligned} |\zeta_i| &\le |x_i| + |z_i| + \sum_{|\mu+\beta|_h=k, \mu>0, \beta>0} |F^i_{\mu,\beta}(u)| x^\mu \cdot z^\beta \\ &\le r^k + \xi^k + \sum_{|\mu+\beta|_h=k} c_i^{\mu\beta}(u) r^{|\mu|_h}\xi^{|\beta|_h} \le (r + C_i(u)\xi)^{\deg X_i}. \end{aligned}$$

Here we assume without loss of generality that $C_i(u), c_i(u) \ge 1$. Denote by $C(u) = \max\limits_i C_i(u)$. From above estimates we obtain

$$d_\infty^u(v,x) = \max_i\{|\zeta_i|^{\deg X_i}\} \le \max_i\{(r + C_i(u)\xi)^{\frac{\deg X_i}{\deg X_i}}\} \le r + C(u)\xi.$$

Since all the $C_i(u)$'s are continuous on u then we may choose $C < \infty$ such that $C(u) \le C$ for all u belonging to a compact neighborhood. The lemma follows. □

Theorem 2.5.2 (Approximation Theorem). *Assume that $u, u', v, w \in \mathcal{U}$. Then the following estimate is valid:*

$$|d_\infty^u(v,w) - d_\infty^{u'}(v,w)| = \begin{cases} \Theta[\rho(u,u')^\alpha d_\infty^u(v,w)]^{\frac{1}{M}} & \text{if } \alpha > 0, \\ \Theta[\omega_\Xi(\rho(u,u'))\rho(v,w)]^{\frac{1}{M}} & \text{if } \alpha = 0. \end{cases} \tag{2.5.1}$$

All these estimates are uniform on $\mathcal{U} \Subset \mathbb{M}$.

Proof. Let $\alpha > 0$. Denote by $z = \exp\Big(\sum\limits_{i=1}^{N} z_i \widehat{X}_i^u\Big)(v)$ and $z' = \exp\Big(\sum\limits_{i=1}^{N} z_i \widehat{X}_i^{u'}\Big)(v)$. If $z \in \mathrm{Box}^u(v, d_\infty^u(v,w))$ then $z' \in \mathrm{Box}^{u'}(v, d_\infty^u(v,w))$ and $z \in \mathrm{Box}^{u'}(z', R(u,u'))$. Here

$$R(u,u') = \sup_{p' \in \mathrm{Box}^{u'}(v, d_\infty^u(v,w))} d_\infty^{u'}(p,p'),$$

where $p=\exp\Big(\sum\limits_{i=1}^{N} p_i\widehat{X}_i^u\Big)(v)$ and $p'=\exp\Big(\sum\limits_{i=1}^{N} p_i\widehat{X}_i^{u'}\Big)(v)$. Using Proposition 2.5.1 we have that

$$\begin{aligned}\operatorname{Box}^u(v,d_\infty^u(v,w))\subset &\bigcup_{x\in\operatorname{Box}^{u'}(v,d_\infty^u(v,w))}\operatorname{Box}^{u'}(x,R(u,u'))\\ &\subset \operatorname{Box}^{u'}(v,d_\infty^u(v,w)+CR(u,u'))\end{aligned}$$

for some $0<C<\infty$. Note that, in view of Theorem 2.3.9 we have

$$R(u,u')=\sup_{p'\in\operatorname{Box}^{u'}(v,d_\infty^u(v,w))}\Theta[\rho(u,u')^\alpha\rho(v,p')]^{\frac{1}{M}}\leq\Theta[\rho(u,u')^\alpha d_\infty^u(v,w)]^{\frac{1}{M}}$$

Thus, we can write

$$\begin{aligned}\operatorname{Box}^u(v,d_\infty^u(v,w))&\subset\operatorname{Box}^{u'}(v,d_\infty^u(v,w)+CR(u,u'))\subset\\ &\operatorname{Box}^{u'}(v,d_\infty^u(v,w)+\Theta[\rho(u,u')^\alpha d_\infty^u(v,w)]^{\frac{1}{M}}).\end{aligned}\tag{2.5.2}$$

If $d_\infty^u(v,w)\leq\Theta[\rho(u,u')^\alpha d_\infty^u(v,w)]^{\frac{1}{M}}$ then the theorem follows:

$$|d_\infty^u(v,w)-d_\infty^{u'}(v,w)|\leq d_\infty^u(v,w)+d_\infty^{u'}(v,w)=\Theta[\rho(u,u')^\alpha d_\infty^u(v,w)]^{\frac{1}{M}}.$$

If $d_\infty^u(v,w)>\Theta[\rho(u,u')^\alpha d_\infty^u(v,w)]^{\frac{1}{M}}$ then applying again Proposition 2.5.1 we obtain

$$\operatorname{Box}^{u'}(v,d_\infty^u(v,w)-\Theta[\rho(u,u')^\alpha d_\infty^u(v,w)]^{\frac{1}{M}})\subset\operatorname{Box}^u(v,d_\infty^u(v,w)).$$

From the latter relation, taking into account the fact that $w\in\partial\operatorname{Box}^u(v,d_\infty^u(v,w))$ and relation (2.5.2), we infer

$$\begin{aligned}d_\infty^u(v,w)-\Theta[\rho(u,u')^\alpha d_\infty^u(v,w)]^{\frac{1}{M}}&\leq d_\infty^{u'}(v,w)\\ &\leq d_\infty^u(v,w)+\Theta[\rho(u,u')^\alpha d_\infty^u(v,w)]^{\frac{1}{M}}.\end{aligned}$$

The case of $\alpha=0$ is proved similarly. The theorem follows. □

Remark 2.5.3. If the derivatives of X_i, $i=1,\dots,N$, are locally Hölder with respect to $\mathfrak{d}$, then we have $\mathfrak{d}(u,u')^\alpha$ instead of $\rho(u,u')^\alpha$ in (2.5.1) (the proof is similar).

Approximation Theorem and local estimates (see Theorem 2.4.1) imply Local Approximation Theorem.

Theorem 2.5.4 (Local Approximation Theorem). *Assume that u, u', v, $w\in\mathcal{U}$. Suppose that $d_\infty(u,u')=C\varepsilon$, $d_\infty(u,v)=\mathcal{C}\varepsilon$ and $d_\infty(u,w)=\mathbb{C}\varepsilon$ for some C, $\mathcal{C}$, $\mathbb{C}<\infty$.*

1. *If $\alpha>0$, then*

$$|d_\infty^u(v,w)-d_\infty^{u'}(v,w)|=\varepsilon[\Theta(C,\mathcal{C},\mathbb{C})]\rho(u,u')^{\frac{\alpha}{M}}.\tag{2.5.3}$$

Moreover, if $u'=v$ and $\alpha>0$, then

$$|d_\infty^u(v,w)-d_\infty(v,w)|=\varepsilon[\Theta(\mathcal{C},\mathbb{C})]\rho(u,v)^{\frac{\alpha}{M}}.$$

2. *If* $\alpha = 0$, *then*
$$|d^u_\infty(v,w) - d^{u'}_\infty(v,w)| = \varepsilon o(1) = o(\varepsilon)$$
as $\varepsilon \to 0$, *where* o *is uniform in* $u, u', v, w \in \mathcal{U} \subset \mathbb{M}$. *Moreover, if* $u' = v$ *and* $\alpha = 0$, *then*
$$|d^u_\infty(v,w) - d_\infty(v,w)| = o(\varepsilon)$$
as $\varepsilon \to 0$, *where* o *is uniform in* $u, v, w \in \mathcal{U} \subset \mathbb{M}$.

The *Proof* follows the same scheme as the proof of Approximation Theorem 2.5.2 with $R(u,u') = \varepsilon[\Theta(C,\mathcal{C},\mathbb{C})]\rho(u,u')^{\frac{\alpha}{M}}$. The latter equality is valid by the uniformity assertion of Theorem 2.4.1.

Remark 2.5.5. If the derivatives of X_i, $i = 1,\dots,N$, are locally Hölder with respect to $\mathfrak{d}$, then we have $\mathfrak{d}(u,u')^{\frac{\alpha}{M}}$ instead of $\rho(u,u')^{\frac{\alpha}{M}}$ in (2.5.3) (the proof is similar).

2.6. Comparison of local geometries of two local homogeneous groups

Proof of Theorem 2.3.1. *First step.* Consider the case of $\alpha > 0$. The case of $Q = 1$ is proved in Theorem 2.4.1.

Second step. Suppose that $Q = 2$. For $\varepsilon = 1$, put $w_2 = w_2^1$ and $w_2' = {w_2^1}'$. Then, we have
$$w_2 = \exp\Bigl(\sum_{i=1}^N \omega_{i,2}\widehat{X}_i^u\Bigr)(w_0) \tag{2.6.1}$$
and
$$w_2' = \exp\Bigl(\sum_{i=1}^N \omega_{i,2}'\widehat{X}_i^{u'}\Bigr)(w_0). \tag{2.6.2}$$
It follows from the formulas of group operation in $\mathcal{G}^u\mathbb{M}$ and $\mathcal{G}^{u'}\mathbb{M}$, that
$$w_2^\varepsilon = \exp\Bigl(\sum_{i=1}^N \omega_{i,2}\varepsilon^{\deg X_i}\widehat{X}_i^u\Bigr)(w_0)$$
and
$${w_2^\varepsilon}' = \exp\Bigl(\sum_{i=1}^N \omega_{i,2}'\varepsilon^{\deg X_i}\widehat{X}_i^{u'}\Bigr)(w_0).$$
To estimate $d^{u'}_\infty(w_2^\varepsilon, {w_2^\varepsilon}')$, consider the auxiliary points
$$w_2'' = \exp\Bigl(\sum_{i=1}^N \omega_{i,2}\widehat{X}_i^{u'}\Bigr)(w_0) \text{ and } {w_2''}^\varepsilon = \exp\Bigl(\sum_{i=1}^N \omega_{i,2}\varepsilon^{\deg X_i}\widehat{X}_i^{u'}\Bigr)(w_0).$$
From the generalized triangle inequality we deduce
$$d^{u'}_\infty(w_2^\varepsilon, {w_2^\varepsilon}') \le c(d^{u'}_\infty(w_2^\varepsilon, {w_2''}^\varepsilon) + d^{u'}_\infty({w_2''}^\varepsilon, {w_2^\varepsilon}')). \tag{2.6.3}$$
In view of Theorem 2.4.1, we have $d^{u'}_\infty(w_2^\varepsilon, {w_2''}^\varepsilon) = \varepsilon\Theta\rho(u,u')^{\frac{\alpha}{M}}$. By the homogeneity of the distance $d^{u'}_\infty$ we have
$$d^{u'}_\infty({w_2''}^\varepsilon, {w_2^\varepsilon}') = \varepsilon d^{u'}_\infty(w_2'', w_2'). \tag{2.6.4}$$

Now, we estimate the value $d_\infty^{u'}(w_2'', w_2')$. For doing this, we use the group operation in the local homogeneous group $\mathcal{G}^{u'}\mathbb{M}$.

First of all, note that $|\omega_{i,2} - \omega_{i,2}'| = \Theta\rho(u,u')^\alpha$. Indeed, if we calculate these values $\omega_{i,2}$ and $\omega_{i,2}'$ via formulas of group operation (2.1.10), we see that all these expressions differ only in values of the function $\{F^j_{\mu,\beta}\}_{j,\mu,\beta}$. Since $|F^j_{\mu,\beta}(u) - F^j_{\mu,\beta}(u')| = \Theta\rho(u,u')^\alpha$ for all j,μ,β, we have also $|\omega_{i,2} - \omega_{i,2}'| = \Theta\rho(u,u')^\alpha$.

Next, note that while applying the group operation (see (2.1.10)), all summands look like $\omega_{i,2} - \omega_{i,2}'$ or $\omega_{i,2} - \omega_{i,2}' + \sum \Theta(\omega_{k,2}\omega_{j,2}' - \omega_{j,2}\omega_{k,2}')$. By (2.1.10), we deduce

$$\begin{aligned}\omega_{k,2}\omega_{j,2}' - \omega_{j,2}\omega_{k,2}'\\ &= \omega_{k,2}(\omega_{j,2} + \Theta\rho(u,u')^\alpha) - \omega_{j,2}(\omega_{k,2} + \Theta\rho(u,u')^\alpha) = \Theta\rho(u,u')^\alpha,\end{aligned}$$

and finally $d_\infty^{u'}(w_2'', w_2') = \Theta(\rho(u,u')^{\frac{\alpha}{M}})$. Here Θ depends on C, $\mathcal{C}$, $Q = 2$ and $\{F^j_{\mu,\beta}(u')\}_{j,\mu,\beta}$.

Taking into account the relations (2.6.3) and (2.6.4), we obtain

$$d_\infty^{u'}(w_2^\varepsilon, w_2^{\varepsilon\prime}) \le \varepsilon\Theta\rho(u,u')^\alpha$$

In view of Local Approximation Theorem 2.5.4, we derive

$$d_\infty^{u}(w_2^\varepsilon, w_2^{\varepsilon\prime}) = \varepsilon\Theta\rho(u,u')^{\frac{\alpha}{M}}.$$

Third step. In the case of $Q = 3$, it is easy to see from the previous case and the group operation, that if

$$w_3 = \exp\Big(\sum_{i=1}^N \omega_{i,3}\widehat{X}_i^u\Big)(w_0)$$

and

$$w_3' = \exp\Big(\sum_{i=1}^N \omega_{i,3}'\widehat{X}_i^{u'}\Big)(w_0),$$

then again $|\omega_{i,3} - \omega_{i,3}'| = \Theta\rho(u,u')^\alpha$. Here Θ depends on C, $\mathcal{C}$, $Q = 3$ and $\{F^j_{\mu,\beta}|\mathcal{U}\}_{j,\mu,\beta}$. (It suffices to apply the group operation (2.1.10) in the local homogeneous groups $\mathcal{G}^u\mathbb{M}$ and $\mathcal{G}^{u'}\mathbb{M}$ to expressions (2.6.1) and (2.6.2) and to points w_3 and w_3', respectively.) From now on, for obtaining estimate (2.3.1) at $Q = 3$, we repeat the arguments of the second step.

Fourth step. It is easy to see similarly to the third step, that the group operation and the induction hypothesis $|\omega_{i,l-1} - \omega_{i,l-1}'| = \Theta\rho(u,u')^\alpha$, $3 < l < Q$, imply $|\omega_{i,l} - \omega_{i,l}'| = \Theta\rho(u,u')^\alpha$. Indeed, it suffices to put $\omega_{i,l}$ and $\omega_{i,l}'$ instead of $\omega_{i,3}$ and $\omega_{i,3}'$, and $\omega_{i,l-1}$ and $\omega_{i,l-1}'$ instead of $\omega_{i,2}$ and $\omega_{i,2}'$ in the third step, and apply arguments from the second step.

The case of $\alpha = 0$ can be proved by applying the similar arguments.

The theorem follows. □

2.7. Comparison of local geometries of a Carnot–Carathéodory space and a local homogeneous group

In this subsection, we compare the local geometry of a Carnot–Carathéodory space manifold with the one of a local homogeneous group.

Theorem 2.7.1. *Let $u, w_0 \in \mathcal{U}$ be such that $d_\infty(u, w_0) = \mathcal{C}\varepsilon$. For a fixed $Q \in \mathbb{N}$, consider points*

$$\widehat{w}_j^\varepsilon = \exp\Bigl(\sum_{i=1}^N w_{i,j}\varepsilon^{\deg X_i}\widehat{X}_i^u\Bigr)(\widehat{w}_{j-1}^\varepsilon), \quad w_j^\varepsilon = \exp\Bigl(\sum_{i=1}^N w_{i,j}\varepsilon^{\deg X_i}X_i\Bigr)(w_{j-1}^\varepsilon),$$

$w_0^\varepsilon = \widehat{w}_0^\varepsilon = w_0$, $j = 1, \dots, Q$. (Here $Q \in \mathbb{N}$ is such that all these points belong to a neighborhood $\mathcal{U} \subset \mathbb{M}$ small enough for all $\varepsilon > 0$.) Then, for $\alpha > 0$ we have

$$\max\{d_\infty^u(\widehat{w}_Q^\varepsilon, w_Q^\varepsilon), d_\infty(\widehat{w}_Q^\varepsilon, w_Q^\varepsilon)\} = \sum_{k=1}^Q \Theta(\mathcal{C}, k, \{F_{\mu,\beta}^j|_\mathcal{U}\}_{j,\mu,\beta}) \cdot \varepsilon^{1+\frac{\alpha}{M}}. \qquad (2.7.1)$$

In the case of $\alpha = 0$ we have

$$\{d_\infty^u(\widehat{w}_Q^\varepsilon, w_Q^\varepsilon), d_\infty(\widehat{w}_Q^\varepsilon, w_Q^\varepsilon)\} = \varepsilon \cdot \Theta(\mathcal{C}, Q, \{F_{\mu,\beta}^j|_\mathcal{U}\}_{j,\mu,\beta})[\omega(\varepsilon)]^{\frac{1}{M}}$$

where $\omega(\varepsilon) \to 0$ as $\varepsilon \to 0$. (Here Θ is a bounded measurable function: $|\Theta| \le C_0 < \infty$, it is uniform in $u, u', w_0 \in \mathcal{U}$ and $\{w_{i,j}\}$, $i = 1, \dots, N$, $j = 1, \dots, Q$, belonging to some compact neighborhood of 0, and it depends on Q and on $\{F_{\mu,\beta}^j|_\mathcal{U}\}_{j,\mu,\beta}$.)

Proof. For simplifying the notation we denote the points $\widehat{w}_j^1$ by $\widehat{w}_j$, and we denote w_j^1 by w_j for $\varepsilon = 1$, $j = 1, \dots, Q$. First, we estimate the distances between the points $\widehat{w}_Q$ and w_Q. To do it, we construct the following sequence of points.

Let

$$\omega_{k,j} = \exp\Bigl(\sum_{i=1}^N w_{i,j}\widehat{X}_i^{w_k}\Bigr)(\omega_{k,j-1}),$$
$$\omega_{k,0} = w_k, \; k = 0, \dots, Q-1, \; j = 1, \dots, Q-k.$$

By another words, for $k = 0$ we obtain such sequence of integral lines that its first fragment coincides with $\exp\Bigl(\sum\limits_{i=1}^N w_{i,1}\varepsilon^{\deg X_i}X_i\Bigr)(w_0)$, and the other ones may diverge. Next, for $k = 1$, we obtain such sequence of integral lines that its first fragment coincides with $\exp\Bigl(\sum\limits_{i=1}^N w_{i,2}\varepsilon^{\deg X_i}X_i\Bigr)(w_1)$, and the other ones may diverge, etc. Hence, $\omega_{Q-1,1} = w_Q$ and

$$d_\infty^u(\widehat{w}_Q, w_Q) = \Theta \cdot \Bigl(d_\infty^u(\widehat{w}_Q, \omega_{0,Q}) + \sum_{k=1}^{Q-1} d_\infty^u(\omega_{k,Q-k}, \omega_{k-1,Q-k+1})\Bigr).$$

If $\alpha > 0$ then Theorem 2.3.1 implies

$$d_\infty^u(\widehat{w}_Q, \omega_{0,Q}) = \Theta(\mathcal{C}, Q, \{F_{\mu,\beta}^j|_\mathcal{U}\}_{j,\mu,\beta})\rho(u, w_0)^{\frac{\alpha}{M}}.$$

Moreover, for each of the summands we have

$$d^u_\infty(\omega_{k,Q-k},\omega_{k-1,Q-k+1})=\Theta(\mathcal{C},Q-k,\{F^j_{\mu,\beta}|\mathcal{U}\}_{j,\mu,\beta})\rho(w_k,w_{k-1})^{\frac{\alpha}{M}}.$$

By the same theorem, if we replace $w_{i,j}$ by $w_{i,j}\varepsilon^{\deg X_i}$, and w_k and $\widehat{w}_k$ by w^ε_k and $\widehat{w}^\varepsilon_k$ respectively in all the above formulas, $i=1,\dots,N$, $j,k=1,\dots,Q$, then it is easy to see using induction by k that firstly $d^u_\infty(w^\varepsilon_k,w^\varepsilon_{k-1})=\mathbb{C}_k\varepsilon$, secondly $d^u_\infty(u,w^\varepsilon_k)=C_k\varepsilon$ and $d^u_\infty(u,\omega^\varepsilon_{k,Q-k})=c_k\varepsilon$ for all $k=0,\dots,Q-1$ (this fact gives possibility to estimate d^u_∞ instead of $d^{w_k}_\infty$, see Local Approximation Theorem 2.5.4), thirdly

$$d^u_\infty(\widehat{w}^\varepsilon_Q,\omega^\varepsilon_{0,Q})=\varepsilon\Theta(\mathcal{C},Q,\{F^j_{\mu,\beta}|\mathcal{U}\}_{j,\mu,\beta})\rho(u,w_0)^{\frac{\alpha}{M}},$$

fourthly,

$$d^{w_k}_\infty(\omega^\varepsilon_{k,Q-k},\omega^\varepsilon_{k-1,Q-k+1})=\varepsilon\Theta(\mathcal{C},Q-k,\{F^j_{\mu,\beta}|\mathcal{U}\}_{j,\mu,\beta})\rho(w^\varepsilon_k,w^\varepsilon_{k-1})^{\frac{\alpha}{M}},$$

and

$$d^u_\infty(\omega^\varepsilon_{k,Q-k},\omega^\varepsilon_{k-1,Q-k+1})=\varepsilon\Theta(\mathcal{C},Q-k,\{F^j_{\mu,\beta}|\mathcal{U}\}_{j,\mu,\beta})\rho(w^\varepsilon_k,w^\varepsilon_{k-1})^{\frac{\alpha}{M}}.$$

Thus we obtain $d^u_\infty(\widehat{w}^\varepsilon_Q,w^\varepsilon_Q)=\sum\limits_{k=1}^{Q}\Theta(\mathcal{C},k,\{F^j_{\mu,\beta}|\mathcal{U}\}_{j,\mu,\beta})\cdot\varepsilon^{1+\frac{\alpha}{M}}$.

Since $d^u_\infty(\widehat{w}^\varepsilon_Q,w^\varepsilon_Q)=O(\varepsilon)$ and $d^u_\infty(u,\widehat{w}^\varepsilon_Q)=O(\varepsilon)$ then, by Local Approximation Theorem 2.5.4, we have

$$d_\infty(\widehat{w}^\varepsilon_Q,w^\varepsilon_Q)=\sum_{k=1}^{Q}\Theta(\mathcal{C},k,\{F^j_{\mu,\beta}|\mathcal{U}\}_{j,\mu,\beta})\cdot\varepsilon^{1+\frac{\alpha}{M}}.$$

If $\alpha=0$ then we repeat the above arguments replacing $\rho(\cdot,\cdot)^{\frac{\alpha}{M}}$ by $o(1)$. The theorem follows. □

Remark 2.7.2. If the derivatives of X_i, $i=1,\dots,N$, are locally Hölder with respect to $\mathfrak{d}$, such that $d_\infty(x,y)\leq\varepsilon$ implies $\mathfrak{d}(x,y)\leq K\varepsilon$, where K is bounded on $\mathcal{U}$, then the same estimate as in (2.7.1) is true (the proof is similar).

A particular case of such $\mathfrak{d}$ is d^z_∞, $d_\infty(u,z)\leq Q\varepsilon$ (see Local Approximation Theorem 2.5.4).

2.8. Applications

In Subsubsections 2.8.1 and 2.8.2 below, we assume that a Carnot–Carathéodory space $\mathbb{M}$ is a Carnot manifold meeting the 4th condition in the definition 2.1.1:

(4) a quotient mapping $[\cdot,\cdot]_0 : H_1\times H_j/H_{j-1}\mapsto H_{j+1}/H_j$ induced by Lie brackets, is an epimorphism for all $1\leq j<M$.

Under Condition (4) a local homogeneous group will be a local Carnot group (a stratified graded nilpotent group Lie).

We will not need this condition only in Remark 2.8.7 and subsubsection 2.8.3.

2.8.1. Rashevskiĭ–Chow theorem.

Definition 2.8.1. An absolutely continuous curve $\gamma : [0, a] \to \mathbb{M}$ is said to be *horizontal* if $\dot{\gamma}(t) \in H_{\gamma(t)}\mathbb{M}$ for almost all $t \in [0, a]$. Its length $l(\gamma)$ equals $\int\limits_0^a |\dot{\gamma}(t)|_{g_{\mathbb{M}}}\, dt$, where the value $|\dot{\gamma}(t)|_{g_{\mathbb{M}}}$ is calculated using the Riemann tensor $g_{\mathbb{M}}$ on $\mathbb{M}$. Analogously, the canonical Riemann tensor $g_{\mathcal{G}^u\mathbb{M}}$ on $\mathcal{G}^u\mathbb{M}$ defines a length $\widehat{l}$ of an absolutely continuous curve $\widehat{\gamma} : [0, a] \to \mathcal{G}^u\mathbb{M}$.

Definition 2.8.2. The *Carnot–Carathéodory distance* between points $x, y \in \mathbb{M}$ is defined as $d_{cc}(x, y) = \inf\limits_{\gamma} l(\gamma)$ where the infimum is taken over all horizontal curves with endpoints x and y.

The *Carnot–Carathéodory distance* $d^u_{cc}(x, y)$ between points $x, y \in \mathcal{G}^u\mathbb{M}$ is defined in the local Carnot group $\mathcal{G}^u\mathbb{M}$ similarly.

Corollary 2.8.3 (of Theorem 2.7.1**).** *Let $\mathbb{M}$ be a Carnot manifold. Suppose that Assumption* 2.1.6 *holds for $\alpha \in (0, 1]$. Let $y \in \mathbb{M}$. Let also ε be small enough to provide $\mathcal{G}^x\mathbb{M} \supset \mathrm{Box}(y, \varepsilon)$ for all $x \in \mathrm{Box}(y, \varepsilon)$, and $u, v, \widehat{w} \in \mathrm{Box}(y, \varepsilon)$. The points $v, \widehat{w} \in \mathrm{Box}(y, \varepsilon)$ can be joined in the local Carnot group $(\mathcal{G}^u\mathbb{M}, d^u_1) \supset \mathrm{Box}(y, \varepsilon)$ by a horizontal curve $\widehat{\gamma}$ composed by at most L segments of integral curves of horizontal fields $\widehat{X}^u_i$, $i = 1, \ldots, \dim H_1$. To the curve $\widehat{\gamma}$ it corresponds a curve γ, horizontal with respect to the initial horizontal distribution $H\mathbb{M}$, constituted by at most L segments of integral curves of the given horizontal fields X_i, $i = 1, \ldots, \dim H_1$. Moreover,*

1) *$\widehat{l}(\widehat{\gamma})$ is equivalent to $d^u_\infty(v, \widehat{w})$: $0 < \alpha \le \frac{\widehat{l}(\widehat{\gamma})}{d^u_\infty(v,\widehat{w})} \le \beta < \infty$;*
2) *the curve γ has endpoints $v, w \in \mathrm{Box}(y, O(\varepsilon))$;*
3) *$|l(\gamma) - \widehat{l}(\widehat{\gamma})| = o(\varepsilon)$;*
4) *$\max\{d^u_\infty(\widehat{w}, w), d_\infty(\widehat{w}, w)\} \le C\varepsilon^{1+\frac{\alpha}{M}}$ where C is independent of $y, u, v, \widehat{w}$ in some compact neighborhood $\mathcal{U} \Subset \mathbb{M}$;*
5) *if $v = u$ then $l(\gamma)$ is equivalent to $d_\infty(v, w)$: $0 < \alpha_1 \le \frac{l(\gamma)}{d_\infty(v,w)} \le \beta_1 < \infty$.*

All these estimates are uniform in $\widehat{w}$, v and y of some compact neighborhood $\mathcal{U} \Subset \mathbb{M}$ as $\varepsilon \to 0$.

Proof. The desired curve comes from those on any Carnot group [47]: given a Carnot group $\mathbb{G}$ with the vector fields $\widehat{X}_1, \ldots, \widehat{X}_N$, each point x can be represented as

$$x = \exp(a_L \widehat{X}_{i_L}) \circ \cdots \circ \exp(a_1 \widehat{X}_{i_1}), \quad i_j \in \{1, \ldots, \dim H_1\},$$

where each $|a_j|$ is controlled by the distance $\widehat{d}_{cc}(0, x)$ (or $\widehat{d}_\infty(0, x)$), $j = 1, \ldots, L$, and L is independent of x. To this composition of exponents it corresponds a horizontal curve $\widehat{\gamma}$ constituted by at most L segments $\widehat{\gamma}_j$, $j = 1, \ldots, L$, of integral curves of the horizontal vector fields $\widehat{X}_{i_1}, \widehat{X}_{i_2}, \ldots, \widehat{X}_{i_L}$ with endpoints 0 and x. For doing this, we set

$$\begin{cases}\widehat{\gamma}_1(t)=\exp(ta_1\widehat{X}_{i_1}),\\ t\in[0,1],\end{cases}$$

$$\begin{cases}\widehat{\gamma}_j(t)=\exp(ta_j\widehat{X}_{i_j})(\widehat{\gamma}_{j-1}(1))=\widehat{\gamma}_{j-1}(1)\cdot\exp(ta_j\widehat{X}_{i_j}),\\ t\in[0,1],\end{cases}$$

$j=2,\dots,L$, and from here we have $x=\widehat{\gamma}_{i_L}(1)$.

Now we carry over a construction described above to the local Carnot group $(\mathcal{G}^u\mathbb{M},d^u_\infty)\supset\operatorname{Box}(y,\varepsilon)$: the given points $\widehat{w},v\in\mathcal{G}^u\mathbb{M}$ can be connected by a horizontal curve $\widehat{\gamma}$ corresponding to the composition:

$$\widehat{w}=\exp(a_L\widehat{X}^u_{i_L})\circ\cdots\circ\exp(a_1\widehat{X}^u_{i_1})(v),\quad i_j\in\{1,\dots,\dim H_1\},\tag{2.8.1}$$

$j=1,\dots,L$. It follows immediately the first statement of the corollary.

Then the curve γ, corresponding to the composition

$$w=\exp(a_LX_{i_L})\circ\cdots\circ\exp(a_1X_{i_1})(v),\quad i_j\in\{1,\dots,\dim H_1\},\tag{2.8.2}$$

and constituted by segments

$$\begin{cases}\gamma_1(t)=\exp(ta_1X_{i_1})(v),\\ t\in[0,1],\end{cases}$$

$$\begin{cases}\gamma_j(t)=\exp(ta_jX_{i_j})(\gamma_{j-1}(1)),\\ t\in[0,1],\end{cases}$$

$j=2,\dots,L$, and from here we have $w=\gamma_{i_L}(1)$, is horizontal and its length equals $\widehat{l}(\widehat{\gamma})+o(\varepsilon)$ due to small difference of Riemann tensors in $\mathbb{M}$ and in $\mathcal{G}^u\mathbb{M}$. It can be verified by a direct estimation of the length integral

$$\int\limits_0^L|\dot{\gamma}(t)|_{g_\mathbb{M}}\,dt=\sum_{j=1}^L\int\limits_0^1|a_jX_{i_j}(\gamma_j(t))|_{g_\mathbb{M}}\,dt$$

taking into account the following evaluations: $g_\mathbb{M}(x)=g_\mathbb{M}(v)+o(1)$ for Riemannian tensor on $\operatorname{Box}(y,\varepsilon)$, and the same behavior of Riemannian tensor $g_{\mathcal{G}^u\mathbb{M}}$ on $\mathcal{G}^u\mathbb{M}$, $|a_j|=O(\varepsilon)$, and the evaluations of Theorem 2.2.9 and Corollary 2.2.11.

The estimate

$$\max\{d^u_\infty(w,w'),d_\infty(w,w')\}\le C\varepsilon^{1+\frac{\alpha}{M}}$$

follows immediately from (2.7.1).

The last statement of the corollary is a consequence of previous ones: we may assume that $d_\infty(v,w)=d^v_\infty(v,w)=\varepsilon$, then

$$\alpha-o(1)\le\frac{\widehat{l}(\widehat{\gamma})}{\varepsilon}-\frac{|l(\gamma)-\widehat{l}(\widehat{\gamma})|}{d_\infty(v,w)}\le\frac{l(\gamma)}{d_\infty(v,w)}\le\frac{\widehat{l}(\widehat{\gamma})}{\varepsilon}+\frac{|l(\gamma)-\widehat{l}(\widehat{\gamma})|}{d_\infty(v,w)}\le\beta+o(1).$$

□

Theorem 2.8.4. *Let $\mathbb{M}$ be a Carnot manifold. Suppose that Assumption* 2.1.6 *holds for some $\alpha \in (0,1]$. Let $y \in \mathbb{M}$. Given two points $w, v \in B(y,\varepsilon)$ where ε is small enough, there exist a curve γ, horizontal with respect to the initial horizontal distribution $H\mathbb{M}$, with endpoints w and v, and a horizontal curve $\widehat{\gamma}$ in the local Carnot group $(\mathcal{G}^y\mathbb{M}, d_\infty^y)$ with the same endpoints, such that*

1) *$\widehat{l}(\widehat{\gamma})$ is equivalent to $d_\infty^y(w,v)$;*
2) *$|l(\gamma) - \widehat{l}(\widehat{\gamma})| = o(\varepsilon)$;*
3) *if $v = y$ then the length $l(\gamma)$ is equivalent to $d_\infty(y,w)$.*

All these estimates are uniform in w, v and y of some compact neighborhood $\mathcal{U} \Subset \mathbb{M}$ as $\varepsilon \to 0$.

Proof. We can choose ε from the condition of the theorem by requests $C^{2+\frac{\alpha}{M}}\varepsilon^{\frac{\alpha^2}{M^2}} \leq 1$ and $\varepsilon \leq \frac{1}{2}$, where C is the constant from Corollary 2.8.3.

Apply Corollary 2.8.3 to the points $u = y$, v and w. It gives a horizontal curve γ_1 $(\widehat{\gamma})$ with respect to the initial horizontal distribution $H\mathbb{M}$ (in the local Carnot group $(\mathcal{G}^y\mathbb{M}, d_\infty^y)$) with endpoints v and w_1 (v and w) constituted by at most L segments of integral curves of given horizontal fields X_i $(\widehat{X}_i^y)$, $i = 1, \dots, \dim H_1$. In view of Corollary 2.8.3, the curve $\widehat{\gamma}$ has length comparable with $d_\infty^y(v,w)$, $|l(\gamma) - \widehat{l}(\widehat{\gamma})| = o(\varepsilon)$ and $\max\{d_\infty^y(w_1,w), d_\infty(w_1,w)\} \leq C\varepsilon^{1+\frac{\alpha}{M}}$.

Next, we apply again Corollary 2.8.3 to the points $u = v = w_1$ and w. It gives a horizontal curve γ_2 with respect to $H\mathbb{M}$ with endpoints w_1 and w_2. Its length is $O(\varepsilon^{1+\frac{\alpha}{M}})$ where O is uniform in $u, v, w \in \operatorname{Box}(y,\varepsilon)$, and $d_\infty(w_2,w) \leq C(C\varepsilon^{1+\frac{\alpha}{M}})^{1+\frac{\alpha}{M}} \leq \varepsilon^{1+\frac{2\alpha}{M}}$.

Assume that we have points $w_1, \dots, w_k$ and horizontal curves γ_l, $l = 2, \dots, k$, with respect to $H\mathbb{M}$ with endpoints w_{l-1} and w_l, such that γ_l has a length not exceeding $C(\varepsilon^{1+\frac{l-1}{M}\alpha})$, and $d_\infty(w_l,w) \leq \varepsilon^{1+\frac{l\alpha}{M}}$.

We continue, by the induction, applying Corollary 2.8.3 to the points $u = v = w_k$ and w. It results a horizontal curve γ_{k+1} with endpoints w_k and w_{k+1}, such that γ_{k+1} has a length $O(\varepsilon^{1+\frac{k\alpha}{M}})$ and $d_\infty(w_{k+1},w) \leq C(C\varepsilon^{1+\frac{k\alpha}{M}})^{1+\frac{\alpha}{M}} \leq \varepsilon^{1+\frac{k+1}{M}\alpha}$.

A curve $\Gamma_m = \gamma_1 \cup \dots \cup \gamma_m$ is horizontal, has endpoints v and w_m, its length does not exceed $l(\gamma_1) + o(\varepsilon) + C\sum\limits_{l=1}^{\infty} \varepsilon^{1+\frac{l\alpha}{M}} \leq l(\gamma_1) + o(\varepsilon)$ and $d_\infty(w_m,w) \to 0$ as $m \to \infty$. Therefore the sequence Γ_m converges to a horizontal curve γ as $m \to \infty$ with Property 2 mentioned in the theorem.

Under $v = y$ we can take $d_\infty(y,w)$ as ε in above estimates: it gives an evaluation $l(\gamma) \leq Cd_\infty(y,w)$. The opposite inequality can be verified directly by means of the above-obtained estimate: indeed, if $d_\infty(y,w) = \varepsilon$ then $d_\infty(y,w) = d_\infty^y(y,w) \leq C\widehat{l}(\widehat{\gamma}) \leq Cl(\gamma) + o(\varepsilon)$; it follows that $d_\infty(y,w) - o\big(d_\infty(y,w)\big) \leq Cl(\gamma)$ and the estimate $d_\infty(y,w) \leq C_1 l(\gamma)$ holds with C_1 independent of y from some compact neighborhood if w is close enough to y. Thus we have obtained the Property 3. $\square$

As an application of Theorem 2.8.4 we obtain a version of Rashevskiĭ–Chow type connectivity theorem.

Theorem 2.8.5. *Let* $\mathbb{M}$ *be a Carnot manifold. Suppose that Assumption* 2.1.6 *holds for* $\alpha \in (0,1]$. *Every two points* v, w *of a connected Carnot manifold can be joined by a rectifiable absolutely continuous horizontal curve* γ *composed by not more than countably many segments of integral lines of given horizontal fields.*

2.8.2. Comparison of metrics, and Ball–Box theorem.

Corollary 2.8.6. *Let* $\mathbb{M}$ *be a Carnot manifold. Suppose that Assumption* 2.1.6 *holds for* $\alpha \in (0,1]$. *In some compact neighborhood* $\mathcal{U} \Subset \mathbb{M}$, *the distance* d_{cc} *is equivalent to the quasimetric* d_∞.

Proof. An estimate $d_{cc}(x,y) \leq C_1 d_\infty(x,y)$ for points x, y from a compact neighborhood $\mathcal{U} \Subset \mathbb{M}$ follows from Theorem 2.8.4. Our next goal is to prove the converse estimate. Fix a compact neighborhood $\mathcal{U} \Subset \mathbb{M}$ and assume the contrary: for any $l \in \mathbb{N}$ there exist points $x_l, y_l \in \mathcal{U}$ such that $d_\infty(x_l, y_l) \geq l d_{cc}(x_l, y_l)$. In this case we have $d_\infty(x_l, y_l) \to 0$ as $l \to \infty$ since otherwise, for some subsequences x_{l_n} and y_{l_n}, we have simultaneously $d_{cc}(x_{l_n}, y_{l_n}) \to 0$ as $n \to \infty$, and $d_\infty(x_{l_n}, y_{l_n}) \geq \alpha > 0$ for all $n \in \mathbb{N}$ what is impossible. We can assume also that $x_l \to x \in \overline{\mathcal{U}}$ as $l \to \infty$ and $x_l \neq y_l$. Setting $d_\infty(x_l, y_l) = \varepsilon_l$ we have $d_\infty\big(x_l, \Delta^{x_l}_{r\varepsilon_l^{-1}} y_l\big) = r$ where $r > 0$ is normalizing factor. Let $\gamma : [0,1] \to \mathbb{M}$ be a Lipschitz horizontal path such that its length equal $d_{cc}(x_l, y_l)$ [22]. Then a length $l(\Gamma_l)$ of the curve Γ_l: $[0,1] \ni t \to \Delta^{x_l}_{r\varepsilon_l^{-1}}(\gamma(t))$, equal $\frac{r}{\varepsilon_l} d_{cc}(x_l, y_l)$ where the length $l(\Gamma_l)$ is measured with respect to the frame $\{X_i^{\varepsilon_l/r}\}$ with pushed-forward Riemannian tensor. Really, if $\dot\gamma(t) = \sum\limits_{i=1}^{\dim H_1} \gamma_i(t) X_j(\gamma(t))$ a. e. in $t \in [0,1]$, then $\dot\Gamma_l(t) = \frac{r}{\varepsilon_l} \sum\limits_{i=1}^{\dim H_1} \gamma_i(t) X^{\varepsilon_l/r}(\Gamma_l(t))$ a. e. in $t \in [0,1]$. It follows directly the equality $l(\Gamma_l) = \frac{r}{\varepsilon_l} d_{cc}(x_l, y_l)$. As far as the vectors $X_i^{\varepsilon_l/r}$, $i = 1, \dots, \dim H_1$, are closed to the corresponding nilpotentized vector fields $\widehat{X}_i^{x_l}$, $i = 1, \dots, \dim H_1$, by Gromov's Theorem (see Corollary 2.2.13), the Riemannian distance $\rho\big(x_l, \Delta^{x_l}_{r\varepsilon_l^{-1}} y_l\big) \to 0$ as $l \to \infty$:

$$\rho\big(x_l, \Delta^{x_l}_{r\varepsilon_l^{-1}} y_l\big) \leq C l(\Gamma_l) = C\frac{r}{\varepsilon_l} d_{cc}(x_l, y_l) \leq C r l^{-1} \frac{d_\infty(x_l, y_l)}{\varepsilon_l} = C r l^{-1},$$

where the constant C is independent of l.

It is in a contradiction with $d_\infty\big(x_l, \Delta^{x_l}_{r\varepsilon_l^{-1}} y_l\big) = r$ for all $l \in \mathbb{N}$ (see Proposition 2.2.2 for a comparison of metrics). □

Remark 2.8.7. Note that, for obtaining the estimate $d_\infty(x,y) \leq C_2 d_{cc}(x,y)$, the value α need not to be strictly greater than zero. Thus, the estimate $d_\infty(x,y) \leq C_2 d_{cc}(x,y)$ is valid not only for $\alpha = 0$ but also in a Carnot–Carathéodory space.

Another corollary is the so-called Ball–Box Theorem proved earlier for smooth vector fields in [116, 70].

Theorem 2.8.8 (Ball–Box Theorem). *Let* $\mathbb{M}$ *be a Carnot manifold. Suppose that Assumption* 2.1.6 *holds for* $\alpha \in (0,1]$. *The shape of a small ball* $B_{cc}(x,r)$ *in the metric* d_{cc} *looks like a box: given compact neighborhood* $\mathcal{U} \subset \mathbb{M}$ *there exist constants* $0 < C_1 \leq C_2 < \infty$ *and* r_0 *independent of* $x \in \mathcal{U}$ *such that*

$$\operatorname{Box}(x, C_1 r) \subset B_{cc}(x,r) \subset \operatorname{Box}(x, C_2 r) \tag{2.8.3}$$

for all $r \in (0, r_0)$.

Theorem 2.8.8 implies

Corollary 2.8.9. *Let* $\mathbb{M}$ *be a Carnot manifold. Suppose that Assumption* 2.1.6 *holds for* $\alpha \in (0,1]$. *The Hausdorff dimension of* $\mathbb{M}$ *equals*

$$\nu = \sum_{i=1}^{M} i(\dim H_i - \dim H_{i-1})$$

where $\dim H_0 = 0$.

This corollary extends Mitchell Theorem [108] to general Carnot–Carathéodory spaces with minimal smoothness of basis vector fields.

Remark 2.8.10. Let Assumption 2.1.6 hold for $\alpha \in (0,1]$. Applying Corollary 2.8.8, we obtain

1) the generalization of Theorem 2.3.9 for points w and w' close enough:

$$\max\{d^u_{cc}(w,w'), d_{cc}(w,w')\} = \Theta[\rho(u,v)\rho(v,w)]^{\frac{1}{M}} \leq \Theta[d_{cc}(u,v)d_{cc}(v,w)]^{\frac{1}{M}};$$

2) the generalization of Theorem 2.4.1:

$$\max\{d^u_{cc}(w_\varepsilon, w'_\varepsilon), d_{cc}(w_\varepsilon, w'_\varepsilon)\} = \varepsilon[\Theta(C,\mathcal{C})]\rho(u,v)^{\frac{\alpha}{M}};$$

3) the generalization of Theorem 2.7.1:

$$\max\{d^u_{cc}(\widehat{w}^\varepsilon_Q, w^\varepsilon_Q), d_{cc}(\widehat{w}^\varepsilon_Q, w^\varepsilon_Q)\} = \sum_{k=1}^{Q} \Theta(\mathcal{C}, k, \{F^j_{\mu,\beta}\}_{j,\mu,\beta}) \cdot \varepsilon^{1+\frac{\alpha}{M}}.$$

Corollary 2.8.6 and [72, Theorem 11.11] imply the following statement containing a result of [67], where only the first assertion is obtained under assumption of higher smoothness of vector fields.

Proposition 2.8.11. *Let* $\mathbb{M}$ *be a Carnot manifold. Let* X *and* Y *be two families of vector fields on* $\mathbb{M}$ *with the same horizontal distribution* $H\mathbb{M}$ *for both of which Assumption* 2.1.6 *holds with some* $\alpha \in (0,1]$. *Then, in some compact neighborhood* $\mathcal{U} \Subset \mathbb{M}$, *the following assertions are equivalent*:

1) *There exists a constant* $C \geq 1$ *such that* $C^{-1}d^X_\infty \leq d^Y_\infty \leq C d^X_\infty$.
2) *There exists a constant* $C \geq 1$ *such that* $C^{-1}|X_H\varphi| \leq |Y_H\varphi| \leq C|X_H\varphi|$ *for all* $\varphi \in C^\infty(\mathbb{M})$.

Here d_∞^X and d_∞^Y are quasimetrics constructed with respect to the bases X and Y, and $X_H\varphi$ and $Y_H\varphi$ are horizontal gradients (i.e., horizontal parts of the gradients) of φ.

Remark 2.8.12. If the derivatives of X_i, $i = 1, \dots, N$, are locally Hölder with respect to $\mathfrak{d}$, where $\mathfrak{d}$ meets conditions of Remark 2.7.2, the statements of Corollary 2.8.3, Theorem 2.8.4, Theorem 2.8.5, Corollary 2.8.6, Theorem 2.8.8, Corollary 2.8.9, Remark 2.8.10 and Proposition 2.8.11 are also true.

2.8.3. Mitchell Type theorem for quasimetric Carnot–Carathéodory spaces. In the papers [129], [130] the theory of convergence for quasimetric spaces is developed. It includes as a particular case the Gromov–Hausdorff theory for metric spaces [22]. Using main results of the present paper, Svetlana Selivanova proves the existence of the tangent cone (with respect to the notion of convergence introduced in her papers) to a quasimetric Carnot–Carathéodory space. The matter of this subsection is taken from [129], [130]. Now we formulate some definitions and statements from those papers. The following notion generalizes definition 2.2.1.

Definition 2.8.13 ([133]**).** A *quasimetric space* (X, d_X) is a topological space X with a quasimetric d_X. A *quasimetric* is a mapping $d_X : X \times X \to \mathbb{R}^+$ with the following properties

(1) $d_X(u, v) \geq 0$; $d_X(u, v) = 0$ if and only if $u = v$;
(2) $d_X(u, v) \leq c_X d_X(v, u)$, where $1 \leq c_X < \infty$ is a constant independent of $u, v \in X$;
(3) $d_X(u, v) \leq Q_X(d_X(u, w) + d_X(w, v))$, where $1 \leq Q_X < \infty$ is a constant independent of $u, v, w \in X$ (generalized triangle inequality);
(4) the function $d_X(u, v)$ is lower semicontinuous on the first argument.

If $c_X = 1$, $Q_X = 1$, then (X, d_X) is a metric space.

The *distortion* of a mapping $\varphi : (X, d_X) \to (Y, d_Y)$ is the value

$$\operatorname{dis}(\varphi) = \sup_{u,v\in X} |d_Y(\varphi(u), \varphi(v)) - d_X(u, v)|.$$

Definition 2.8.14 ([129]**).** The *distance* $d_{qm}(X, Y)$ between quasimetric spaces (X, d_X) and (Y, d_Y) is defined as the infimum is taken over $\rho > 0$ for which there exist (not necessarily continuous) mappings $\varphi : X \to Y$ and $\psi : Y \to X$ such that

$$\max\bigl\{\operatorname{dis}(\varphi),\, \operatorname{dis}(\psi),\, \sup_{x\in X} d_X(x, \psi(\varphi(x))),\, \sup_{y\in Y} d_Y(y, \varphi(\psi(y)))\bigr\} \leq \rho.$$

Note that for bounded quasimetric spaces the introduced distance is obviously finite.

Proposition 2.8.15 ([130]**).** *The distance d_{qm} possesses the following properties*:

1) *if quasimetric spaces X and Y are isometric, then $d_{qm}(X, Y) = 0$; if X and Y are compact and $d_{qm}(X, Y) = 0$, then X and Y are isometric* (*nondegeneracy*).
2) $d_{qm}(X, Y) = d_{qm}(Y, X)$ (*symmetricity*).
3) $d_{qm}(X, Y) \leq (Q_Z + 1)(d_{qm}(X, Z) + d_{qm}(Z, Y))$ (*analog of the triangle inequality*).

Definition 2.8.16 ([129])**.** A sequence of compact quasimetric spaces (X_n, d_{X_n}), $n \in \mathbb{N}$, *converges* to a compact quasimetric space (X, d_X), if $\lim\limits_{n\to\infty} d_{qm}(X_n, X) = 0$.

Proposition 2.8.15 implies

Proposition 2.8.17 ([130])**.** *If compact quasimetric spaces* (X, d_X)*,* (Y, d_Y) *are obtained as limits of the same sequence of compact spaces* (X_n, d_{X_n}) *such that* $|Q_{X_n}| \leq C < \infty$ *for all* $n \in \mathbb{N}$*, then* X *and* Y *are isometric.*

For noncompact spaces the following more general notion of convergence is introduced. A *pointed* (*quasi*)*metric space* is a pair (X, p) consisting of a (quasi)metric space X and a point $p \in X$. Whenever we want to emphasize what kind of (quasi)metric is on X, we shall write the pointed space as a triple (X, p, d_X).

Definition 2.8.18 ([129])**.** A sequence (X_n, p_n, d_{X_n}) of pointed quasimetric spaces *converges* to the pointed space (X, p, d_X), if there exists a sequence of reals $\delta_n \to 0$ such that for each $r > 0$ there exist mappings $\varphi_{n,r} : B^{d_{X_n}}(p_n, r + \delta_n) \to X$, $\psi_{n,r} : B^{d_X}(p, r + 2\delta_n) \to X_n$ such that

1) $\varphi_{n,r}(p_n) = p$, $\psi_{n,r}(p) = p_n$;
2) $\operatorname{dis}(\varphi_{n,r}) < \delta_n$, $\operatorname{dis}(\psi_{n,r}) < \delta_n$;
3) $\sup\limits_{x \in B^{d_{X_n}}(p_n, r+\delta_n)} d_{X_n}(x, \psi_{n,r}(\varphi_{n,r}(x))) < \delta_n$.

Recall that a quasimetric space X is *boundedly compact*, if all closed bounded subsets of X are compact. Two pointed quasimetric spaces (X, p) and (Y, q) are called *isometric*, if there exists an isometry $\eta : Y \to X$ such that $\eta(q) = p$.

The following theorem (see [129, 130] for details) informally states that, for boundedly compact spaces, the limit is unique up to isometry.

Theorem 2.8.19. *Let* (X, p)*,* (Y, q) *be two complete pointed quasimetric spaces obtained as limits* (*in the sense of Definition* 2.8.18) *of the same sequence* (X_n, p_n) *such that* $|Q_{X_n}| \leq C$ *for all* $n \in \mathbb{N}$*. If* X *is boundedly compact then* (X, p) *and* (Y, q) *are isometric.*

Definition 2.8.20. Let X be a boundedly compact (quasi)metric space, $p \in X$. If the limit of pointed spaces $\lim\limits_{\lambda\to\infty}(\lambda X, p) = (T_pX, e)$ (in the sense of Definition 2.8.18) exists, then T_pX is called the *tangent cone* to X at p. Here $\lambda X = (X, \lambda \cdot d_X)$; the symbol $\lim\limits_{\lambda\to\infty}(\lambda X, p)$ means that, for any sequence $\lambda_n \to \infty$, there exists $\lim\limits_{\lambda_n\to\infty}(\lambda_n X, p)$ which is independent of the choice of sequence $\lambda_n \to \infty$ as $n \to \infty$.

Remark 2.8.21 ([130])**.** According to theorem 2.8.19, the tangent cone is unique up to isometry, i.e., one should treat the tangent cone from definition 2.8.20 as a class of pointed quasimetric spaces isometric to each other. Note also that the tangent cone is completely defined by any (arbitrarily small) neighborhood of the point.

In [129], [130] the introduced definitions are then compared with their counterparts for metric spaces [22]. Recall that the *Hausdorff distance* between subsets

A and B of a metric space (X, d_X) is the value $d_H(A, B) = \inf\{r > 0 \,|\, A \subset U_r(B), B \subset U_r(A)\}$, where $U_r(A) = \bigcup\limits_{a \in A} B^{d_X}(a, r)$ is the r-neighborhood of the set A.

Definition 2.8.22. The *Gromov–Hausdorff distance* $d_{GH}(X, Y)$ between metric spaces X and Y is the infimum over $r > 0$ for which there exists a metric space Z and its subspaces X' and Y', isometric to X and Y, respectively, such that $d_H(X', Y') < r$.

Proposition 2.8.23 ([130]**).** *Let (X, d_X), (Y, d_Y) be metric spaces and $\rho > 0$.*

1) *If there exist (not necessarily continuous) mappings $\varphi : X \to Y$ and $\psi : Y \to X$ such that* $\operatorname{dis}(\varphi) \leq \rho$, $\operatorname{dis}(\psi) \leq \rho$, $\sup\limits_{x \in X} d_X(x, \psi(\varphi(x))) \leq \rho$, *then* $d_{GH}(X, Y) \leq \rho$.
2) *If $d_{GH}(X, Y) \leq \rho$, then for each $\varepsilon > 0$ there exist mappings $\varphi : X \to Y$ and $\psi : Y \to X$ such that* $\operatorname{dis}(\varphi) \leq 2\rho + \varepsilon$, $\operatorname{dis}(\psi) \leq 2\rho + \varepsilon$, $\sup\limits_{x \in X} d_X(x, \psi(\varphi(x))) \leq 2\rho + \varepsilon$.

As noted in [130], the conditions on the behaviour of $\sup\limits_{x \in X} d_X(x, \psi(\varphi(x)))$ (missing in [16], where an analog of Proposition 2.8.23 is formulated) are indeed necessary for validity of this proposition. From Proposition 2.8.23 it follows

Proposition 2.8.24 ([129, 130]**).** *Distances d_{qm} and d_{GH} are equivalent. More precisely, the following inequality holds: $d_{GH}(X, Y) \leq d_{qm}(X, Y) \leq 2d_{GH}(X, Y)$.*

Definition 2.8.25 ([22, 69, 71]**).** A sequence of compact metric spaces $\{X_n\}_{n=1}^{\infty}$ *converges in the Gromov–Hausdorff sense* to a compact metric space X, if

$$\lim_{n \to \infty} d_{GH}(X_n, X) = 0.$$

According to Proposition 2.8.24, one obtains

Proposition 2.8.26 ([129, 130]**).** *For metric spaces, Definitions* 2.8.16 *and* 2.8.25 *are equivalent.*

Definition 2.8.27 ([22]**).** A sequence (X_n, p_n) of pointed metric spaces *converges in the Gromov–Hausdorff sense* to a pointed metric space (X, p), if for all $r > 0$, $\varepsilon > 0$ there is a number n_0 such that for all $n > n_0$ there exist mappings $\varphi_{n,r} : B^{d_{X_n}}(p_n, r) \to X$ such that

1) $\varphi_{n,r}(p_n) = p$;
2) $\operatorname{dis}(\varphi_{n,r}) < \varepsilon$;
3) $U_\varepsilon(\varphi_{n,r}(B^{d_{X_n}}(p_n, r)) \supset B^{d_X}(p, r - \varepsilon)$.

Using known criteria of convergence in the Gromov–Hausdorff sense [22], it is not difficult to show [130] the equivalence of Definitions 2.8.18 and 2.8.27 for metric spaces.

Thus, taking in account Theorem 2.8.19 and its metric analog, one derives

Corollary 2.8.28 ([130]). *Tangent cones to the metric space X at $p \in X$, obtained using convergences from Definitions* 2.8.18 *and* 2.8.27*, respectively, are isometric.*

Consider again Carnot–Carathéodory spaces. The main result of [129] is the following statement.

Theorem 2.8.29. *Let $\mathbb{M}$ be a Carnot–Carathéodory space. The quasimetric space $(\mathbb{G}^u\mathbb{M}, 0, d_\infty^u)$ is the tangent cone at the point $u \in \mathbb{M}$ to the quasimetric space $(\mathcal{U}, u, d_\infty)$, $\mathcal{U} \subset \mathbb{M}$.*

Recall that the quasidistance d_∞^u for points $x, y \in \mathbb{G}^u\mathbb{M}$ such that $x = \exp\Big(\sum\limits_{i=1}^{N} x_i (\widehat{X}_i^u)'\Big)(y)$, is defined as $d_\infty^u(x,y) = \max\limits_i\{|x_i|^{\frac{1}{\deg X_i}}\}$. Theorem 2.8.29 follows from Local Approximation Theorem 2.5.4. The dilations Δ_{λ_n}, $\Delta^{\psi}_{\lambda_n^{-1}}$ are taken as the mappings $\varphi_{n,r}$, $\psi_{n,r}$ in Definition 2.8.18.

Remark 2.8.30 ([129]). The metric version of Theorem 2.8.29 (in the case when the Carnot–Carathéodory space $\mathbb{M}$ can be equipped with the intrinsic metric d_{cc}) is proved in [70, 108] with the help of Gromov's criterion (based on finite ε-nets) of convergence for compact metric spaces. Moreover, the convergence of noncompact pointed spaces is defined as convergence of the corresponding balls in the sense of Definition 2.8.25. The definition of the tangent cone obtained in this way is equivalent to the definition with respect to convergence from Definition 2.8.27 only for length metric spaces. Thus, this approach is not applicable to the situation under consideration.

Remark 2.8.31. In fact, one can consider an abstract quasimetric space with dilations (which generalizes Carnot–Carathéodory spaces) and prove that the tangent cone exists at each point satisfying some condition similar to the Local Approximation Theorem [130] (compare with [21]). Moreover, assuming additionally a certain regularity condition one can show that the tangent cone has the structure of a Lie group, the Lie algebra of which is graded and nilpotent. The proof of this fact is based on the well-known theorem due to A.I. Mal'cev [103] that provides necessary and sufficient conditions for a local topological group to be locally isomorphic to a topological group, and on results of [132] concerning the structure of contractible groups. The last result will appear in a forthcoming paper [131].

3. Differentiability on a Carnot–Carathéodory spaces

3.1. Primitive calculus

Recall that the *dilation group* δ_ε^u is defined in the local homogeneous group $\mathcal{G}^u\mathbb{M}$: to an element $x = \exp\Big(\sum\limits_{i=1}^{N} x_i \widehat{X}_i^u\Big)(u)$, it assigns $\delta_\varepsilon^u x = \exp\Big(\sum\limits_{i=1}^{N} x_i \varepsilon^{\deg X_i} \widehat{X}_i^u\Big)(u)$ in the cases where the right-hand side makes sense.

Further, we extend the dilations δ_t^u to negative t by setting $\delta_t^u x = \delta_{|t|}^u(x^{-1})$ for $t < 0$. The convenience of this definition is seen from the comparison of different kinds of differentiability.

3.1.1. Definition.

Notation 3.1.1. Let $\mathbb{M}, \widetilde{\mathbb{M}}$ be two Carnot–Carathéodory spaces. We denote the vector fields on $\widetilde{\mathbb{M}}$ by $\widetilde{X}_i$. We label the remaining objects on $\widetilde{\mathbb{M}}$ (the distance, the tangent cone etc.) with the same symbols as on $\mathbb{M}$ but with a tilde ˜ excluding the cases where the objects under consideration are obvious: for example, for a given mapping $\varphi : E \to \widetilde{\mathbb{M}}$, it is clear that $\mathcal{G}^u\mathbb{M}$ is the tangent cone at a point $u \in \mathbb{M}$ and $\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}$ is the tangent cone at the point $\varphi(u) \in \widetilde{\mathbb{M}}$; d_∞^u is Carnot–Carathéodory metric in the cone $\mathcal{G}^u\mathbb{M}$, $\tilde{d}_\infty^{\varphi(u)}$ is Carnot–Carathéodory metric in $\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}$, etc.

Recall that a *horizontal homomorphism* of homogeneous groups $L : \mathbb{G}$ and $\widetilde{\mathbb{G}}$ is a continuous homomorphism $L : \mathbb{G} \to \widetilde{\mathbb{G}}$ such that

1) $DL(0)(H\mathbb{G}) \subset H\widetilde{\mathbb{G}}$.

The notion of a horizontal homomorphism $L : \big(\mathcal{G}^u\mathbb{M}, d_\infty^u\big) \to \big(\mathcal{G}^q\widetilde{\mathbb{M}}, \tilde{d}_\infty^q\big)$, $u \in \mathbb{M}$, $q \in \widetilde{\mathbb{M}}$, of local homogeneous groups is different from this only in that the inclusion $L(\mathcal{G}^u\mathbb{M} \cap \exp H\mathcal{G}^u\mathbb{M}) \subset \mathcal{G}^q\widetilde{\mathbb{M}} \cap \exp\, H\mathcal{G}^q\widetilde{\mathbb{M}}$ holds only for $v \in \mathcal{G}^u\mathbb{M} \cap \exp H\mathcal{G}^u\mathbb{M}$ such that $L(v) \in \mathcal{G}^q\widetilde{\mathbb{M}}$.

Since a homomorphism of Lie groups is continuous, it is easy to prove that a horizontal homomorphism $L : \mathbb{G} \to \widetilde{\mathbb{G}}$ also has the property

2) $L(\delta_t v) = \tilde{\delta}_t L(v)$ for all $v \in \mathbb{G}$ and $t > 0$ (in the case of a horizontal homomorphism $L : \big(\mathcal{G}^u\mathbb{M}, d_\infty^u\big) \to \big(\mathcal{G}^q\widetilde{\mathbb{M}}, \tilde{d}_\infty^q\big)$ of local homogeneous groups, the equality $L(\delta_t v) = \tilde{\delta}_t L(v)$ is fulfilled only for $v \in \mathcal{G}^u\mathbb{M}$ and $t > 0$ such that $\delta_t v \in \mathcal{G}^u\mathbb{M}$ and $\tilde{\delta}_t L(v) \in \mathcal{G}^q\widetilde{\mathbb{M}}$).

Definition 3.1.2. Given two Carnot–Carathéodory spaces $\mathbb{M}$ and $\widetilde{\mathbb{M}}$, and a set $E \subset \mathbb{M}$, a mapping $\varphi : E \to \widetilde{\mathbb{M}}$ is called *hc-differentiable* at a point $u \in E$ if there exists a horizontal homomorphism $L : \big(\mathcal{G}^u\mathbb{M}, d_\infty^u\big) \to \big(\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}, \tilde{d}_\infty^{\varphi(u)}\big)$ of the local homogeneous groups such that

$$\tilde{d}_\infty^{\varphi(u)}(\varphi(v), L(v)) = o\big(d_\infty^u(u, v)\big) \quad \text{as } E \cap \mathcal{G}^u\mathbb{M} \ni v \to u. \tag{3.1.1}$$

A horizontal homomorphism $L : \big(\mathcal{G}^u\mathbb{M}, d_\infty^u\big) \to \big(\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}, \tilde{d}_\infty^{\varphi(u)}\big)$ satisfying condition (3.1.1), is called an *hc-differential* of the mapping $\varphi : E \to \widetilde{\mathbb{M}}$ at $u \in E$ on E and is denoted by $\widehat{D}\varphi(u)$. It can be proved [141] that if u is a density point of E then the *hc*-differential is unique.

Moreover, it is easy to verify that the *hc*-differential commutes with the one-parameter dilation group:

$$\tilde{\delta}_t^{\varphi(u)} \circ \widehat{D}\varphi(u) = \widehat{D}\varphi(u) \circ \delta_t^u. \tag{3.1.2}$$

Proposition 3.1.3 ([141]). *Definition* 3.1.2 *is equivalent to each of the following assertion:*

1) $\tilde{d}_\infty^{\varphi(u)}\bigl(\Delta_{t^{-1}}^{\varphi(u)}\varphi\bigl(\delta_t^u(v)\bigr), L(v)\bigr) = o(1)$ *as* $t \to 0$, *where* $o(\cdot)$ *is uniform in the points* v *of any compact part of* $\mathcal{G}^u\mathbb{M}$;
2) $d_\infty\bigl(\Delta_{t^{-1}}^{\varphi(u)}\varphi\bigl(\delta_t^u(v)\bigr), L(v)\bigr) = o(1)$ *as* $t \to 0$, *where* $o(\cdot)$ *is uniform in the points* v *of any compact part of* $\mathcal{G}^u\mathbb{M}$;
3) $\tilde{d}_\infty(\varphi(v), L(v)) = o\bigl(d_\infty^u(u,v)\bigr)$ *as* $E \cap \mathcal{G}^u\mathbb{M} \ni v \to u$;
4) $\tilde{d}_\infty(\varphi(v), L(v)) = o\bigl(d_\infty(u,v)\bigr)$ *as* $E \cap \mathcal{G}^u\mathbb{M} \ni v \to u$.

Proof. First, we prove item 1. Consider a point v of a compact part of $\mathcal{G}^u\mathbb{M}$ and a sequence $\varepsilon_i \to 0$ as $i \to \infty$ such that $\delta_{\varepsilon_i}^u v \in E$ for all $i \in \mathbb{N}$. From (3.1.1) we have $\tilde{d}_\infty^{\varphi(u)}\bigl(\varphi\bigl(\delta_{\varepsilon_i}^u v\bigr), L\bigl(\delta_{\varepsilon_i}^u v\bigr)\bigr) = o\bigl(d_\infty^u\bigl(u, \delta_{\varepsilon_i}^u v\bigr)\bigr) = o(\varepsilon_i)$. In view of (3.1.2), we infer

$$\tilde{d}_\infty^{\varphi(u)}\bigl(\Delta_{\varepsilon_i}^{\varphi(u)}\bigl(\Delta_{\varepsilon_i^{-1}}^{\varphi(u)}\varphi\bigl(\delta_{\varepsilon_i}^u v\bigr)\bigr), \delta_{\varepsilon_i}^{\varphi(u)}L(v)\bigr) = o(\varepsilon_i) \quad \text{uniformly in } v.$$

From here and properties of quasimetric $\tilde{d}_\infty^{\varphi(u)}$ item 1 follows. Obviously, the argument is reversible.

Further we have $\alpha\rho_{\mathcal{G}^{\varphi(u)}}(x,y) \le \tilde{d}_\infty^{\varphi(u)}(x,y) \le \beta\rho_{\mathcal{G}^{\varphi(u)}}(x,y)^{\frac{1}{M}}$, where $\rho_{\mathcal{G}^{\varphi(u)}}$ is the Riemannian metric on $\mathcal{G}^{\varphi(u)}$, $\rho(x,y) \le d_\infty(x,y) \le \rho(x,y)^{\frac{1}{M}}$ (α and β depend on the choice of the compact part), and metrics $\rho_{\mathcal{G}^{\varphi(u)}}(x,y)$ and $\rho(x,y)$ are equivalent on any compact neighborhood part of point u. It gives the equivalence of item 1 to item 2.

By Local Approximation Theorem 2.5.4, we obtain the equivalence of (3.1.1) to the items 3 and 4. Really, it is enough to apply the following relations:

$$\begin{aligned}
\bigl|\tilde{d}_\infty^{\varphi(u)}(\varphi(v), L(v)) - \tilde{d}_\infty(\varphi(v), L(v))\bigr| &= o\bigl(\tilde{d}_\infty^{\varphi(u)}(L(v), \varphi(u)\bigr) \\
&= o\bigl(d_\infty^{\varphi(u)}(v,u)\bigr) \\
&= o\bigl(\tilde{d}_\infty(L(v), \varphi(u)\bigr) \\
&= o(d_\infty(v,u)).
\end{aligned}$$

□

3.1.2. Chain rule. In this subsubsection, we prove the chain rule.

Theorem 3.1.4 (The Chain Rule [141]**).** *Suppose that* $\mathbb{M}, \widetilde{\mathbb{M}}, \widehat{\mathbb{M}}$ *are Carnot–Carathéodory spaces,* E *is a set in* $\mathbb{M}$, *and* $\varphi : E \to \widetilde{\mathbb{M}}$ *is a mapping from* E *into* $\widetilde{\mathbb{M}}$ *hc-differentiable at a point* $u \in E$. *Suppose also that* F *is a set in* $\widetilde{\mathbb{M}}$, $\varphi(E) \subset F$ *and* $\psi : F \to \widehat{\mathbb{M}}$ *is a mapping from* F *into* $\widehat{\mathbb{M}}$ *hc-differentiable at* $p = \varphi(u) \in \widetilde{\mathbb{M}}$. *Then the composition* $\psi \circ \varphi : E \to \widehat{\mathbb{M}}$ *is hc-differentiable at* u *and*

$$\widehat{D}(\psi \circ \varphi)(u) = \widehat{D}\psi(p) \circ \widehat{D}\varphi(u).$$

Proof. By hypothesis, $d_\infty^{\varphi(u)}(\varphi(v), \widehat{D}\varphi(u)[v]) = o\big(d_\infty^u(u,v)\big)$ as $v \to u$, $v \in E$, and also $d_\infty^{\psi(p)}(\psi(w), \widehat{D}\psi(p)[w]) = o\big(d_\infty^p(p,w)\big)$ as $w \to p$, $w \in F$. We now infer

$$\begin{aligned}
&d_\infty^{\psi(p)}((\psi\circ\varphi)(v), (\widehat{D}\psi(p)\circ\widehat{D}\varphi(u))[v])\\
&\quad\le Q\big(d_\infty^{\psi(p)}(\psi(\varphi(v)), \widehat{D}\psi(p)[\varphi(v)]) + d_\infty^{\psi(p)}(\widehat{D}\psi(p)[\varphi(v)], \widehat{D}\psi(p)[\widehat{D}\varphi(u)(v)])\big)\\
&\quad\le o\big(d_\infty^p(p,\varphi(v))\big) + O\big(d_\infty^p\big(\varphi(v), \widehat{D}\varphi(u)[v]\big)\big)\\
&\quad\le o\big(d_\infty^u(u,v)\big) + O\big(o\big(d_\infty^u(u,v)\big)\big) = o\big(d_\infty^u(u,v)\big) \quad \text{as } v\to u,
\end{aligned}$$

since $p = \varphi(u)$ and

$$\begin{aligned}
d_\infty^p\big(p,\varphi(v)\big) &\le Q\big(d_\infty^p\big(p, \widehat{D}\varphi(u)[v]\big) + d_\infty^p\big(\varphi(v), \widehat{D}\varphi(u)[v]\big)\big)\\
&= O\big(d_\infty^u(u,v)\big) + o\big(d_\infty^u(u,v)\big) = O\big(d_\infty^u(u,v)\big) \quad \text{as } v\to u.
\end{aligned}$$

The estimate $d_\infty^p\big(p, \widehat{D}\varphi(u)[v]\big) = O\big(d_\infty^u(u,v)\big)$ as $v \to u$ follows from the continuity of the homomorphism $\widehat{D}\varphi(u)$ (see [151]) and (3.1.2). □

3.2. hc-differentiability of curves on Carnot–Carathéodory spaces

3.2.1. Coordinate hc-differentiability criterion. Recall that a mapping $\gamma : E \to \mathbb{M}$, where $E \subset \mathbb{R}$ is an arbitrary set, is called a *Lipschitz mapping* if there exists a constant L such that the inequality $d_\infty(\gamma(y), \gamma(x)) \le L|y-x|$ holds for all $x, y \in E$.

Definition 3.2.1. A mapping $\gamma : E \to \mathbb{M}$, where $E \subset \mathbb{R}$ is an arbitrary set, is called *hc-differentiable at a limit point $s \in E$ of E* if there exists a horizontal vector $a = \sum\limits_{i=1}^{\dim H_1} \alpha_i \widehat{X}_i^{\gamma(s)}(\gamma(s)) \in H_{\gamma(s)}\mathbb{M}$ such that the local homomorphism $\tau \mapsto \exp\Big(\tau \sum\limits_{i=1}^{\dim H_1} \alpha_i \widehat{X}_i^{\gamma(s)}\Big)(\gamma(s)) \in \mathcal{G}^{\gamma(s)}\mathbb{M}$ is the hc-differential of the mapping $\gamma : E \to \mathbb{M}$, i.e., $d_\infty^{\gamma(s)}\big(\gamma(s+\tau), \delta_\tau^{\gamma(s)} a\big) = o(\tau)$ for $\tau \to 0$, $s+\tau \in E$. The point $\exp\Big(\sum\limits_{i=1}^{\dim H_1} \alpha_i \widehat{X}_i^{\gamma(s)}\Big)(\gamma(s)) \in \mathcal{G}^{\gamma(s)}\mathbb{M}$ is called the *hc-derivative*[1].

Some properties of the introduced notion of hc-differentiability can be obtained from Proposition 3.1.3. For instance, the coefficients α_i are defined uniquely: if, in the normal coordinates, $\gamma(s+\tau) = \exp\Big(\sum\limits_{i=1}^{N} \gamma_i(\tau)\widehat{X}_i^{\gamma(s)}\Big)(\gamma(s))$, $s+\tau \in E$, for sufficiently small τ then Proposition 3.1.3 implies:

Property 3.2.2 ([141]**).** A mapping $\gamma : [a,b] \to \mathbb{M}$ is hc-differentiable at a point $s \in (a,b)$ if and only if one of the following assertions holds:

(1) $\gamma_i(\tau) = \alpha_i\tau + o(\tau)$, $i = 1, \dots, \dim H_1$, and $\gamma_i(\tau) = o(\tau^{\deg X_i})$, $i > \dim H_1$, as $\tau \to 0$, $s+\tau \in E$;

[1] See Remark 2.1.23 for the cases of $C^{1,\alpha}$-smooth basis vector fields, $\alpha \in [0,1]$.

(2) the vector $\sum\limits_{i=1}^{\dim H_1} \alpha_i \widehat{X}_i^{\gamma(s)}(\gamma(s)) \in H_{\gamma(s)}\mathbb{M}$ is the Riemannian derivative of $\gamma : [a,b] \to \mathbb{M}$ at a point $s \in (a,b)$, and $\gamma_i(\tau) = o(\tau^{\deg X_i})$, $i > \dim H_1$, as $\tau \to 0$, $s+\tau \in E$.

Proof. Really, consider an arbitrary real number τ of a compact neighborhood of 0 such that $s+\tau \in E$ and any sequence t_n going to 0 as $n \to \infty$ such that $s+t_n\tau \in E$. Then, by 3.1.3, $\Delta^{\gamma(s)}_{t_n^{-1}}\gamma(s+t_n\tau) = \exp\Bigl(\sum\limits_{i=1}^{N} \frac{\gamma_i(t_n\tau)}{t_n^{\deg X_i}} X_i^{t_n}\Bigr)(\gamma(s))$, $s+t_n\tau \in E$, has to go to $\exp\Bigl(\sum\limits_{i=1}^{\dim H_1} \tau\alpha_i \widehat{X}_i^{\gamma(s)}\Bigr)(\gamma(s))$, $s+\tau \in E$. As far as, by Corollary 2.2.13, $X_i^{t_n}$ converges uniformly to $\widehat{X}_i^{\gamma(s)}$ we derive $\dot\gamma_i(s) = \alpha_i$ for $i = 1,\dots,\dim H_1$ and $\gamma_i(\tau) = o(\tau^{\deg X_i})$ as $\tau \to 0$, $s+\tau \in E$, for $i > \dim H_1$. □

3.2.2. ***hc*-differentiability of absolutely continuous curves.** If a curve $\gamma : [a,b] \to \mathbb{M}$ is absolutely continuous in Riemannian sense then all coordinate functions $\gamma_i(t)$ are absolutely continuous on the closed interval $[a,b]$ (it is clear that this property is independent of the choice of the coordinate system). Therefore the tangent vector $\dot\gamma(t)$ is defined almost everywhere on $[a,b]$. If, moreover, $\dot\gamma(t) \in H_{\gamma(t)}\mathbb{M}$ at the points $t \in [a,b]$ of Riemannian differentiability then the curve $\gamma : [a,b] \to \mathbb{M}$ is called *horizontal*.

It is well known that almost all points t of a closed interval $E = [a,b]$ are Lebesgue points of the derivatives of the horizontal components, that is, if, in the normal coordinates $\gamma(t+\tau) = \exp\Bigl(\sum\limits_{j=1}^{N} \gamma_j(\tau)X_j\Bigr)(\gamma(t))$, $t+\tau \in E$, $\tau \in (-\varepsilon,\varepsilon)$ for some $\varepsilon > 0$, then the *horizontal* components $\gamma_j(\sigma)$, $j = 1,\dots,\dim H_1$, have the property

$$\int\limits_{\{\sigma\in(\alpha,\beta)\,|\,t+\sigma\in E\}} |\dot\gamma_j(\sigma) - \dot\gamma_j(0)|\,d\sigma = o(\beta-\alpha) \quad \text{as } \beta-\alpha \to 0 \tag{3.2.1}$$

on intervals $(\alpha,\beta) \ni 0$. Note that property (3.2.1) is independent of the choice of the coordinate system in a neighborhood of $\gamma(t)$.

Below we formulate some statements on *hc*-differentiability of curves on Carnot–Carathéodory space.

Theorem 3.2.3 (see also [141]**).** *Let a curve $\gamma : [a,b] \to \mathbb{M}$ on a Carnot–Carathéodory space be absolutely continuous in the Riemannian sense and horizontal. Then $\gamma : [a,b] \to \mathbb{M}$ is hc-differentiable almost everywhere: any point $t \in [a,b]$ which is a Lebesgue point of the derivatives of its horizontal components is also a point at which γ is hc-differentiable. If $\gamma(t+\tau) = \exp\bigl(\sum\limits_{j=1}^{N} \gamma_j(\tau)X_j\bigr)(\gamma(t))$, where $\tau \in$*

$(-\varepsilon,\varepsilon)$ for ε small enough, then hc-derivative $\dot\gamma(t)$ equals

$$\exp\Bigl(\sum_{j=1}^{\dim H_1}\dot\gamma_j(0)\widehat{X}_j^{\gamma(t)}\Bigr)(\gamma(t))=\exp\Bigl(\sum_{j=1}^{\dim H_1}\dot\gamma_j(0)X_j\Bigr)(\gamma(t)).$$

Proof. Fix a Lebesgue point $t_0\in(a,b)$ of the derivatives of the horizontal components of the mapping $\gamma(t_0+\tau)=\exp\Bigl(\sum\limits_{j=1}^{N}\gamma_j(\tau)X_j\Bigr)(\gamma(t_0))$. Put $u=\gamma(t_0)$. In this proof, we also fix a normal coordinate system θ_u at u. To simplify the notation, we write the vector fields $X_i'^u=(\theta_u^{-1})_*X_i$ and $\widehat{X}_i'^u=(\theta_u^{-1})_*\widehat{X}_i^u$ defined in a neighborhood of $0\in\mathbb{R}^N$ without the superscript u: $X_i'=(\theta_u^{-1})_*X_i$ and $\widehat{X}_i'=(\theta_u^{-1})_*\widehat{X}_i^u$ respectively.

For proving the hc-differentiability of the mapping γ at t_0, we need to establish the estimate $\gamma_j(\tau)=o(\tau^{\deg X_j})$ as $\tau\to0$ for all $j>\dim H_1$, $t_0+\tau\in[a,b]$ (see Property 3.2.2). The proof given below is new with those on Carnot groups [118] and on Carnot manifolds with some restrictions on a system of given vector fields [105]. The main goal of the proof is to obtain the following behaviour of the derivative $\dot\gamma_j(\tau)$: if $\deg X_j=l>1$ then $\dot\gamma_j(\tau)=o(\tau^{l-1})$ as $\tau\to0$ (it follows from (3.2.11)). After that the Newton–Leibnitz formula does job. Partition the proof of the desired estimate into several steps.

First step. Here we show that the hypothesis implies the Riemannian differentiability of the mapping γ at t_0, and $\dot\gamma(t_0)\in H_u\mathbb{M}$. Put $\Gamma(\tau)=\theta_u^{-1}(\gamma(t_0+\tau))=(\gamma_1(\tau),\dots,\gamma_N(\tau))$. The curve $\Gamma(\tau)$ is absolutely continuous, and its tangent vector $\dot\Gamma(\tau)$ is horizontal in a neighborhood of $0\in T_u\mathbb{M}$ with respect to the vector fields $\{X_i'\}$: $\dot\Gamma(\tau)\in(\theta_u^{-1})_*\langle H_{\gamma(t_0+\tau)}\mathbb{M}\rangle$ for almost all τ since γ is horizontal. From here, for almost all τ sufficiently close to 0, we infer

$$\dot\Gamma(\tau)=\sum_{j=1}^{N}\dot\gamma_j(\tau)\frac{\partial}{\partial x_j}=\sum_{i=1}^{\dim H_1}a_i(\tau)X_i'(\Gamma(\tau)).\tag{3.2.2}$$

The Riemann tensor pulled back from the manifold $\mathbb{M}$ onto a neighborhood of $0\in T_u\mathbb{M}$ is continuous at zero. Therefore, using this continuity and taking into account the horizontality of γ, we see that, for any τ such that $t_0+\tau\in[a,b]$, (3.2.1) implies

$$\begin{aligned}d_{cc}(\gamma(t_0),\gamma(t_0+\tau))&\le c_1\int\limits_{(0,\tau)}|\dot\Gamma(\sigma)|_r\,d\sigma\\&\le c_2\sum_{j=1}^{\dim H_1}\int\limits_{(0,\tau)}(|\dot\gamma_j(\sigma)-\dot\gamma_j(0)|+|\dot\gamma_j(0)|)\,d\sigma=O(\tau)\end{aligned}$$

as $\tau\to0$, where $|\dot\Gamma(\sigma)|_r$ stands for the length of the tangent vector in the pulled-back Riemannian metric. By Proposition 2.8.6 and Remark 2.8.7, we have

$d_\infty(\gamma(t_0), \gamma(t_0+\tau)) = O\big(d_{cc}(\gamma(t_0), \gamma(t_0+\tau))\big)$ as $\tau \to 0$. Therefore the coordinate components $\gamma_j(\tau)$ of the mapping γ satisfy

$$\gamma_j(\tau) = O(\tau^{\deg X_j}) \quad \text{as } \tau \to 0 \text{ for all } j \geq 1. \tag{3.2.3}$$

It follows that the curve $\Gamma(\tau)$ is differentiable at 0 and

$$\dot{\Gamma}(0) = (\dot{\gamma}_1(0), \dots, \dot{\gamma}_{\dim H_1}(0), 0, \dots, 0).$$

Hence, the curve γ is differentiable in the Riemannian sense at t_0 and $\dot{\gamma}(t_0) \in H_u\mathbb{M}$. From (3.2.3) we also obtain $\gamma(\tau) \in \mathrm{Box}(u, O(\tau))$.

Second step. Corollary 2.2.11 and the fact that $\gamma(\tau) \in \mathrm{Box}(u, O(\tau))$ imply that, in a neighborhood of 0, each vector field X_i' can be expressed via $\{\widehat{X}_k'\}_{k=1}^N$ so that

$$X_i'(\Gamma(\tau)) = \sum_{k=1}^N \alpha_{ik}(\tau)\widehat{X}_k'(\Gamma(\tau)), \text{ where } \alpha_{ik}(\tau) = \begin{cases} o(\tau^{\deg X_k - \deg X_i}) & \text{if} \\ \qquad \deg X_k > \deg X_i, \\ \delta_{ik} + O(\tau) & \text{otherwise} \end{cases}$$

as $\tau \to 0$ (here $\alpha_{ik}(\tau) = a_{i,k}(\Gamma(\tau))$ from Corollary 2.2.11). Now, using expansion (2.1.11) of the vector fields $\widehat{X}_i'$ in the standard Euclidean basis, for all points τ sufficiently close to 0, from (3.2.2) we now obtain

$$\begin{aligned} \sum_{j=1}^N \dot{\gamma}_j(\tau)\frac{\partial}{\partial x_j} &= \sum_{i=1}^{\dim H_1} a_i(\tau)\widetilde{X}_i(\Gamma(\tau)) = \sum_{k=1}^N \sum_{i=1}^{\dim H_1} a_i(\tau)\alpha_{ik}(\tau)\widehat{X}_k'(\Gamma(\tau)) \\ &= \sum_{j=1}^N \sum_{k=1}^j \sum_{i=1}^{\dim H_1} a_i(\tau)\alpha_{ik}(\tau) z_k^j(u, \Gamma(\tau))\frac{\partial}{\partial x_j}. \end{aligned} \tag{3.2.4}$$

Third step. For $1 \leq j \leq \dim H_1$, we have $\deg X_j = 1$. Then from (2.1.11) and (3.2.3) we conclude that $z_k^j(u, \Gamma(\tau)) = \delta_{jk} + O(\tau)$. Therefore, from (3.2.4) we infer

$$\dot{\gamma}_j(\tau) = \sum_{k=1}^j \sum_{i=1}^{\dim H_1} a_i(\tau)(\delta_{ik} + O(\tau))(\delta_{jk} + O(\tau)) = \sum_{i=1}^{\dim H_1} a_i(\tau)\tilde{\alpha}_{ij}(\tau),$$

where, as before, $\tilde{\alpha}_{ij}(\tau) = \delta_{ij} + O(\tau)$.

Hence,

$$a_i(\tau) = \sum_{q=1}^{\dim H_1} \dot{\gamma}_q(\tau)\beta_{qi}(\tau), \tag{3.2.5}$$

where $\{\beta_{qi}(\tau)\}$, $q, i = 1, \dots, \dim H_1$, is a matrix inverse to $\{\tilde{\alpha}_{ij}(\tau)\}$, has the elements $\beta_{qi}(\tau) = \delta_{qi} + O(\tau)$. Consequently,

$$a_i(\tau) = \sum_{i=1}^{\dim H_1} \dot{\gamma}_q(\tau)\beta_{qi}(\tau) = \sum_{q=1}^{\dim H_1} \dot{r}_q(\tau)\beta_{qi}(\tau) + \sum_{q=1}^{\dim H_1} \dot{\gamma}_q(0)\beta_{qi}(\tau),$$

$$\text{where} \qquad r_q(\tau) = \int_0^\tau (\dot{\gamma}_q(\sigma) - \dot{\gamma}_q(0))\, d\sigma. \tag{3.2.6}$$

Fourth step. Fix $\dim H_{l-1} < j \le \dim H_l$, $1 < l \le M$. For estimating $\dot{\gamma}_j(\tau)$, we replace $a_i(\tau)$ in (3.2.4) by (3.2.5) to come to

$$\dot{\gamma}_j(\tau) = \sum_{k,i,q=1}^{\dim H_1} \dot{\gamma}_q(\tau)\beta_{qi}(\tau)\alpha_{ik}(\tau)z_k^j(u, \Gamma(\tau)) \tag{3.2.7}$$

$$+ \sum_{k=\dim H_1+1}^{j} \sum_{i,q=1}^{\dim H_1} \dot{\gamma}_q(\tau)\beta_{qi}(\tau)\alpha_{ik}(\tau)z_k^j(u, \Gamma(\tau)) = I_j + II_j,$$

where

$$I_j = \sum_{k,i,q=1}^{\dim H_1} \dot{\gamma}_q(\tau)\beta_{qi}(\tau)\alpha_{ik}(\tau)z_k^j(u, \Gamma(\tau))$$

and

$$II_j = \sum_{k=\dim H_1+1}^{j} \sum_{i,q=1}^{\dim H_1} \dot{\gamma}_q(\tau)\beta_{qi}(\tau)\alpha_{ik}(\tau)z_k^j(u, \Gamma(\tau)).$$

From the one hand, since in this case we have $\deg X_k > \deg X_i$, then consequently $\alpha_{ik}(\tau) = o(\tau^{\deg X_k - \deg X_i})$. Next,

$$z_k^j(u, \Gamma(\tau)) = O(\tau^{\deg X_j - \deg X_k}) \tag{3.2.8}$$

in view of the fact that $\Gamma(\tau) \in \mathrm{Box}(0, O(\tau))$. From here, taking into account that $\deg X_j = l$, we deduce that all the components in the double sum II_j have a factor $o(\tau^{l-1})$. Therefore

$$II_j = \sum_{q=1}^{\dim H_1} \dot{\gamma}_q(\tau)o(\tau^{l-1}). \tag{3.2.9}$$

From the other hand, in view of Corollary 2.2.11, (2.1.11) and (3.2.8) we infer

$$I_j = \sum_{q=1}^{\dim H_1} \dot{\gamma}_q(\tau)z_q^j(u, \Gamma(\tau)) + \sum_{k,q=1}^{\dim H_1} \dot{\gamma}_q(\tau)o(1)z_k^j(u, \Gamma(\tau)) \tag{3.2.10}$$

$$= \sum_{q=1}^{\dim H_1} \dot{\gamma}_q(0)z_q^j(u, \Gamma(\tau)) + \sum_{q=1}^{\dim H_1} \dot{r}_q(\tau)z_q^j(u, \Gamma(\tau)) + \sum_{k,q=1}^{\dim H_1} \dot{\gamma}_q(\tau)o(1)z_k^j(u, \Gamma(\tau))$$

$$= \sum_{q=1}^{\dim H_1} \dot{\gamma}_q(0) \sum_{|\mu+e_q|_h=\deg X_j,\, \mu>0} F^j_{\mu,e_q}(u)\Gamma(\tau)^\mu$$
$$+ \sum_{q=1}^{\dim H_1} \dot{r}_q(\tau)O(\tau^{l-1}) + \sum_{q=1}^{\dim H_1} \dot{\gamma}_q(\tau)o(\tau^{l-1}).$$

In the estimation of the increment of $\gamma_j(\tau)$ on $[0,\tau]$ by the Newton–Leibnitz formula, the components of (3.2.9) and the last two summands in (3.2.10) have order $o(\tau^l)$. Indeed, for all $1 \le q \le \dim H_1$ and $s > 0$, from (3.2.1) and (3.2.6) we have $|\dot{\gamma}_q(\tau)| \le |\dot{\gamma}_q(0)| + |\dot{r}_q(\tau)|$ from (3.2.6), $|r_q(\tau)| \le \int\limits_0^\tau |\dot{\gamma}_q(\sigma) - \dot{\gamma}_q(0)|\, d\sigma = o(\tau)$ and

$$\biggl|\int\limits_0^\tau \dot{r}_q(\sigma)O(\sigma^s)\, d\sigma\biggr| \le |O(\tau^s)| \int\limits_0^\tau |\dot{\gamma}_q(\sigma) - \dot{\gamma}_q(0)|\, d\sigma = o(\tau^{s+1}).$$

Fifth step. In the remaining double sum in (3.2.10), the summands with index μ for which $|\mu + e_q| < \deg X_j$ contain the factor $\Gamma(\tau)^\mu = o(\tau^{l-1})$, since, in this case, the product $\Gamma(\tau)^\mu$ necessarily contains the factor $\gamma_j(\tau) = \dot{\gamma}_j(0)\tau + o(\tau) = o(\tau)$, $j > \dim H_1$. Therefore, expression (3.2.10) for $\dot{\gamma}_j(\tau)$ is reduced to the following:

$$\dot{\gamma}_j(\tau) = \sum_{q=1}^{\dim H_1} \dot{\gamma}_q(0) \sum_{\substack{|\mu+e_q|_h=\deg X_j,\\ |\mu+e_q|=\deg X_j}} F^j_{\mu,e_q}(u)\Gamma(\tau)^\mu + o(\tau^{l-1}). \tag{3.2.11}$$

Since also $\Gamma(\tau) = \dot{\Gamma}(0)\tau + o(\tau)$, we see that each summand in (3.2.11) is equal to $\dot{\gamma}_q(0)F^j_{\mu,e_q}(u)\Gamma(\tau)^\mu = \tau^{l-1}\dot{\gamma}_q(0)F^j_{\mu,e_q}(u)\dot{\Gamma}(0)^\mu + o(\tau^{l-1})$. Consequently, (3.2.11) can be written as

$$\dot{\gamma}_j(\tau) = \sum_{q=1}^{\dim H_1} \tau^{l-1} \sum_{|\mu|=|\mu|_h=l-1} \dot{\gamma}_q(0)F^j_{\mu,e_q}(u)\dot{\Gamma}(0)^\mu + o(\tau^{l-1}). \tag{3.2.12}$$

Similarly, the second summand in the estimation of the increment of $\gamma_j(\tau)$ is equal to $o(\tau^l)$. Consequently, for the validity of the theorem, it is necessary and sufficient that the double sum in (3.2.12) equals zero. This was established in Lemma 2.1.25.

Thus, we have proved that $\gamma_j(\tau) = o(\tau^{\deg X_j})$ for all $j > \dim H_1$. Since the horizontal components of γ are differentiable at t_0, by Property 3.2.2, the estimate $\gamma_j(\tau) = o(\tau^{\deg X_j})$ for all $j > \dim H_1$ yields the hc-differentiability of γ at t_0. $\square$

The method of proving Theorem 3.2.3 is applicable to a wider class of mappings and makes it possible to make additional conclusions about the nature of hc-differentiability.

Corollary 3.2.4. *Suppose that a curve $\gamma : [a,b] \to \mathbb{M}$ on a Carnot–Carathéodory space is Lipschitz with respect to the Riemannian metric and horizontal, i.e.,*

$\dot\gamma(s) \in H_{\gamma(s)}\mathbb{M}$ for almost every $s \in [a,b]$. Then the curve $\gamma : [a,b] \to \mathbb{M}$ is hc-differentiable almost everywhere[2]*.*

Proof. Every Lipschitz curve with respect to the Riemannian metric is also absolutely continuous in the Riemannian sense. Thus all conditions of Theorem 3.2.3 hold. □

Corollary 3.2.5. *Suppose that we have a family of curves $\gamma : [a,b] \times F \to \mathbb{M}$ on a Carnot–Carathéodory space $\mathbb{M}$ that is bounded and continuous in the totality of its variables, where F is a locally compact metric space. Suppose that, for each fixed $u \in F$, the curve $\gamma(\cdot,u)$ is differentiable in the Riemannian sense at all points of $[a,b]$ and horizontal, i.e., $\frac{d}{ds}\gamma(s,u) \in H_{\gamma(s,u)}\mathbb{M}$ for all $s \in [a,b]$. If the Riemannian derivative $\frac{d}{ds}\gamma(s,u)$ is bounded and continuous in the totality of its variables s and u then its hc-derivative is also bounded and continuous on $[a,b] \times F$. Furthermore, the convergence $\Delta^{\gamma(s)}_{\tau^{-1}}\gamma(s+\tau,u)$ to $\dot\gamma(s,u) \in \mathcal{G}^{\gamma(s,u)}\mathbb{M}$ is locally uniform in the totality of $s \in [a,b]$ and $u \in F$.*

Proof. It suffices to prove in all items of the proof of Theorem 3.2.3 that the smallness of all quantities converging to zero is locally uniform on $[a,b] \times F$. □

Corollary 3.2.6. *Suppose that a curve $\gamma : [a,b] \to \mathbb{M}$ on a Carnot–Carathéodory space $\mathbb{M}$ belongs to C^1 and its Riemannian tangent vector $\dot\gamma_i(t)$ is horizontal for all $t \in [a,b]$. Then the curve $\gamma : [a,b] \to \mathbb{M}$ is hc-differentiable at all $t \in [a,b]$. Furthermore, the convergence of $\Delta^{\gamma(s)}_{\tau^{-1}}\gamma(s+\tau)$ to $\dot\gamma(s) \in \mathcal{G}^{\gamma(s)}\mathbb{M}$ is uniform in $s \in [a,b]$.*

Proof. For any $x, y \in [a,b]$, the length $L(\gamma|_{[x,y]})$ of the curve $\gamma : [x,y] \to \mathbb{M}$ is defined; moreover, $d_\infty(\gamma(y),\gamma(x)) \le c_1 L(\gamma|_{[x,y]}) \le c_1 C|y-x|$, where $C = \max\limits_{t\in[a,b]} |\dot\gamma(t)|$. Thus, the curve $\gamma : [a,b] \to \mathbb{M}$ meets the conditions of Theorem 3.2.3 at all points of $[a,b]$ and, therefore, is uniformly *hc*-differentiable by Corollary 3.2.5 (in this case F can be considered as one-point set). The corollary follows. □

Lemma 3.2.7. *Let $\mathbb{M}$ be a Carnot–Carathéodory space. Every Lipschitz (with respect to d_∞) mapping $\gamma : E \to \mathbb{M}$, $E \subset \mathbb{R}$, is differentiable almost everywhere in the Riemannian sense, and $\dot\gamma(t) \in H_{\gamma(t)}\mathbb{M}$ at the points of the Riemannian differentiability of γ.*

Proof. In the normal coordinates at a point $u = \gamma(t)$, we have

$$\gamma(t+\tau) = \exp\biggl(\sum_{j=1}^{N} \gamma_j(\tau)\, X_j\biggr)(u), \quad t+\tau \in E.$$

The Lipschitzity with respect to d_∞ of the mapping $\gamma : E \to \mathbb{M}$ and the properties of d_∞ imply the estimate $\gamma_j(\tau) = O(\tau^{\deg X_j})$ for all $j \ge 1$, $t+\tau \in E$. Since $\deg X_j \ge 2$ for $j > \dim H_1$, the derivative $\dot\gamma_j(0)$ exists and equals zero for all

[2]In papers [141, 142], a wrong Corollary 3.1 is formulated instead of this.

such j. Consequently, the Riemannian differentiability of γ at t is equivalent to the differentiability of the horizontal components γ_j, $j = 1, \dots, n$, of γ at 0.

Now, the Lipschitz mapping $\gamma : E \to \mathbb{M}$ is also Lipschitz with respect to the Riemannian metric (see Proposition 2.2.2). Thus, by Rademacher's classical theorem, the Riemannian derivative $\dot{\gamma}(t) \in T_{\gamma(t)}\mathbb{M}$ exists for almost every $t \in [a, b]$. The above implies that, at every such point, $\dot{\gamma}(t) \in H_{\gamma(t)}\mathbb{M}$. □

Since a Lipschitz with respect to d_∞ mapping $\gamma : [a, b] \to \mathbb{M}$ is absolutely continuous in the Riemannian sense (see the comparison of the metrics in Proposition 2.2.2), from Lemma 3.2.7 and Theorem 3.2.3 we infer

Corollary 3.2.8. *Let $\mathbb{M}$ be a Carnot–Carathéodory space. Every Lipschitz with respect to d_∞ mapping $\gamma : [a, b] \to \mathbb{M}$ is hc-differentiable almost everywhere on $[a, b]$: if $t \in [a, b]$ is a Lebesgue point of the derivatives of its horizontal components then this point is its hc-differentiability point.*

3.2.3. *hc*-differentiability of scalar Lipschitz mappings. In this subsubsection, we establish the hc-differentiability of Lipschitz mappings $\gamma : E \to \mathbb{M}$, where $E \subset \mathbb{R}$ is an arbitrary set. We assume that a Carnot-Carathéodory space $\mathbb{M}$ is a Carnot manifold meeting the 4th condition in the definition 2.1.1.

Recall that $x \in A$, where $A \subset \mathbb{R}$ is a measurable set, is the density point of A if

$$|A \cap (\alpha, \beta)|_1 = \beta - \alpha + o(\beta - \alpha) \quad \text{for } \beta - \alpha \to 0,\ x \in (\alpha, \beta) \tag{3.2.13}$$

(here $|\cdot|_1$ stands for the one-dimensional Lebesgue measure). It is known that almost all points of a measurable set A are its density points (see, for example, [42]).

It is explicitly seen from the above proof of Lemma 3.2.7 that the answer to the question on hc-differentiability of Lipschitz with respect to d_∞ mapping depends on the differentiability of its horizontal components. If a Lipschitz with respect to d_∞ mapping $\gamma : E \to \mathbb{M}$ (we may assume that $E \subset \mathbb{R}$ is closed) is written in the normal coordinates: $\gamma(t + \tau) = \exp\Big(\sum\limits_{j=1}^{N} \gamma_j(\tau)\, X_j\Big)(\gamma(t))$, $t \in E$ is a fixed number, $t + \tau \in E$, then, by Lemma 3.2.7, its components $\gamma_j(\tau)$, $j = 1, \dots, N$, are differentiable almost everywhere on E. It is known that almost all density points of E are Lebesgue points of the derivative of the horizontal components (see, for example, Lemma 3.2.7 and [42]), i.e., for intervals $(\alpha, \beta) \ni \tau$, $t + \tau \in E$, we infer

$$\int\limits_{\{\sigma \in (\alpha, \beta) \mid t + \sigma \in E\}} |\dot{\gamma}_j(\sigma) - \dot{\gamma}_j(\tau)|\, d\sigma = o(\beta - \alpha) \quad \text{for } \beta - \alpha \to 0 \tag{3.2.14}$$

for all $j = 1, \dots, \dim H_1$. Note that Property (3.2.14) does not depend on the choice of the coordinate system in a neighborhood of the point $u = \gamma(t)$.

Theorem 3.2.9 ([141]). *Let $\mathbb{M}$ be a Carnot manifold. Every Lipschitz with respect to d_∞ mapping $\gamma : E \to \mathbb{M}$, where $E \subset \mathbb{R}$ is closed, is hc-differentiable almost everywhere on E. Namely, the mapping $\gamma : E \to \mathbb{M}$ is hc-differentiable at every* *point* $t \in E$ *such that*

1) t *is a density point of* E;
2) *there exist derivatives* $\dot\gamma_j(0)$, $j = 1, \dots, \dim H_1$, *of the horizontal components of* γ, *where* $\gamma(t+\tau) = \exp\Bigl(\sum\limits_{j=1}^{N} \gamma_j(\tau) X_j\Bigr)(\gamma(t))$, $t + \tau \in E$;
3) *condition* (3.2.14) *is fulfilled at the point* $\tau = 0$.

The hc-derivative $\widehat{D}\gamma(t)$ *equals*

$$\exp\Biggl(\sum_{j=1}^{\dim H_1} \dot\gamma_j(0)\widehat{X}_j^{\gamma(t)}\Biggr)(\gamma(t)) = \exp\Biggl(\sum_{j=1}^{\dim H_1} \dot\gamma_j(0)X_j\Biggr)(\gamma(t)).$$

Proof. On a Carnot group $\mathbb{G}$ the theorem is proved in [137]. The method used in [137] is based on an extension of a given Lipschitz mapping $\gamma : E \to \mathbb{G}$, where $E \subset \mathbb{R}$ is closed, to a Lipschitz mapping $\widetilde{\gamma} : [a, b] \to \mathbb{G}$ where $[a, b] \supset E$ is a closed segment. On this way we have to apply Rashevskiĭ–Chow Theorem on a horizontal path with given endpoints. Under condition of Theorem 3.2.9 we do not have this opportunity. Therefore the sketch of the proof, given in [137], is essentially modified: instead of horizontal paths in the Carnot manifolds we connect given two points by a horizontal path in a local Carnot group associated with one point of every pair (because of this, local Carnot groups are different one from another). It is clear that this construction does not give a Lipschitz curve with respect to d_∞-quasimetric (therefore we are not able to reduce the proof to Theorem 3.2.3). But it will be a Lipschitz curve with respect to Riemannian metric, and the Newton–Leibnitz formula can be applied. Estimates similar to those in Theorem 3.2.3 provide the *hc*-differentiability of the given curve in a prescribed point.

Suppose that $t \in E$ is a point at which conditions 1–3 of the theorem hold and $u = \gamma(t)$. Since the result is local, we may also assume that E is included in an interval $[a, b] \subset \mathbb{R}$, $t \in [a, b]$, $a, b \in E$, whose image is included in $\mathcal{G}^u\mathbb{M}$ (we may assume by diminishing the interval $[a, b]$ if necessary that $\gamma([a, b] \cap E) \subset \mathcal{G}^{\gamma(\eta)}\mathbb{M}$ for every $\eta \in [a, b] \cap E$).

First step. The open bounded set $Z = (a, b) \setminus E$ is representable as the union of an at most countable collection of disjoint intervals: $Z = \bigcup_j (\alpha_j, \beta_j)$, where, for convenience of the subsequent estimates, we put $\alpha_j < \beta_j$ if $t \le \alpha_j$ and $\beta_j < \alpha_j$ if $\alpha_j < t$. It is known (for example, see [47]), that, in $\mathcal{G}^{\gamma(\alpha_j)}\mathbb{M}$, there exists a horizontal (with respect to the basis $\{\widehat{X}_l^{\gamma(\alpha_j)}\}_{l=1}^N$) curve $\tilde\sigma_j : [0, b_j] \to \mathcal{G}^{\gamma(\alpha_j)}\mathbb{M}$ joining the points $\tilde\sigma_j(0) = \gamma(\alpha_j)$ and $\tilde\sigma_j(b_j) = \gamma(\beta_j)$ and parameterized by the arc length; moreover, $b_j = d_{cc}^{\gamma(\alpha_j)}(\gamma(\alpha_j), \gamma(\beta_j)) \le C d_\infty^{\gamma(\alpha_j)}(\gamma(\alpha_j), \gamma(\beta_j)) = C d_\infty(\gamma(\alpha_j), \gamma(\beta_j)) \le CL|\beta_j - \alpha_j|$ (since γ is Lipschitz with respect to d_∞), where C is independent of j. Consequently, the mapping $\sigma_j : [\alpha_j, \beta_j] \to \mathbb{M}$ defined by the rule

$$[\alpha_j, \beta_j] \ni \eta \mapsto \sigma_j(\eta) = \tilde\sigma_j\Bigl(\frac{b_j}{|\beta_j - \alpha_j|}|\eta - \alpha_j|\Bigr) \in \mathcal{G}^{\gamma(\alpha_j)}\mathbb{M}$$

is Lipschitz in the metric $d_{cc}^{\gamma(\alpha_j)}$ with the Lipschitz constant cL for all $j \in \mathbb{N}$.

Define now the extension $f : [a, b] \to \mathbb{M}$ as follows:

$$f(\eta) = \begin{cases} \gamma(\eta), & \text{if } \eta \in E, \\ \sigma_j(\eta), & \text{if } \eta \in (\alpha_j, \beta_j). \end{cases}$$

Second step. The mapping $f : [a, b] \to \mathbb{M}$ has the following properties (justifications are given below):

(1) $f : [\alpha, \beta] \to \mathbb{M}$ is a Lipschitz mapping with respect to the Riemannian metric;
(2) the Riemannian derivative of f exists for almost every $\eta \in [a, b]$ and is bounded;
(3) the vector $\dot{f}(\eta)$ belongs to the horizontal space $H_{\gamma(\eta)}\mathbb{M}$ for almost every $\eta \in E$;
(4) the mapping $f : [a, b] \to \mathbb{M}$ has a Riemannian derivative at t equal to $\dot{\gamma}(t)$;

if $f(t + \tau) = \exp\Big(\sum\limits_{j=1}^{N} f_j(\tau)\, X_j\Big)(u)$, $t + \tau \in [a, b]$, then

(5) $f_j(\tau) = O(\tau^{\deg X_j})$ as $\tau \to 0$ for all $j \geq 1$;
(6) 0 is a Lebesgue point for the derivatives $\dot{f}_j(\tau)$, $j = 1, \dots, \dim H_1$.

Indeed, if $t \leq \alpha_j < \eta_1 < \beta_j < \alpha_k < \eta_2 < \beta_k \leq b$ then, taking the relations between the metrics into account, we obtain the estimates $\rho(f(\eta_1), f(\eta_2)) \leq C_1(\rho(f(\eta_1), \gamma(\beta_j)) + \rho(\gamma(\beta_j), \gamma(\alpha_k)) + \rho(\gamma(\alpha_k), f(\eta_2))) \leq C_2((\beta_j - \eta_1) + (\alpha_k - \beta_j) + (\eta_2 - \alpha_k)) = C_2|\eta_2 - \eta_1|$. The other cases of mutual disposition of η_1 and η_2 with respect to t are considered similarly. Hence we obtain properties (1) and (2).

Next, if $t \leq \alpha_j < t+\tau < \beta_j$ then $d_\infty(f(t+\tau), f(t)) \leq Q(d_\infty(f(t+\tau), \gamma(\alpha_j)) + d_\infty(\gamma(\alpha_j), \gamma(t))) \leq Q_1\big(d_\infty^{\gamma(\alpha_j)}(f(t+\tau), \gamma(\alpha_j)) + (\alpha_j - t)\big) = Q_2((t+\tau-\alpha_j) + (\alpha_j - t)) = Q_2\tau$ by the generalized triangle inequality, the construction of f, Lipschitzity of γ, and the relations between the metrics. From this we obtain Property (5) and, hence, the differentiability of all components f_j at 0, $j > \dim H_1$: $\dot{f}_j(0) = 0$.

Since the derivatives of Lipschitz functions are bounded and t is the density point of E, for intervals $(r, s) \ni 0$ we have

$$\int\limits_{(r,s)} |\dot{f}_j(\sigma) - \dot{\gamma}_j(0)|\, d\sigma = \int\limits_{\{\sigma\in(r,s) \,|\, t+\sigma\in E\cap[a,b]\}} |\dot{\gamma}_j(\sigma) - \dot{\gamma}_j(0)|\, d\sigma + \int\limits_{\{\sigma\in(r,s) \,|\, t+\sigma\notin E\cap[a,b]\}} |\dot{f}_j(\sigma) - \dot{\gamma}_j(0)|\, d\sigma = o(|s - r|) \tag{3.2.15}$$

as $s-r \to 0$ for all $j = 1, \dots, \dim H_1$ (the first (second) integral is $o(|s-r|)$ because of (3.2.14) ((3.2.13))). Hence, $\int\limits_0^\tau (\dot{f}_j(\sigma) - \dot{\gamma}_j(0))\, d\sigma = f_j(\tau) - \dot{\gamma}_j(0)\tau = o(\tau)$ and $\frac{df_j}{d\tau}(0) = \dot{\gamma}_j(0)$ for all $j = 1, \dots, \dim H_1$ (for negative τ, such estimate is obtained similarly). Thus, we have proved properties (4) and (6).

Note that the preceding arguments are independent of the coordinate system. They are based on the following principle: if η is the density point for E, the mapping $f|_E$ has a Riemannian derivative at $\eta \in E$, and $\eta \in E$ is a Lebesgue point for the horizontal coordinate functions of $f|_E$ then, in view of Lemma 3.2.7 and what has been proved above, f has a Riemannian derivative at η; moreover, the Riemannian tangent vector belongs to the horizontal space $H_{\gamma(\eta)}\mathbb{M}$. This proves Property (3).

Third step. Since the Riemannian derivative $\dot{f}(\eta)$ of the mapping $f : [a, b] \to \mathbb{M}$ belongs to the horizontal space $H_{f(\eta)}\mathbb{M}$ only at almost every point $\eta \in E$, and E is an arbitrary measurable set but not necessarily a closed interval, a direct application of Theorem 3.2.3 is impossible. However, granted the fact that the complement $[a, b] \setminus E$ has density zero at t, the method of its proof can be adapted also to this case. We now indicate the changes to the proof of Theorem 3.2.3 necessary for obtaining the hc-differentiability of f at the point t fixed above.

Introduce the notation

$$\Gamma(\tau) = \begin{cases} (\gamma_1(\tau), \dots, \gamma_N(\tau)), & \text{if } t + \tau \in E, \\ (f_1(\tau), \dots, f_N(\tau)), & \text{if } t + \tau \notin E. \end{cases}$$

It has been proved above that $\dot{\Gamma}(0) = (\dot{\gamma}_1(0), \dots, \dot{\gamma}_{\dim H_1}(0), 0, \dots, 0)$. Deduce a representation like (3.2.2) for the points τ sufficiently close to 0 and such that $t + \tau \in E$. At the points $t + \tau \in (\alpha_j, \beta_j)$, we have (due to the construction of f)

$$\dot{\Gamma}(\tau) = \sum_{j=1}^{N} \dot{f}_j(\tau) \frac{\partial}{\partial x_j} = \sum_{i=1}^{\dim H_1} a_i(\tau) (\widehat{X}_i^{f(\alpha_j)})'(\Gamma(\tau)). \tag{3.2.16}$$

According to Theorem 2.2.8, at the points $t+\tau \in (\alpha_j, \beta_j)$ the relation $f(\tau) \in B(u, O(\tau))$ implies that, in a neighborhood of 0, the vector fields $(\widehat{X}_i^{f(\alpha_j)})'$ are expressed via the vector fields $\widehat{X}'_k$ (here we write $\widehat{X}'_k$ instead of $(\widehat{X}_k^u)'$) in the form

$$(\widehat{X}_i^{f(\alpha_j)})'(\Gamma(\tau)) = \sum_{k=1}^{N} \gamma_{ik}(\tau) \widehat{X}'_k(\Gamma(\tau)),$$

where

$$\gamma_{ik}(\tau) = \begin{cases} o(\tau^{\deg X_k - \deg X_i}), \text{ if } \deg X_k > \deg X_i, \\ \delta_{ik} + O(\tau) \text{ otherwise} \end{cases}$$

as $\tau \to 0$. Indeed, by (2.2.7), we have $\widehat{X}_i'^{f(\alpha_j)}(\Gamma(\tau)) = \sum\limits_{l=1}^{N} \beta_{il}(\tau) X'_l(\Gamma(\tau))$ at points $f(\tau) \in B(u, O(\tau))$, where

$$\beta_{il}(\tau) = \begin{cases} o(\tau^{\deg X_l - \deg X_i}) & \text{if } \deg X_l > \deg X_i, \\ \delta_{il} + O(\tau) & \text{otherwise} \end{cases} \tag{3.2.17}$$

as $\tau \to 0$, and $X_l'(\Gamma(\tau)) = \sum\limits_{k=1}^{N} \alpha_{lk}(\tau)\widehat{X}_k'(\Gamma(\tau))$ where

$$\alpha_{lk}(\tau) = \begin{cases} o(\tau^{\deg X_k - \deg X_l}) & \text{if } \deg X_k > \deg X_l, \\ \delta_{ik} + O(\tau) & \text{otherwise} \end{cases} \tag{3.2.18}$$

as $\tau \to 0$. It follows $\widehat{X}_i'^{f(\alpha_j)}(\Gamma(\tau)) = \sum\limits_{k=1}^{N}\sum\limits_{l=1}^{N} \beta_{il}(\tau)\alpha_{lk}(\tau)\widehat{X}_k'(\Gamma(\tau))$. Now taking into account (3.2.17) and (3.2.18), we consider two cases for getting the desired asymptotic behaviour of $\gamma_{ik}(\tau)$ as $\tau \to 0$.

Case I: $\deg X_k \leq \deg X_i$. Then

$$\begin{aligned} \gamma_{ik}(\tau) &= \sum_{l=1}^{N} \beta_{il}(\tau)\alpha_{lk}(\tau) = \sum_{l:\deg X_l < \deg X_k} \beta_{il}(\tau)\alpha_{lk}(\tau) \\ &\quad + \sum_{l:\deg X_k \leq \deg X_l \leq \deg X_i} \beta_{il}(\tau)\alpha_{lk}(\tau) + \sum_{l:\deg X_i < \deg X_l} \beta_{il}(\tau)\alpha_{lk}(\tau) \\ &= \sum_{l:\deg X_l < \deg X_k} (\delta_{il} + O(\tau)) \cdot o(\tau^{\deg X_k - \deg X_l}) \\ &\quad + \sum_{l:\deg X_k \leq \deg X_l \leq \deg X_i} (\delta_{il} + O(\tau)) \cdot (\delta_{lk} + O(\tau)) \\ &\quad + \sum_{l:\deg X_i < \deg X_l} o(\tau^{\deg X_l - \deg X_i}) \cdot (\delta_{lk} + O(\tau)) = \delta_{ik} + O(\tau). \end{aligned} \tag{3.2.19}$$

Case II: $\deg X_k > \deg X_i$. Then representing the sum for $\gamma_{ik}(\tau)$ in (3.2.19) as $\sum\limits_{l:\deg X_l \leq \deg X_k} + \sum\limits_{l:\deg X_i < \deg X_l < \deg X_k} + \sum\limits_{l:\deg X_i \leq \deg X_l}$ we obtain

$$\begin{aligned} \gamma_{ik}(\tau) &= \sum_{l:\deg X_l \leq \deg X_i} o(\tau^{\deg X_k - \deg X_l}) + \sum_{l:\deg X_k \leq \deg X_l} o(\tau^{\deg X_l - \deg X_i}) \\ &\quad + \sum_{l:\deg X_i < \deg X_l < \deg X_k} o(\tau^{\deg X_l - \deg X_i}) \cdot o(\tau^{\deg X_k - \deg X_l}) \\ &= o(\tau^{\deg X_k - \deg X_i}) \end{aligned}$$

as $\tau \to 0$. Consequently, we have just qualitative situation similar to those on the third step of the proof of Theorem 3.2.3. Thus, the further proof repeats verbatim the 3RD, the 4TH and the 5TH Steps of the proof of Theorem 3.2.3 with f instead of γ. Thus, the theorem follows. □

3.2.4. ***hc*-differentiability of rectifiable curves.** In this section, we in particular prove that, in a Carnot manifold, rectifiable curves are *hc*-differentiable almost everywhere. We obtain this result as a corollary to the more general assertion about the *hc*-differentiability of a mapping $\varphi : E \to \mathbb{M}$ from a measurable set

$E \subset \mathbb{R}$ that satisfies the condition

$$\overline{\lim_{y\to x,\, y\in E}} \frac{d_\infty(\varphi(y),\varphi(x))}{|y-x|} < \infty \qquad (3.2.20)$$

for almost all $x \in E$.

Theorem 3.2.10 ([141]). *Let $\mathbb{M}$ be a Carnot manifold. Every mapping $\varphi : E \to \mathbb{M}$, where $E \subset \mathbb{R}$ is a measurable set, satisfying* (3.2.20) *is hc-differentiable almost everywhere in E.*

This theorem is a particular case of Theorem 3.3.6 (see its proof below).

Now we can prove the *hc*-differentiability of rectifiable curves. Consider a curve (continuous mapping) $\gamma : [a,b] \to \mathbb{M}$. By a partition $I_n = I_n([a,b])$ of the segment $[a,b]$ we mean any finite sequence of points $\{s_1,\dots,s_n\}$ with $a = s_1 < \dots < s_n = b$. To every partition $I_n([a,b])$, we assign a number $M(I_n)$ by setting

$$M(I_n) = \sum_{i=1}^{n-1} d_\infty(\gamma(s_i),\gamma(s_{i+1})).$$

Put $m_n = \max\{s_{i+1} - s_i \mid i = 1,\dots,n-1\}$.

Definition 3.2.11 ([22]). A curve $\gamma : [a,b] \to \mathbb{M}$ is called *rectifiable* if

$$L([a,b]) = \lim_{m_n\to 0} \sup_{I_n} M_n < \infty.$$

Making use of standard arguments (see, for instance, [22]), we may prove:

Property 3.2.12. Suppose that a sequence of curves $\gamma_q : [a,b] \to \mathbb{M}$, $q \in \mathbb{N}$, converges pointwise to a curve $\gamma : [a,b] \to \mathbb{M}$: $\gamma_q(s) \to \gamma(s)$ for every $s \in [a,b]$. Then the lengths $L_q([a,b])$ of γ_q possess the semicontinuity property:

$$L([a,b]) \le \underline{\lim}_{q\to\infty} L_q([a,b]).$$

If we have an usual metric but not the quasimetric d_∞, the above-mentioned property is a well-known classical result. Its proof given in [22] can be generalized straightforward to our situation. We notice that the length of Definition 3.2.11 is not additive set function since d_∞ does not meet the triangle inequality. Nevertheless the following statement holds.

Proposition 3.2.13. *Every rectifiable curve $\gamma : [a,b] \to \mathbb{M}$ meets* (3.2.20).

Proof. Consider the following set function Φ defined on intervals included in $[a,b]$: the value $\Phi(\alpha,\beta)$ at an interval $(\alpha,\beta) \subset [a,b]$ equals $L([\alpha,\beta])$, the length of the curve $\gamma : [\alpha,\beta] \to \mathbb{M}$. The set function Φ is quasiadditive: the inequality

$$\sum_i \Phi(\alpha_i,\beta_i) \le \Phi(\alpha,\beta)$$

holds for every finite collection of pairwise disjoint intervals (α_i, β_i) with $(\alpha_i, \beta_i) \subset (\alpha, \beta)$, where $(\alpha, \beta) \subset [a, b]$ is some interval. It is known (see, for example, [150]), that Φ has a finite derivative

$$\Phi'(x) = \lim_{\substack{(\alpha,\beta)\ni x,\\ \beta-\alpha\to 0}} \frac{\Phi(\alpha, \beta)}{\beta - \alpha} = \lim_{\substack{(\alpha,\beta)\ni x,\\ \beta-\alpha\to 0}} \frac{L([\alpha, \beta])}{\beta - \alpha}$$

almost everywhere in $[a, b]$. Hence,

$$\varlimsup_{y\to x} \frac{d_\infty(\gamma(y), \gamma(x))}{|y - x|} \leq \varlimsup_{\substack{(\alpha,\beta)\ni x,\\ \beta-\alpha\to 0}} \frac{d_\infty(\gamma(\alpha), \gamma(\beta))}{L([\alpha, \beta])} \cdot \lim_{\substack{(\alpha,\beta)\ni x,\\ \beta-\alpha\to 0}} \frac{L([\alpha, \beta])}{\beta - \alpha} \leq \Phi'(x) < \infty$$

for almost all $x \in [a, b]$ if $L([\alpha, \beta]) \neq 0$ for any interval $(\alpha, \beta) \ni x$. Otherwise, $d_\infty(\gamma(y), \gamma(x)) = 0$ in a neighborhood of x, and it is evident that the hc-derivative of γ at x equals 0. □

Theorem 3.2.10 and Proposition 3.2.13 imply

Proposition 3.2.14. *Let $\mathbb{M}$ be a Carnot manifold. Every rectifiable curve $\gamma : [a, b] \to \mathbb{M}$ is hc-differentiable almost everywhere.*

Remark 3.2.15. If the Carnot manifold is a Carnot group then our definition of the hc-differentiability of curves coincides with the $\mathcal{P}$-differentiability of curves given by P. Pansu in [121]. He proved also [121, Proposition 4.1] the $\mathcal{P}$-differentiability almost everywhere of rectifiable curves on Carnot groups using a different method.

3.3. hc-differentiability of smooth mappings on Carnot manifolds

In this subsection we prove hc-differentiability of some classes of mappings $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$ (or $\varphi : E \to \widetilde{\mathbb{M}}$ where $E \subset \mathbb{M}$) assuming that $\mathbb{M}$ is a Carnot manifold, i.e., $\mathbb{M}$ meets the 4th condition in the definition 2.1.1.

Only in the Corollary 3.3.3 we assume that both $\mathbb{M}$ and $\widetilde{\mathbb{M}}$ are Carnot manifolds. We recall that a local tangent cone of a Carnot manifold is a local Carnot group (a stratified graded nilpotent group Lie) properties of which are essentially used in the proofs below.

The approach to the subject is based on methods of papers [137, 139, 140, 141, 142]

3.3.1. Continuity of horizontal derivatives and hc-differentiability. In this subsubsection, we generalize the classical property that the continuity of the partial derivatives of a function defined on a Euclidean space guarantees its differentiability.

In what follows, we repeatedly use the following correspondence: to an arbitrary element $a = \exp\Big(\sum\limits_{i=1}^{N} a_i \widehat{X}_i^u\Big)(u) \in \mathcal{G}^u\mathbb{M}$ and point $w \in \mathcal{G}^u$, assign the element

$$\Delta_\varepsilon^w a = \exp\bigg(\sum_{j=1}^{N} a_j \varepsilon^{\deg X_j} X_j\bigg)(w) \tag{3.3.1}$$

for those ε for which the right-hand side of (3.3.1) exists. Note that, by Property 2.2.5, we have $\Delta^u_\varepsilon a = \delta^u_\varepsilon a$ for all $a \in \mathcal{G}^u\mathbb{M}$.

Theorem 3.3.1. *Suppose that $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$ is a Lipschitz with respect to d_∞ and $\widetilde{d}_\infty$ mapping of a Carnot manifold $\mathbb{M}$ to a Carnot–Carathéodory space $\widetilde{\mathbb{M}}$ such that, at each point $u \in \mathbb{M}$, there exist horizontal derivatives $X_i\varphi(u) \in H_{\varphi(u)}\widetilde{\mathbb{M}}$ continuous on $\mathbb{M}$, $i = 1, \dots, \dim H_1$. Then φ is hc-differentiable at every point of $\mathbb{M}$. The homomorphism of Lie algebras, corresponding to $\widehat{D}\varphi(u) : \mathcal{G}^u\mathbb{M} \to \mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}$, is uniquely defined by the mapping*

$$H_u\mathbb{M} \ni X_i(u) \mapsto X_i\varphi(u) = \frac{d}{dt}\varphi(\exp tX_i(u))|_{t=0} = \sum_{j=1}^{\dim \widetilde{H}_1} b_{ij}\widetilde{X}_j(\varphi(u)) \in H_{\varphi(u)}\widetilde{\mathbb{M}}$$

of the basis horizontal vectors $X_i(u)$, $i = 1, \dots, \dim H_1$, to horizontal vectors in $H_{\varphi(u)}\widetilde{\mathbb{M}}$:

$$H\mathcal{G}^u\mathbb{M} \ni \widehat{X}^u_i \mapsto \sum_{j=1}^{\dim \widetilde{H}_1} b_{ij}\widehat{\widetilde{X}}{}^{\varphi(u)}_j \in H\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}.$$

Proof. First step. Fix a point $u \in \mathbb{M}$ and a compact neighborhood $F \subset \mathcal{G}^u\mathbb{M}$ of the local Carnot group $\mathcal{G}^u\mathbb{M}$. For each horizontal vector field X_i, a family of curves $\gamma : [-\varepsilon, \varepsilon] \times F \to \widetilde{\mathbb{M}}$ is defined: for $u \in F$, put $\gamma_i(s, u) = \varphi(\exp(s\alpha_i X_i)(u))$, where $\alpha_i \in A$, $A \subset \mathbb{R}$ is a bounded neighborhood of $0 \in \mathbb{R}$. This family of curves meets the conditions of Corollary 3.2.5 since $\frac{d}{ds}\gamma_i(s, u) = \alpha_i X_i\varphi(\gamma_i(s, u))$ is bounded and continuous. Hence, the convergence

$$\Delta^{\varphi(u)}_{s^{-1}}\gamma_i(s, u) \to \delta^{\varphi(u)}_{\alpha_i}\exp([X_i\varphi](u))(\varphi(u)) \in \mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}} \quad \text{as } s \to 0 \tag{3.3.2}$$

is uniform on $F \times A$ and the hc-derivative $\delta^{\varphi(u)}_{\alpha_i}\exp(X_i\varphi(u))(u)$ is continuous with respect to $(u, \alpha_i) \in F \times A$. Denote by x_i the "horizontal basis element" $\exp(X_i)(u) = \exp(\widehat{X}^u_i)(u) \in \mathcal{G}^u\mathbb{M}$ and, for all $1 \le i \le \dim H_1$, denote by a_i the hc-derivative $\exp(X_i\varphi(u))(\varphi(u)) = \frac{d}{ds}\varphi(\gamma_i(s, u))|_{s=0}$.

It is known [47] that any element $v \in F$ can be represented (non-uniquely) in the form

$$\delta^u_{\alpha_1}x_{j_1} \cdot \dots \cdot \delta^u_{\alpha_S}x_{j_S}, \quad 1 \le j_i \le \dim H_1, \tag{3.3.3}$$

where S is independent of the choice of the point, and the numbers α_i are bounded by a common constant. Together with the mapping

$$[0, \varepsilon) \ni t \mapsto \hat{v}_i(t) = \delta^u_{t\alpha_1}x_{j_1} \cdot \dots \cdot \delta^u_{t\alpha_i}x_{j_i}, \quad 1 \le j_k \le \dim H_1, \quad 1 \le k \le i \le S,$$

consider the mapping (see (3.3.1))

$$[0, \varepsilon) \ni t \mapsto v_i(t) = \Delta^{v_{i-1}(t)}_{t\alpha_i}x_{j_i} = \exp(t\alpha_i X_{j_i})(v_{i-1}(t)),\ 2 \le i \le S,\ \text{where}$$
$$v_1(t) = \Delta^u_{t\alpha_1}x_{j_1} = \exp(t\alpha_1 X_{j_1})(u).$$

By Theorem 2.7.1, $d_\infty(v_i(t), \hat{v}_i(t)) = o(t)$ as $t \to 0$ uniformly in $u \in F$ and $\alpha_i \in A$, $i \le S$. Since the mapping φ is Lipschitz on F, the limits $\lim\limits_{t\to 0}\Delta^{\varphi(u)}_{t^{-1}}\varphi(\hat{v}_S(t))$

and $\lim\limits_{t\to 0}\Delta^{\varphi(u)}_{t^{-1}}\varphi(v_S(t))$ exist simultaneously. According to Proposition 3.1.3 we have to prove that $\Delta^{\varphi(u)}_{t^{-1}}\varphi(\hat v_S(t))$ converges uniformly to homomorphism of the local homogeneous group $\mathcal{G}^u\mathbb{M}$ to $\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}$. Taking into account above-mentioned observation, it suffices to prove the existence of the limit $\lim\limits_{t\to 0}\Delta^{\varphi(u)}_{t^{-1}}\varphi(v_S(t))$.

Second step. For proving this, by (3.3.2), we infer that

$$w_1(t)=\varphi(v_1(t))=\exp\biggl(\sum_{k=1}^{\widetilde N}\xi^1_k(t)\widetilde X_k\biggr)(\varphi(u))$$
$$\text{has } hc\text{-derivative}\quad \delta^{\varphi(u)}_{\alpha_1}a_{j_1}\in\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}\quad\text{at } t=0.$$

Here $\widetilde X_k$, $k=1,\dots,\widetilde N$, is a local basis on $\widetilde{\mathbb{M}}$ around the point $\varphi(u)$. Assume that the mapping

$$t\mapsto w_i(t)=\varphi(v_i(t))=\exp\biggl(\sum_{k=1}^{\widetilde N}\xi^i_k(t)\widetilde X_k\biggr)(\varphi(v_{i-1}(t)))$$
$$\text{has } hc\text{-derivative}\quad \delta^{\varphi(u)}_{\alpha_1}a_{j_1}\cdot\dots\cdot\delta^{\varphi(u)}_{\alpha_i}a_{j_i}\in\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}},\quad\text{at } t=0,\quad 2\le i<S.$$

Our next goal is to show that the hc-derivative of the mapping $t\mapsto w_{i+1}(t)=\varphi(v_{i+1}(t))=\exp\Bigl(\sum\limits_{k=1}^{\widetilde N}\xi^{i+1}_k(t)\widetilde X_k\Bigr)(\varphi(v_i(t)))$ equals $\delta^{\varphi(u)}_{\alpha_1}a_{j_1}\cdot\dots\cdot\delta^{\varphi(u)}_{\alpha_i}a_{j_i}\cdot\delta^{\varphi(u)}_{\alpha_{i+1}}a_{j_{i+1}}$. Together with the mapping $w_{i+1}(t)$, consider the mapping

$$t\mapsto\widehat w_{i+1}(t)=\exp\biggl(\sum_{k=1}^{\widetilde N}z^{i+1}_k(t)\widehat{\widetilde X}{}^u_k\biggr)(\varphi(v_i(t))).$$

By Theorem 2.7.1 we have $d^{\varphi(u)}_\infty(w_{i+1}(t),\widehat w_{i+1}(t))=o(t)$ as $t\to 0$. Therefore, the relation $d^{\varphi(u)}_\infty\bigl(w_{i+1}(t),\delta^{\varphi(u)}_t\bigl(\delta^{\varphi(u)}_{\alpha_1}a_{j_1}\cdot\dots\cdot\delta^{\varphi(u)}_{\alpha_{i+1}}a_{j_{i+1}}\bigr)\bigr)=o(t)$ as $t\to 0$ holds if and only if $d^{\varphi(u)}_\infty\bigl(\widehat w_{i+1}(t)),\delta^{\varphi(u)}_t\bigl(\delta^{\varphi(u)}_{\alpha_1}a_{j_1}\cdot\dots\cdot\delta^{\varphi(u)}_{\alpha_{i+1}}a_{j_{i+1}}\bigr)\bigr)=o(t)$ as $t\to 0$. On the local homogeneous group $\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}$ the last property is equivalent to the relation

$$d^{\varphi(u)}_\infty\bigl(\delta^u_{t^{-1}}\widehat w_{i+1}(t)),\delta^{\varphi(u)}_{\alpha_1}a_{j_1}\cdot\dots\cdot\delta^{\varphi(u)}_{\alpha_{i+1}}a_{j_{i+1}}\bigr)=o(1)\quad\text{as}\quad i\to\infty.$$

Note that, by the continuity of the group operation in $\mathcal{G}^u\mathbb{M}$, we always have the convergence

$$\delta^{\varphi(u)}_{t^{-1}}(\widehat w_{i+1}(t))\to\delta^{\varphi(u)}_{\alpha_1}a_{j_1}\cdot\dots\cdot\delta^{\varphi(u)}_{\alpha_{i+1}}a_{j_{i+1}}\quad\text{as } t\to 0.$$

Thus, by induction, the hc-derivative of the mapping $[0,\varepsilon)\ni t\mapsto\varphi(v_S(t))$ at 0 is equal to $\delta^{\varphi(u)}_{\alpha_1}a_{j_1}\cdot\dots\cdot\delta^{\varphi(u)}_{\alpha_S}a_{j_S}$; moreover, the convergence is uniform in $v\in F$ and α_i, $1\le i\le S$. Consequently, granted the equality $v_S(t)=\delta^u_t v$, we infer

$$d^{\varphi(u)}_\infty\bigl(\varphi\bigl(\delta^u_t v\bigr),L\bigl(\delta^u_t v\bigr)\bigr)=o\bigl(d^u_\infty\bigl(u,\delta^u_t v\bigr)\bigr)=o(t)\tag{3.3.4}$$

uniformly in $v \in F$, where L stands for the correspondence

$$\mathcal{G}^u\mathbb{M} \ni v = \delta^u_{\alpha_1} x_{j_1} \cdot \dots \cdot \delta^u_{\alpha_S} x_{j_S} \mapsto \delta^{\varphi(u)}_{\alpha_1} a_{j_1} \cdot \dots \cdot \delta^{\varphi(u)}_{\alpha_S} a_{j_S} \in \mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}. \qquad (3.3.5)$$

For finishing the proof, it remains to check that the correspondence $L : \mathcal{G}^u\mathbb{M} \to \mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}$ is a homomorphism of these local homogeneous groups.

Third step. Note that $L(v)$ is the *hc*-derivative at 0 of the mapping $t \mapsto \varphi\big(\delta^u_t v\big)$ for a fixed $v \in \mathcal{G}^u\mathbb{M}$ (see (3.3.4)), which is obviously independent of representation (3.3.3) (since the path $t \mapsto \varphi\big(\delta^u_t v\big)$ depends only on v). Consequently, $L : \mathcal{G}^u\mathbb{M} \to \mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}$ is a mapping of the local groups. By definition (3.3.5) of L, it is clear this mapping is continuous. Demonstrate that it is a group homomorphism. Consider a second element $\overline{v} = \delta^u_{\beta_1} x_{j_1} \cdot \dots \cdot \delta^u_{\beta_S} x_{j_S}$, $1 \leq j_i \leq \dim H_1$, such that

$$v\overline{v} = \delta^u_{\alpha_1} x_{j_1} \cdot \dots \cdot \delta^u_{\alpha_S} x_{j_S} \cdot \delta^u_{\beta_1} x_{j_1} \cdot \dots \cdot \delta^u_{\beta_S} x_{j_S} \in \mathcal{G}^u\mathbb{M} \quad \text{and} \quad L(v) \cdot L(\overline{v}) \in \mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}. \qquad (3.3.6)$$

By (3.3.4) and (3.3.5), the value $L(v\overline{v})$ is independent of the representation of an element $v\overline{v}$ as the product (3.3.6). Hence, applying the above-mentioned conclusions to $v\overline{v}$ and its representation (3.3.6), we see that

$$L(v\overline{v}) = \delta^{\varphi(u)}_{\alpha_1} a_{j_1} \cdot \dots \cdot \delta^{\varphi(u)}_{\alpha_S} a_{j_S} \cdot \delta^{\varphi(u)}_{\beta_1} a_{j_1} \cdot \dots \cdot \delta^{\varphi(u)}_{\beta_S} a_{j_S} = L(v) \cdot L(\overline{v}).$$

Thus, the mapping $L : \mathcal{G}^u\mathbb{M} \to \mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}$ is a continuous group homomorphism. By the well-known properties of the Lie group theory [151], the mapping L is a homomorphism of the local homogeneous groups.

Now, from (3.3.5) it follows directly that L commutes with a dilation, $L \circ \delta^u_t = \delta^{\varphi(u)}_t \circ L$, $t > 0$. Furthermore, since $X_i\varphi(u) \in H_{\varphi(u)}\widetilde{\mathbb{M}}$, the homomorphism L is the *hc-differential of the mapping* $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$ at u. The Lie algebra homomorphism corresponding to L is a mapping of horizontal subspaces. The theorem follows. □

Corollary 3.3.2 ([141]). *Assume that we have a basis $\{X_i\}$, $i = 1, \dots, N$, on a Carnot manifold $\mathbb{M}$ for which Assumption* 2.1.6 *or conditions of Remark* 2.7.2 *hold with some $\alpha \in (0, 1]$. Suppose that $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$ is a mapping of the Carnot manifold $\mathbb{M}$ to a Carnot–Carathéodory space $\widetilde{\mathbb{M}}$ such that, at each point $u \in \mathbb{M}$, there exist horizontal derivatives $X_i\varphi(u) \in H_{\varphi(u)}\widetilde{\mathbb{M}}$ continuous on $\mathbb{M}$, $i = 1, \dots, \dim H_1$. Then φ is hc-differentiable at every point of $\mathbb{M}$. The homomorphism of Lie algebras, corresponding to $\widehat{D}\varphi(u) : \mathcal{G}^u\mathbb{M} \to \mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}$, is defined uniquely by the*

$$H_u\mathbb{M} \ni X_i(u) \mapsto X_i\varphi(u) = \frac{d}{dt}\varphi(\exp tX_i(u))|_{t=0} = \sum_{j=1}^{\dim \widetilde{H}_1} b_{ij}\widetilde{X}_j(\varphi(u)) \in H_{\varphi(u)}\widetilde{\mathbb{M}}$$

of the basis horizontal vectors $X_i(u)$, $i = 1, \dots, \dim H_1$, to horizontal vectors in $H_{\varphi(u)}\widetilde{\mathbb{M}}$:

$$H\mathcal{G}^u\mathbb{M} \ni \widehat{X}^u_i \mapsto \sum_{j=1}^{\dim \widetilde{H}_1} b_{ij}\widehat{\widetilde{X}}{}^{\varphi(u)}_j \in H\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}.$$

Proof. The hypothesis implies that φ is a locally Lipschitz mapping:

$$\tilde{d}_\infty(\varphi(x), \varphi(y)) \leq C d_\infty(x, y),$$

x, y belong to some compact neighborhood of $\mathcal{U}$. To verify this, it suffices to join points $x, y \in \mathcal{U}$ by the horizontal curve γ of Subsection 2.8 whose length is controlled by the quasidistance $d_\infty(x, y)$ and observe that $\varphi \circ \gamma$ is a horizontal curve whose length is controlled by the length of the initial curve. From this, Corollary 2.8.6 and Remark 2.8.7 we infer $\tilde{d}_\infty(\varphi(x), \varphi(y)) \leq C_1 L(\varphi \circ \gamma) \leq C_2 L(\gamma) \leq C_3 d_\infty(x, y)$. □

3.3.2. Functorial property of tangent cones. The definition of the tangent cone depends on the local basis. The question arises on the connection between two tangent cones found from two different bases. Theorem 3.3.1 implies:

Corollary 3.3.3 ([140, 141])**.** *Suppose that we have two local bases $\{X_i\}$ and $\{\widetilde{X}_i\}$, $i = 1, \dots, N$, on a Carnot manifold $\mathbb{M}$ for both of which Assumption* 2.1.6 *or conditions of Remark* 2.7.2 *hold with some $\alpha \in (0, 1]$, and that two collections $X_1, \dots, X_{\dim H_1}$ and $\widetilde{X}_1, \dots, \widetilde{X}_{\dim H_1}$ generate the same horizontal subbundle H_1. Then the local Carnot group $\mathcal{G}^u\mathbb{M}$ defined by the $\{X_i\}$'s is isomorphic to the local Carnot group $\widetilde{\mathcal{G}}^u\mathbb{M}$, determined by the $\{\widetilde{X}_i\}$'s: $(\widetilde{\delta}^u_{t^{-1}} \circ \delta^u_t)(v)$ converges to an isomorphism $\widehat{D}i(u)$ of local Carnot groups $\mathcal{G}^u\mathbb{M}$ and $\widetilde{\mathcal{G}}^u\mathbb{M}$ as $t \to 0$ uniformly in $v \in \mathcal{G}^u\mathbb{M}$. (Here $\widetilde{\delta}^u_t$ is the one-parameter dilation group associated with the vector fields $\{\widetilde{X}_i\}$.)*

The isomorphism of Lie algebras, corresponding to $\widehat{D}i(u)$, is defined uniquely by giving the mapping

$$H_u\mathbb{M} \ni X_i(u) \mapsto X_i(u) = \sum_{j=1}^{\dim H_1} b_{ij} \widetilde{X}_j(u) \in H_u\mathbb{M}$$

of the basis vectors $X_i(u)$, $i = 1, \dots, \dim H_1$, of the horizontal space $H_u\mathbb{M}$ to horizontal vectors of the space $H_u\mathbb{M}$:

$$H\mathcal{G}^u\mathbb{M} \ni \widehat{X}^u_i \mapsto \sum_{j=1}^{\dim H_1} b_{ij} \widehat{\widetilde{X}}{}^u_j \in H\widetilde{\mathcal{G}}^u\mathbb{M}.$$

Proof. Denote by $\mathbb{M}^X$ the Carnot manifold $\mathbb{M}$ with the local basis $\{X_i\}$ and denote by $\mathbb{M}^{\widetilde{X}}$ the Carnot manifold $\mathbb{M}$ with the local basis $\{\widetilde{X}_i\}$, $i = 1, \dots, N$. Let also the symbol $i : \mathbb{M}^X \to \mathbb{M}^{\widetilde{X}}$ stand for the identity mapping from $\mathbb{M}$ into $\mathbb{M}$. Clearly, i meets the conditions of Corollary 3.3.2 since the collections $X_1, \dots, X_{\dim H_1}$ and $\widetilde{X}_1, \dots, \widetilde{X}_{\dim H_1}$ generate the same horizontal subbundle H_1. Then i is *hc*-differentiable at u and, by Corollary 3.3.2 and Proposition 3.1.3, the "difference ratios" $\widetilde{\delta}^u_{t^{-1}}(\delta^u_t(w))$ converge uniformly to a homomorphism $Di(u) : \mathcal{G}^u\mathbb{M}^X \to \widetilde{\mathcal{G}}^u\mathbb{M}^{\widetilde{X}}$ as $t \to 0$. Applying the same argument to the inverse mapping i^{-1} and Theorem 3.1.4, we infer that $Di(u)$ is an isomorphism of the local Carnot groups (of the local tangent cones at u with respect to different local bases). □

Remark 3.3.4. In [4, 16, 67, 106] above statement is proved by other methods under additional assumptions on the smoothness of the basis vector fields.

3.3.3. Rademacher Theorem. The aim of this part is to formulate Rademacher type theorems on the *hc*-differentiability of Lipschitz with respect to d_∞ and $\widetilde{d}_\infty$ mappings of a Carnot manifold to a Carnot–Carathéodory space. This theorem was proved in [141] by means of the theory expounded above. The way of proving this result generalizes essentially the methods of [137], where the $\mathcal{P}$-differentiability of Lipschitz mappings of Carnot groups defined on measurable sets was proved in details: additional arguments are needed since a tangent cone has different metric properties with respect to a given Carnot–Carathéodory space.

Let $\mathbb{M}$, $\widetilde{\mathbb{M}}$ be two Carnot–Carathéodory spaces and let $E \subset \mathbb{M}$ be an arbitrary set. A mapping $\varphi : E \to \widetilde{\mathbb{M}}$ is called a Lipschitz mapping if

$$\tilde{d}_\infty(\varphi(x), \varphi(y)) \leq C d_\infty(x, y), \quad x, y \in E,$$

for some constant C independent of x and y. The least constant in this relation is denoted by $\operatorname{Lip}\varphi$.

The following result extends the theorems on the $\mathcal{P}$-differentiability on Carnot groups [121, 137, 149] (see also [98]) to the case of Carnot–Carathéodory spaces.

Theorem 3.3.5 ([141])**.** *Let E be a set in a Carnot manifold $\mathbb{M}$ and let $\varphi : E \to \widetilde{\mathbb{M}}$ be a Lipschitz mapping from E into a Carnot–Carathéodory space $\widetilde{\mathbb{M}}$. Then φ is hc-differentiable on E.*

The homomorphism of the Lie algebras corresponding to the hc-differential is defined uniquely by the mapping

$$H_u\mathbb{M} \ni X_i(u) \mapsto X_i\varphi(u) = \frac{d}{dt}\varphi(\exp tX_i(u))|_{t=0} = \sum_{j=1}^{\dim \widetilde{H}_1} a_{ij}\widetilde{X}_j(\varphi(u)) \in H_{\varphi(u)}\widetilde{\mathbb{M}}$$

of the horizontal basis vectors $X_i(u)$, $i = 1, \dots, \dim H_1$, to horizontal vectors of the space $H_{\varphi(u)}\widetilde{\mathbb{M}}$:

$$H\mathcal{G}^u\mathbb{M} \ni \widehat{X}_i^u \mapsto \sum_{j=1}^{\dim \widetilde{H}_1} a_{ij}\widehat{\widetilde{X}}_j^{\varphi(u)} \in H\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}.$$

3.3.4. Stepanov Theorem. As a corollary to Theorem 3.3.5, we obtain a generalization of Stepanov's theorem:

Theorem 3.3.6 ([141])**.** *Let $E \subset \mathbb{M}$ be a measurable set in a Carnot manifold $\mathbb{M}$ and let $\varphi : E \to \widetilde{\mathbb{M}}$, where $\widetilde{\mathbb{M}}$ is a Carnot–Carathéodory space, be a mapping such that the relation*

$$\varlimsup_{x\to a, x\in E} \frac{\tilde{d}_\infty(\varphi(a), \varphi(x))}{d_\infty(a, x)} < \infty \tag{3.3.7}$$

holds for almost all $a \in E$. Then φ is hc-differentiable almost everywhere on E and the hc-differential is unique.

The homomorphism of the Lie algebras corresponding to the hc-differential is defined uniquely by the mapping

$$H_u\mathbb{M} \ni X_i(u) \mapsto X_i\varphi(u) = \frac{d}{dt}\varphi(\exp tX_i(u))|_{t=0} = \sum_{j=1}^{\dim \widetilde{H}_1} a_{ij}\widetilde{X}_j(\varphi(u)) \in H_{\varphi(u)}\widetilde{\mathbb{M}}$$

of the basis horizontal vectors $X_i(u)$, $i = 1,\dots,\dim H_1$, *to horizontal vectors of the space* $H_{\varphi(u)}\widetilde{\mathbb{M}}$:

$$H\mathcal{G}^u\mathbb{M} \ni \widehat{X}_i^u \mapsto \sum_{j=1}^{\dim \widetilde{H}_1} a_{ij}\widehat{\widetilde{X}}_j^{\varphi(u)} \in H\mathcal{G}^{\varphi(u)}\widetilde{\mathbb{M}}.$$

The proof written below is based on a sketch of the proof of Stepanov theorem given in [42] but is different of it in some details (a reader can look also at [88] for the proof of metric differentiability of mappings meeting the condition (3.3.7) in which the quasimetric $\tilde{d}_\infty$ is replaced by an usual metric, and d_∞ is replaced by the Euclidean one).

Proof. Since the result is local, we may assume that E is bounded. Since, in view of (3.3.7), the "upper derivative" is finite almost everywhere, it follows that every point $x \in E \setminus \Sigma$, where $\Sigma \subset E$ is some set of measure zero, belongs at least to one of the sets

$$A_k = \left\{x \in E : \frac{\tilde{d}_\infty(\varphi(x),\varphi(y))}{d_\infty(x,y)} \leq k \quad \text{for all} \quad y \in \operatorname{Box}(x,k^{-1}) \cap E\right\},\ k \in \mathbb{N}. \tag{3.3.8}$$

Note that the sequence of sets A_k is monotone: $A_k \subset A_{k+1}$, $k \in \mathbb{N}$. Suppose that the measure of A_k is nonzero for some $k \in \mathbb{N}$. Up to a set of measure zero, we represent A_k as the union of a disjoint family of sets $A_{k,1}, A_{k,2},\dots$ of nonzero measure whose diameters are at most $1/k$:

$$A_k = Z_k \cup A_{k,1} \cup A_{k,2} \cup \dots, \quad |Z_k| = 0.$$

Then the restriction $\varphi_{k,j} = \varphi|_{A_{k,j}}$ meets a Lipschitz condition for all j; therefore, it is extendable by continuity to a Lipschitz mapping $\tilde{\varphi}_{k,j} : \overline{A}_{k,j} \to \widetilde{\mathbb{M}}$.

Verify that if $(E\setminus\Sigma)\cap(\overline{A}_{k,j}\setminus A_{k,j}) \neq \emptyset$ then $\tilde{\varphi}_{k,j} : (E\setminus\Sigma)\cap\overline{A}_{k,j} \to \widetilde{\mathbb{M}}$ coincides with $\varphi : (E \setminus \Sigma) \cap \overline{A}_{k,j} \to \widetilde{\mathbb{M}}$. In other words, if $x \in (E \setminus \Sigma) \cap (\overline{A}_{k,j} \setminus A_{k,j})$ then the extension of $\varphi : A_{k,j} \to \widetilde{\mathbb{M}}$ by continuity to the point x equals $\varphi(x)$. Indeed, if the fixed point x belongs $E \setminus \Sigma$ then $x \in A_l$ for some $l > k$. Consequently, the inequality described in (3.3.8) (with l instead of k) holds for $y \in E \cap \operatorname{Box}(x,l^{-1})$. Since $A_l \cap \operatorname{Box}(x,l^{-1}) \supset A_{k,j} \cap \operatorname{Box}(x,l^{-1})$, we have

$$\varphi(x) = \lim_{y\to x,\, y\in A_l} \varphi(y) = \lim_{y\to x,\, y\in A_{k,j}} \varphi(y) = \tilde{\varphi}_{k,j}(x).$$

By Theorem 3.3.5 (Theorem 3.2.9 in the case of $E \subset \mathbb{R}$), the mapping $\tilde{\varphi}_{k,j} : \overline{A}_{k,j} \to \widetilde{\mathbb{M}}$ is hc-differentiable almost everywhere in $\overline{A}_{k,j}$. We are left with checking the hc-differentiability of the mapping $\varphi : E\setminus\Sigma \to \widetilde{\mathbb{M}}$ at the points of hc-differentiability of the mapping $\tilde{\varphi}_{k,j} : \overline{A}_{k,j} \to \widetilde{\mathbb{M}}$ having density one with respect to $\overline{A}_{k,j}$.

For brevity, denote the set $A_{k,j}$ by A and denote the mapping $\varphi_{k,j}$ by φ. Extend the Lipschitz mapping $\varphi : A \to \widetilde{\mathbb{M}}$ by continuity to a Lipschitz mapping $\tilde{\varphi} : \overline{A} \to \widetilde{\mathbb{M}}$. By the definition of A, the inequality $d_\infty(\varphi(y), \varphi(z)) \le kd_\infty(y, z)$ holds for all $y \in A$ and all $z \in \mathrm{Box}(y, k^{-1}) \cap E$. Note that this inequality is extendable to $\overline{A}$ by continuity. Consequently, the inequality

$$\tilde{d}_\infty(\tilde{\varphi}(y), \varphi(z)) \le kd_\infty(y, z)$$

holds for all $y \in \overline{A}$ and all $z \in \mathrm{Box}(y, k^{-1}) \cap E$.

Suppose now that a point $a \in \overline{A}$ is a point of hc-differentiability for $\tilde{\varphi}$ and the point density of $\overline{A}$. If $z \in E$ belongs to the neighborhood $B(x, k^{-1}) \cap E$ of a then, by the well-known property of a density point (see, for example, [133, 150]), there exists a point $y \in \overline{A}$ such that

$$d_\infty(z, y) = o(d_\infty(z, a)) \quad \text{as} \quad z \to a.$$

Let $\widehat{D}\varphi(a)$ be the hc-differential of $\tilde{\varphi} : \overline{A} \to \widetilde{\mathbb{M}}$ at the point a. Then, in a sufficiently small neighborhood of a, from what was said above, Proposition 3.1.3 and the Local Approximation Theorem 2.5.4 we have

$$\begin{aligned}
\tilde{d}_\infty(\varphi(z), \widehat{D}\varphi(a)[z]) &\le Q^2(\tilde{d}_\infty(\varphi(z), \tilde{\varphi}(y)) + \tilde{d}_\infty(\tilde{\varphi}(y), \widehat{D}\varphi(a)[y]) \\
&\quad + \tilde{d}_\infty(\widehat{D}\varphi(a)[y], \widehat{D}\varphi(a)[z])) \\
&\le Q^2\big(kd_\infty(z, y) + o\big(d_\infty(a, y)\big) \\
&\quad + \tilde{d}_\infty^{\varphi(a)}(\widehat{D}\varphi(a)[y], \widehat{D}\varphi(a)[z]) + \tilde{d}_\infty^{\varphi(a)}(\varphi(a), \widehat{D}\varphi(a)[z])\big) \\
&= o\big(d_\infty(a, z)\big) + \|\widehat{D}\varphi(a)\| \cdot \big(d_\infty^a(y, z) + d_\infty^a(a, z)\big) \\
&= o\big(d_\infty(a, z)\big)
\end{aligned}$$

as $z \to a$, $z \in E$. Here

$$\|\widehat{D}\varphi(a)\| = \sup_{y \in \mathcal{G}^a\mathbb{M}} \frac{\tilde{d}_\infty^{\varphi(a)}(\widehat{D}\varphi(a)[y])}{d_\infty^a(y)},$$

and we have used again the Local Approximation Theorem 2.5.4: $|d_\infty^a(y, z) - d_\infty(y, z)| = o(d_\infty(a, z))$.

Hence, by Proposition 3.1.3, the mapping $\varphi : E \to \widetilde{\mathbb{M}}$ is hc-differentiable at a. Thus we have just proved the hc-differentiability of φ at almost all points of $A_{k,j} \cap (E \setminus \Sigma)$ for arbitrary k and j. Since the collection of sets $A_{k,j}$ covers $E \setminus \Sigma$ the theorem follows. □

4. Application: The coarea and area formulas

4.1. Notations

All the above results on geometry and differentiability are applied in proving the sub-Riemannian analogs of the well-known coarea and area formulas for some classes of contact mappings of Carnot manifolds.

Notation 4.1.1. Denote by N ($\widetilde{N}$) the topological dimensions of $\mathbb{M}$ ($\widetilde{\mathbb{M}}$) and denote by ν ($\tilde{\nu}$) its Hausdorff dimension (see Corollary 2.8.9). Assume that

$$T\mathbb{M} = \bigoplus_{j=1}^{M}(H_j/H_{j-1}),\ H_0 = \{0\},\ \text{and}\ T\widetilde{\mathbb{M}} = \bigoplus_{j=1}^{\widetilde{M}}(\widetilde{H}_j/\widetilde{H}_{j-1}),\ \widetilde{H}_0 = \{0\},$$

where $H_1 \subset T\mathbb{M}$ and $\widetilde{H}_1 \subset T\widetilde{\mathbb{M}}$ are *horizontal* subbundles. The subspace $H_j \subset T\mathbb{M}$ ($\widetilde{H}_j \subset T\widetilde{\mathbb{M}}$) is spanned by H_1 ($\widetilde{H}_1$) and all commutators of order not exceeding $j-1$, $j = 2, \dots, M$ ($\widetilde{M}$).

Denote the dimension of H_j/H_{j-1} ($\widetilde{H}_j/\widetilde{H}_{j-1}$) by n_j ($\tilde{n}_j$), $j = 1, \dots, M$ ($\widetilde{M}$).

Here the number M ($\widetilde{M}$) is such that

$$H_M/H_{M-1} \neq \{0\}, \quad (\widetilde{H}_{\widetilde{M}}/\widetilde{H}_{\widetilde{M}-1} \neq \{0\}),$$

and

$$H_{M+1}/H_M = \{0\}\ (\widetilde{H}_{\widetilde{M}+1}/\widetilde{H}_{\widetilde{M}} = \{0\}).$$

The number M ($\widetilde{M}$) is called the *depth* of $\mathbb{M}$ ($\widetilde{\mathbb{M}}$).

4.2. Lay-out of the proof of the coarea formula

The key point in proving the non-holonomic coarea formula is to investigate the interrelation of "Riemannian" and Hausdorff measures on level sets (see below). The research on the comparison of "Riemannian" and Hausdorff dimensions of submanifolds of Carnot groups can be found in paper by Z.M. Balogh, J.T. Tyson and B. Warhurst [15]. See other results on sub-Riemannian geometric measure theory in works by L. Ambrosio, F. Serra Cassano and D. Vittone [13], L. Capogna, D. Danielli, S.D. Pauls and J.T. Tyson [30], D. Danielli, N. Garofalo and D.-M. Nhieu [36], B. Franchi, R. Serapioni and F. Serra Cassano [56, 57], B. Kirchheim and F. Serra Cassano [92], V. Magnani [99], S.D. Pauls [122] and many other.

The purpose of Section 4 is to explain some ideas and methods of a proof of the coarea formula for sufficiently smooth contact mappings $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$ of Carnot manifolds from [146] and its forthcoming complete version. Note that, all the obtained results are new even for mappings of Carnot groups.

Assumption 4.2.1. Suppose that

1) $N \geq \widetilde{N}$;
2) $n_i \geq \tilde{n}_i$, $i = 1, \dots, M$;
3) the basis vector fields $X_1, \dots, X_N$ (in the preimage) are $C^{1,\alpha}$-smooth, $\alpha > 0$, and $\widetilde{X}_1, \dots, \widetilde{X}_{\widetilde{N}}$ (in the image) are $C^{1,\varsigma}$-smooth, $\varsigma > 0$, or conditions of Remark 2.7.2 hold for $\alpha > 0$ in the preimage and $\varsigma > 0$ in the image.

Remark 4.2.2. Note that, if there exists at least one point where the hc-differential $\widehat{D}\varphi$ is non-degenerate, then the condition $\dim H_1 \geq \dim \widetilde{H}_1$ implies $n_i \geq \tilde{n}_i$, $i = 2, \ldots, M$ (compare with the above assumption).

Notation 4.2.3. Denote by Z the set of points $x \in \mathbb{M}$ such that

$$\operatorname{rank}(D\varphi(x)) < \widetilde{N}.$$

Remark 4.2.4. For proving Theorems 4.2.9, 4.2.11, 4.2.13, and 4.2.18, the smoothness C^1 (in Riemannian sense) for mappings $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$ is sufficient. For proving Theorem 4.2.17, the (Riemannian) smoothness $C^{2,\varpi}$, $\varpi > 0$, of φ is sufficient.

As it is mentioned above, for the first time, a non-holonomic analogue of the coarea formula is proved in paper of P. Pansu [118]. The main idea of this work (which is used in many other ones) is to prove the coarea formula via the Riemannian one:

$$(1.0.4) \Longrightarrow \int\limits_U \mathcal{J}^{Sb}_{\widetilde{N}}(\varphi, x)\, d\mathcal{H}^{\nu}(x) \tag{4.2.1}$$

$$= \int\limits_{\widetilde{\mathbb{M}}} d\mathcal{H}^{\tilde{\nu}}(z) \int\limits_{\varphi^{-1}(z)} \frac{\mathcal{J}^{Sb}_{\widetilde{N}}(\varphi, u)}{\mathcal{J}_{\widetilde{N}}(\varphi, x)}\, d\mathcal{H}^{N-\widetilde{N}}(u) \stackrel{?}{=} \int\limits_{\widetilde{\mathbb{M}}} d\mathcal{H}^{\tilde{\nu}}(z) \int\limits_{\varphi^{-1}(z)} d\mathcal{H}^{\nu-\tilde{\nu}}(u)$$

Here N, $\widetilde{N}$ are topological dimensions, and ν, $\tilde{\nu}$ are Hausdorff dimensions of preimage and image, respectively; $\mathcal{J}^{Sb}_{\widetilde{N}}$ is a sub-Riemannian coarea factor introduced below in Definition 4.2.12 and $\mathcal{J}_{\widetilde{N}}(\varphi, x)$ is the Riemannian coarea factor; $\mathcal{H}^{N-\widetilde{N}}$ is the well-known Riemannian Hausdorff measure with respect to the distance ρ in $\mathbb{M}$, $\mathcal{H}^{\nu-\tilde{\nu}}$ is a sub-Riemannian Hausdorff measure defined below in Definition 4.2.12; it is well known that, in sub-Riemannian case, topological and Hausdorff dimensions differ. It easily follows from (4.2.1), that the key point in this problem is to investigate the interrelation of “Riemannian” and Hausdorff measures on Carnot manifolds theirselves and on level sets of φ, and of Riemannian and sub-Riemannian coarea factors. It is well known that the question on interrelation of measures on Carnot manifolds is quite easy, while both the investigation of geometry of level sets and the calculation of sub-Riemannian coarea factor are non-trivial. The main problems are connected with peculiarities of a sub-Riemannian metric. In particular, the non-equivalence of Riemannian and sub-Riemannian metrics can be seen in the fact that “Riemannian” radius of a sub-Riemannian ball of a radius r varies from r to r^M, $M > 1$, where the constant M depends on the Carnot manifold structure. Thus, a question arises immediately on how “sharp” the approximation of a level by its tangent plain is (since the “usual” order of tangency $o(r)$ is obviously insufficient here: a level may “jump” from a ball earlier then it is expected). Also a question arises on existence of a such sub-Riemannian metric suitable for the description of the geometry of an intersection of a ball and a level set. But even if we answer these questions, one more question appears: what is the relation of the Hausdorff dimension of the image and measure of the intersection of a ball and a level set.

We have solved all the above problems. First of all, the points in which the differential is non-degenerate, are divided into two sets: regular and characteristic.

Definition 4.2.5. The set

$$\chi = \{x \in \mathbb{M} \setminus Z : \operatorname{rank} \widehat{D}\varphi(x) < \widetilde{N}\}$$

is called the *characteristic* set. The points of χ are called *characteristic.*

Definition 4.2.6. The set

$$\mathbb{D} = \{x \in \mathbb{M} : \operatorname{rank} \widehat{D}\varphi(x) = \widetilde{N}\}$$

is called the *regular* set. If $x \in \mathbb{D}$, then we say that, x is a *regular* point.

We define a number $\nu_0(x)$ depending on $x \in \mathbb{M}$ that shows whether a point is regular or characteristic.

Definition 4.2.7. Consider the number ν_0 such that

$$\nu_0(x) = \min\Bigl\{\nu = \sum_{j=1}^{\widetilde{N}} \deg X_{i_j} : \exists\{X_{i_1}, \dots, X_{i_{\widetilde{N}}}\} \ \bigl(\operatorname{rank}([X_{i_j}\varphi](x))_{j=1}^{\widetilde{N}} = \widetilde{N}\bigr)\Bigr\}.$$

It is clear that $\nu_0|_\chi > \tilde{\nu}$ and $\nu_0|_{\mathbb{D}} = \tilde{\nu}$.

We also define such sub-Riemannian quasimetric d_2, that makes the calculation of measure of the intersection of a sub-Riemannian and a tangent plain to a level set possible:

Definition 4.2.8. Let $\mathbb{M}$ be a Carnot manifold of topological dimension N and of depth M, and let $x = \exp\Bigl(\sum\limits_{i=1}^{N} x_i X_i\Bigr)(g)$. Define the distance $d_2(x, g)$ as follows:

$$d_2(x, g) = \max\Bigl\{ \Bigl(\sum_{j=1}^{n_1} |x_j|^2\Bigr)^{\frac{1}{2}}, \\ \Bigl(\sum_{j=n_1+1}^{n_1+n_2} |x_j|^2\Bigr)^{\frac{1}{4}}, \dots, \Bigl(\sum_{j=N-n_M+1}^{N} |x_j|^2\Bigr)^{\frac{1}{2M}}\Bigr\}.$$

The similar metric d_2^u is introduced in the local Carnot group $\mathbb{G}_u\widetilde{\mathbb{M}}$.

The construction of d_2 is based on the fact that a ball in this quasimetric Box_2 asymptotically equals a Cartesian product of Euclidean balls:

$$\mathrm{Box}_2(x, r) \approx B^{n_1}(x, r) \times B^{n_2}(x, r^2) \times \dots \times B^{n_M}(x, r^M),\ M > 1,$$

where N, n_i, $i = 1, \dots, M$, are (topological) dimensions of balls. The latter fact makes the calculation of above-mentioned measure possible (while the geometry of boxes is rather complicated for an estimation of a measure of sections since the sections have different shapes).

Using properties of this quasimetric, we calculate the $\mathcal{H}^{N-\widetilde{N}}$-measure of the intersection of a tangent plain to a level set and a sub-Riemannian ball in the quasimetric d_2.

Theorem 4.2.9. *Fix $x \in \varphi^{-1}(t)$. Then, the Riemannian Hausdorff measure $\mathcal{H}^{N-\widetilde{N}}$-measure of the intersection $T_0[(\varphi \circ \theta_x)^{-1}(t)] \cap \mathrm{Box}_2(0, r)$ is equivalent to*

$$C(1 + o(1))r^{\nu-\nu_0(x)}$$

where C is independent of r, and $o(1) \to 0$ as $r \to 0$.

While investigating the approximation of a surface by its tangent plain, we introduce a "mixed" metric possessing some Riemannian and sub-Riemannian properties.

Definition 4.2.10. For $v, w \in \mathrm{Box}_2(0, r)$ put $d^0_{2E}(v, w) = d^0_2(0, w - v)$, where $w - v$ denotes the Euclidean difference.

This definition implies that $\mathrm{Box}_2(0, r)$ coincides with a ball $\mathrm{Box}_{2E}(0, r)$ centered at 0 of radius r in the metric d^0_{2E}.

We prove that in regular points the tangent plain approximates the level set quite sharp with respect to this metric, and from here we deduce the possibility of calculation of the Riemannian measure of a level set and a sub-Riemannian ball intersection. Notable is the fact that this measure can be expressed via Hausdorff dimensions of the preimage and the image: it is equivalent to $r^{\nu-\tilde{\nu}}$ (see below):

Theorem 4.2.11. *Suppose that $x \in \varphi^{-1}(t)$ is a regular point. Then:*

(I) *In the neighborhood of $0 = \theta_x^{-1}(x)$, there exists a mapping from*

$$T_0[(\varphi \circ \theta_x)^{-1}(t)] \cap \mathrm{Box}_2(0, r(1 + o(1))) \quad \text{to} \quad \psi^{-1}(t) \cap \mathrm{Box}_2(0, r),$$

such that both d_2- and ρ-distortions with respect to 0 equal $1 + o(1)$, where $o(1)$ is uniform on $\mathrm{Box}_2(0, r)$;

(II) *The $\mathcal{H}^{N-\widetilde{N}}$-measure of the intersection $\varphi^{-1}(t) \cap \mathrm{Box}_2(x, r)$ equals*

$$\sqrt{\det(g_{\mathbb{M}}|_{\ker D\varphi}(x) g_{\mathbb{M}}|_{\ker D\varphi}{}^*(x))} \cdot \prod_{k=1}^{M} \omega_{n_k - \tilde{n}_k} \cdot \frac{\sqrt{\det(D\varphi(x) D\varphi^*(x))}}{\sqrt{\det(\widehat{D}\varphi(x)\widehat{D}\varphi^*(x))}} r^{\nu-\tilde{\nu}}(1+o(1)),$$

where $g_{\mathbb{M}}$ is a Riemann tensor, $\widehat{D}\varphi$ is the hc-differential of φ, and $o(1) \to 0$ as $r \to 0$.

Definition 4.2.12. The (spherical) Hausdorff $\mathcal{H}^s$-measure, $s > 0$, of a set $A \subset \mathbb{M}$ is defined as

$$\mathcal{H}^\nu(A) = \omega_\nu \lim_{\delta \to 0} \inf \Big\{ \sum_{i \in \mathbb{N}} r_i^s : \bigcup_{i \in \mathbb{N}} \mathrm{Box}_2(x_i, r_i) \supset A,\ x_i \in A,\ r_i \leq \delta \Big\}.$$

From these results and obtained properties, using a result of [150], we deduce the interrelation of two measures in regular points of a level sets.

Theorem 4.2.13 (Measure Derivative on Level Sets). *Hausdorff measure* $\mathcal{H}^{\nu-\tilde{\nu}}$ *of the intersection* $\operatorname{Box}_2(x,r)\cap\varphi^{-1}(\varphi(x))$*, where* x *is a regular point, and*

$$\operatorname{dist}(\operatorname{Box}_2(x,r)\cap\varphi^{-1}(\varphi(x)),\chi)>0,$$

asymptotically equals $\omega_{\nu-\tilde{\nu}}r^{\nu-\tilde{\nu}}$*. The derivative* $D_{\mathcal{H}^{N-\widetilde{N}}}\mathcal{H}^{\nu-\tilde{\nu}}(x)$ *equals*

$$\frac{1}{\sqrt{\det(g_{\mathbb{M}}|_{\ker D\varphi}(x)g_{\mathbb{M}}|_{\ker D\varphi}{}^{*}(x))}}\cdot\frac{\omega_{\nu-\tilde{\nu}}}{\prod\limits_{k=1}^{M}\omega_{n_k-\tilde{n}_k}}\cdot\frac{\sqrt{\det(\widehat{D}\varphi(x)\widehat{D}\varphi^{*}(x))}}{\sqrt{\det(D\varphi(x)D\varphi^{*}(x))}}.$$

Finally, we introduce the notion of the sub-Riemannian coarea factor via the values of the *hc*-differential of φ.

Definition 4.2.14. The *sub-Riemannian coarea factor* equals

$$\mathcal{J}^{SR}_{\widetilde{N}}(\varphi,x)=\sqrt{\det(\widehat{D}\varphi(x)\widehat{D}\varphi^{*}(x))}\cdot\frac{\omega_N}{\omega_\nu}\frac{\omega_{\tilde{\nu}}}{\omega_{\widetilde{N}}}\frac{\omega_{\nu-\tilde{\nu}}}{\prod\limits_{k=1}^{M}\omega_{n_k-\tilde{n}_k}}.$$

We consider and solve problems connected with the characteristic set. The case of characteristic points is a little more complicated since in characteristic points a surface may jump from a sub-Riemannian ball, consequently, we cannot estimate the measure of the intersection of the surface and the ball via the one of the tangent plain and the ball. Note also that in all the other works on sub-Riemannian coarea formula, the preimage has a group structure, which is essentially used in proving the fact that the Hausdorff measure of characteristic points on each level set equals zero (see also the paper [14] by Z.M. Balogh, dedicated to properties of the characteristic set). In the case of a mapping of two Carnot manifolds, there is no group structure neither in image, nor in preimage. Moreover, the approximation of Carnot manifold by its local Carnot group is insufficient for generalization of methods developed before. That is why we construct new "intrinsic" method of investigation of properties of the characteristic set. First of all, in all the characteristic points the *hc*-differential is degenerate. We solve this problem with the following assumption.

Property 4.2.15. Suppose that $x\in\chi$, and $\operatorname{rank}\widehat{D}\varphi(x)=\widetilde{N}-m$. Let also $\widehat{D}\varphi(x)$ equals zero on $n_1-\tilde{n}_1+m_1$ horizontal (linearly independent) vectors, $n_2-\tilde{n}_2+m_2$ (linearly independent) vectors from H_2/H_1, $n_k-\tilde{n}_k+m_k$ (linearly independent) vectors from H_k/H_{k-1}, $k=3,\dots,\widetilde{M}$. Then, on the one hand, since $\operatorname{rank}\widehat{D}\varphi(x)=\widetilde{N}-m$, we have $\sum\limits_{i=1}^{M}m_k=m$. On the other hand, $\operatorname{rank}D\varphi(x)=\widetilde{N}$. Consequently, there exist m (linearly independent) vectors $Y_1,\dots,Y_m$ of degrees $l_1,\dots,l_{\widetilde{M}}$ (which are minimal) from the kernel of the *hc*-differential $\widehat{D}\varphi$, such that $D\varphi(x)(\operatorname{span}\{H_{\widetilde{M}},Y_1,\dots,Y_m\})=T_{\varphi(x)}\widetilde{\mathbb{M}}$.

In this subsection, we will assume that, among the vectors $Y_1,\dots,Y_m$, m_1 of them of the degree l_1 have the horizontal image, m_2 of them of the degree $l_2\geq l_1$

have image belonging to $\widetilde{H}_2$, and m_k of them of the degree l_k, $l_k \geq l_{k-1}$, have image belonging to $\widetilde{H}_k$, $k = 3, \dots, \widetilde{M}$.

By another words, the "extra" vector fields on which the hc-differential of φ is degenerate in characteristic points, possess the following property: if in $H_k/H_{k-1}(x)$ the quantity of such "extra" vectors equals $m_k > 0$, then there exist m_k vectors from $H_{l_k}/H_{l_k-1}(x)$ such that their images have the degree k, they are linearly independent with each other and with the images of $H_{l_k-1}(x)$, $l_k \geq l_{k-1}$. We develop new "intrinsic" method of investigation of the properties of the characteristic set.

Example 4.2.16. The condition described in Assumption 4.2.15, is always valid for the following $\mathbb{M}$ and $\widetilde{\mathbb{M}}$:

1. $\mathbb{M}$ is an arbitrary Carnot manifold, and $\widetilde{\mathbb{M}} = \mathbb{R}$;
2. M is an arbitrary Carnot manifold of the topological dimension $2m+1$, $\mathcal{G}^u\mathbb{M} = \mathbb{H}^m$ for all $u \in \mathbb{M}$, $\widetilde{\mathbb{M}} = \mathbb{R}^k$, $k \leq 2m$;
3. $M = \widetilde{M}$, $\dim H_1 \geq \dim \widetilde{H}_1$, $\dim(H_i/H_{i-1}) = \dim(\widetilde{H}_i/\widetilde{H}_{i-1})$, $i = 2, \dots, M$;
4. $M = \widetilde{M} + 1$, $\dim H_i = \dim \widetilde{H}_i$, $i = 1, \dots, \widetilde{M}$.

In particular, in Theorem 4.2.9 it is shown, that in the characteristic points $\mathcal{H}^{N-\tilde{N}}$-measure of the intersection of a sub-Riemannian ball and the tangent plain to the level set is equivalent to r to the power $\nu - \nu_0(x) < \nu - \tilde{\nu}$. Next, we show, that $\mathcal{H}^{N-\tilde{N}}$-measure of the intersection of the level set and the sub-Riemannian ball centered at a characteristic point is infinitesimally big in comparison with $r^{\nu-\tilde{\nu}}$, i.e., is equivalent to $\frac{r^{\nu-\tilde{\nu}}}{o(1)}$ (but it is not necessarily equivalent to $r^{\nu-\nu_0(x)}$). From here we deduce that, the intersection of the characteristic set with each level set has zero $\mathcal{H}^{\nu-\tilde{\nu}}$-measure.

Theorem 4.2.17 (Size of the characteristic set). *The Hausdorff measure*

$$\mathcal{H}^{\nu-\tilde{\nu}}(\chi \cap \varphi^{-1}(t)) = 0 \quad \textit{for all } z \in \widetilde{\mathbb{M}}.$$

We also show that the degenerate set of the differential does not influence both parts of the coarea formula.

Theorem 4.2.18. *For $\mathcal{H}^{\tilde{\nu}}$-almost all $t \in \widetilde{\mathbb{M}}$, we have*

$$\mathcal{H}^{\nu-\tilde{\nu}}(\varphi^{-1}(t) \cap Z) = 0.$$

Finally, we deduce the sub-Riemannian coarea formula.

Theorem 4.2.19. *For any smooth contact mapping $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$ possessing Property* 4.2.15*, the coarea formula holds:*

$$\int\limits_{\mathbb{M}} \mathcal{J}^{Sb}_{\tilde{N}}(\varphi, x)\, d\mathcal{H}^{\nu}(x) = \int\limits_{\widetilde{\mathbb{M}}} d\mathcal{H}^{\tilde{\nu}}(t) \int\limits_{\varphi^{-1}(t)} d\mathcal{H}^{\nu-\tilde{\nu}}(u). \tag{4.2.2}$$

Using the result of the paper by R. Monti and F. Serra Cassano [113, Theorem 4.2] for Lip-functions defined on a Carnot–Carathéodory space $\mathbb{M}$ of the Hausdorff dimension ν, we deduce that the De Giorgi perimeter coincides with $\mathcal{H}^{\nu-1}$-measure on almost every level of a smooth function $\varphi : \mathbb{M} \to \mathbb{R}$.

Theorem 4.2.20. *For $C^{2,\alpha}$-functions $\varphi : \mathbb{M} \to \mathbb{R}$, $\alpha > 0$, where $\dim_{\mathcal{H}} \mathbb{M} = \nu$, the De Giorgi perimeter coincides with $\mathcal{H}^{\nu-1}$-measure on almost every level.*

In paper [91], under assumption that vector fields are C^∞-smooth on $\mathbb{M}$ and $\widetilde{\mathbb{M}}$, Maria Karmanova proves the following

Theorem 4.2.21. *Let $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$ be an arbitrary smooth contact mapping of two Carnot manifolds. Then, on every level, $\mathcal{H}^{\nu-\tilde{\nu}}$-measure of characteristic points equals zero.*

Remark 4.2.22. Notice, Theorem 4.2.21 is proved in [91] without Property 4.2.15. As a consequence, under the same assumptions, the coarea formula (4.2.2) holds.

4.3. Lay-out of the proof of the area formula

In this Subsection we exhibit a proof of the area formula for contact C^1-mappings $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$ of Carnot manifolds:

$$\int\limits_{\mathbb{M}} f(y)\sqrt{\det(\widehat{D}\varphi(y)^*\widehat{D}\varphi(y))}\, d\mathcal{H}^\nu(y) = \int\limits_{\widetilde{\mathbb{M}}} \sum_{y:\, y\in\varphi^{-1}(x)} f(y)\, d\mathcal{H}^\nu(x).$$

Recall that we denote the quasimetric of Definition 4.2.8 in the preimage (in the image) by the symbol d_2 ($\widetilde{d}_2$).

Assumption 4.3.1. We suppose that

1) $N \le \widetilde{N}$, $n_i \le \tilde{n}_i$, $i = 1, \dots, M$;
2) the basis vector fields in the preimage and in the image belong to the class C^2, and φ is a contact (i.e., $D\varphi(H_1) \subset \widetilde{H}_1$) C^1-mapping.

Under these assumptions the mapping $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$ of Carnot manifolds has both Riemannian $D\varphi(x)$ and sub-Riemannian $\widehat{D}\varphi(x)$) differentials at all points $x \in \mathbb{M}$.

Put $Z = \{x \in \mathbb{M} : \operatorname{rank}(\widehat{D}\varphi(x)) < N\}$.

Theorem 4.3.2 ([148])**.** *Fix $y \in \mathbb{M} \setminus Z$, $y \in \mathcal{U}$, and $x = \varphi(y)$. Then, $\mathcal{H}^N$-measure of the intersection $\varphi(\mathcal{U}) \cap \operatorname{Box}_2(x, r)$ equals*

$$\prod_{k=1}^{M} \omega_{n_k} \cdot \sqrt{\det(\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}{}^*(x)\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}(x)} \cdot \frac{\sqrt{\det(D\varphi^*(y)D\varphi(y))}}{\sqrt{\det(\widehat{D}\varphi^*(y)\widehat{D}\varphi(y))}} \cdot r^\nu \cdot (1 + o(1)),$$

where ω_l is the volume of the unit ball in $\mathbb{R}^l$, $\tilde{g}|_{\widetilde{\mathbb{M}}}$ is the Riemann tensor in $\widetilde{\mathbb{M}}$, and $o(1) \to 0$ as $r \to 0$.

Definition 4.3.3. The (spherical) Hausdorff $\mathcal{H}^\nu$-measure of a set $A \subset \varphi(\mathbb{M})$ is defined as

$$\mathcal{H}^\nu(A) = \omega_\nu \lim_{\delta\to 0} \inf\Bigl\{\sum_{i\in\mathbb{N}} r_i^\nu : \bigcup_{i\in\mathbb{N}} \mathrm{Box}_2(x_i, r_i) \supset A,\, x_i \in A,\, r_i \le \delta\Bigr\}.$$

The next result is established by the application of the above result on "Riemann" measure of the intersection $\mathrm{Box}_2(x,r) \cap \varphi(\mathcal{U})$. Recall that the symbol $D_{\mu_1}\mu_2(y)$ denotes the derivative of a measure μ_2 with respect to a measure μ_1 at y: $D_{\mu_1}\mu_2(y) = \lim\limits_{r\to 0} \frac{\mu_2(B(y,r))}{\mu_1(B(y,r))}$.

Theorem 4.3.4 ([148]). *Hausdorff $\mathcal{H}^\nu$-measure of the intersection $\mathrm{Box}_2(x,r)\cap\varphi(\mathbb{M})$ is asymptotically equal to $\omega_\nu r^\nu$, and the derivative $D_{\mathcal{H}^\nu}\mathcal{H}^N(x)$ equals*

$$\frac{\prod\limits_{k=1}^{M} \omega_{n_k} \cdot \sqrt{\det(\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}{}^*(x)\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}(x)}}{\omega_\nu} \cdot \frac{\sqrt{\det(D\varphi^*(\varphi^{-1}(x))D\varphi(\varphi^{-1}(x)))}}{\sqrt{\det(\widehat{D}\varphi^*(\varphi^{-1}(x))\widehat{D}\varphi(\varphi^{-1}(x)))}},$$

and for each set $A \subset \varphi(\mathcal{U})$ we have

$$\mathcal{H}^N(A) = \int\limits_A D_{\mathcal{H}^\nu}\mathcal{H}^N(x)\, d\mathcal{H}^\nu(x).$$

Since outside the set Z we have $|D_{\mathcal{H}^\nu}\mathcal{H}^N(x)| \ge \beta > 0$ locally, and the measure $\mathcal{H}^N$ is locally doubling on balls, then in view of Lebesgue Differentiability Theorem we have

$$\begin{aligned}\mathcal{H}^\nu(A) &= \int\limits_A D_{\mathcal{H}^N}\mathcal{H}^\nu(x)\, d\mathcal{H}^N(x)\\ &= \int\limits_A \frac{\omega_\nu}{\prod\limits_{k=1}^{M} \omega_{n_k} \cdot \bigl|\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}(x)\bigr|} \cdot \frac{\sqrt{\det(\widehat{D}\varphi^*(\varphi^{-1}(x))\widehat{D}\varphi(\varphi^{-1}(x)))}}{\sqrt{\det(D\varphi^*(\varphi^{-1}(x))D\varphi(\varphi^{-1}(x)))}}\, d\mathcal{H}^N(x),\end{aligned}$$

where $\bigl|\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}(x)\bigr| = \sqrt{\det(\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}{}^*(x)\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}(x)}$.

The latter result motivates the following

Definition 4.3.5. The *sub-Riemannian Jacobian* at y equals

$$\mathcal{J}^{SR}(\varphi, y) = \sqrt{\det(\widehat{D}\varphi(y)^*\widehat{D}\varphi(y))}.$$

Theorem 4.3.6 ([148]). *We have $\mathcal{H}^\nu(\varphi(Z)) = 0$, where*

$$Z = \{y \in \mathbb{M} :\ \mathrm{rank}\, \widehat{D}\varphi(y) < N\}.$$

The above results and the Riemannian area formula imply the sub-Riemannian area formula.

Theorem 4.3.7 (**[148] The area formula**). *Let $\varphi : \mathbb{M} \to \widetilde{\mathbb{M}}$ be a contact C^1-mapping. Let also $N \leq \widetilde{N}$ and $\dim H_i - \dim H_{i-1} \leq \dim \widetilde{H}_i - \widetilde{H}_{i-1}$, $i = 1, \dots, M$. Then, the formula*

$$\int\limits_{\mathbb{M}} f(y)\mathcal{J}^{SR}(\varphi, y)\, d\mathcal{H}^\nu(y) = \int\limits_{\widetilde{\mathbb{M}}} \sum_{y:\, y\in\varphi^{-1}(x)} f(y)\, d\mathcal{H}^\nu(x) \tag{4.3.1}$$

holds, where $f : \mathbb{M} \to \mathbb{E}$ ($\mathbb{E}$ is an arbitrary Banach space) is such that the function $f(y)\mathcal{J}^{SR}(\varphi, y)$ is integrable. Here, Hausdorff measures are constructed with respect to d_2 and $\widetilde{d}_2$ with the normalizing multiple ω_ν.

Proof. Consider the neighborhood $\mathcal{U} \subset \mathbb{M}$, on which φ is bi-Lipschitz on its image $\varphi(\mathcal{U})$, and $\operatorname{rank} \widehat{D}\varphi(y) = N$ for all $y \in \mathcal{U}$. For simplicity, put $f(y) \equiv 1$. Assuming $y = \varphi^{-1}(x)$, $|\widehat{D}\varphi(y)| = \sqrt{\det(\widehat{D}\varphi(y)^*\widehat{D}\varphi(y))}$, $|g(y)| = \sqrt{\det(g^*(y)g(y))}$, $\bigl|\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}(x)\bigr| = \sqrt{\det(\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}{}^*(x)\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}(x)}$, $|D\varphi(y)| = \sqrt{\det(D\varphi(y)^*D\varphi(y))}$, we have

$$\begin{aligned}
&\int\limits_{\mathbb{M}} \mathcal{J}^{SR}(\varphi, y)\, d\mathcal{H}^\nu(y)\\
&\quad= \int\limits_{\mathbb{M}} |\widehat{D}\varphi(y)| \frac{\omega_\nu}{\prod\limits_{k=1}^{M} \omega_{n_k}} \frac{1}{|g(y)|}\, d\mathcal{H}^N(y) = \int\limits_{\widetilde{\mathbb{M}}} \frac{|\widehat{D}\varphi(y)|}{\mathcal{J}(\varphi, y)} \frac{\omega_\nu}{\prod\limits_{k=1}^{M} \omega_{n_k}} \frac{1}{|g(y)|}\, d\mathcal{H}^N(x)\\
&\quad= \int\limits_{\widetilde{\mathbb{M}}} \frac{|\widehat{D}\varphi(y)|}{\mathcal{J}(\varphi, y)} \frac{\omega_\nu}{\prod\limits_{k=1}^{M} \omega_{n_k}} \frac{1}{|g(y)|} D_{\mathcal{H}^\nu}\mathcal{H}^N(x)\, d\mathcal{H}^\nu(x)\\
&\quad= \int\limits_{\widetilde{\mathbb{M}}} \frac{|\widehat{D}\varphi(y)|}{\mathcal{J}(\varphi, y)} \frac{\omega_\nu}{\prod\limits_{k=1}^{M} \omega_{n_k}} \frac{\prod\limits_{k=1}^{M} \omega_{n_k} \cdot \bigl|\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}(x)\bigr|}{\omega_\nu} \cdot \frac{|D\varphi(y)|}{|\widehat{D}\varphi(y)|} \frac{1}{|g(y)|}\, d\mathcal{H}^\nu(x)\\
&\quad= \int\limits_{\widetilde{\mathbb{M}}} \frac{|D\varphi(y)| \cdot \bigl|\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}(x)\bigr|}{\mathcal{J}(\varphi, y)|g(y)|}\, d\mathcal{H}^\nu(x) = \int\limits_{\widetilde{\mathbb{M}}} d\mathcal{H}^\nu(x),
\end{aligned}$$

since the Riemannian Jacobian $\mathcal{J}(\varphi, y)$ equals $|D\varphi(y)|\frac{\bigl|\tilde{g}_{\widetilde{\mathbb{M}}}|_{T\varphi(\mathbb{M})}(x)\bigr|}{|g(y)|}$. □

Remark 4.3.8. In the paper [90], Maria Karmanova proves the area formula for Lipschitz mappings of Carnot manifolds with respect to sub-Riemannian metrics. Note that a proof of the formula 4.3.1 under these assumptions requires essentially new methods of investigation of Lipschitz mappings. The methods of proving are new even in the case of Lipschitz mappings of Euclidean spaces.

Remark 4.3.9. In the paper [122], the area formula is proved for Lipschitz with respect to sub-Riemannian metrics mappings of Carnot groups, and the Jacobian is defined as

$$\lim_{r\to 0}\frac{\mathcal{H}^{\nu}_{cc}(\varphi(B_{cc}(y,r)))}{\mathcal{H}^{\nu}_{cc}(B_{cc}(y,r))}, \tag{4.3.2}$$

where B_{cc} is a ball in Carnot–Carathéodory metric [70], and Hausdorff measures $\mathcal{H}^{\nu}_{cc}$ are constructed also with respect to d_{cc}. If we consider in (4.3.2) Hausdorff measures $\mathcal{H}^{\nu}$ constructed with respect to metrics d_2 and $\widetilde{d}_2$, then, in view of Theorem 4.3.7, we infer $\lim\limits_{r\to 0}\frac{\mathcal{H}^{\nu}(\varphi(B_{cc}(y,r)))}{\mathcal{H}^{\nu}(B_{cc}(y,r))} = \sqrt{\det(\widehat{D}\varphi(y)^{*}\widehat{D}\varphi(y))}$ for C^1-mappings $\varphi:\mathbb{M}\to\widetilde{\mathbb{M}}$.

Theorem 4.3.10 ([148]). *Suppose that the mapping $\varphi:\mathbb{G}\to\widetilde{\mathbb{G}}$ is Lipschitz with respect to sub-Riemannian metrics on Carnot groups. Then the area formula*

$$\int\limits_{\mathbb{G}} f(y)\mathcal{J}^{SR}(\varphi,y)\,d\mathcal{H}^{\nu}(y) = \int\limits_{\widetilde{\mathbb{G}}}\sum_{y:\,y\in\varphi^{-1}(x)} f(y)\,d\mathcal{H}^{\nu}(x), \tag{4.3.3}$$

holds, where $\mathcal{J}^{SR}(\varphi,x)$ is defined in the Definition 4.3.5, *and $f:\mathbb{G}\to\mathbb{E}$ ($\mathbb{E}$ is an arbitrary Banach space) is such that the function $f(y)\mathcal{J}^{SR}(\varphi,y)$ is integrable. Here, Hausdorff measures Definition* 4.3.3 *are constructed with respect to d_2 and $\widetilde{d}_2$ with the normalizing multiple ω_{ν}.*

Proof. It is easy to show using arguments of the papers [138, 122], that in a neighborhood of almost every point $g\in\mathbb{G}\setminus Z$ we have

$$\widetilde{d}_2(\varphi(v),\varphi(w)) = \widetilde{d}_2(\widehat{D}\varphi(y)[v],\widehat{D}\varphi(g)[w])\cdot(1+o(1)),$$

where $o(1)\to 0$ as $v,w\to y$. Consequently, the distortions of $\mathcal{H}^{\nu}$-measure under φ and under the mapping $w\mapsto\widehat{D}\varphi(y)[w]$ are asymptotically equal (see, for example, [138, 122]). Since the mapping $w\mapsto\widehat{D}\varphi(y)[w]$ is smooth and contact, then, its sub-Riemannian Jacobian and, consequently, the sub-Riemannian Jacobian of φ coincides:

$$\lim_{r\to 0}\frac{\mathcal{H}^{\nu}(\varphi(B_{cc}(y,r)))}{\mathcal{H}^{\nu}(B_{cc}(y,r))} = \lim_{r\to 0}\frac{\mathcal{H}^{\nu}(\widehat{D}\varphi(y)(B_{cc}(y,r)))}{\mathcal{H}^{\nu}(B_{cc}(y,r))} = |\widehat{D}\varphi(y)|$$

almost everywhere, and in view of the $\mathcal{N}$-property of Lipschitz mappings and results of [122], the area formula (4.3.3) is valid.

5. Appendix

5.1. Proof of Lemma 2.1.16

Proof. It is well known, that the solution $y(t,u)$ of the ODE (2.1.12) equals $y(t,u) = \lim\limits_{n\to\infty} y_n(t,u)$, where

$$y_0(t,u) = \int\limits_0^t f(y(0),u)\,d\tau, \quad \text{and} \quad y_n(t,u) = \int\limits_0^t f(y_{n-1}(\tau,u),u)\,d\tau.$$

This convergence is uniform in u, if u belongs to some compact set.

From the definition of this sequence it follows, that $y_n(t) \to y(t)$ as $n \to \infty$ in C^1-norm.

1. We show, that every $y_n(t,u) \in H^\alpha(u)$ for each $t \in [0,1]$. We have

$$\begin{aligned}
&\max_t |y_n(t,u_1) - y_n(t,u_2)| \\
&\quad \le \int\limits_0^1 |f(y_{n-1}(\tau,u_1),u_1) - f(y_{n-1}(\tau,u_2),u_2)|\,d\tau \\
&\quad \le \int\limits_0^1 |f(y_{n-1}(\tau,u_1),u_1) - f(y_{n-1}(\tau,u_1),u_2)|\,d\tau \\
&\qquad + \int\limits_0^1 |f(y_{n-1}(\tau,u_1),u_2) - f(y_{n-1}(\tau,u_2),u_2)|\,d\tau \\
&\quad \le H(f)|u_1 - u_2|^\alpha + L \max_t |y_{n-1}(t,u_1) - y_{n-1}(t,u_2)| \\
&\quad \le H(f) \sum_{m=0}^{n-1} L^m |u_1 - u_2|^\alpha + L^n \max_t |y_0(t,u_1) - y_0(t,u_2)| \\
&\quad \le H(f) \sum_{m=0}^{\infty} L^m |u_1 - u_2|^\alpha,
\end{aligned}$$

where $H(f)$ is a constant, such that $|f(u_1) - f(u_2)| \le H(f)|u_1 - u_2|^\alpha$. Note that the constant $H = H(f) \sum\limits_{m=0}^{\infty} L^m < \infty$ since $L < 1$, and it does not depend on $n \in \mathbb{N}$.

Suppose that u belongs to some compact set $\mathcal{U}$. Then

$$\begin{aligned}
&|y(t,u_1) - y(t,u_2)| \\
&\quad \le |y(t,u_1) - y_n(t,u_1)| + |y_n(t,u_1) - y_n(t,u_2)| + |y(t,u_2) - y_n(t,u_2)| \\
&\quad \le H|u_1 - u_2|^\alpha + 2\varepsilon
\end{aligned}$$

for every $\varepsilon = \varepsilon(n) > 0$. Since the convergence is uniform in $u \in \mathcal{U}$, and $\varepsilon(n) \to 0$ as $n \to \infty$, then $|y(t,u_1) - y(t,u_2)| \le H|u_1 - u_2|^\alpha$, and $y \in H^\alpha(u)$ locally.

To show, that $\frac{\partial y}{\partial v_i}(t,v,u) \in H^\alpha(u)$ locally, $i = 1,\dots,N$, we obtain our estimates in the simplest case of $N = 1$.

2. Note that the mappings $\{y_n\}_{n\in\mathbb{N}}$ converge to y in C^1-norm, and this convergence is uniform, if u belongs to some compact set $\mathcal{U}$.

Let $u \in \mathcal{U}$, $v \in W(0) \subset \mathbb{R}^N$. Then similarly to the case **1**, we see, that if the Hölder constant of y'_n does not depend on $n \in \mathbb{N}$, then $y' \in H^\alpha(u)$.

$$\begin{aligned}
&\max_{t,v}\Big|\frac{dy_n}{dv}(t,v,u_1) - \frac{dy_n}{dv}(t,v,u_2)\Big| \qquad (5.1.1)\\
&\le \max_{t,v}\Big|\frac{d}{dv}\int_0^t f(y_{n-1}(\tau,v,u_1),v,u_1) - f(y_{n-1}(\tau,v,u_2),v,u_2)\,d\tau\Big|\\
&\le \max_{t,v}\Big|\frac{d}{dv}\int_0^t f(y_{n-1}(\tau,v,u_1),v,u_1) - f(y_{n-1}(\tau,v,u_2),v,u_1)\,d\tau\Big|\\
&\quad + \max_{t,v}\Big|\frac{d}{dv}\int_0^t f(y_{n-1}(\tau,v,u_2),v,u_1) - f(y_{n-1}(\tau,v,u_2),v,u_2)\,d\tau\Big|.
\end{aligned}$$

For the first summand we have

$$\begin{aligned}
&\max_{t,v}\Big|\frac{d}{dv}\int_0^1 f(y_{n-1}(\tau,v,u_1),v,u_1) - f(y_{n-1}(\tau,v,u_2),v,u_1)\,d\tau\Big|\\
&\le \max_{v}\int_0^1\Big|\frac{d}{dv}(f(y_{n-1}(\tau,v,u_1),v,u_1) - f(y_{n-1}(\tau,v,u_2),v,u_1))\Big|\,d\tau\\
&\le \max_{v}\int_0^1\Big|\frac{df}{dy}\frac{dy_{n-1}}{dv}(\tau,v,u_1) - \frac{df}{dy}\frac{dy_{n-1}}{dv}(\tau,v,u_2)\Big|\,d\tau\\
&\quad + \max_{v}\int_0^1\Big|\frac{\partial f}{\partial v}(y_{n-1}(\tau,v,u_1)) - \frac{\partial f}{\partial v}(y_{n-1}(\tau,v,u_2))\Big|\,d\tau. \qquad (5.1.2)
\end{aligned}$$

Then, we get

$$\max_{v}\int_0^1\Big|\frac{\partial f}{\partial v}(y_{n-1}(\tau,v,u_1)) - \frac{\partial f}{\partial v}(y_{n-1}(\tau,v,u_2))\Big|\,d\tau \le C(f)H(y)|u_1-u_2|^\alpha,$$

since each y_m is Hölder. The first summand in (5.1.2) is evaluated in the following way:

$$\max_{v}\int_0^1\Big|\frac{df}{dy}\frac{dy_{n-1}}{dv}(\tau,v,u_1) - \frac{df}{dy}\frac{dy_{n-1}}{dv}(\tau,v,u_2)\Big|\,d\tau$$

$$
\begin{aligned}
&\le \max_v \int_0^1 \Big| \frac{df}{dy}(u_1)\frac{dy_{n-1}}{dv}(\tau,v,u_1) - \frac{df}{dy}(u_1)\frac{dy_{n-1}}{dv}(\tau,v,u_2)\Big|\,d\tau \\
&\quad + \max_v \int_0^1 \Big| \frac{df}{dy}(u_1)\frac{dy_{n-1}}{dv}(\tau,v,u_2) - \frac{df}{dy}(u_2)\frac{dy_{n-1}}{dv}(\tau,v,u_2)\Big|\,d\tau \\
&\le L\max_{t,v}\Big|\frac{dy_{n-1}}{dv}(t,v,u_1) - \frac{dy_{n-1}}{dv}(t,v,u_2)\Big| \\
&\quad + \max_{u,v}\int_0^1 \Big|\frac{dy_{n-1}}{dv}(\tau,v,u)\Big|\,d\tau\cdot H(Df)|u_1-u_2|^\alpha. \qquad (5.1.3)
\end{aligned}
$$

Next, we estimate

$$
\begin{aligned}
\max_{u,v}\int_0^1 \Big|\frac{dy_m}{dv}(\tau,v,u)\Big|\,d\tau &\le \max_{t,u,v}\Big|\frac{dy_m}{dv}(t,v,u)\Big| \\
&= \max_{t,u,v}\Big[L\Big|\frac{dy_{m-1}}{dv}\Big| + \Big|\frac{\partial f}{\partial v}\Big|\Big] \le \max_{t,u,v}\Big|\frac{\partial f}{\partial v}\Big|\Big[\sum_{k=0}^{\infty} L^k\Big] < \infty.
\end{aligned}
$$

Thus, in the first summand of (5.1.1) we have

$$
L\max_{t,v}\Big|\frac{dy_{n-1}}{dv}(t,v,u_1) - \frac{dy_{n-1}}{dv}(t,v,u_2)\Big| + C|u_1-u_2|^\alpha,
$$

where $0 < C < \infty$ does not depend on $n \in \mathbb{N}$. The second summand in (5.1.1) is

$$
\max_{t,v}\Big|\frac{d}{dv}\int_0^t f(y_{n-1}(\tau,v,u_2),v,u_1) - f(y_{n-1}(\tau,v,u_2),v,u_2)\,d\tau\Big|
$$

$$
\max_v \int_0^1 \Big|\frac{\partial f}{\partial v}(y_{n-1},v,u_1) - \frac{\partial f}{\partial v}(y_{n-1},v,u_2)\Big|\,d\tau \le C(f)|u_1-u_2|.
$$

Thus,

$$
\begin{aligned}
&\max_{t,v}\Big|\frac{dy_n}{dv}(t,v,u_1) - \frac{dy_n}{dv}(t,v,u_2)\Big| \\
&\quad\le L\max_{t,v}\Big|\frac{dy_{n-1}}{dv}(t,v,u_1) - \frac{dy_{n-1}}{dv}(t,v,u_2)\Big| + K|u_1-u_2|^\alpha \\
&\quad\le k\sum_{k=0}^{\infty} L^k|u_1-u_2|^\alpha,
\end{aligned}
$$

and $\frac{dy_n}{dv} \in H^\alpha(u)$ locally. Hence, $\frac{dy}{dv} \in H^\alpha(u)$ locally. □

Acknowledgment

We are grateful to Valeriĭ Berestovskiĭ, for his collaboration in proving of Gromov's theorem on nilpotentization of vector fields (see Remark 2.2.14). We also want to thank Pierre Pansu for fruitful discussion of our results concerning the geometry of Carnot–Carathéodory spaces and the sub-Riemannian coarea formula, and Martin Reimann for interesting discussions of differentiability theorems.

References

[1] A.A. Agrachev, *Compactness for sub-Riemannian length minimizers and subanalyticity.* Rend. Semin. Mat. Torino, **56** (1998).

[2] A.A. Agrachev and R. Gamkrelidze. *Exponential representation of flows and chronological calculus.* Math. USSR-Sb. **107** (4) (1978), 487–532 (in Russian).

[3] A.A. Agrachev and J.-P. Gauthier. *On subanalyticity of Carnot–Carathéodory distances.* Ann. Inst. H. Poincaré Anal. Non Linéaire, **18** (3), 2001.

[4] A.A. Agrachev, A. Marigo, *Nonholonomic tangent spaces: intrinsic construction and rigid dimensions.* Electron. Res. Announc. Amer. **9** (2003), 111–120.

[5] A.A. Agrachev, Yu.L. Sachkov, *Control theory from the geometric viewpoint.* Springer, Berlin, 2004.

[6] A.A. Agrachev and A.V. Sarychev, *Filtrations of a Lie algebra of vector fields and nilpotent approximations of control systems.* Dokl. Akad. Nauk SSSR **285** (1987), 777–781.

[7] A.A. Agrachev and A.V. Sarychev, *Strong minimality of abnormal geodesics for 2-distributions.* J. Dyn. Control Syst. **1** (2) (1995).

[8] A.A. Agrachev and A.V. Sarychev, *Abnormal sub-Riemannian geodesics: Morse index and rigidity.* Ann. Inst. Henri Poincaré, Analyse Non Linéaire **13** (6) (1996), 635–690.

[9] A.A. Agrachev and A.V. Sarychev, *On abnormal extremals for Lagrange variational problems.* J. Math. Syst. Estim. Cont. **8** (1) (1998), 87–118.

[10] A.A. Agrachev and A.V. Sarychev, *Sub-Riemannian metrics: Minimality of abnormal geodesics versus subanalyticity.* ESAIM Control Optim. Calc. Var. **4** (1999).

[11] H. Airault, P. Malliavin, *Intégration géometrique sur l'espace de Wiener*, Bull. Sci. Math. **112** (1988), 3–52.

[12] L. Ambrosio, B. Kirchheim, *Rectifiable sets in metric and Banach spaces.* Math. Ann. **318** (2000), 527–555.

[13] L. Ambrosio, F. Serra Cassano and D. Vittone, *Intrinsic regular hypersurfaces in Heisenberg groups.* J. Geom. Anal. **16** (2) (2006), 187–232.

[14] Z.M. Balogh, *Size of characteristic sets and functions with prescribed gradients*, Crelle's Journal **564** (2003), 63–83.

[15] Z.M. Balogh, J.T. Tyson, B. Warhurst, *Sub-Riemannian Vs. Euclidean Dimension Comparison and Fractal Geometry on Carnot Groups*, preprint.

[16] A. Bellaïche, *Tangent Space in Sub-Riemannian Geometry.* Sub-Riemannian geometry, Birkhäuser, Basel, 1996, 1–78.

[17] M. Biroli, U. Mosco, *Forme de Dirichlet et estimationes structurelles dans les milieux discontinues.* C. R. Acad. Sci. Paris, **313** (1991), 593–598.

[18] M. Biroli, U. Mosco, *Sobolev inequalities on homogeneous spaces.* Pot. Anal. **4** (1995), 311–324.

[19] A. Bonfiglioli, E. Lanconelli, F. Uguzzoni, *Stratified Lie Groups and Potential Theory for Their Sub-Laplacians.* Springer, 2007.

[20] S.M. Buckley, P. Koskela, G. Lu, *Subelliptic Poincaré inequalities: the case* $p < 1$. Publ. Mat. **39** (1995), 313–334.

[21] M. Buliga, *Dilatation structures I. Fundamentals.* J. Gen. Lie Theory and Appl. **2** (1) (2007), 65–95.

[22] D.Yu. Burago, Yu.D. Burago, S.V. Ivanov, *A Course in Metric Geometry.* Graduate Studies in Mathematics, **33**, American Mathematical Society, Providence, RI, 2001.

[23] L. Capogna, *Regularity of quasi-linear equations in the Heisenberg group*, Comm. Pure Appl. Math. **50** (9) (1997) 867–889.

[24] L. Capogna, *Regularity for quasilinear equations and* 1*-quasiconformal maps in Carnot groups*, Math. Ann. **313** (2) (1999), 263–295.

[25] L. Capogna, D. Danielli, N. Garofalo, *An imbedding theorem and the Harnack inequality for nonlinear subelliptic equations.* Comm. Partial Diff. Equations **18** (1993), 1765–1794.

[26] L. Capogna, D. Danielli, N. Garofalo, *Subelliptic mollifiers and characterization of Rellich and Poincaré domains.* Rend. Sem. Mat. Univ. Polit. Torino **54** (1993), 361–386.

[27] L. Capogna, D. Danielli, N. Garofalo, *The geometric Sobolev embedding for vector fields and the isoperimetric inequality.* Comm. Anal. Geom. **2** (1994), 203–215.

[28] L. Capogna, D. Danielli, N. Garofalo, *Subelliptic mollifiers and a basic pointwise estimate of Poincaré type.* Math. Zeit. **226** (1997), 147–154.

[29] L. Capogna, D. Danielli, N. Garofalo, *Capacitary estimates and the local behavior of solutions to nonlinear subelliptic equations.* Amer. J. Math. **118** (1996), 1153–1196.

[30] L. Capogna, D. Danielli, S.D. Pauls and J.T. Tyson, *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem.* Progress in Mathematics **259**. Birkhäuser, 2007.

[31] V.M. Chernikov, S.K. Vodop′yanov, *Sobolev Spaces and hypoelliptic equations I, II.* Siberian Advances in Mathematics. **6** (3) (1996) 27–67; **6** (4), 64–96. Translation from: Trudy In-ta matematiki RAN. Sib. otd-nie. 29 (1995), 7–62.

[32] W.L. Chow, *Über Systeme von linearen partiellen Differentialgleichungen erster Ordung*, Math. Ann. **117** (1939), 98–105.

[33] G. Citti, N. Garofalo, E. Lanconelli, *Harnack's inequality for sum of squares of vector fields plus a potential*, Amer. J. Math. **115** (3) (1993) 699–734.

[34] G. Citti, A. Sarti, *A cortical based model of perceptual completion in the roto-translation space.* Lecture Notes of Seminario Interdisciplinare di Matematica **3** (2004), 145–161.

[35] D. Danielli, N. Garofalo, D.-M. Nhieu, *Trace inequalities for Carnot–Carathéodory spaces and applications to quasilinear subelliptic equations*, preprint.

[36] D. Danielli, N. Garofalo and D.-M. Nhieu, *Non-doubling Ahlfors measures, perimeter measures, and the characterization of the trace spaces of Sobolev functions in Carnot–Carathéodory spaces.* Mem. Amer. Math. Soc. 182 **857** (2006).

[37] Ya. Eliashberg, *Contact 3-Manifolds Twenty Years Since J. Martinet's Work.* Ann. Inst Fourier (Grenoble) **42** (1992), 1–12.

[38] Ya. Eliashberg, *New Invariants of Open Symplectic and Contact Manifolds.* J. Amer. Math. Soc. **4** (1991), 513–520.

[39] Ya. Eliashberg, *Classification of overtwisted contact structures on 3-manifolds.* Invent. Math. **98** (1989), 623–637.

[40] L.C. Evans, R.F. Gariepy, *Measure theory and fine properties of functions.* CRC Press, Boca Raton, 1992.

[41] H. Federer, *Curvature measures.* Trans. Amer. Math. Soc. **93** (1959), 418–491.

[42] H. Federer, *Geometric Measure Theory.* NY: Springer, 1969.

[43] H. Federer, W.H. Fleming *Normal and Integral Currents.* Ann. Math. **72** (2) (1960), 458–520.

[44] C. Fefferman, D.H. Phong, *Subelliptic eigenvalue problems.* Proceedings of the conference in harmonic analysis in honor of Antoni Zygmund, Wadsworth Math. Ser., Wadsworth, Belmont, California, 1981, 590–606.

[45] G.B. Folland, *A fundamental solution for a subelliptic operator*, Bull. Amer. Math. Soc. **79** (1973), 373–376.

[46] G.B. Folland, *Subelliptic estimates and function spaces on nilpotent Lie groups*, Ark. Mat. **13** (2) (1975) 161–207.

[47] G.B. Folland, E.M. Stein, *Hardy spaces on homogeneous groups.* Princeton Univ. Press, 1982.

[48] B. Franchi, *Weighted Sobolev–Poincaré inequalities and pointwise inequalities for a class of degenerate elliptic equations.* Trans. Amer. Math. Soc. **327** (1991), 125–158.

[49] B. Franchi, S. Gallot, R. Wheeden, *Sobolev and isoperimetric inequalities for degenerate metrics.* Math. Ann. **300** (1994), 557–571.

[50] B. Franchi, C.E. Gutiérrez, R.L. Wheeden, *Weighted Sobolev–Poincaré inequalities for Grushin type operators.* Comm. Partial Differential Equations **19** (1994), 523–604.

[51] B. Franchi, E. Lanconelli, *Hölder regularity theorem for a class of non uniformly elliptic operators with measurable coefficients.* Ann. Scuola Norm. Sup. Pisa **10** (1983), 523–541.

[52] B. Franchi, E. Lanconelli, *An imbedding theorem for Sobolev spaces related to non smooth vector fields and Harnack inequality.* Comm. Partial Differential Equations **9** (1984), 1237–1264.

[53] B. Franchi, G. Lu, R. Wheeden, *Representation formulas and weighted Poincaré inequalities for Hörmander vector fields.* Ann. Inst. Fourier (Grenoble) **45** (1995), 577–604.

[54] B. Franchi, G. Lu, R. Wheeden, *A relationship between Poincaré type inequalities and representation formulas in spaces of homogeneous type.* Int. Mat. Res. Notices (1) (1996), 1–14.

[55] B. Franchi, R. Serapioni, *Pointwise estimates for a class of strongly degenerate elliptic operators: a geometric approach.* Ann. Scuola Norm. Sup. Pisa **14** (1987), 527–568.

[56] B. Franchi, R. Serapioni and F. Serra Cassano, *Regular hypersurfaces, intrinsic perimeter and implicit function theorem in Carnot groups.* Comm. Anal. Geom. **11** (5) (2003), 909–944.

[57] B. Franchi, R. Serapioni and F. Serra Cassano, *Regular submanifolds, graphs and area formula in Heisenberg groups*, Adv. Math. **211** (1) (2007), 152–203.

[58] B. Franchi, R. Serapioni, F. Serra Cassano, *Rectifiability and Perimeter in the Heisenberg group*, Math. Ann. **321** (3) (2001), 479–531.

[59] B. Franchi, R. Serapioni, F. Serra Cassano, *On the structure of finite perimeter sets in step* 2 *Carnot groups*, J. Geom. Anal. **13** (3) (2003), 421–466.

[60] R. Garattini, *Harnack's inequality on homogeneous spaces*, Annali di Matematica Pura ed Applicata **179** (1) (2001), 1–16.

[61] N. Garofalo, *Analysis and Geometry of Carnot–Carathéodory Spaces, With Applications to PDE's*, Birkhäuser, in preparation.

[62] N. Garofalo, E. Lanconelli, *Existence and nonexistence results for semilinear equations on the Heisenberg group.* Indiana Univ. Math. J. **41** (1992), 71–98.

[63] N. Garofalo, D.-M. Nhieu, *Isoperimetric and Sobolev Inequalities for Carnot–Carathéodory Spaces and the Existence of Minimal Surfaces*, Comm. Pure Appl. Math. **49** (1996), 1081–1144.

[64] N. Garofalo, D.-M. Nhieu, *Lipschitz continuity, global smooth approximation and extension theorems for Sobolev functions in Carnot–Carathéodory spaces*, Jour. Anal. Math., **74** (1998), 67–97.

[65] M. Giaquinta, G. Modica, J. Souček, *Cartesian currents in the calculus of variations. V. I, II.* Springer-Verlag, Berlin, 1998.

[66] R.W. Goodman, *Nilpotent Lie groups: structure and applications to analysis.* Springer-Verlag, Berlin-Heidelberg-New York, 1976. Lecture Notes in Mathematics, vol. 562.

[67] A.V. Greshnov, *Metrics of Uniformly Regular Carnot–Carathéodory Spaces and Their Tangent Cones.* Sib. Math. Zh. **47** (2) (2006), 259–292.

[68] A.V. Greshnov, *Local Approximation of Equiregular Carnot–Carath'eodory Spaces by its Tangent Cones.* Sib. Math. Zh. **48** (2) (2007), 290–312.

[69] M. Gromov, *Groups of polynomial growth and expanding maps.* Inst. Hautes Etudes Sci. Publ. Math. **53** (1981), 53–73.

[70] M. Gromov, *Carnot–Carathéodory Spaces Seen From Within.* Sub-Riemannian geometry, Birkhäuser, Basel, 1996, 79–318.

[71] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces.* Birkhäuser, 2001.

[72] P. Hajłasz, P. Koskela, *Sobolev Met Poincaré.* Memoirs of the American Mathematical Society **145** (2000), no. 688.

[73] P. Hajłasz, P. Strzelecki, *Subelliptic p-harmonic maps into spheres and the ghost of Hardy spaces*, Math. Ann. **312** (1998), 341–362.

[74] J. Heinonen, *Calculus on Carnot groups.* Fall school in analysis, (Jyväskylä, 1994). Jyväskylä, University of Jyväskylä, 1994, pp. 1–32.

[75] R.K. Hladky, S.D. Pauls, *Minimal surfaces in the roto-translation group with applications to a neuro-biological image completion model.* arXiv:math.DG/0509636, 27 Sep. 2005.

[76] L. Hörmander, *Hypoelliptic second-order differential equations.* Acta Math. **119** (1967), 147–171.

[77] F. Jean, *Uniform estimation of sub-riemannian balls.* Journal on Dynamical and Control Systems **7** (4) (2001), 473–500.

[78] D. Jerison, *The Poincaré inequality for vector fields satisfying Hörmander's condition.* Duke Math. J. **53** (1986), 503–523.

[79] J. Jost, *Equilibrium maps between metric spaces.* Calc. Var. **2** (1994), 173–205.

[80] J. Jost, *Generalized harmonic maps between metric spaces*, in: Geometric analysis and the calculus of variations (J. Jost, ed.), International Press, 1966, 143–174.

[81] J. Jost, *Generalized Dirichlet forms and harmonic maps.* Calc. Var. **5** (1997), 1–19.

[82] J. Jost, *Nonlinear Dirichlet forms*, preprint.

[83] J. Jost, C. J. Xu, *Subelliptic harmonic maps.* Trans. Amer. Math. Soc. **350** (1998), 4633–4649.

[84] V. Jurdjevic, *Geometric Control Theory*, Cambridge Studies in Mathematics **52**. Cambridge University Press, 1997.

[85] M.B. Karmanova, *Metric Differentiability of Mappings and Geometric Measure Theory.* Doklady Mathematics **71** (2) (2005), 224–227.

[86] M.B. Karmanova, *Rectifiable Sets and the Coarea Formula for Metric-Valued Mappings.* Doklady Mathematics **73** (3) (2005), 323–327.

[87] M. Karmanova, *Geometric Measure Theory Formulas on Rectifiable Metric Spaces.* Contemporary Mathematics **424** (2007), 103–136.

[88] M.B. Karmanova, *Area and coarea formulas for the mappings of Sobolev classes with values in a metric space.* Sib. Math. J. **48** (4) (2007), 621–628.

[89] M. Karmanova, *Rectifiable Sets and Coarea Formula for Metric-Valued Mappings*, Journal of Functional Analysis **254** (5) (2008), 1410–1447.

[90] M. Karmanova *An Area Formula for Lipschitz Mappings of Carnot–Carathódory Spaces*, Doklady Mathematics **78** (3) (2008), 901–906.

[91] M. Karmanova *Characteristic Set of Smooth Contact Mappings of Carnot–Carathéodory Spaces*, Doklady Mathematics **79** (2008) (to appear).

[92] B. Kirchheim and F. Serra Cassano, *Rectifiability and parameterization of intrinsic regular surfaces in the Heisenberg group.* Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **3** (4) (2004), 871–896.

[93] A.S. Kronrod, *On functions of two variables.* Uspekhi Matematicheskikh Nauk (N. S.) **5** (1950), 24–134.

[94] G.P. Leonardi, S. Rigot, *Isoperimetric sets on Carnot groups*, Houston Jour. Math. **29** (3) (2003), 609–637.

[95] F. Lin, X. Yang, *Geometric measure theory – an introduction.* Science Press, Beijing a. o., 2002.

[96] W. Liu, H.J. Sussman, *Shortest paths for sub-Riemannian metrics on rank-two distributions*, Mem. Amer. Math. Soc. **118** (564) (1995).

[97] G. Lu, *Weighted Poincaré and Sobolev inequalities for vector fields satisfying Hörmander's condition and applications*, Rev. Mat. Iberoamericana **8** (3) (1992), 367–439.

[98] V. Magnani, *The coarea formula for real-valued Lipschitz maps on stratified groups.* Math. Nachr. **27** (2) (2001), 297–323.

[99] V. Magnani, *A blow-up theorem for regular hypersurfaces on nilpotent groups.* Man. math. **110** (2003), 55–76.

[100] V. Magnani, *Elements of Geometric Measure Theory on sub-Riemannian groups.* Tesi di Perfezionamento. Pisa: Scuola Normale Superiore (Thesis), 2002.

[101] V. Magnani, *Blow-up of regular submanifolds in Heisenberg groups and applications.* Cent. Eur. J. Math. **4** (1) (2006), 82–109.

[102] P. Malliavin, *Stochastic Analysis*, Springer, NY, 1997.

[103] A.I. Mal'cev, Doklady Akademii Nauk [Russian] (1941).

[104] S. Marchi, *Hölder continuity and Harnack inequality for De Giorgi classes related to Hörmander vector fields.* Ann. Mat. Pura Appl. **168** (1995), 171–188.

[105] G.A. Margulis, G.D. Mostow, *The differential of quasi-conformal mapping of a Carnot–Carathéodory spaces.* Geometric and Functional Analysis **5** (2) (1995), 402–433.

[106] G.A. Margulis, G.D. Mostow, *Some remarks on the definition of tangent cones in a Carnot–Carathéodory space.* Journal D'Ánalyse Math. **80** (2000), 299–317.

[107] G. Metivier, *Fonction spectrale et valeurs propres d'une classe d'opérateurs non elliptiques.* Commun. Partial Differential Equations **1** (1976), 467–519.

[108] J. Mitchell, *On Carnot–Carathéodory metrics.* J. Differential Geometry **21** (1985), 35–45.

[109] R. Montgomery, *Abnormal minimizers*, SIAM J. Control Optim. **32** (6) (1994), 1605–1620.

[110] R. Montgomery, *A Tour of Subriemannian Geometries, Their Geodesics and Applications.* Providence, AMS, 2002.

[111] R. Monti, *Some properties of Carnot–Carathéodory balls in the Heisenberg group*, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Mem. s.9 **11** (2000), 155–167.

[112] R. Monti, *Distances, boundaries and surface measures in Carnot–Carathéodory spaces*, PhD thesis.

[113] R. Monti and F. Serra Cassano, *Surface measures in Carnot–Carathédory spaces*, Calc. Var. Partial Differential Equations **13** (3) (2001), 339–376.

[114] A. Nagel, F. Ricci, E.M. Stein, *Fundamental solutions and harmonic analysis on nilpotent groups*, Bull. Amer. Math. Soc. (N.S.) **23** (1) (1990), 139–144.

[115] A. Nagel, F. Ricci, E.M. Stein, *Harmonic analysis and fundamental solutions on nilpotent Lie groups*, Analysis and partial differential equations, 249–275, Lecture Notes in Pure and Appl. Math. **122**, Dekker, New York, 1990.

[116] A. Nagel, E.M. Stein, S. Wainger, *Balls and metrics defined by vector fields I: Basic properties.* Acta Math. **155** (1985), 103–147.

[117] M. Ohtsuka, *Area Formula.* Bull. Inst. Math. Acad. Sinica **6** (2) (2) (1978), 599–636.

[118] P. Pansu, *Géometrie du group d'Heisenberg.* Univ. Paris VII, 1982.

[119] P. Pansu, *Une inégalité isopérimetrique sur le groupe de Heisenberg*, C.R. Acad. Sc. Paris, **295** Série I (1982), 127–130.

[120] P. Pansu, *Croissance des boules et des géodésiques fermées dans les nilvariété*, Ergod. Dinam. Syst. **3** (1983) 415–445.

[121] P. Pansu, *Métriques de Carnot–Carathéodory et quasiisométries des espaces symétriques de rang un.* Ann. Math. (2) **129** (1) (1989), 1–60.

[122] S.D. Pauls, *A Notion of Rectifiability Modeled on Carnot Groups*, Indiana University Mathematics Journal **53** (2004) 49–82.

[123] J. Petitot, *Neurogéométrie de la vision. Modèles mathématiques et physiques des architectures fonctionelles.* Les Éditions de l'École Polytechnique, 2008.

[124] M.M. Postnikov, *Lectures in Geometry. Semester V: Lie Groups and Lie Algebras.* Moscow, "Nauka", 1982.

[125] L.S. Pontryagin, *Continuous Groups* [Russian]. Moscow, "Nauka", 1984.

[126] P.K. Rashevsky, *Any two point of a totally nonholonomic space may be connected by an admissible line*, Uch. Zap. Ped. Inst. im. Liebknechta. Ser. Phys. Math. **2** (1938), 83–94.

[127] L.P. Rothschild, E. M. Stein, *Hypoelliptic differential operators and nilpotent groups.* Acta Math. **137** (1976), 247–320.

[128] A. Sánchez-Calle, *Fundamental solutions and geometry of sums of squares of vector fields.* Invent. Math. **78** (1984), 143–160.

[129] S.V. Selivanova, *The tangent cone to a regular quasimetric Carnot–Carathéodory space.* Doklady Mathematics (2009) (to appear).

[130] S.V. Selivanova, *The tangent cone to a quasimetric space with dilations.* Sib. Math. J. (2009) (to appear).

[131] S.V. Selivanova, S.K. Vodop′yanov, *The structure of the tangent cone to a quasimetric space with dilations* (in preparation).

[132] E. Siebert *Contractive automorphisms on locally compact groups.* Mat. Z. **191** (1986), 73–90.

[133] E.M. Stein, *Harmonic analysis: real-variables methods, orthogonality, and oscillatory integrals.* Princeton, NJ, Princeton University Press, 1993.

[134] R.S. Strichartz, *Sub-Riemannian geometry*, J. Diff. Geom. **24** (1986) 221–263. Corrections: J. Diff. Geom. **30** (1989), 595–596.

[135] K.T. Sturm, *Analysis on local Dirichlet spaces III. The parabolic Harnack inequality.* J. Math. Pures Appl. **75** (1996), 273–297.

[136] A.M. Vershik, V.Ya. Gershkovich, *Nonholonomic dynamical systems, geometry of distributions and variational problems.* Dynamical systems. VII. Encycl. Math. Sci., vol. 16. 1994, pp. 1–81 (English Translation from: Itogi Nauki Tekh., Ser. Sovrem. Probl. Mat., Fundam. Napravleniya vol. 16, Moscow, VINITI 1987, pp. 7–85).

[137] S.K. Vodop′yanov, *$\mathcal{P}$-differentiability on Carnot groups in different topologies and related topics.* Proceedings on Analysis and Geometry (S.K. Vodop′yanov, ed.) Novosibirsk, Sobolev Institute Press, 2000, pp. 603–670.

[138] S.K. Vodop′yanov, *Theory of Lebesgue Integral: Lecture Notes on Analysis.* Novosibirsk: NSU Publishing, 2003.

[139] S.K. Vodop′yanov, *Differentiability of Curves in Carnot Manifold Category.* Doklady Mathematics **74** (3) (2006), 799–804.

[140] S.K. Vodop′yanov, *Differentiability of mappings of Carnot Manifolds and Isomorphism of Tangent Cones.* Doklady Mathematics **74** (3) (2006), 844–848.

[141] S.K. Vodop′yanov, *Geometry of Carnot–Carathéodory Spaces and Differentiability of Mappings.* Contemporary Mathematics **424** (2007), 247–302.

[142] S.K. Vodop′yanov, *Differentiability of Mappings in Carnot Manifold Geometry.* Sib. Mat. J. **48** (2) (2007), 251–271.

[143] S.K. Vodop′yanov, A.V. Greshnov, *On the Differentiability of Mappings of Carnot–Carathéodory Spaces.* Doklady Mathematics **67** (2) (2003), 246–250.

[144] S.K. Vodop′yanov, M.B. Karmanova, *Local Geometry of Carnot Manifolds Under Minimal Smoothness.* Doklady Mathematics **75** (2) (2007), 240–246.

[145] S.K. Vodop′yanov, M.B. Karmanova, *Sub-Riemannian geometry under minimal smoothness of vector fields.* Doklady Mathematics **78** (2) (2008), 737–742.

[146] S.K. Vodop′yanov, M.B. Karmanova, *Coarea Formula for Smooth Contact Mappings of Carnot Manifolds.* Doklady Mathematics **76** (3) (2007), 908–912.

[147] S.K. Vodop′yanov, M.B. Karmanova, *An Area Formula for C^1-Smooth Contact Mappings of Carnot Manifolds.* Doklady Mathematics **78** (2) (2008), 655–659.

[148] S.K. Vodop′yanov, M.B. Karmanova, *An Area Formula for Contact C^1-Mappings of Carnot Manifolds.* Complex Variables and Elliptic Equations (2009) (accepted).

[149] S.K. Vodop′yanov, A.D. Ukhlov, *Approximately differentiable transformations and change of variables on nilpotent groups.* Siberian Math. J. **37** (1) (1996), 62–78.

[150] S.K. Vodop′yanov, A.D. Ukhlov, *Set functions and their applications in the theory of Lebesgue and Sobolev spaces. I.* Siberian Adv. Math. **14** (4) (2004), 78–125. *II.* Siberian Adv. Math. **15** (1) (2005), 91–125.

[151] F.W. Warner, *Foundations of differentiable manifolds and Lie groups.* New York a. o., Springer-Verlag, 1983. Graduate Texts in Mathematics, vol. 94.

[152] C.J. Xu, C. Zuily, *Higher interior regularity for quasilinear subelliptic systems*, Calc. Var. Partial Differential Equations **5** (4) (1997) 323–343.

Maria Karmanova
Sobolev Institute of Mathematics
Pr-t Akademika Koptyuga 4
Novosibirsk 630090, Russia
e-mail: `maryka@math.nsc.ru`

Sergey Vodop′yanov
Sobolev Institute of Mathematics
Pr-t Akademika Koptyuga 4
Novosibirsk 630090, Russia
e-mail: `vodopis@math.nsc.ru`

Analysis and Mathematical Physics
Trends in Mathematics, 337–347
© 2009 Birkhäuser Verlag Basel/Switzerland

Fourier Transforms of UD Integrals

Igor Kondrashuk and Anatoly Kotikov

Abstract. UD integrals published by N. Usyukina and A. Davydychev in 1992–1993 are integrals corresponding to ladder-type Feynman diagrams. The results are UD functions $\Phi^{(L)}$, where L is the number of loops. They play an important role in $\mathcal{N}=4$ supersymmetric Yang-Mills theory. The integrals were defined and calculated in the momentum space. In this paper the position space representation of UD functions is investigated. We show that Fourier transforms of UD functions are UD functions of space-time intervals but this correspondence is indirect. For example, the Fourier transform of second UD integral is the second UD integral.

Mathematics Subject Classification (2000). 81Q30.

Keywords. UD integrals, UD functions.

1. Introduction

As has been shown in Refs. [1]–[12], Slavnov-Taylor identity predicts that the correlators of dressed mean fields for $\mathcal{N}=4$ supersymmetric Yang-Mills theory in the position space can be represented in terms of UD integrals. The UD integrals correspond to the momentum representation of ladder diagrams and were calculated in Refs. [15, 16] in the momentum space, and the result can be written in terms of certain functions (UD functions) of conformally invariant ratios of momenta. Indeed, the Lcc correlator in the position space in Wess-Zumino-Landau gauge of maximally supersymmetric Yang-Mills theory is a function of Davydychev integral $J(1,1,1)$ at two loop level[1] [5, 6, 7]. By using Slavnov-Taylor identity one can

The work of I.K. was supported in part by Ministry of Education (Chile) under grant Mecesup FSM9901 and by DGIP UTFSM, by Fondecyt (Chile) project #1040368, and by Departamento de Ciencias Básicas de la Universidad del Bío-Bío, Chillán (Chile). The work of A.K. is supported by Fondecyt international cooperation project #7070064.

[1]In the position space Feynman diagrams contain integrations over coordinates of internal vertices. Integration over internal vertices appears in dual representation of the momentum diagrams too (see below and [13, 14]).

represent correlators of dressed mean gluons in terms of this integral at two-loop level. What kind of integrals will contribute to the scale-independent *Lcc* correlator at all loops is unclear at present. However, using a method of Ref. [7] one can suggest that at higher loop orders in the position space the UD integrals will survive only. Conformal invariance of the effective action of dressed mean fields in the position space, predicted in Refs. [1, 2, 3, 4] corresponds to the property of conformal invariance of UD integrals.

In the momentum space it was shown that UD functions are the only contributions (at least up to three loops) to off-shell four-point correlator of gluons that corresponds to four gluon amplitude [17, 18]. The conformal invariance of UD functions was used in the momentum space to calculate four point amplitude and to classify all possible contributions to it [19, 20]. Later, the conformal symmetry in the momentum space appeared on the string side in the Alday-Maldacena approach [21] in the limit of strong coupling.

The purpose of this paper is to find the position space representation of the ladder diagrams that produce UD functions in the momentum space. In this paper we show that Fourier transform of the second UD integral is the second UD integral and that Fourier transform of the first UD function can be related to the second UD function. We consider three-point ladder UD integrals and comment four-point ladder UD integrals. In Section 2 we illustrate the idea of the method on an example of the simplest diagram. The most important point is that the problem is solved diagrammatically via conformal transformation. Two other solutions to this problem are given in Section 3 and Section 4.

2. First UD triangle diagram

First UD triangle diagram is depicted on the l.h.s. of Fig. 1. All the notation used in this paper is the notation of Ref. [5]. To calculate it we use conformal transformation. The conformal transformation for each vector of the integrand

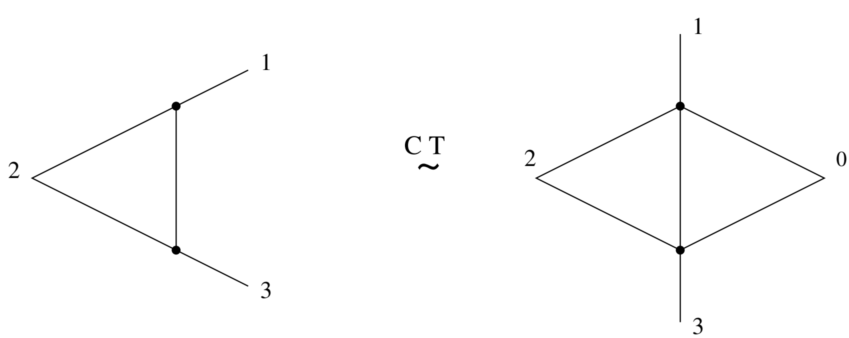


FIGURE 1. Initial conformal transformation in the position space.

(including the external vectors) is

$$y_\mu = \frac{y'_\mu}{y'^2}, \;\; z_\mu = \frac{z'_\mu}{z'^2}. \tag{2.1}$$

On the level of equations, the chain of conformal transformations of the l.h.s. of Fig. 1 according to Eq. (2.1) is

$$\begin{aligned}
&\int d^4y\; d^4z \frac{1}{[2y][yz][2z][3z][1y]} \\
&= [2']^2[3'][1'] \int d^4y'\; d^4z' \frac{1}{[2'y'][y'z'][2'z'][3'z'][1'y'][y'][z']} \\
&= [2']^2[3'][1'] \frac{1}{[3'1'][2']^2} \Phi^{(2)} \left(\frac{[1'2'][3']}{[3'1'][2']}, \frac{[1'][2'3']}{[3'1'][2']} \right) \\
&= \frac{[3'][1']}{[3'1']} \Phi^{(2)} \left(\frac{[1'2'][3']}{[3'1'][2']}, \frac{[1'][2'3']}{[3'1'][2']} \right) = \frac{1}{[31]} \Phi^{(2)} \left(\frac{[12]}{[31]}, \frac{[23]}{[31]} \right).
\end{aligned}$$

The second row of this chain of transformations looks like the second UD integral in the dual representation of Ref. [17]. It corresponds to the r.h.s. of Fig. 1. The last line is the conformal transformation back to the initial variables. Thus, we have proved the formula

$$\int d^4y\; d^4z \frac{1}{[2y][1y][3z][yz][2z]} = \frac{1}{[31]} \Phi^{(2)} \left(\frac{[12]}{[31]}, \frac{[23]}{[31]} \right). \tag{2.2}$$

After making Fourier transformation, we have the following representation for the l.h.s. (definitions of momenta are indicated on Fig. 2):

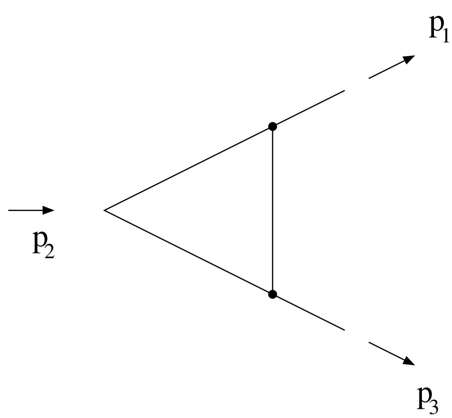


FIGURE 2. One loop diagram in the momentum space.

$$
\begin{aligned}
&\int d^4y\ d^4z \frac{1}{[2y][1y][3z][yz][2z]} \\
&= 4\pi^2 \int d^4p_1 d^4p_2 d^4p_3 d^4r\ \delta(p_1 - p_2 + p_3) \\
&\quad \times \frac{1}{p_1^2\ p_3^2} e^{ip_2x_2} e^{-ip_1x_1} e^{-ip_3x_3} \frac{1}{(r+p_1)^2 r^2 (r-p_3)^2} \\
&= 4\pi^2 \int d^4p_1 d^4p_2 d^4p_3\ \delta(p_1 - p_2 + p_3) e^{ip_2x_2} e^{-ip_1x_1} e^{-ip_3x_3} \\
&\quad \times \frac{1}{p_2^2\ p_1^2\ p_3^2} \Phi^{(1)}\left(\frac{p_1^2}{p_2^2}, \frac{p_3^2}{p_2^2}\right).
\end{aligned}
$$

Thus, comparing with Eq.(2.2), we can derive the first relation:

$$
\begin{aligned}
\frac{1}{[31]}\Phi^{(2)}\left(\frac{[12]}{[31]}, \frac{[23]}{[31]}\right) &= 4\pi^2 \int d^4p_1 d^4p_2 d^4p_3\ \delta(p_1 - p_2 + p_3) \\
&\quad \times e^{ip_2x_2} e^{-ip_1x_1} e^{-ip_3x_3} \frac{1}{p_2^2\ p_1^2\ p_3^2} \Phi^{(1)}\left(\frac{p_1^2}{p_2^2}, \frac{p_3^2}{p_2^2}\right).
\end{aligned} \tag{2.3}
$$

However, looking at the definition of the UD integrals in Refs. [15, 16], we can write from Eq. (2.3) another relation:

$$
\begin{aligned}
\frac{1}{[31]^2}\Phi^{(2)}\left(\frac{[12]}{[31]}, \frac{[23]}{[31]}\right) &= \frac{1}{16\pi^4} \int d^4p_1 d^4p_2 d^4p_3\ \delta(p_1 - p_2 + p_3) \\
&\quad \times e^{ip_2x_2} e^{-ip_1x_1} e^{-ip_3x_3} \frac{1}{(p_2^2)^2} \Phi^{(2)}\left(\frac{p_1^2}{p_2^2}, \frac{p_3^2}{p_2^2}\right).
\end{aligned} \tag{2.4}
$$

The next two sections demonstrate how to derive the formula (2.2) by other two different methods.

3. Graphical identity

First of all, we show validity of the graphical identity of Fig. 3. This is identity in the position space. We assume integration over internal vertices. This identity can be proved in two ways.

1. First way to prove Fig. 3 is to use a relation on Fig. 4. This is a graphical representation of the equation (rules of the integration are taken from Ref. [5])

$$
\partial^2_{(y)} \frac{1}{[1y]^{1-\epsilon}} = k(\epsilon)\delta^{(4-2\epsilon)}(1y)
$$

from Ref [7]. The coefficient between the l.h.s. and the r.h.s. of Fig. 4 is $k = -4$ in the number of dimensions $d = 4$. On the other side, d'Alembertian can travel along

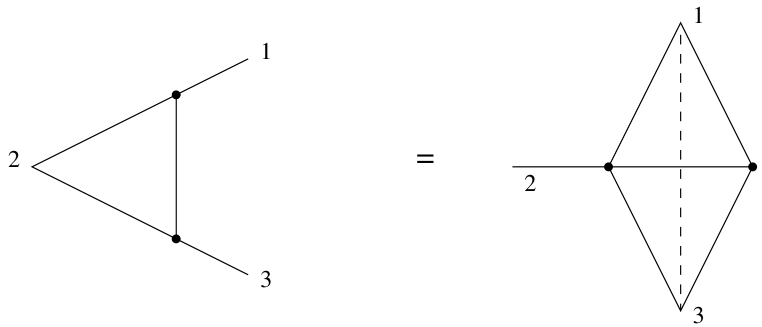


FIGURE 3. Graphical identity.

the propagator. Computer program of Ref. [7], written in *Mathematica*, produces the equation[2]

$$\partial^2_{(2)} \int Dy\, Dz \frac{1}{[2y][1y][3z][yz][2z]} = -\frac{4[31]}{[12][23]}\, J(1,1,1). \tag{3.1}$$

This identity is depicted on Fig. 5. The dash lines correspond to the inverse

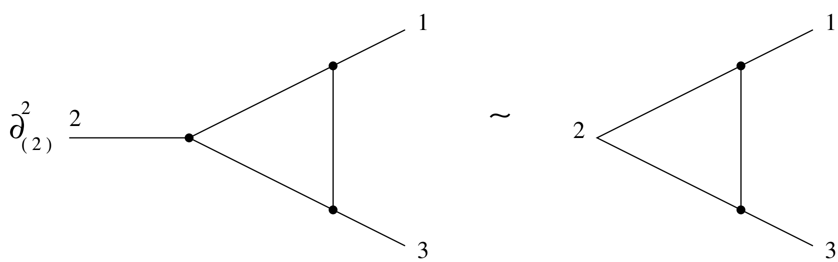


FIGURE 4. Use of d'Alembertian.

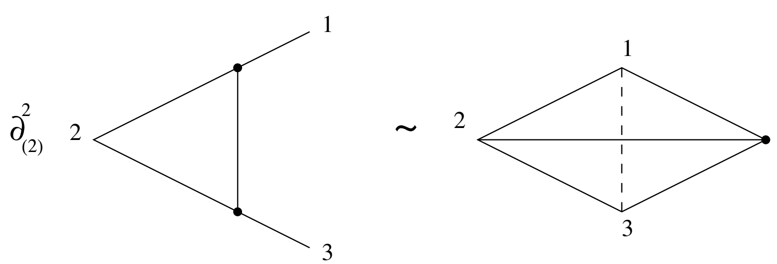


FIGURE 5. Use of d'Alembertian.

[2]The l.h.s. of (3.1) appears on page 24 of Ref. [7]

propagator. The coefficient between the l.h.s. and the r.h.s. of Fig. 5 is $k = -4$. Combining Fig. 4 and Fig. 5 we reproduce the graphical identity of Fig. 3. Integration in the internal vertices includes powers of π due to the definition of Ref. [5]. Since both parts of Fig. 3 contain two internal vertices, the identity of Fig. 3 is valid for a usual four-dimensional measure of integration. Thus, we have proved the formula:

$$\int d^4y\, d^4z \frac{1}{[2y][1y][3z][yz][2z]} = [31] \int d^4y\, d^4z \frac{1}{[2y][yz][1y][3y][1z][3z]}. \qquad (3.2)$$

This formula corresponds to the graphical identity of Fig. 3.

2. Another way to show validity of Fig. 3 is a useful identity of Ref. [17] in the position space which can be obtained from the property that $\Phi^{(2)}$ function depends on two conformally invariant ratios of spacetime intervals. This representation is valid in the position space. The turning identity is re-presented in Fig. 6. Historically it appeared in Ref. [17] as a "dual" representation of the momentum two-loop UD integral which is not exactly the same as a position representation (the position representation is usual Feynman ladder diagram integrated over coordinates of internal vertices). Internal vertices correspond to the momenta that run into the loops. However, in the dual representation the integrations are done over "coordinates" of the internal vertices too and thus the dual diagram can be considered as another Feynman diagram in the position space. By multiplying both parts of the

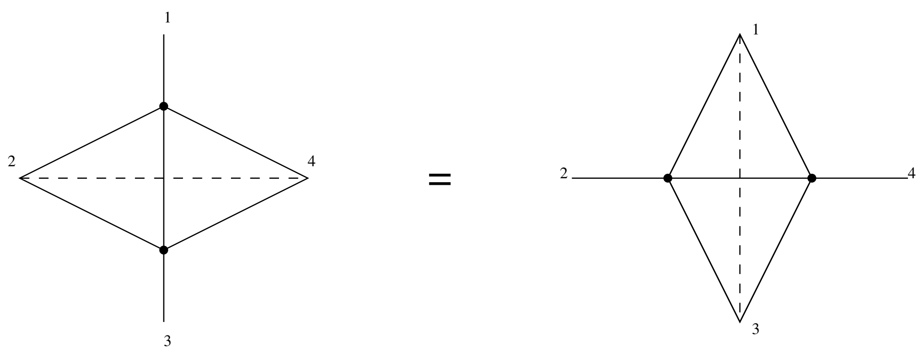


FIGURE 6. Turning identity.

turning identity by propagator [24], we have another graphical identity depicted on Fig. 7. Integrating both parts over the variable x_4 in $d = 4 - 2\epsilon$ dimensions, we obtain the relation re-presented on Fig. 8, where

$$A(1, 1, 2 - 2\epsilon) = \frac{\Gamma^2(1-\epsilon)\Gamma(\epsilon)}{\Gamma(2-2\epsilon)}.$$

Cancelling the coefficient $A(1, 1, 2 - 2\epsilon)$ in the both parts and taking the limit $\epsilon \to 0$, we reproduce Fig. 3.

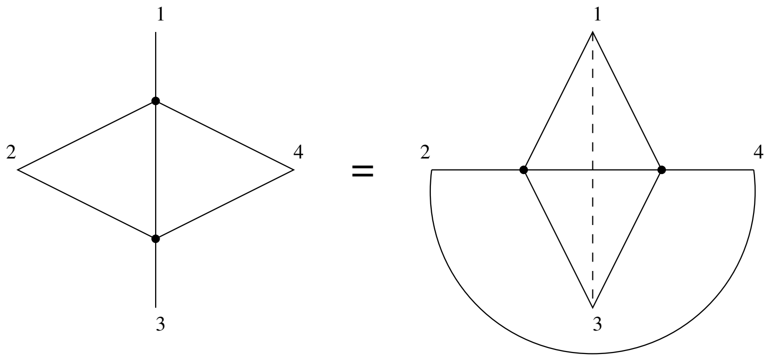


FIGURE 7. Multiplication by propagator.

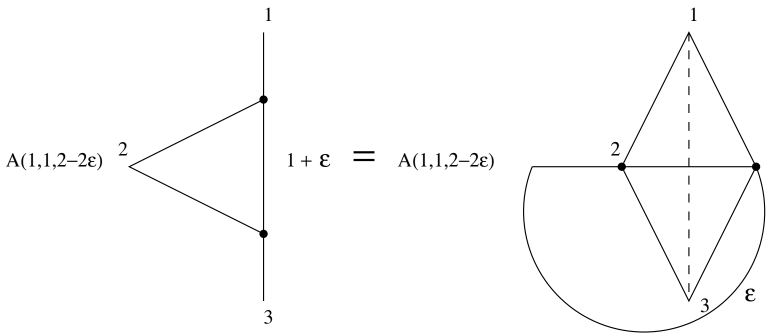


FIGURE 8. Integration over x_4.

4. Relation between graphical identity and UD integral

After proving Fig. 3, we can relate its r.h.s. to UD integrals. This can be done in two ways.

1. First way is by conformal transformation of the r.h.s. of Fig. 3. Indeed, the integral that corresponds to the r.h.s. of Fig. 3 is

$$\int d^4y\ d^4z \frac{1}{[2y][yz][1y][3y][1z][3z]}.$$

According to the conformal substitution of Eq. (2.1) the integral of the r.h.s. of Fig. 3 can be transformed to

$$\begin{aligned}\int d^4y\ d^4z \frac{1}{[2y][yz][1y][3y][1z][3z]} &= \int \frac{d^4y'\ d^4z'}{[y']^4[z']^4} \frac{[2'][3']^2[1']^2[y']^4[z']^3}{[2'y'][y'z'][1'y'][3'y'][1'z'][3'z']} \\ &= \int d^4y'\ d^4z' \frac{[2'][3']^2[1']^2}{[2'y'][y'z'][1'y'][3'y'][1'z'][3'z'][z']}.\end{aligned} \tag{4.1}$$

The transformation is presented on Fig. 9. The r.h.s. of Fig. 9 looks like the second UD integral in the dual representation of Ref. [17]. Thus, we can represent Eq. (4.1)

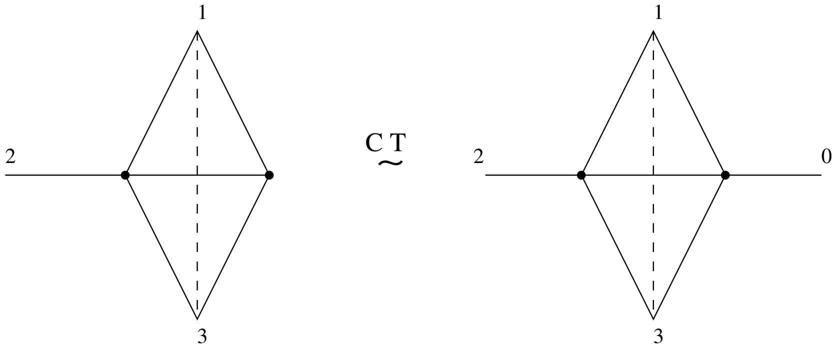


FIGURE 9. Conformal transformation of the r.h.s. of Fig. 3.

as

$$\int d^4y'\ d^4z' \frac{[2'][3']^2[1']^2}{[2'y'][y'z'][1'y'][3'y'][1'z'][3'z'][z']}$$
$$= [2'][3']^2[1']^2 \frac{1}{[3'1']^2[2']} \Phi^{(2)} \left(\frac{[1'2'][3']}{[3'1'][2']}, \frac{[1'][2'3']}{[3'1'][2']} \right)$$
$$= \frac{[3']^2[1']^2}{[3'1']^2} \Phi^{(2)} \left(\frac{[1'2'][3']}{[3'1'][2']}, \frac{[1'][2'3']}{[3'1'][2']} \right) = \frac{1}{[31]^2} \Phi^{(2)} \left(\frac{[12]}{[31]}, \frac{[23]}{[31]} \right).$$

The last line is the conformal transformation back to the initial variables of Eq. (2.1). Thus, we have demonstrated that

$$\int d^4y\ d^4z \frac{1}{[2y][yz][1y][3y][1z][3z]} = \frac{1}{[31]^2} \Phi^{(2)} \left(\frac{[12]}{[31]}, \frac{[23]}{[31]} \right).$$

Taking into account Eq. (3.2), we have proved the formula of Eq. (2.2)

$$\int d^4y\ d^4z \frac{1}{[2y][1y][3z][yz][2z]} = \frac{1}{[31]} \Phi^{(2)} \left(\frac{[12]}{[31]}, \frac{[23]}{[31]} \right).$$

2. Second way to demonstrate validity of Eq. (3.2) does not require the conformal transformation. The dual representation for the two-loop diagram on Fig. 10 in the momentum space is given on Fig. 11. Thus, for the case $\alpha = \beta = 0$ we have the structure of the dual diagram depicted on Fig. 12. However, using the definition of UD functions of Refs. [15, 16] and taking into account the relation for dual momenta, one can see that the r.h.s. of Fig. 12 is

$$\frac{1}{[31]^2} \Phi^{(2)} \left(\frac{[12]}{[31]}, \frac{[23]}{[31]} \right). \tag{4.2}$$

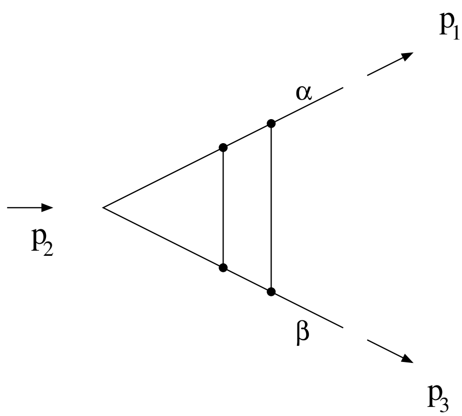


Figure 10. Two-loop diagram.

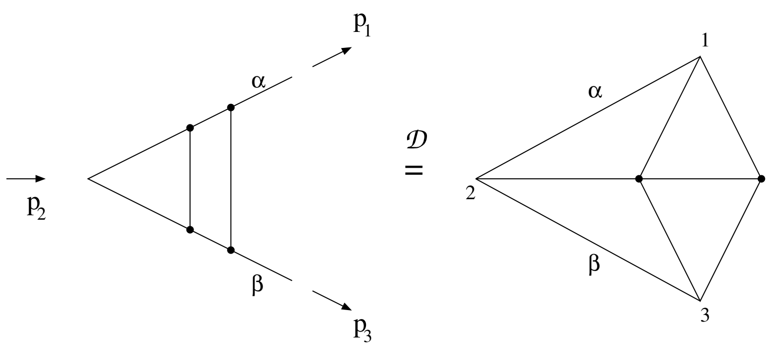


Figure 11. Dual representation.

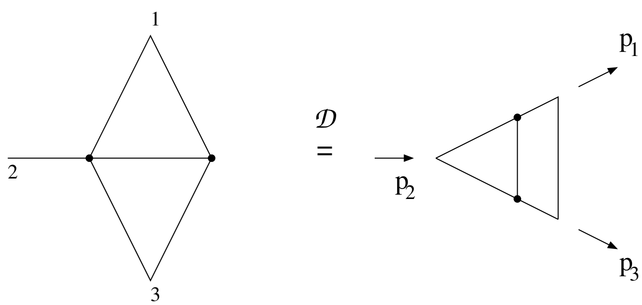


Figure 12. $\alpha = \beta = 0$.

Thus, comparing the l.h.s. of Fig. 12 with the graphical identity Fig. 3 and taking into account Eq. (4.2) we have in the position space for the l.h.s. of Fig. 3 the result

$$\frac{1}{[31]}\Phi^{(2)}\left(\frac{[12]}{[31]},\frac{[23]}{[31]}\right).$$

This coincides with the result in Eq. (2.2).

5. Conclusion

We have shown that Fourier transform of the second UD integral is the second UD integral, and that Fourier transforms of the first and the second UD functions are related. Apart from pure academic interest, this conclusion allows to investigate correlators of $\mathcal{N}=4$ supersymmetric Yang-Mills theory in the position space. It is useful from the point of view of Slavnov-Taylor identity. This identity will allow to find even maybe yet unknown relations between different types of UD integrals. We hope that the property exists above the present consideration, that is for four-point ladder diagrams and at higher loops too. We plan to consider this in future.

Acknowledgment

We thank Alvaro Vergara for his work with computer graphics for this paper. I.K. is grateful to organizers of conference "New trends in complex and harmonic analysis" for giving him wonderful opportunity to participate in it and for providing financial support of his stay in Bergen.

References

[1] G. Cvetič, I. Kondrashuk and I. Schmidt, "Effective action of dressed mean fields for $N = 4$ super-Yang-Mills theory," Mod. Phys. Lett. A **21** (2006) 1127 [hep-th/0407251].

[2] I. Kondrashuk and I. Schmidt, "Finiteness of $n = 4$ super-Yang-Mills effective action in terms of dressed $n = 1$ superfields," hep-th/0411150

[3] K. Kang and I. Kondrashuk, "Semiclassical scattering amplitudes of dressed gravitons," hep-ph/0408168.

[4] G. Cvetič, I. Kondrashuk and I. Schmidt, "On the effective action of dressed mean fields for $N = 4$ super-Yang-Mills theory," in *Symmetry, Integrability and Geometry: Methods and Applications,* SIGMA (2006) 002 [math-ph/0601002].

[5] G. Cvetič, I. Kondrashuk, A. Kotikov and I. Schmidt, "Towards the two-loop Lcc vertex in Landau gauge," Int. J. Mod. Phys. A **22** (2007) 1905 [arXiv:hep-th/0604112].

[6] G. Cvetič and I. Kondrashuk, "Further results for the two-loop Lcc vertex in the Landau gauge," J. High Energy Phys. **2** (2008) 023 [arXiv:hep-th/0703138]

[7] G. Cvetic and I. Kondrashuk, "Gluon self-interaction in the position space in Landau gauge," Int. J. Mod. Phys. A **23** (2008) 4145 [arXiv:0710.5762[hep-th]].

[8] I. Kondrashuk, "An approach to solve Slavnov-Taylor identity in D4 $N = 1$ supergravity," Mod. Phys. Lett. A **19** (2004) 1291 [gr-qc/0309075].

[9] I. Kondrashuk, G. Cvetič, and I. Schmidt, “Approach to solve Slavnov-Taylor identities in nonsupersymmetric non-Abelian gauge theories,” Phys. Rev. D **67** (2003) 065006 [hep-ph/0203014].

[10] G. Cvetič, I. Kondrashuk and I. Schmidt, “QCD effective action with dressing functions: Consistency checks inperturbative regime,” Phys. Rev. D **67** (2003) 065007 [hep-ph/0210185].

[11] I. Kondrashuk, “The solution to Slavnov-Taylor identities in D4 $N = 1$ SYM,” JHEP **0011**, 034 (2000) [hep-th/0007136].

[12] I. Kondrashuk, “Renormalizations in softly broken $N = 1$ theories: Slavnov-Taylor identities,” J. Phys. A **33** (2000) 6399 [hep-th/0002096].

[13] D.I. Kazakov and A.V. Kotikov, “The method of uniqueness “Multiloop calculation in QCD” Theor. Math. Phys. **73** (1988) 1264 [Teor. Mat. Fiz. **73** (1987) 348];

[14] D.I. Kazakov and A.V. Kotikov, “Total alpha-s correction to deep inelastic scattering cross-section ratio, R = sigma-L / sigma-t in qcd. Calculation of longitudinal structure function,” Nucl. Phys. B **307** (1988) 721 [Erratum-ibid. B **345** (1990) 299].

[15] N.I. Usyukina and A.I. Davydychev, “An approach to the evaluation of three and four point ladder diagrams,” Phys. Lett. B **298** (1993) 363.

[16] N.I. Usyukina and A.I. Davydychev, “Exact results for three and four point ladder diagrams with an arbitrary number of rungs,” Phys. Lett. B **305** (1993) 136.

[17] J.M. Drummond, J. Henn, V.A. Smirnov and E. Sokatchev, “Magic identities for conformal four-point integrals,” JHEP **0701** (2007) 064 [arXiv:hep-th/0607160].

[18] Z. Bern, L.J. Dixon and V.A. Smirnov, “Iteration of planar amplitudes in maximally supersymmetric Yang-Mills theory at three loops and beyond,” Phys. Rev. D **72** (2005) 085001 [hep-th/0505205].

[19] Z. Bern, M. Czakon, L.J. Dixon, D.A. Kosower and V.A. Smirnov, “The Four-Loop Planar Amplitude and Cusp Anomalous Dimension in Maximally Supersymmetric Yang-Mills Theory,” Phys. Rev. D **75** (2007) 085010 [arXiv:hep-th/0610248].

[20] D. Nguyen, M. Spradlin and A. Volovich, “New Dual Conformally Invariant Off-Shell Integrals,” arXiv:0709.4665 [hep-th].

[21] L.F. Alday and J.M. Maldacena, “Gluon scattering amplitudes at strong coupling,” JHEP **0706** (2007) 064 [arXiv:0705.0303 [hep-th]].

Igor Kondrashuk
Departamento de Ciencias Básicas, Universidad del Bío-Bío
Campus Fernando May, Casilla 447, Chillán, Chile;
Centro de Estudios Subatómicos y Departamento de Física,
Universidad Técnica Federico Santa María, Casilla 110-V, Valparaíso, Chile
e-mail: `igor.kondrashuk@ubiobio.cl`

Anatoly Kotikov
Departamento de Ciencias Básicas, Universidad del Bío-Bío
Campus Fernando May, Casilla 447, Chillán, Chile;
Bogoliubov Laboratory of Theoretical Physics
Joint Institute for Nuclear Research, Dubna, Russia
e-mail: `anatoly.kotikov@mail.desy.de`

Analysis and Mathematical Physics
Trends in Mathematics, 349–368
© 2009 Birkhäuser Verlag Basel/Switzerland

Fredholm Eigenvalues of Jordan Curves: Geometric, Variational and Computational Aspects

Samuel Krushkal

Abstract. The Fredholm eigenvalues of closed Jordan curves L on the Riemann sphere $\widehat{\mathbb{C}}$ (especially their least nontrivial values $\rho_L = \rho_1$) are intrinsically connected with conformal and quasiconformal maps and have various applications. These values have been investigated by many authors from different points of view.

We provide completely different quantitative and qualitative approaches which involve the complex Finsler geometry of the universal Teichmüller space, the metrics of generalized negative curvature and holomorphic motions.

Mathematics Subject Classification (2000). Primary: 30C35, 30C62, 30C75, 32G15; Secondary: 30F60, 32F45, 45C05, 53A35.

Keywords. Fredholm eigenvalue, univalent function, quasiconformal, Grunsky inequalities, universal Teichmüller space, Teichmüller metric, Kobayashi metric, Gaussian curvature, variation, generalized Laplacian.

1. Introductory remarks

The Fredholm eigenvalues of closed Jordan curves L on the Riemann sphere $\widehat{\mathbb{C}}$ have various applications and have been investigated by many authors from different points of view (see, e.g., the classical works of Ahlfors, Gaier, Schiffer, Warschawski et al.). These values, especially their least nontrivial values $\rho_L = \rho_1$, are intrinsically connected with conformal and quasiconformal maps.

Though many interesting problems still remain open. One of the reasons is that there are no appropriate variational methods for the related maps. The variational approach is very important and requires further investigations. There are also important computational problems, first of all to find the general algorithms for exact and approximate calculation of these values. We will see that this is important, for example, in quasiconformal classification of curves.

In this talk, we provide completely different quantitative and qualitative approaches. They rely on the deep results of complex Finsler geometry of universal Teichmüller space $\mathbf{T}$ and accordingly involve the metrics of generalized negative curvature and holomorphic motions. It became possible in this way to solve, for example, certain the old problem of Kühnau on inversion of classical Ahlfors and Grunsky inequalities, certain extremal problems for Fredholm eigenvalues going back to Schiffer and obtain many other results.

We briefly survey here the main recently obtained results in these topics.

2. Fredholm eigenvalues

2.1. The **Fredholm eigenvalues** ρ_n of a smooth closed Jordan curve $L \subset \widehat{\mathbb{C}}$ are the eigenvalues of its double-layer potential, or equivalently, of the integral equation

$$u(z) = \frac{\rho}{\pi} \int\limits_L u(\zeta) \frac{\partial}{\partial n_\zeta} \log \frac{1}{|\zeta - z|} ds_\zeta. \tag{1}$$

This equation has many applications, in particular, in two-dimensional potential theory, in the theory of boundary value problems (solving the Dirichlet and Neumann problems either in bounded or in unbounded domains), in approximate construction of conformal maps, etc. (see, e.g., [1], [8], [23], [27], [34], [35], [36], [40] and references cited there). Note that

$$-\frac{\partial}{\partial n_\zeta} \log |\zeta - z| ds_\zeta = d_\zeta \arg(\zeta - z);$$

this equality is useful, for example, in numerical applications.

The smallest eigenvalue $\rho_L = \rho_1 > 1$ plays a crucial role. One of the reasons is that by applying to the equation (1) the standard approximation method, the speed of approximation is equal to $O(1/\rho_L)$.

This value ρ_L can be defined for any oriented closed Jordan curve L on the Riemann sphere $\widehat{\mathbb{C}}$ by

$$\frac{1}{\rho_L} = \sup \frac{|\mathcal{D}_G(u) - \mathcal{D}_{G^*}(u)|}{\mathcal{D}_G(u) + \mathcal{D}_{G^*}(u)},$$

where G and G^* are, respectively, the interior and exterior of L; $\mathcal{D}$ denotes the Dirichlet integral, and the supremum is taken over all functions u continuous on $\widehat{\mathbb{C}}$ and harmonic on $G \cup G^*$. In particular, $\rho_L = \infty$ only for the circle.

A basic ingredient for estimating ρ_L is the well-known Ahlfors inequality

$$\frac{1}{\rho_L} \le q_L, \tag{2}$$

where q_L is the reflection coefficient of L, i.e., the minimal dilatation of quasiconformal reflections across L (that is, of the orientation reversing quasiconformal homeomorphisms of $\widehat{\mathbb{C}}$ which preserve the curve L point-wise); see, e.g., [2], [20], [27]. This inequality remains invariant under the action of the group $\mathbf{PSL}(2,\mathbb{C})$.

The inequality (2) is a basis for all known algorithms for the evaluation of exact or approximate values of ρ_L for specified curves L.

3. Ahlfors and Grunsky inequalities

3.1. It suffices to take the images $L = f^\mu(S^1)$ of the unit circle under quasiconformal self-maps of $\widehat{\mathbb{C}}$ with Beltrami coefficients $\mu = \partial_{\bar{z}} f/\partial_z f$ supported in the unit disk Δ. Then q_L equals the minimal dilatation $k(f^\mu) = \|\mu\|_\infty$ of such maps, and the inequality (2) is reduced to a strengthened Grunsky inequality.

The classical Grunsky theorem states that a holomorphic function $f(z) = z + \text{const} + O(z^{-1})$ in a neighborhood U_0 of $z = \infty$ is extended to a univalent holomorphic function on the disk

$$\Delta^* = \{z \in \widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\} : \ |z| > 1\}$$

if and only if its **Grunsky coefficients** α_{mn} satisfy the inequalities

$$\Big| \sum_{m,n=1}^{\infty} \sqrt{mn}\ \alpha_{mn} x_m x_n \Big| \le 1. \tag{3}$$

These coefficients are generated by

$$\log \frac{f(z) - f(\zeta)}{z - \zeta} = - \sum_{m,n=1}^{\infty} \alpha_{mn} z^{-m} \zeta^{-n}, \quad (z,\zeta) \in (\Delta^*)^2, \tag{4}$$

where $\mathbf{x} = (x_n)$ runs over the unit sphere $S(l^2)$ of the Hilbert space l^2 with $\|\mathbf{x}\| = \Big(\sum_1^\infty |x_n|^2\Big)^{1/2}$, and the principal branch of logarithmic function is chosen (cf. [12]). In particular, this assumes that $f(z) \neq 0$ on Δ^*. The quantity

$$\varkappa(f) := \sup \Big\{ \Big| \sum_{m,n=1}^{\infty} \sqrt{mn}\ \alpha_{mn} x_m x_n \Big| : \mathbf{x} = (x_n) \in S(l^2) \Big\} \tag{5}$$

is called the **Grunsky norm** of f.

Note that the function $g(z) = 1/f(1/z) = z + a_2 z^2 + \cdots$, which is univalent in Δ, has the same Grunsky coefficients α_{mn} as $f(z)$.

Consider the class Σ of univalent holomorphic functions $f(z) = z + b_0 + b_1 z^{-1} + \cdots$ mapping the disk Δ^* into $\widehat{\mathbb{C}} \setminus \{0\}$, and let $\Sigma(k)$ be its subclass of f with k-quasiconformal extensions to the unit disk $\Delta = \{|z| < 1\}$ so that $f(0) = 0$, and $\Sigma^0 = \bigcup_k \Sigma(k)$.

Grunsky's theorem has been essentially strengthened for the functions with quasiconformal extensions, for which we have instead of (3) a stronger inequality (see [25])

$$\Big|\sum_{m,n=1}^{\infty} \sqrt{mn}\ \alpha_{mn} x_m x_n\Big| \le k = k(f), \tag{6}$$

where

$$k(f) := \inf\{k(w^\mu) = \|\mu\|_\infty : \ w^\mu|\partial\Delta^* = f\}$$

is the **minimal dilatation** of (quasiconformal extensions of) f.

Here $\|\mu\|_\infty = \operatorname{ess\,sup}_{\mathbb{C}} |\mu(z)|$.

This inequality implies that the Grunsky and Teichmüller norms of $f \in \Sigma^0$ are related as follows:

$$\varkappa(f) \le k(f). \tag{7}$$

On the other hand, by a theorem of Pommerenke and Zhuravlev, any $f \in \Sigma$ with $\varkappa(f) \le k < 1$, has k_1-quasiconformal extensions to $\widehat{\mathbb{C}}$ with $k_1 = k_1(k) \ge k$ (see [31], [42]; [23, pp. 82–84]).

3.2. The eigenvalues ρ_L are intrinsically connected with the Grunsky matrix $(\alpha_{mn}(f))$, which is qualitatively expressed by Kühnau-Schiffer theorem which states that ρ_L is reciprocal to the Grunsky norm $\varkappa(f)$ of the Riemann mapping function of the exterior domain of L (cf. [27], [35]). This is one of the key results in the investigation and estimation of Fredholm eigenvalues by applying the Grunsky inequalities technique.

Our first main theorem relates to Moser's conjecture that each function $f \in \Sigma$ can be approximated locally uniformly on Δ^* by functions $f_n \in \Sigma^0$ with $\varkappa(f_n) < k(f_n)$, which sheds light to deep geometric features of Fredholm eigenvalues and is recently proved in [26]. To formulate it, we shall need certain results concerning the universal Teichmüller space; we present them in the next section.

3.3. Note that the set of $f \in \Sigma^0$ with $\varkappa(f) = k(f)$, or equivalently, with $1/\rho_{f(S^1)} = k(f)$, is rather sparse in Σ^0, but these functions play a crucial role in various applications of the Grunsky inequalities technique.

To describe these functions, denote

$$A_1(\Delta) = \{\psi \in L_1(\Delta) : \ \psi \text{ holomorphic}\},$$
$$A_1^2 = \{\psi \in A_1(\Delta) : \ \psi = \omega^2, \ \omega \text{ holomorphic}\}$$

and put

$$\langle \mu, \psi\rangle_\Delta = \iint_\Delta \mu(z)\psi(z)dxdy \quad (\mu \in L_\infty(\Delta), \ \psi \in L_1(\Delta), \ z = x + iy).$$

Then we have the following basic result.

Theorem 3.1. [15], [20] *The equality* $1/\rho_L = \varkappa(f) = k(f)$ *for the Riemann mapping function of the exterior* (*or interior*) *of* L *holds if and only if the function* f *is the restriction to* Δ^* *of a quasiconformal self-map* w^{μ_0} *of* $\widehat{\mathbb{C}}$ *with Beltrami coefficient* μ_0 *satisfying the condition*

$$\sup |\langle \mu_0, \varphi\rangle_\Delta| = \|\mu_0\|_\infty, \tag{8}$$

where the supremum is taken over holomorphic functions $\varphi \in A_1^2(\Delta)$ *with* $\|\varphi\|_{A_1(\Delta)} = 1$.

If, in addition, the equivalence class of f *(the collection of maps equal* f *on* S^1*) is a Strebel point, i.e., contains an extremal Teichmüller map, then* μ_0 *is necessarily of the form*

$$\mu_0(z) = \|\mu_0\|_\infty |\psi_0(z)|/\psi_0(z) \quad \textit{with } \psi_0 \in A_1^2 \textit{ in } \Delta. \tag{9}$$

The condition (8) has a geometric nature and equality (9) holds, for example, for all f with asymptotically conformal values on the unit circle. For analytic curves $f(S^1)$ the equality (9) was obtained by a different method in [28]. Note also that Strebel points are dense in Teichmüller spaces (cf. [9], [39]).

4. Universal Teichmüller space and its Finsler metrics

Recall that the universal Teichmüller space $\mathbf{T}$ is the space of quasisymmetric homeomorphisms of the unit circle $S^1 = \partial\Delta$ factorized by Möbius maps.

This space admits a complex Banach structure defined by factorization of the unit ball of Beltrami coefficients

$$\mathrm{Belt}(\Delta)_1 = \{\mu \in L_\infty(\mathbb{C}) : \ \mu|\Delta^* = 0, \ \|\mu\|_\infty < 1\}, \tag{10}$$

letting $\mu, \nu \in \mathrm{Belt}(\Delta)_1$ be equivalent if the corresponding maps $w^\mu, w^\nu \in \Sigma^0$ coincide on S^1 (and hence on the closed disk $\overline{\Delta^*}$). The equivalence class of a map w^μ will be denoted by $[w^\mu]$.

Using the functions $f \in \Sigma^0$, the universal Teichmüller space can be modelled (holomorphically embedded) as a bounded domain in the Banach space $\mathbf{B}$ of hyperbolically bounded holomorphic functions in Δ^* with finite norm $\|\varphi\|_{\mathbf{B}} = \sup_{\Delta^*}(|z|^2 - 1)^2|\varphi(z)|$.

All $\varphi \in \mathbf{B}$ can be regarded as the **Schwarzian derivatives**

$$S_f(z) = \Big(\frac{f''(z)}{f'(z)}\Big)' - \frac{1}{2}\Big(\frac{f''(z)}{f'(z)}\Big)^2, \quad z \in \Delta^*,$$

of locally univalent holomorphic functions in Δ^*. The points of $\mathbf{T}$ represent $f \in \Sigma^0$, i.e., the univalent functions in the whole disk Δ^* with quasiconformal extensions to $\widehat{\mathbb{C}}$.

The defining projection $\phi_{\mathbf{T}} : \mu \to S_{w^\mu}$ is a holomorphic map from $L_\infty(\Delta)$ to $\mathbf{B}$.

An intrinsic complete metric on the space $\mathbf{T}$ is the **Teichmüller metric** defined by

$$\tau_{\mathbf{T}}(\phi_{\mathbf{T}}(\mu), \phi_{\mathbf{T}}(\nu)) = \frac{1}{2} \inf\big\{\log K\big(w^{\mu_*} \circ \big(w^{\nu_*}\big)^{-1}\big) : \ \mu_* \in \phi_{\mathbf{T}}(\mu), \nu_* \in \phi_{\mathbf{T}}(\nu)\big\}. \tag{11}$$

It is generated by the **Finsler structure** on the tangent bundle $\mathcal{T}(\mathbf{T}) = \mathbf{T} \times \mathbf{B}$ of $\mathbf{T}$ defined by

$$\begin{aligned} F_{\mathbf{T}}(\phi_{\mathbf{T}}(\mu), \phi'_{\mathbf{T}}(\mu)\nu) = &\inf\{\|\nu_*(1-|\mu|^2)^{-1}\|_\infty : \\ &\phi'_{\mathbf{T}}(\mu)\nu_* = \phi'_{\mathbf{T}}(\mu)\nu;\ \mu \in \mathrm{Belt}(\Delta)_1;\ \nu, \nu_* \in L_\infty(\mathbb{C})\}. \end{aligned} \tag{12}$$

The space $\mathbf{T}$ admits also invariant metrics. The largest of those is the **Kobayashi metric** $d_{\mathbf{T}}$. It is contracted by holomorphic maps $h:\ \Delta \to \mathbf{T}$ so that for any two points $\psi_1,\ \psi_2 \in \mathbf{T}$, we have

$$d_{\mathbf{T}}(\psi_1, \psi_2) \le \inf\{d_\Delta(0,t):\ h(0) = \psi_1,\ h(t) = \psi_2\},$$

where d_Δ is the **hyperbolic Poincaré metric** on Δ of Gaussian curvature -4, with the differential form

$$ds = \lambda_{\mathrm{hyp}}(z)|dz| := \frac{|dz|}{1-|z|^2}. \tag{13}$$

The following key theorem on plurisubharmonicity is a strengthened version for universal Teichmüller space of the Gardiner-Royden theorem on the equality of Kobayashi and Teichmüller metrics on Teichmüller spaces (cf. [6], [7], [9], [32]).

Theorem 4.1. *The differential Kobayashi metric $\mathcal{K}_{\mathbf{T}}(\varphi, v)$ on the tangent bundle $\mathcal{T}(\mathbf{T})$ of the universal Teichmüller space $\mathbf{T}$ is logarithmically plurisubharmonic in $\varphi \in \mathbf{T}$, equals the canonical Finsler structure $F_{\mathbf{T}}(\varphi, v)$ on $\mathcal{T}(\mathbf{T})$ generating the Teichmüller metric of $\mathbf{T}$ and has constant holomorphic sectional curvature $\kappa_{\mathcal{K}}(\varphi, v) = -4$ on $\mathcal{T}(\mathbf{T})$.*

The proof of Theorem 2.1 essentially involves the technique of Grunsky coefficient inequalities. In fact, these inequalities give that all invariant metrics on $\mathbf{T}$ are equal, but this will not be used here.

Recall that the **generalized Gaussian curvature** κ_λ of a upper semicontinuous Finsler metric $ds = \lambda(t)|dt|$ in a domain $\Omega \subset \mathbb{C}$ is defined by

$$\kappa_\lambda(t) = -\frac{\mathbf{\Delta} \log \lambda(t)}{\lambda(t)^2}, \tag{14}$$

where $\mathbf{\Delta}$ is the **generalized Laplacian**

$$\mathbf{\Delta}\lambda(t) = 4 \liminf_{r\to 0} \frac{1}{r^2}\Big\{\frac{1}{2\pi}\int_0^{2\pi} \lambda(t + re^{i\theta})d\theta - \lambda(t)\Big\} \tag{15}$$

(provided that $-\infty \le \lambda(t) < \infty$), and the **sectional holomorphic curvature** of a Finsler metric on $\mathbf{T}$ is the supremum of curvatures (14) over appropriate collections of holomorphic maps from the disk into $\mathbf{T}$.

Similar to C^2 functions, for which $\mathbf{\Delta}$ coincides with the usual Laplacian, one obtains that λ is subharmonic on Ω if and only if $\mathbf{\Delta}\lambda(t) \ge 0$; hence, at the points t_0 of local maxima of λ with $\lambda(t_0) > -\infty$, we have $\mathbf{\Delta}\lambda(t_0) \le 0$ (cf., e.g., [5], [14]).

5. Main geometric theorems

The desired geometric theorem on Fredholm eigenvalues mentioned in the end of the previous section states:

Theorem 5.1. *The set of quasiconformal curves L, for which Ahlfors' inequality (2) is fulfilled in the strict form $1/\rho_L > q_L$, is open and dense in the strongest topology determined by the norm of the space* $\mathbf{B}$.

We derive this theorem from the following theorem which proves a much stronger version of Moser's conjecture.

Theorem 5.2. *The set of points $\varphi = S_f$, which represent the maps $f \in \Sigma^0$ with $\varkappa(f) < k(f)$, is open and dense in the space* $\mathbf{T}$.

The proof of Theorem 5.2 is given in [22]; it is complicate and relies on deep results concerning the Finsler geometry of the universal Teichmüller space.

A basic fact is that Grunsky coefficients $\alpha_{mn}(f)$ depend holomorphically on the Schwarzians S_f and hence generate the holomorphic maps

$$h_{\mathbf{x}}(\varphi) = \sum_{m,n=1}^{\infty} \sqrt{mn}\, \alpha_{mn}(\varphi) x_m x_n : \ \mathbf{T} \to \Delta \tag{16}$$

for each $\mathbf{x} = (x_n) \in S(l^2)$. Using these functions, we define a new complex Finsler metric on $\mathbf{T}$ of holomorphic curvature at most -4.

Let Ω be a holomorphic disk in $\mathbf{T}$; then $\Omega = \phi_{\mathbf{T}}(\Omega^0)$, where Ω^0 is a holomorphic disk in the ball $\mathrm{Belt}(\Delta)_1$. Let us consider in the tangent bundle $\mathcal{T}(\mathbf{T}) = \mathbf{T} \times \mathbf{B}$ the holomorphic disks $\widetilde{\Omega}$ covering Ω. Their points are pairs (φ, v), where $v = \phi_{\mathbf{T}}'[\varphi]\mu \in \mathbf{B}$ is a tangent vector to $\mathbf{T}$ at the point φ, and μ runs over the ball

$$\mathrm{Belt}(D_\varphi)_1 = \{\mu \in L_\infty(\mathbb{C}) : \ \mu|D_\varphi^* = 0, \ \|\mu\|_\infty < 1\}.$$

Here D_φ and D_φ^* denote the images of Δ and Δ^* under $f = f_\varphi \in \Sigma^0$ with $S_f = \varphi$.

To get the maps $\Delta \to \mathbf{T}$ preserving the origins, we transform the functions (16) by the chain rule for Beltrami coefficients $w^\nu = w^{\sigma(\nu)} \circ (f^{\nu_0})^{-1}$, where

$$\sigma(\nu) \circ f^{\nu_0} = \frac{\nu - \nu_0}{1 - \overline{\nu}_0 \nu} \frac{\partial_z f^{\nu_0}}{\overline{\partial_z f^{\nu_0}}}; \tag{17}$$

denote the composed maps by $g_{\mathbf{x}}[\sigma_\varphi]$. Using the form

$$H_{\mathbf{x}}(\varphi, \varphi_0) = \frac{h_{\mathbf{x}}(\varphi) - h_{\mathbf{x}}(\varphi_0)}{1 - \overline{h_{\mathbf{x}}(\varphi_0)} h_{\mathbf{x}}(\varphi)},$$

one defines on $\mathcal{T}(\mathbf{T})$ a new Finsler structure

$$F_\varkappa(\varphi_0, v) = \sup\{|dH_{\mathbf{x}}(\varphi_0; \varphi_0)v| : \mathbf{x} \in S(l^2)\}. \tag{18}$$

It is dominated by the canonical Finsler structure (12).

The structure (18) determines in a standard way on embedded holomorphic disks $\gamma(\Delta)$, where γ are injective holomorphic maps $\Delta \to \mathbf{T}$, the Finsler metrics $\lambda_\gamma(t) = F_\varkappa(\gamma(t), \gamma'(t))$ and then the corresponding distance

$$d_\gamma(\varphi_1, \varphi_2) = \inf \int_\beta \lambda_\gamma(t) ds_t.$$

The infimum over C^1 is taken over smooth curves $\beta : [0,1] \to \mathbf{T}$ joining the points φ_1 and φ_2.

This structure reproduces the Grunsky norm at least on the Teichmüller extremal disks:

Lemma 5.3. *On any extremal disk* $\Delta(\mu_0) = \{\phi_{\mathbf{T}}(t\mu_0) : \ t \in \Delta\}$ (*and on its isometric images in* $\mathbf{T}$), *we have the equality*

$$\tanh^{-1}[\varkappa(S_{f^{r\mu_0}})] = \int_0^r \lambda_\varkappa(t) dt. \tag{19}$$

Using this structure, we pull back via the maps (16) the hyperbolic metric (13) onto suitable holomorphic disks $\widetilde{\Omega}$ in the tangent bundle $\mathcal{T}(\mathbf{T})$ over $\mathbf{T}$ and get on these disks the conformal subharmonic metrics $ds = \lambda_{g_{\mathbf{x}}[\sigma_\varphi(\nu)]}(t)|dt|$ on G_0, with

$$\lambda_{g_{\mathbf{x}}[\sigma_\varphi(\nu)]} = g_{\mathbf{x}}[\sigma_\varphi(\nu)]^*(\lambda_{\mathrm{hyp}}) = \frac{|g'_{\mathbf{x}}[\sigma_\varphi(\nu)]|}{1 - |g_{\mathbf{x}}[\sigma_\varphi(\nu)]|^2},$$

whose curvature at nonsingular points is equal to -4.

The upper envelope of these metrics $\widehat{\lambda}_\varkappa = \sup \widehat{\lambda}_{g_{\mathbf{x}}[\sigma_\varphi]}$, where the supremum is taken over all $\mathbf{x} \in S(l^2)$ and all $\sigma_\varphi \in \mathrm{Belt}(\Delta)_1$, generates a subharmonic metric on these disks, which descends to a subharmonic metric $\lambda_\varkappa$ on the underlying holomorphic disks Ω in $\mathbf{T}$. The generalized Gaussian curvature of $\lambda_\varkappa$ satisfies $k_{\lambda_\varkappa} \le -4$, or equivalently, $\mathbf{\Delta} u_\varkappa \ge 4e^{2u_\varkappa}$.

We can compare this metric with the restrictions of the Kobayashi differential metric $\mathcal{K}$ to Ω by applying Minda's maximum principle for upper semicontinuous solutions of the differential inequality $\mathbf{\Delta} u \ge Ku$.

Lemma 5.4. [30] *If a function* $u : \ \Omega \to [-\infty, +\infty)$ *is upper semicontinuous in a domain* $\Omega \subset \mathbb{C}$ *and its generalized Laplacian satisfies the inequality* $\mathbf{\Delta} u(z) \ge Ku(z)$ *with some positive constant* K *at any point* $z \in \Omega$, *where* $u(z) > -\infty$, *and if* $\limsup_{z \to \zeta} u(z) \le 0$ *for all* $\zeta \in \partial\Omega$, *then either* $u(z) < 0$ *for all* $z \in \Omega$ *or else* $u(z) = 0$ *for all* $z \in \Omega$.

This lemma is applied to the function $u = \log(\lambda_\varkappa/\lambda_{\mathcal{K}})$ and implies that both metrics $\lambda_\varkappa$ and $\lambda_{\mathcal{K}}$ must coincide on Ω_a. Thereafter, one derives the assertion of Theorem 5.2

6. Inversion of Ahlfors inequality

6.1. An important question, which actually arose after the seminal Ahlfors paper [1] is whether the reflection coefficient of a quasicircle L can be estimated by its Fredholm eigenvalue ρ_L. An equivalent problem on the sharp estimation of the dilatation $k(f)$ by the Grunsky norm of f was stated by Kühnau in 1981.

He established a not sharp estimate of $k(f)$ by an explicit function of $\varkappa(f)$. His bound relates to known estimates of dilatations of quasiconformal extensions of k-quasisymmetric homeomorphisms of the real axes (see [29]).

The following theorem proved in [22] solves this problem completely and has many other applications.

Theorem 6.1. *For $f \in \Sigma^0$ we have the estimate*

$$k(f) \le \frac{3}{2\sqrt{2}} \, \frac{1}{\rho_{f(S^1)}} = \frac{1.07\ldots}{\rho_{f(S^1)}} \tag{20}$$

which is asymptotically sharp as $\rho \to \infty$. The equality holds for the map

$$f_{3,t}(z) = \begin{cases} z(1 + t/z^3)^{2/3} & \text{if } \ |z| > 1 \\ z[1 + t(|z|/z)^3]^{2/3} & \text{if } \ |z| \le 1 \end{cases} \tag{21}$$

with $t = \mathrm{const} \in (0, 1)$.

Note that the Beltrami coefficient of $f_{3,t}$ in Δ is $\mu_3(z) = t|z|/z$. This map was the first (though the simplest) example of functions with $\varkappa(f) < k(f)$ found by Kühnau.

We again can consider only $f \in \Sigma^0$ represented in $\mathbf{T}$ by Strebel points, i.e., with Teichmüller coefficients $\mu_f = k|\varphi|/\varphi$, where $k = \mathrm{const} \in (0, 1)$ and $\varphi \in A_1(\Delta)$. Put $\mu^*(z) = \mu(z)/\|\mu\|_\infty$.

We apply the following improvement of Theorem 3.1.

Theorem 6.2. *For every function $f \in \Sigma^0$ with unique extremal extension f^{μ_0} to Δ, we have the sharp bound*

$$k(f^{\mu_0}) \le \frac{1}{\alpha(f^{\mu_0})} \min_{|t|=1} \varkappa(f^{t\mu_0})$$

with

$$a(f^{\mu_0}) = \sup\{|\langle \mu_0^*, \varphi \rangle_\Delta| : \ \varphi \in A_1^2, \ \|\varphi\|_{A_1} = 1\}. \tag{22}$$

The proof of Theorem 6.2 is geometric and relies on properties of conformal metrics $ds = \lambda(z)|dz|$ on the disk Δ with $\lambda(z) \ge 0$ of **negative integral curvature** bounded from above. The curvature is understood in the classical supporting sense of Ahlfors or, more generally, in the potential sense introduced by Royden [33] in his investigation of the case of equality in the Ahlfors-Schwarz lemma (cf. [13]).

Namely, a **metric λ has curvature at most K in the potential sense** at a point z_0 if there is a disk U about z_0 in which the function

$$\log \lambda + K \, \mathrm{Pot}_U(\lambda^2),$$

where Pot_U denotes the logarithmic potential

$$\mathrm{Pot}_U\, h = \frac{1}{2\pi} \int\limits_U h(\zeta) \log|\zeta - z| d\xi d\eta \quad (\zeta = \xi + i\eta),$$

is subharmonic. One can replace U by any open subset $V \subset U$, because the function $\mathrm{Pot}_U(\lambda^2) - \mathrm{Pot}_V(\lambda^2)$ is harmonic on U. This is equivalent to condition that λ satisfies the inequality $\mathbf{\Delta} \log \lambda \geq K\lambda^2$ in the sense of distributions; here $\mathbf{\Delta} = 4\partial\overline{\partial}$. For such metrics, we have the following Royden's lemma.

Lemma 6.3. *If a circularly symmetric conformal metric $\lambda(|z|)|dz|$ in Δ has curvature at most -4 in the potential sense, then $\lambda(r) \geq a/(1 - a^2r^2)$, where $a = \lambda(0)$.*

We consider the extremal disk $\Delta(\mu_0^*) = \{\phi_{\mathbf{T}}(t\mu_0^*) : \ t \in \Delta\} \subset \mathbf{T}$, on which the differential Kobayashi metric $\lambda_{\mathcal{K}}$ coincides with hyperbolic metric (13), and construct the corresponding maps (16), getting similar to Theorem 5.2 a logarithmically subharmonic metric $\lambda_{\varkappa}(t)$ on Δ whose curvature in both supporting and in the potential senses is less than or equal -4. Its circular mean

$$\mathcal{M}[\lambda_{\varkappa}](|t|) = (2\pi)^{-1} \int_0^{2\pi} \lambda_{\varkappa}(re^{i\theta}) d\theta$$

is a circularly symmetric metric with curvature also at most -4 in the potential sense. The needed value of this mean at zero can be calculated by applying the standard quasiconformal variations. Together with Lemma 5.3, this implies Theorem 6.2.

To get (20), we have to estimate the quantity (22) from below. Applying Theorems 3.1 and 6.2, we can restrict ourselves by finding the minimal value of the functionals $l_\mu(\psi) = |\langle \mu^*, \varphi \rangle_\Delta|$ on the set $\{\varphi \in A_1^2 : \ \|\varphi\|_1 = 1\}$ for $\mu^* = |\psi|/\psi$ defined by integrable holomorphic functions in Δ of the form

$$\psi(z) = z^m(c_0 + c_1 z + \cdots), \quad m = 1, 3, 5, \dots \ .$$

The rather long calculations imply that this minimum equals $\frac{2\sqrt{2}}{3}$ and is attained on the map (21).

The equality in (20) is attained by the map (21) only asymptotically as $\rho \to \infty$ (see [22]).

6.2. The inequalities (2) and (20) result in

$$\frac{1}{\rho_L} \leq q_L \leq \frac{3}{2\sqrt{2}} \, \frac{1}{\rho_L} \tag{23}$$

(and in the equivalent inequalities for Grunsky norm).

Since the universal Teichmüller space is a homogeneous Banach domain, so that any two its points are connected by a map from the universal Teichmüller modular group Mod **1** preserving the invariant distances, one obtains from (23) the following interesting geometric estimates.

Theorem 6.4. *Every contractable separately plurisubharmonic metric* $r_{\mathbf{T}}(\varphi, \psi)$ *on* $\mathbf{T}$ *is deviated from the Kobayashi metric at most as follows*

$$r_{\mathbf{T}}(\varphi_1, \varphi_2) \le d_{\mathbf{T}}(\varphi_1, \varphi_2) \le \tanh\Big[\frac{3}{2\sqrt{2}} \tanh^{-1} r_{\mathbf{T}}(\varphi_1, \varphi_2)\Big].$$

7. Quasireflections and Finsler metrics

We mention here an application of geometric method exploited above to somewhat other kind of problems.

7.1. As is well known, a topological circle admitting quasiconformal reflections is a quasicircle and, according to [Ah2], it is characterized by uniform boundedness of the cross-ratios of ordered quadruples of its points. Moreover, it was established in [16] that any set $E \subset S^2$, which admits quasireflections, is necessarily located on a quasicircle with the same reflection coefficient.

For each mirror E, its **reflection coefficient** is defined as $q_E = \inf \|\partial_z f / \partial_{\bar{z}} f\|_\infty$ (where the infimum is over all quasireflections across E) and the quasiconformal dilatation is defined by

$$Q_E = (1 + q_E)/(1 - q_E) \ge 1.$$

Due to [2], [16], [27], we have

$$Q_E = (1 + k_E)^2/(1 - k_E)^2,$$

where $k_E = \inf \|\partial_{\bar{z}} f / \partial_z f\|_\infty$, taking the infimum over all quasicircles $L \supset E$ and all orientation preserving quasiconformal homeomorphisms $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ with $f(\widehat{\mathbb{R}}) = L$.

One of the important problems in this theory and its applications is developing general algorithms for exactly, or at least approximately, calculating of the reflection coefficients of particular curves and their subarcs.

One of the standard ways of determining the reflection coefficients q_L and the Fredholm eigenvalues ρ_L for a given curve L consists of verifying whether we have for this curve the case of equality in (7). This is unknown even for the rectilinear quadrilaterals.

7.2. Already more than 20 years ago, Kühnau stated a question whether for any rectangle $\mathcal{P}$, the equality $q_{\mathcal{P}} = 1/\rho_{\mathcal{P}}$ holds, which was completely answered only recently. It was established in [27], [41] (by explicit constructing the extremal reflections) that the answer is affirmative for the rectangles $\mathcal{R}$ sufficiently close to the square so that their moduli $m(\mathcal{R})$ satisfy $1 \le m(\mathcal{R}) < 1.037$; for such rectangles, we have $q_L = 1/\rho_L = 1/2$.

The above geometric approach provides also a complete answer to this question as well as to certain other related problems. Namely, *the equalities*

$$q_{\partial\mathcal{P}} = 1/\rho_{\partial\mathcal{P}} = \varkappa(f^*) = k(f^*) \tag{24}$$

hold for every rectangle $\mathcal{P}$; here f^* denotes the Riemann mapping function $\Delta^* \to \mathcal{P}^* = \widehat{\mathbb{C}} \setminus \overline{\mathcal{P}}$.

In this case, the elliptic integral

$$f_t(\zeta) = C_t \int_0^\zeta \frac{d\xi}{\sqrt{(1-\xi^2)(1-t^2\xi^2)}} = \mathbf{F}(\zeta, t),$$

defines a conformal map of the upper half-plane $U = \{z : \ \operatorname{Im} z > 0\}$ onto the rectangle $\mathcal{P}$ so that the points points ± 1 and $\pm t$ $(0 < t < 1)$ go to the vertices $(\pm a, 0), (\pm a, \pm a + ib)$ of $\mathcal{P}$. The corresponding map $\mathbf{b}(t) = S_{f_t}(\cdot, t) : \ (-1, 1) \to \mathbf{T}$ defines a real curve Γ in $\mathbf{T}$ which represents all rectangles $\mathcal{P}$ (up to linear equivalence). It is located on a non-geodesic holomorphic disk (planar domain) Ω_0 in a tubular neighborhood V_Γ of Γ in $\mathbf{T}$.

Now, composing the solutions to the Schwarz equation $S_f = \varphi$, $\varphi \in \mathbf{T}$, with a suitable Möbius map $\Delta^* \to U$, one obtains the functions $F \in \Sigma^0$, for which we have by (16) the corresponding holomorphic maps $h_{\mathbf{x}}(\varphi) : \ \mathbf{T} \to \Delta$. This implies, similar to the Theorem 5.2, a subharmonic Finsler metric $\lambda_\varkappa$ on the disk Ω_0 of generalized Gaussian curvature $k_{\lambda_\varkappa} \le -4$, which is the restriction of a corresponding Finsler metric determined on the whole neighborhood V_Γ.

The equality $q_{\partial\mathcal{P}} = 1/\rho_{\partial\mathcal{P}}$, already known for rectangles $\mathcal{P}$ close to the square, allows us to conclude that this metric coincides at the corresponding points $\varphi \in \Omega_0$ with the dominant differential Teichmüller-Kobayashi metric $\lambda_\mathcal{K}$ on $\mathbf{T}$. Then, by Lemma 5.4, $\lambda_\varkappa$ and $\lambda_\mathcal{K}$ are equal on the whole disk Ω_0, which implies the equalities (24). For details see [21].

8. Variational problems for Fredholm eigenvalues

8.1. In this section, we consider certain basic important variational problems concerning the Grunsky eigenvalues and quasiconformal maps.

Some extremal problems for Fredholm eigenvalues have been investigated in pioneering papers of Schiffer (see, e.g., [34]), and applied to obtaining the existence of conformal maps onto canonical domains.

Note that the smallest positive Fredholm eigenvalue ρ_L regarded as a curve functional is upper semicontinuous with respect to uniform convergence on curves. The discontinuity causes the difficulties in solving the extremal problems for Fredholm eigenvalues of arbitrary Jordan curves. Another source of difficulties is provided by non-compactness of many of the intrinsically involved collections of maps.

The situation is different when the maps are convergent in the Teichmüller distance, because the function $\varkappa(f)$ is continuous on $\mathbf{T}$ (cf. [37]).

From this point of view, it is natural to consider the extremal problems on the classes of conformal maps f with Fredholm eigenvalues $\rho_{f(S^1)}$ bounded from below.

Such problems have not been fully investigated. One of the points is that there are no variational methods for these classes. We will apply the complex geometry of universal Teichmüller space.

8.2. Consider the subclass $\Sigma\langle k\rangle$ of Σ which consists of $f \in \Sigma$ having k-quasiconformal extensions to $\widehat{\mathbb{C}}$ and with the eigenvalues $\rho_{f(S^1)} \le 1/k$ (or equivalently, of f with the Grunsky norm $\varkappa(f) \le k$). Theorem 5.2 implies that this class is much wider than $\Sigma(k)$.

Let $F(f)$ be a holomorphic functional on the class Σ, which means that F is continuous and Gateaux $\mathbb{C}$-differentiable, i.e., we have for any $f \in \Sigma$ and small $t \in \mathbb{C}$ the equality

$$F(f + th) = F(f) + tF'_f(h) + O(t^2), \quad t \to 0, \tag{25}$$

in the topology of uniform convergence on compact sets in Δ^*. Here $F'_f(h)$ is a $\mathbb{C}$-linear functional. The restriction of F to Σ^0 can be lifted to the ball $\mathrm{Belt}(\Delta)_1$, using the equality $\widehat{F}(\mu) = F(f^\mu)$. We require that this lifting is holomorphic on $\mathrm{Belt}(\Delta)_1$. This yields that the functional $F'_f(h)$ in (25) is a strong (Fréchet) derivative.

We shall assume that the derivative on the identity map $\mathrm{id}(z) = z$,

$$\varphi_0(z) = F'_{\mathrm{id}}(g(\mathrm{id}, z)) \tag{26}$$

is a meromorphic functions on $\mathbb{C}$, which is holomorphic and integrable on the unit disk Δ.

This rather natural assumption holds, for example, for the general distortion functionals F of the form

$$F(f) = F(f(z_1), f'(z_1), \dots, f^{(m_1)}(z_1); \dots; f(z_p), f'(z_p), \dots, f^{(m_p)}(z_1)),$$

where $z_1, \dots, z_p$ are the distinct fixed points in $\overline{\Delta^*}$ with assigned orders $m_1, \dots, m_p$, respectively. In this case, the function (26) is rational.

The following general theorem solves the maximization problem for any functional F on $\Sigma\langle k\rangle$ of the above form, under an appropriate restriction to dilatation k.

Theorem 8.1. *Suppose that the range domain of F on Σ^0 has more than two boundary points. If a function φ_0 defined by* (26) *has in Δ only zeros of even order, then there exists a number $k_0(F) > 0$ such that for all $k \le k_0(F)$, the extremal functions for F on the class $\Sigma\langle k\rangle$ are the same as in the smaller class Σ_k; in other words,*

$$\max_{\varkappa(f)\le k} |F(f) - F(\mathrm{id})| = \max_{k(f)\le k} |F(f) - F(\mathrm{id})| = \max_{|t|=k} |F(f^{t|\varphi_0|/\varphi_0}) - F(\mathrm{id})|. \tag{27}$$

A similar result has been established earlier by the author for the classes $\Sigma(k)$ using the equality of Carathéodory and Kobayashi metrics on the disk

$$\Delta(\varphi_0) = \{\phi_{\mathbf{T}}(t|\varphi_0|/\varphi_0) : t \in \Delta\} \subset \mathbf{T}$$

following from Theorem 3.1 (see, e.g., [19]). The basic ingredient of the proof is the existence of a projector with norm one, which is ensured by the mentioned above equality of Carathéodory and Kobayashi metrics. Also the standard variational formula for the maps $f^\mu \in \Sigma(k)$,

$$f^\mu(z) = z + \frac{1}{2\pi i} \iint\limits_{\Delta} \frac{\mu(\zeta) d\zeta \wedge d\overline{\zeta}}{\zeta - z} + O(\|\mu\|^2), \quad \|\mu\| \to 0,$$

where the ratio $O(\|\mu\|^2)/\|\mu\|^2$ is uniformly bounded on compact sets of $\mathbb{C}$, was essentially used.

This method does not work for the maps of class $\Sigma\langle k\rangle$, where no variational formulas are known.

Theorem 8.1 sounds rather surprisingly and shows that the underlying features are generated by hyperbolicity, like in the classes $\Sigma(k)$. We derive this theorem as a consequence of the following result which is of independent interest.

Theorem 8.2. *If a holomorphic functional F on Σ^0 with $F(\mathrm{id}) = 0$ satisfies*

$$\sup_{\|\mu\|_\infty < 1} |F(f^\mu)| = 1,$$

and if its derivative $F'_{\mathrm{id}}(g(\mathrm{id}, z)) = \varphi_0$ has in Δ only zeros of even order, then, for any $k \in [0,1]$ and any t with $|t| = k$, we have

$$\max_{\varkappa(f) \le k} |F(f)| = |F(f^{t|\varphi_0|/\varphi_0})| = k. \tag{28}$$

Theorem 3.1 plays a crucial role in the proof of Theorems 8.1 and 8.2. We again use the density of Strebel points in $\mathbf{T}$ and establish that the maps $f^{t|\psi|/\psi}$ with $\psi \in A_1(\Delta)$ cannot be extremals F unless $\psi = \varphi_0$.

8.3. For the bounded functionals $F: \Sigma \to \mathbb{C}$ with $F(\mathrm{id}) = 0$, there is a useful lower estimate for $k_0(F)$, which allows us to apply this theorem effectively. Namely (cf. [19]),

$$k_0(F) \ge a \frac{\|F'_{\mathrm{id}}\|}{\|F'_{\mathrm{id}}\| + M(F) + 1} =: k'_0(a), \tag{29}$$

where a is any number satisfying $0 \le a \le 1/2, M(F) := \sup_\Sigma |F(f)| < \infty$ and

$$\|F'_{\mathrm{id}}\| = \frac{1}{\pi} \iint_\Delta |F'_{\mathrm{id}}(g(\mathrm{id}, z))| dx dy. \tag{30}$$

8.4. Specifying the functional (25), one gets the new sharp distortion estimates. For example, taking $F(f) = b_{2n-1},\ n \ge 1$, one obtains

Theorem 8.3. *For each $f(z) = z + \sum\limits_0^\infty b_n z^{-n} \in \Sigma\langle k\rangle$, we have the sharp bound*

$$\max_{\varkappa(f) \le k} |b_{2n-1}| \le \frac{k}{n}, \quad n = 1, 2, \dots, \tag{31}$$

provided that

$$k < k_n = \frac{1}{n + \frac{n}{\sqrt{2n-1}} + 1}. \tag{32}$$

The equality in (30) *is attained on the maps*

$$f_{n,t}(z) = \begin{cases} z\left(1 + \frac{t}{z^{n+1}}\right)^{2/(n+1)}, & |z| \le 1, \\ z\left[1 + t\left(\frac{\overline{z}}{z}\right)^{(n+1)/2}\right]^{2/(n+1)}, & |z| > 1, \end{cases}$$

with Beltrami coefficients

$$\mu_n(z) = t(\overline{z}/z)^{n-1}, \quad |t| = k,$$

as well as for their admissible translations $h_{n,t}(z) = f_{n,t}(z) - a$ *with* $a \in \mathbb{C}$, *i.e., such that* $h_{n,t}$ *also remains in the class* $\Sigma\langle k\rangle$.

The bound (32) follows from the known corresponding best estimate for the maps of class $\Sigma(k)$.

8.5. The above arguments work well also for the classes $S\langle k\rangle$ of univalent functions $f(z) = z + \sum_2^\infty a_n z^n$ in the unit disk, which have quasiconformal extensions to $\widehat{\mathbb{C}}$ satisfying $f(\infty) = \infty$ and such that their Grunsky norms satisfy $\varkappa(f) \le k$.

In this case, one obtains, by applying the bound (28), that for every function $f \in S\langle k\rangle$ its **odd** coefficients a_{2n-1} satisfy the inequality

$$|a_{2n-1}| \le \frac{k}{n-1}, \quad n = 2, 3, \dots ,$$

provided that

$$\varkappa(f) < k_n = 1/[(2n-1)^2 + 1].$$

In contrast to classes $\Sigma\langle k\rangle$, an additional normalization condition for the functions of $f \in S\langle k\rangle$ (for example, at infinity) essentially reflects upon the growth order of the coefficients.

Such classes of univalent functions in the disk without an additional normalization were investigated by Grinshpan and Pommerenke, also by applying the Grunsky inequalities (see [10], [11]); they established the exact growth order of coefficients a_n in n.

8.6. The following question bridges the maximization problem studied above with the problems for Fredholm eigenvalues considered by Schiffer.

Let $\Sigma(b_1^0, b_2^0, \dots , b_m^0)$ denote the collection of functions $f \in \Sigma$ with the same first m coefficients $b_1^0, b_2^0, \dots , b_m^0$, which are fixed. Find the Jordan curves $L = f(S^1)$ with maximal Fredholm eigenvalue ρ_L.

In other words, one has to find the curves whose eigenvalue is the closest to the eigenvalue of the circle.

The case $m = 1$ is rather simple; then $1/\max \rho_L = |b_1^0|$, with equality only for the maps

$$f_{1,t}(z) = \begin{cases} z + t/z, & |z| \ge 1, \\ z + t\overline{z}, & |z| < 1, \end{cases}$$

with $|t| = |b_1^0|$. These maps are extremal for many problems.

A partial answer for $m > 1$ follows from Theorem 8.1; in the general case, the problem is open.

9. Computation of Fredholm eigenvalues

9.1. Another approach to estimating the reflection coefficients and Fredholm eigenvalues of curves relies on the properties of holomorphic motions, i.e., of holomorphic isotopies depending on complex parameter (see, e.g., [4], [38]). We illustrate it on the following theorem proved in [17].

Theorem 9.1. *Let a function f map conformally the upper half-plane U into $\mathbb{C}$, and let the equation*

$$w''(\zeta) = tb_f(\zeta)w'(\zeta), \quad \zeta \in U, \tag{33}$$

have univalent solutions on U for all $t \in [0, t_0]$, $t_0 > 1$. Then the image $f(U)$ is a quasidisc, and the reflection coefficient of its boundary $L = f(\widehat{\mathbb{R}}) = \partial f(U)$ satisfies

$$q_L \leq \frac{1}{t_0}. \tag{34}$$

This bound cannot be improved in the general case. The equality in (34) is attained by any quasicircle which contains two $C^{1+\epsilon}$ smooth subarcs ($\epsilon > 0$) with the interior intersection angle $\alpha\pi$, where $\alpha = 1 - 1/t_0$ (under above univalence assumption for logarithmic derivative $\mathbf{b}_f = f''/f'$). In this case,

$$q_L = \frac{1}{\rho_L} = \frac{1}{t_0}. \tag{35}$$

The exact bound for the reflection coefficient q_L follows from (34) by choosing the maximal value of t_0 admitting the indicated univalence property for all $t \in [0, t_0]$. The corresponding solution w_{t_0} of (33) for this value is also univalent on U, but the domain $w_{t_0}(U)$ is not a quasidisc.

To prove the theorem, we use the conformal maps σ_t of the disks $\Delta_t = \{|z| < t\}$, $t < t_0$ onto U and construct the desired holomorphic motions $w(z, t) : \Delta \times \Delta_t \to \widehat{\mathbb{C}}$ as the ratios u_2/u_1 of normalized independent solutions of the equation

$$u'' = tb_{f\circ\sigma^{-1}}u', \quad 0 \leq t < t_0.$$

The well-known properties of holomorphic motions provide the estimate (34). To examine the case of equality, one must combine (34) with the angle inequality of Kühnau [27] which asserts that if a closed curve L contains two analytic arcs with the interior intersection angle $\pi\alpha$, then its reflection coefficient satisfies $q_L \geq |1-\alpha|$ and use approximation.

Theorem 9.1 concerns the Ahlfors conjecture that conformal maps of U onto the interior of quasicircles are analytically characterized by their logarithmic derivatives (see [2]). It was investigated by many authors.

9.2. The following theorem shows that the bound (35) given by Theorem 9.1 is attained on the maps onto unbounded convex or concave domains and on their fractional linear images.

Theorem 9.2. *For every unbounded convex domain $D \subset \mathbb{C}$ with piecewise $C^{1+\epsilon}$-smooth boundary L ($\epsilon > 0$), the equalities*

$$q_L = 1/\rho_L = \varkappa(g) = \varkappa(g^*) = k_0(g) = k_0(g^*) = 1 - |\alpha| \tag{36}$$

hold, where g and g^* denote the appropriately normalized conformal maps $\Delta \to D$ and $\Delta^* \to \widehat{\mathbb{C}} \setminus \overline{D}$, respectively; $k_0(g)$ and $k_0(g^*)$ are the minimal dilatations of their quasiconformal extensions to $\widehat{\mathbb{C}}$, and $\pi|\alpha|$ is the opening of the least interior angle between the boundary arcs $L_j \subset L$. Here $0 < \alpha < 1$ if the corresponding vertex is finite and $-1 < \alpha < 0$ for the angle at the vertex at infinity.

The same is true for the unbounded concave domains which do not contain ∞; for those one must replace the last term by $|\beta| - 1$, where $\pi|\beta|$ is the opening of the largest interior angle of D.

The univalence of solutions to the equation (33) for $0 \le t \le 1/(1 - |\alpha|)$ is established for such domains approximating $f : U \to D$ by conformal maps f_n of the half-plane onto the rectilinear polygons P_n which are chosen to be also unbounded and convex or concave simultaneously with the original domain D. These maps f_n are represented by the Schwarz-Christoffel integral.

The main point is that if the least interior angle of P_n equals $|\alpha_n|$ (taking the negative sign for the angle at infinity), then for any $t \in [0, 1/(1 - |\alpha_n|)]$ the corresponding fiber map $w_t(z)$ is again a conformal map of the half-plane onto a well-defined rectilinear polygon. After the needed approximation, the basic equalities $q_{\partial P_n} = 1/\rho_{\partial P_n} = 1 - |\alpha_n|$ follow from Theorem 9.1, while the remaining equalities in (36) are obtained by combining this theorem with the Kühnau-Schiffer theorem mentioned above.

9.3. Theorems 9.1 and 9.2 have various important consequences. For example, for any closed unbounded curve L with the convex interior, which is $C^{1+\epsilon}$-smooth at all finite points and has at infinity the asymptotes approaching the interior angle $\pi\alpha < 0$, we have

$$q_L = 1/\rho_L = 1 - |\alpha|. \tag{37}$$

More generally, let $L = \gamma_1 \cup \gamma_2 \cup \gamma_3$, where

$$\gamma_1 = [a_1, \infty],\ \gamma_2 = e^{i\pi\alpha}[a_2, \infty],\ a_1 \ge a_2 > 0,\ 0 < \alpha \le 1/2,$$

and

$$\gamma_3 = \{(x, y) :\ y = h(x),\ a_2 \cos \pi\alpha \le x \le a_1\}$$

with a decreasing convex piecewise $C^{1+\epsilon}$-smooth function h such that

$$h(a_2 \cos \pi\alpha) = a_2 \sin \pi|\alpha|,\ h(a_1) = 0.$$

Equalities (37) hold for any such curve.

The above geometric assumptions on the domains are essential. In particular, the assertion of Theorem 9.2 extends neither to the arbitrary unbounded nonconvex and nonconcave domains nor to the arbitrary bounded convex domains, without some additional assumptions. This shows also that univalence of solutions w_t for all $t \in [0, t_0]$ in Theorem 9.1 is essential.

For a few special curves, similar equalities were established in [26], [27], [41]) by applying geometric constructions giving explicitly the extremal quasireflections.

9.4. Note that Theorem 9.1 closely relates to the structure of holomorphic embeddings of the universal Teichmüller space. The space $\mathbf{T}$ is not starlike with respect to any of its points, and there exist points $\varphi \in \mathbf{T}$ for which the line interval $I_\varphi = \{t\varphi : 0 < t < 1\} \subset \mathbf{B}$ contains the points v_t for which the corresponding functions f with $S_f = \varphi_t$ are only locally univalent on Δ^* (see, e.g., [19]).

The logarithmic derivatives $\beta_f = f''/f'$ determine the Becker embedding of $\mathbf{T}$ as a domain $\mathbf{T}_1$ in the Banach space $\mathbf{B}_1$ of holomorphic functions ψ on Δ^* with the norm $\|\psi\| = \sup_{\Delta^*}(|z|^2 - 1)|z\psi(z)|$; see [3]. This space also is not starlike with respect to its origin, and therefore the general intervals in $\mathbf{T}_1$ are not connected.

References

[1] L.V. Ahlfors, *Remarks on the Neumann-Poincaré integral equation*, Pacific J. Math. **2** (1952), 271–280.

[2] L.V. Ahlfors, *Lectures on Quasiconformal Mappings*, Van Nostrand, Princeton, 1966.

[3] J. Becker, *Conformal mappings with quasiconformal extensions*, Aspects of Contemporary Complex Analysis, Proc. Confer. Durham 1979 (D.A. Brannan and J.G. Clunie, eds.), Academic Press, New York, 1980, pp. 37–77.

[4] E.M. Chirka, *On the extension of holomorphic motions*, Doklady Mathematics **70** (2004), 516–519.

[5] S. Dineen, *The Schwarz Lemma*, Clarendon Press, Oxford, 1989.

[6] C.J. Earle, I. Kra and S.L. Krushkal, *Holomorphic motions and Teichmüller spaces*, Trans. Amer. Math. Soc. **944** (1994), 927–948.

[7] C.J. Earle and S. Mitra, *Variation of moduli under holomorphic motions*, In the tradition of Ahlfors and Bers (Stony Brook, NY, 1998), Contemp. Math. **256**, Amer. Math. Soc., Providence, RI, 2000, pp. 39–67.

[8] D. Gaier, *Konstruktive Methoden der konformen Abbildung*, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1964.

[9] F.P. Gardiner and N. Lakic, *Quasiconformal Teichmüller Theory*, Amer. Math. Soc., 2000.

[10] A.Z. Grinshpan, *Univalent functions and regularly measurable mappings*, Siberian Math. J. **27** (1986), 825–837.

[11] A.Z. Grinshpan and Ch. Pommerenke, *The Grunsky norm and some coefficient estimates for bounded functions*, Bull. London Math. Soc. **29** (1997), 705–712.

[12] H. Grunsky, *Koeffizientenbedingungen für schlicht abbildende meromorphe Funktionen*, Math. Z. **45** (1939), 29–61.

[13] M. Heins, *A class of conformal metrics*, Nagoya Math. J. **21** (1962), 1–60.

[14] S. Kobayashi, *Hyperbolic Complex Spaces*, Springer, New York, 1998.

[15] S.L. Krushkal, *Grunsky coefficient inequalities, Carathéodory metric and extremal quasiconformal mappings*, Comment. Math. Helv. **64** (1989), 650–660.

[16] S.L. Krushkal, *Quasiconformal reflections across arbitrary planar sets*, Sci. Ser. A Math. Sci. **8** (2002), 57–62.

[17] S.L. Krushkal, *A bound for reflections across Jordan curves*, Georgian Math. J. **10** (2003), 561–572.

[18] S.L. Krushkal, *Plurisubharmonic features of the Teichmüller metric*, Publications de l'Institut Mathématique-Beograd, Nouvelle série **75(89)** (2004), 119–138.

[19] S.L. Krushkal, *Univalent holomorphic functions with quasiconformal extensions* (*variational approach*), Ch. 5 in: Handbook of Complex Analysis: Geometric Function Theory, Vol. II (R. Kühnau, ed.), Elsevier Science, Amsterdam, 2005, pp. 165–241.

[20] S.L. Krushkal, *Quasiconformal extensions and reflections*, Ch. 11 in: Handbook of Complex Analysis: Geometric Function Theory, Vol. II (R. Kühnau, ed.), Elsevier Science, Amsterdam, 2005, pp. 507–553.

[21] S.L. Krushkal, *Complex Finsler metrics and quasiconformal characteristics of maps and curves*, Revue Roumaine Pure et Appl. **51** (2006), 665–682.

[22] S.L. Krushkal, *Strengthened Moser's conjecture, geometry of Grunsky coefficients and Fredholm eigenvalues*, Central European J. Math **5(3)** (2007), 551–580.

[23] S.L. Krushkal and R. Kühnau, *Quasikonforme Abbildungen – neue Methoden und Anwendungen*, Teubner-Texte zur Math., Bd. 54, Teubner, Leipzig, 1983.

[24] S.L. Krushkal and R. Kühnau, *Grunsky inequalities and quasiconformal extension*, Israel J. Math. **152** (2006), 49–59.

[25] R. Kühnau, *Verzerrungssätze und Koeffizientenbedingungen vom Grunskyschen Typ für quasikonforme Abbildungen*, Math. Nachr. **48** (1971), 77–105.

[26] R. Kühnau, *Zur Berechnung der Fredholmschen Eigenwerte ebener Kurven*, ZAMM **66** (1986), 193–201.

[27] R. Kühnau, *Möglichst konforme Spiegelung an einer Jordankurve*, Jber. Deutsch. Math. Verein. **90** (1988), 90–109.

[28] R. Kühnau, *Wann sind die Grunskyschen Koeffizientenbedingungen hinreichend für Q-quasikonforme Fortsetzbarkeit*? Comment. Math. Helv. **61** (1986), 290–307.

[29] R. Kühnau, *Über die Grunskyschen Koeffizientenbedingungen*, Ann. Univ. Mariae Curie-Sklodowska, sect. A **54** (2000), 53–60.

[30] D. Minda, *The strong form of Ahlfors' lemma*, Rocky Mountain J. Math., **17** (1987), 457–461.

[31] Chr. Pommerenke, *Univalent Functions*, Vandenhoeck & Ruprecht, Göttingen, 1975.

[32] H.L. Royden, *Automorphisms and isometries of Teichmüller space*, Advances in the Theory of Riemann Surfaces (Ann. of Math. Stud., vol. 66), Princeton Univ. Press, Princeton, 1971, pp. 369–383.

[33] H.L. Royden, *The Ahlfors-Schwarz lemma: the case of equality*, J. Anal. Math. **46** (1986), 261–270.

[34] M. Schiffer, *Extremum problems and variational methods in conformal mapping*, 1958 Internat. Congress Math., Cambridge Univ. Press, New York, 1958, pp. 211–231.

[35] M. Schiffer, *Fredholm eigenvalues and Grunsky matrices*, Ann. Polon. Math. **39** (1981), 149–164.

[36] G. Schober, *Estimates for Fredholm eigenvalues based on quasiconformal mapping*, Numerische, insbesondere approximationstheoretische Behandlung von Funktionsgleichungen. Lecture Notes in Math. **333**, Springer-Verlag, Berlin, 1973, pp. 211–217.

[37] Y.L. Shen, *Pull-back operators by quasisymmetric functions and invariant metrics on Teichmüller spaces*, Complex Variables **42** (2000), 289–307.

[38] Z. Slodkowski, *Holomorphic motions and polynomial hulls*, Proc. Amer. Math. Soc. **111** (1991), 347–355.

[39] K. Strebel, *On the existence of extremal Teichmueller mappings*, J. Anal. Math **30** (1976), 464–480.

[40] S.E. Warschawski, *On the effective determination of conformal maps*, Contribution to the Theory of Riemann surfaces (L. Ahlfors, E. Calabi et al., eds.), Princeton Univ. Press, Princeton, 1953, pp. 177–188.

[41] S. Werner, *Spiegelungskoeffizient und Fredholmscher Eigenwert für gewisse Polygone*, Ann. Acad.Sci. Fenn. Ser.AI. Math. **22** (1997), 165–186.

[42] I.V. Zhuravlev, *Univalent functions and Teichmüller spaces*, Inst. of Mathematics, Novosibirsk, preprint, 1979, 23 pp. (Russian).

Samuel Krushkal
Department of Mathematics, Bar-Ilan University,
52900 Ramat-Gan, Israel

and

Department of Mathematical Sciences
University of Memphis
Memphis, TN 38152, USA

Analysis and Mathematical Physics
Trends in Mathematics, 369–376
© 2009 Birkhäuser Verlag Basel/Switzerland

A Note on Life-span of Classical Solutions to the Hele–Shaw Problem

Alexander Kuznetsov

Abstract. The Hele–Shaw flow with a source and with a simply connected initial domain is considered. It is shown that, if the solution exists for a sufficiently long time, then it is close to the identical map, and hence, it is starlike and exists infinitely long time. An estimate for this time is given.

Mathematics Subject Classification (2000). Primary 30C45, 76D27, 76S05; Secondary 35Q35, 30C35.

Keywords. Free boundary problem, Hele–Shaw flow, univalent function, Loewner-Kufarev equation.

1. Introduction

Hele–Shaw flow is a difficult problem of mathematical physics. A powerful tool for the description of this process can be provided by the theory of univalent functions.

Let $D(t)$ be a simply connected domain in the complex plane, occupied with a viscous fluid. Without loss of generality, we can assume, that the source is situated at the origin and it is of unit strength. We define a parametric function $f(z,t), f(0,t) = 0, f'(0,t) > 0$, which maps the unit disk $\mathbb{D} = \{z : |z| < 1\}$ conformally onto the domain $D(t)$, and which satisfies (see, e.g., [1, 6]) the following equation

$$\dot{f}(z,t) = zf'(z,t)\frac{1}{2\pi}\int_0^{2\pi} \frac{1}{|f'(e^{i\theta},t)|^2}\frac{e^{i\theta}+z}{e^{i\theta}-z}\,d\theta, \tag{1.1}$$

in the case of negligible surface tension. In (1.1) dot and prime denote derivatives with respect to t and z, respectively. A nice introduction to Hele–Shaw flows is given in [7].

It is known that some geometric properties of the function $f(z,t)$ are preserved as long as the solution to (1.1) exists, in particulars, starlikeness with respect to the origin. Based on this fact, it was shown [2], that if the initial domain is starlike and has an analytic boundary, then the solution to the equation (1.1) exists infinitely long.

On the other hand, if solution exists infinitely long, then the normalized domain $\frac{1}{\sqrt{2t}}f(\mathbb{D},t)$ tends to the unit disk in the Hausdorff metric [3]. Thus, it is reasonable to assume that, if the solution exists long enough, then it is close to the identical map at some moment, and hence, is starlike and exists infinitely long. This paper is devoted to the proof of this fact.

Let

$$M = \max_{|z|\le 1}|f'(z,0)|^2 = ||f'(z,0)||_1^2,$$
$$||f(z)||_r = \max_{|z|\le r}|f(z)|,$$

and S be the area of the domain $D(0)$. Let

$$D = \sqrt{M} - \sqrt{S/\pi},$$
$$d = \sqrt{\pi M - S},$$

and

$$T = \max\Big\{2D + 4d, \frac{(D+2d)(3D+2d+2\sqrt{S/\pi}) - S/\pi}{D+\sqrt{S/\pi}}\Big\} + \frac{5}{2}D.$$

Theorem 1.1. *If the solution to (1.1) exists for all*

$$t \in \Big[0, \max\Big\{\Big(\frac{2}{9}\frac{2d+D+(\frac{2}{9}-x)T}{x}\Big)^2 - S/\pi, \frac{(3D+2d+2\sqrt{S/\pi})^2 - S/\pi}{2}\Big\}\Big],$$

then it exists infinitely long.

(The number $x = 0.0010686995709770337\ldots$ will be defined below as a solution to a transcendental equation.)

2. Estimate for internal and external radii

The fact that the normalized domain tends to the unit disk was proved in [4], and it is based on the following inequality $R_e(t) = \max_{|z|=1}|f(z,t)|$, and $R_i(t) = \min_{|z|=1}|f(z,t)|$, where the ring $r_t < |z-\xi_t| < R_t$ is the smallest ring containing $\partial\Omega(t)$. First we need the following inequality (see [4] p. 57 or [5] p. 22 and p. 19)

$$R_e(t) \le \sqrt{2t+M}.$$

From the inequality

$$\sqrt{2t+M} - \sqrt{2t+S/\pi} \le D, \tag{2.1}$$

we obtain

$$R_e(t) \le \sqrt{2t+S/\pi} + D. \tag{2.2}$$

Also we need an estimate for $R_i(t)$, that can be obtain as a consequence of the following inequalities (see paper [4] p. 78 or [5] p. 30):

$$R_t \ge \sqrt{2t+S/\pi}, \qquad r_t \ge R_t - \sqrt{\pi M - S}.$$

Therefore,

$$r_t \geq \sqrt{2t + S/\pi} - \sqrt{\pi M - S}.$$

For $|\xi_t|$, we have

$$|\xi_t| \leq \sqrt{2t + M} - \sqrt{2t + S/\pi} + \sqrt{\pi M - S}.$$

Using (2.1) we obtain $|\xi_t| \leq D + d$. Taking into account, that $R_i \geq r_t - |\xi_t|$, we have

$$R_i(t) \geq \sqrt{2t + S/\pi} - 2d - D. \tag{2.3}$$

3. Extension of solution

Proposition 3.1. *If the solution $f(z,t)$ to the equation* (1.1) *exists for t satisfying the inequality*

$$\sqrt{2t + S/\pi} \geq 3D + 2d + 2\sqrt{S/\pi}, \tag{3.1}$$

then the function $f(z,t)$ has an analytic extension to the disk $\mathbb{D}_2 = \{z : |z| \leq 2\}$ and satisfies the following inequality

$$||f(z,t)||_2 \leq \frac{9}{2}\sqrt{2t + S/\pi} + T.$$

Proof. Let $r(t) = \max_{z \in \partial\Omega(0)} |f^{-1}(z,t)|$. The function $f(z,t)$ (see, e.g., [2]) can be analytically extended to the disk $|z| < 1/r$ by the following equality:

$$f(1/\overline{\xi}, t) = \overline{S(f(\xi,t),t)}.$$

Where

$$S(z,t) = \overline{z} - \chi_{\Omega(t)}(z) + \chi_{\Omega(0)}(z) + \frac{2t}{z} \tag{3.2}$$

and

$$\chi_\Omega(z) = -\frac{1}{\pi} \iint\limits_{\Omega} \frac{d\sigma_w}{z - w},$$

where σ_w stands for the area measure in the w-plane.

By the Schwartz lemma $r(t) \leq \frac{R_e(0)}{R_i(t)}$. Therefore from (2.2) and (2.3) we obtain, that inequality (3.1) means $R_i(t) > 2R_e(0)$. Thus, the function $f(z,t)$ is analytic in the disk $|z| < 2$.

It is easy to see that

$$||f(z,t)||_2 \leq \max_{\phi \in \mathbb{R}} |S(f(e^{i\phi}/2, t)), t)|.$$

Since $||f(z,t)||_1 \leq R_e(t)$, then

$$|f(e^{i\phi}/2, t)| \leq R_e(t)/2. \tag{3.3}$$

Noting that the preimage of the disk $|z| \leq R_i(t)$ lies in $\mathbb{D}$, and using the Schwartz lemma, we have that the preimage of the disk $|z| \leq R_i(t)/2$ lies in the disk $|z| \leq 1/2$. From this we obtain

$$|f(e^{i\phi}/2, t)| \geq R_i(t)/2. \tag{3.4}$$

From (3.2), (3.3) and (3.4) we have

$$|S(f(e^{i\phi}/2,t)),t)| \le R_e(t)/2 + \frac{4t}{R_i(t)} + \frac{1}{\pi}\iint\limits_{\Omega(t)/\Omega(0)} \frac{d\sigma_w}{|z-w|},\ z := f(e^{i\phi}/2,t).$$

Again using (3.3) and (3.4) we obtain

$$|S(f(e^{i\phi}/2,t)),t)| \le R_e(t)/2 + \frac{4t}{R_i(t)} + \frac{1}{\pi}\iint\limits_{|w|\le R_e(t)} \frac{d\sigma_w}{|w|} = \frac{5}{2}R_e(t) + \frac{4t}{R_i(t)}.$$

Thus,

$$|S(f(e^{i\phi}/2,t)),t)| \le \frac{9}{2}\sqrt{2t+S/\pi} + \frac{5}{2}D + \frac{4t}{\sqrt{2t+S/\pi}-2d-D} - 2\sqrt{2t+S/\pi}.$$

Latter two expressions are equivalent to

$$\frac{2(D+2d)\sqrt{2t+S/\pi}-2S/\pi}{\sqrt{2t+S/\pi}-2d-D}.$$

Using the monotonicity of this function, the inequality (3.3), and assuming, that

$$T = \max\left\{2D+4d, \frac{(D+2d)(3D+2d+2\sqrt{S/\pi})-S/\pi}{D+\sqrt{S/\pi}}\right\} + \frac{5}{2}D,$$

we finish the proof. □

4. Sufficient condition of starlikeness

Fist we are going to prove the following statement

Lemma 4.1. *For an analytic function $f:\mathbb{D}\to\mathbb{D}$, $||f(z)||_r = \varepsilon, 0<r<1$, we have*

$$||f'(z)||_r \le F(\varepsilon,r) = \begin{cases} \frac{\sqrt{1-\log_\varepsilon^2 r}\,\log_r\varepsilon}{2r}\varepsilon^{\frac{\log r\sqrt{\frac{1+\log_\varepsilon r}{1-\log_\varepsilon r}}}{\log r}}, & \varepsilon \le r^{\frac{1+r^2}{1-r^2}} \\ \frac{1}{1-r^2}, & \varepsilon > r^{\frac{1+r^2}{1-r^2}}. \end{cases}$$

Proof. Using the Hadamard three-circle theorem, we have

$$||f(z)||_{r_1} \le \varepsilon^{\frac{\log r_1}{\log r}}, r<r_1\le 1.$$

Using the invariant version of the Schwartz lemma

$$|g'(z)| \le \frac{1-|g(z)|^2}{1-|z|^2} \le \frac{1}{1-|z|^2},\ |z|<1, |g(z)|<1$$

for the function

$$\frac{f(z/r_1)}{\varepsilon^{\frac{\log r_1}{\log r}}},$$

we conclude that

$$||f'(z)||_r \le \psi(r_1) = \frac{\varepsilon^{\frac{\log r_1}{\log r}}}{r_1(1-(\frac{r}{r_1})^2)}.$$

Let us find the minimum of the right-hand side of this inequality. In order to do this we calculate the derivative

$$\psi'(r_1) = \varepsilon^{\frac{\log r_1}{\log r}} \frac{(r_1^2 - r^2)\log\varepsilon - (r_1^2 + r^2)\log r}{(r_1^2 - r^2)^2 \log r}.$$

The equation $\psi'(r_1) = 0$ is equivalent to

$$r_1^2(\log\varepsilon - \log r) = r^2(\log\varepsilon + \log r). \tag{4.1}$$

In the case $\varepsilon \geq r$, the equation (4.1) has no solutions. If $0 < \varepsilon < r$, the equation (4.1) has one positive solution

$$r_1 = r\sqrt{\frac{\log\varepsilon + \log r}{\log\varepsilon - \log r}} = r\sqrt{\frac{1 + \log_\varepsilon r}{1 - \log_\varepsilon r}}.$$

Solving the inequality $r_1 \leq 1$, we have that if

$$\varepsilon \leq r^{\frac{1+r^2}{1-r^2}} < r,$$

then the solution belongs to the interval $(r, 1]$. So we have

$$||f'(z)||_r \leq F(\varepsilon, r) = \begin{cases} \frac{\sqrt{1-\log_\varepsilon^2 r}\,\log_r \varepsilon}{2r}\varepsilon^{\frac{\log r\sqrt{\frac{1+\log_\varepsilon r}{1-\log_\varepsilon r}}}{\log r}}, & \varepsilon \leq r^{\frac{1+r^2}{1-r^2}} \\ \frac{1}{1-r^2}, & \varepsilon > r^{\frac{1+r^2}{1-r^2}}. \end{cases} \qquad \square$$

$\square$

Proposition 4.2. *An analytic function $f(z) : \mathbb{D} \to \mathbb{D}$, $f(0) = 0, f(z) = \alpha z + \phi(z)$ is starlike in the disk $\mathbb{D}_r, 0 < r < 1$ if $||\phi(z)||_r \leq x$, where x is a unique solution to the equation*

$$\arcsin\frac{x}{r\alpha} + \arcsin\frac{F(x, r)}{\alpha} = \frac{\pi}{2}.$$

Proof. Due to the necessary and sufficient condtion of starlikeness, we have to prove that

$$\Re e\frac{\alpha + \phi'(z)}{\alpha + \phi(z)/z} \geq 0, \ |z| = r. \tag{4.2}$$

Using lemma 4.1, we find the estimate for the argument of the numerator θ of the fraction in (4.2)

$$-\arcsin\frac{F(||\phi(z)||_r, r)}{\alpha} \leq \theta \leq \arcsin\frac{F(||\phi(z)||_r, r)}{\alpha}.$$

Similarly, we derive an estimate for the argument γ of the denominator

$$-\arcsin\frac{||\phi(z)||_r}{r\alpha} \leq \gamma \leq \arcsin\frac{||\phi(z)||_r}{r\alpha}.$$

Condition (4.2) can be written as

$$\frac{\pi}{2} \geq \theta - \gamma \geq -\frac{\pi}{2},$$

or

$$-\arcsin\frac{||\phi(z)||_r}{r\alpha}-\arcsin\frac{F(||\phi(z)||_r,r)}{\alpha}$$
$$\geq\theta-\gamma\geq\arcsin\frac{||\phi(z)||_r}{r\alpha}+\arcsin\frac{F(||\phi(z)||_r,r)}{\alpha}$$

Thus, if the inequality

$$\arcsin\frac{||\phi(z)||_r}{r\alpha}+\arcsin\frac{F(||\phi(z)||_r,r)}{\alpha}\leq\frac{\pi}{2},$$

holds, then the function $f(z)$ is starlike.

Using the monotonicity of $F(x,r)$ for a fixed r, we arrive at the equation

$$\arcsin\frac{x}{r\alpha}+\arcsin\frac{F(x,r)}{\alpha}=\frac{\pi}{2}.$$

It has a unique solution x, and from $||\phi(z)||_r\leq x$, we have (4.2). □

5. Proof of the theorem

Let

$$g(z/2,t)=\frac{f(z,t)}{\frac{9}{2}\sqrt{2t+S/\pi}+T}.$$

From Proposition 3.1 we have that if

$$\sqrt{2t+S/\pi}\geq 3D+2d+2\sqrt{S/\pi}, \tag{5.1}$$

then the function $g(z,t)$ maps $\mathbb{D}$ into itself.

Proposition 4.2 implies that, if

$$\Big|\Big|g(z,t)-\frac{4}{9}z\Big|\Big|_{1/2}\leq\delta=0.03898585230688595\ldots,$$

where δ is a unique solution to the equation

$$\arcsin\frac{9\delta}{2}+\arcsin\frac{9F(\delta,1/2)}{4}=\frac{\pi}{2},$$

then the function $f(z,t)$ is starlike in $\mathbb{D}$ and the solution to (1.1) exists infinitely long time [2]. We have

$$\Big|\Big|\,|g(z,t)|-\Big|\frac{4}{9}z\Big|\,\Big|\Big|_{1/2}\leq\max\Big\{\Big|\frac{R_e(t)}{\frac{9}{2}\sqrt{2t+S/\pi}+T}-\frac{2}{9}\Big|,\Big|\frac{R_i(t)}{\frac{9}{2}\sqrt{2t+S/\pi}+T}-\frac{2}{9}\Big|\Big\}$$
$$=x,$$

also

$$\frac{2R_i(t)}{\frac{9}{2}\sqrt{2t+S/\pi}+T}\leq g'(0,t)\leq\frac{2R_e(t)}{\frac{9}{2}\sqrt{2t+S/\pi}+T}.$$

So we have

$$\Big|g'(0,t)-\frac{4}{9}\Big|\leq 2x. \tag{5.2}$$

We need the following statement.

Lemma 5.1. *Suppose that* $f(z) = \beta z + zq(z)$, $\beta > 0$, $q(z) = a_1 z + a_2 z^2 + \cdots$, *is an analytic function in* $\mathbb{D}$ *satisfying the inequality* $|f(z)| \le 1$, $|z| < 1$. *Let* $r \in (0,1)$, $\varepsilon > 0$, *and* $\alpha > 0$, $|\beta - \alpha| \le \varepsilon$. *If* $\big\| \, |f(z)| - |\alpha z| \, \big\|_r \le r\varepsilon$, *then*

$$\|zq(z)\|_r \le 4\varepsilon r \log_r 4\varepsilon + \frac{4\varepsilon r}{1-r}.$$

Proof. For the function $f(z)/z$ we have

$$\|f(z)/z\|_r = ||\beta + q(z)||_r \le \alpha + \varepsilon.$$

Taking into account the inequality $|\alpha - \beta| \le \varepsilon$ we conclude that

$$\Re e\, q(z) \le 2\varepsilon, |z| \le r.$$

Hence the function

$$p(z) = \frac{2\varepsilon - q(rz)}{2\varepsilon} = 1 + p_1 z + p_2 z^2 + \ldots$$

has positive real part in $\mathbb{D}$. It follows that $|p_n| \le 2$ for all $n > 0$. Consequently, $|a_n| \le 4\varepsilon r^{-n}$, $n > 0$. Since $|f(z)| \le 1$, $|z| < 1$, we have $|a_n| \le 1$, $n > 0$. Therefore,

$$|a_n| \le \begin{cases} 4\varepsilon r^{-n}, & 0 < n \le \log_r 4\varepsilon, \\ 1, & n > \log_r 4\varepsilon. \end{cases}$$

It follows that

$$||q(z)||_r \le 4\varepsilon \log_r 4\varepsilon + \frac{4\varepsilon}{1-r},$$

which finishes the proof of the lemma. □

From this lemma and inequality (5.2) we obtain

$$\left|\left| g(z,t) - \frac{4}{9} z \right|\right|_{1/2} \le 4x \log_{1/2} 8x + 9x.$$

From this it follows that if

$$\max\left\{ \left| \frac{R_e(t)}{\frac{9}{2}\sqrt{2t + S/\pi + T}} - \frac{2}{9} \right|, \left| \frac{R_i(t)}{\frac{9}{2}\sqrt{2t + S/\pi + T}} - \frac{2}{9} \right| \right\} \le x,$$

then the solution is starlike, where $x = 0.0010686995709770337\ldots$ is the minimal positive solution to the equation

$$4x \log_{1/2} 8x + 9x = \delta.$$

Taking into account the inequality $R_e(t) > R_i(t)$, we find that is sufficient to show that

$$\frac{R_e(t)}{\frac{9}{2}\sqrt{2t + S/\pi + T}} - \frac{2}{9} \le x, \quad \text{and} \quad \frac{R_i(t)}{\frac{9}{2}\sqrt{2t + S/\pi + T}} - \frac{2}{9} \ge -x.$$

Using estimates (2.3) and (2.2) for $R_e(t)$ and $R_i(t)$, we have

$$\frac{D - \frac{2}{9}T}{\frac{9}{2}\sqrt{2t + S/\pi} + T} \leq x \quad \text{and} \frac{2d + D + \frac{2}{9}T}{\frac{9}{2}\sqrt{2t + S/\pi} + T} \leq x.$$

As D, d, T are positive, first inequality implies second. And it is sufficient to show that

$$\sqrt{2t + S/\pi} \geq \frac{2}{9}\frac{2d + D + (\frac{2}{9} - x)T}{x}.$$

Tacking into account (4.2) we have that, if a solution exists for

$$t \geq \max\Big\{\Big(\frac{2}{9}\frac{2d + D + (\frac{2}{9} - x)T}{x}\Big)^2 - S/\pi, \frac{(3D + 2d + 2\sqrt{S/\pi})^2 - S/\pi}{2}\Big\},$$

then it exists infinitely long.

Acknowledgments

The author acknowledges the reviewer for many helpful remarks and observations regarding my paper. This work is partially supported by the Russian Foundation for Basic Research grant 07-01-00120.

References

[1] L.A. Galin, *Unsteady filtration with a free surface*, Dokl. Akad. Nauk USSR, **47** (1945), 246–249. (in Russian)

[2] B. Gustafsson, D. Prokhorov and A. Vasil'ev, *Infinite lifetime for the starlike dynamics in Hele–Shaw cells*, Proc. of the American Math. Soc. 2004 **132** no. 9, 2661–2669.

[3] O.S. Kuznetsova and V.G. Tkachev, *Asymptotic Properties of Solutions to the Hele–Shaw Equation*, Dokl.Akad. Nauk USSR, **60** (1998), no. 1, 35–37. (in Russian)

[4] O.S. Kuznetsova, *Geometrical and functional properties of solutions of Hele–Shaw problem*, Ph.D. Thesis, Volgograd, 2000. (in Russian)

[5] O. Kuznetsova, *Invariant families in the Hele–Shaw problem*, Preprint TRITA-MAT-2003-07, Royal Institute of Technology, Stockholm, Sweden, 2003, available in the Internet: http://www.math.kth.se/math/forskningsrapporter/Kuznetsova.pdf.

[6] P.Ya. Polubarinova-Kochina, *On a problem of the motion of the contour of a petroleum shell*, Dokl. Akad. Nauk USSR, **47** (1945), no. 4, 254–257. (in Russian)

[7] B. Gustafsson, A. Vasil'ev, *Conformal and potential analysis in Hele–Shaw cells.* Birkhäuser Verlag, 2006, 231 pp.

Alexander Kuznetsov
Department of Mathematics and Mechanics
Saratov State University
Astrakhanskay Str. 83
410012, Saratov, Russia
e-mail: `Alexander.A.Kuznetsov@gmail.com`

Analysis and Mathematical Physics
Trends in Mathematics, 377–395
© 2009 Birkhäuser Verlag Basel/Switzerland

The Fourier Transforms of General Monotone Functions

E. Liflyand and S. Tikhonov

Abstract. Extending the notion of the general monotonicity for sequences to functions, we exploit it to investigate integrability problems for Fourier transforms. The problem of controlling integrability properties of the Fourier transform separately near the origin and near infinity is examined. We then apply the obtained results to the problems of integrability of trigonometric series.

Mathematics Subject Classification (2000). Primary 42A38; Secondary 26A48, 26A45, 42A32.

Keywords. Fourier transform, monotone/general monotone functions, bounded variation, trigonometric series.

1. Introduction

We consider the cosine and sine Fourier transforms

$$F(x) = \int_0^\infty f(t)\cos xt\,dt \tag{1.1}$$

and

$$G(x) = \int_0^\infty g(t)\sin xt\,dt, \tag{1.2}$$

respectively, with $x \geq 0$.

We assume that f and g are locally absolutely continuous on $(0, +\infty)$ and $\lim\limits_{t\to\infty} f(t), g(t) = 0$. We call such functions *admissible.* For applications, we are interested in bounded functions, but some of our results are valid for functions tending to infinity near the origin. Our functions can be non-integrable on the whole half-axis $\mathbb{R}_+ = [0, \infty)$, hence the integrals are understood as improper integrals.

While studying conditions for integrability of the Fourier transform, those for the integrability near infinity are usually of the most interest and of crucial importance in applications. One of the reasons is that for L^1 functions integrability of the Fourier transform on a compact set is an obvious fact. But since we allow the functions we deal with to be not in L^1, integrability near the origin is of interest as well. For instance, the example of an even monotone function with non-integrable Fourier transform given in [14, 6.11, Theorem 125] is the one where the Fourier transform is not integrable on $[0, 1]$. The reader can find many important results in more recent books [4, 16].

Conditions for integrability of the Fourier transform near the origin are also helpful for applications to trigonometric series (see Section 5 below).

The paper is organized as follows. Section 2 is devoted to monotone functions. The results presented in that section are a starting point for further study and one of the model cases on which the sharpness (or lack of it) of general results to be checked. In Section 3, we introduce our main concept: a two-parameter family of classes of general monotone functions. Then, in Section 4, we study integrability conditions for the Fourier transform of a function that either itself or its derivative is from certain class of the mentioned family. Finally, we obtain applications of the results from the previous sections to trigonometric series.

Throughout the paper C will denote absolute positive constant which may differ from line to line. We will use $f \ll g$ if $f \le Cg$ and $f \gg g$ if $f \ge Cg$. Also, $f \asymp g$ means that $f \ll g$ and $f \gg g$ simultaneously.

2. Monotone functions

Let us start with a few examples. First, if f and g are $t^{-1/2}$ each, their Fourier transforms F and G both are equal to $(\pi/2)^{1/2}x^{-1/2}$. The latter is integrable over $(0, 1)$ but not near infinity. If both f and g are e^{-t}, then we have that $F(x) = 1/(1+x^2)$ and $G(x) = x/(1+x^2)$. In this case F is integrable over $(0, +\infty)$ while G is integrable only on finite intervals. Finally, if both f and g are t^{-1}, we obtain (for all these relations, see [2]) $G(x) = \pi/2$ for all $x > 0$ while $F(x)$ merely does not exist as improper integral. These examples give us a general idea what to expect near infinity and near the origin.

The results we give below are probably known but we failed to find them in that form. Clear hints are delivered by trigonometric series with monotone coefficients.

Theorem 2.1. *For f and g admissible and monotone,*

$$\int_0^\pi |F(x)|\,dx \le \pi \int_0^1 |f(t)|\,dt + 3\int_1^\infty t^{-1}|f(t)|\,dt \tag{2.1}$$

and

$$\int_0^1 t|g(t)|\,dt + (1/12)\int_1^\infty t^{-1}|g(t)|\,dt \le \int_0^\pi |G(x)|\,dx$$
$$\le (\pi^2/2)\int_0^1 t|g(t)|\,dt + 2\int_1^\infty t^{-1}|g(t)|\,dt. \tag{2.2}$$

Proof. Let us begin with the cosine transform. We have

$$\int_0^\infty f(t)\cos xt\,dt = \int_0^{\pi/x} f(t)\cos xt\,dt - x^{-1}\int_{\pi/x}^\infty f'(t)\sin xt\,dt. \tag{2.3}$$

Integrating the absolute value of the first integral on the right, we obtain

$$\int_0^\pi \Big|\int_0^{\pi/x} f(t)\cos xt\,dt\Big|\,dx \le \int_0^1 |f(t)| \int_0^\pi |\cos xt|\,dx\,dt$$
$$+\int_1^\infty |f(t)| \int_0^{\pi/t} |\cos xt|\,dx\,dt \le \pi\int_0^1 |f(t)|\,dt + 2\int_1^\infty t^{-1}|f(t)|\,dt. \tag{2.4}$$

In the second integral on the right-hand side of (2.3) we merely use the monotonicity of f and rough estimates. By this we arrive at (2.1).

It is clear that in the above examples only $f(t) = t^{-1}$ does not satisfy (2.1).

For the sine transform, G is of the same sign as g is (see, e.g., [14, 6.10, Theorem 123]). Let, for simplicity, g be monotone decreasing with, consequently, non-negative Fourier transform. We have

$$\int_0^\pi |G(x)|\,dx = \int_0^\pi \int_0^\infty g(t)\sin xt\,dt\,dx = 2\int_0^\infty t^{-1}g(t)\sin^2(\pi t/2)\,dt. \tag{2.5}$$

The right-hand side of (2.5) is

$$2\Big(\int_0^1 + \int_1^\infty\Big) t^{-1}g(t)\sin^2(\pi t/2)\,dt \le (\pi^2/2)\int_0^1 tg(t)\,dt + 2\int_1^\infty t^{-1}g(t)\,dt.$$

The estimate from below is derived as follows.

$$\int_0^\pi |G(x)|\,dx \ge 2\int_0^1 t^{-1}g(t)\sin^2(\pi t/2)\,dt + 2\int_{5/2}^\infty t^{-1}g(t)\sin^2(\pi t/2)\,dt$$
$$\ge 2\int_0^1 tg(t)\,dt + 2\sum_{k=1}^\infty \int_{2k+1/2}^{2k+1} t^{-1}g(t)\sin^2(\pi t/2)\,dt.$$

To estimate the sum on the right, we observe that on each $(2k+1/2, 2k+1)$ there holds $\sin^2(\pi t/2) = \sin^2\Big((t-2k)\pi/2\Big) \ge 1/2$ and, by the monotonicity of $t^{-1}g(t)$,

$$3\int_{2k+1/2}^{2k+1} t^{-1}g(t)\,dt \ge \int_{2k+1}^{2k+2+1/2} t^{-1}g(t)\,dt. \tag{2.6}$$

We get

$$\int_0^\pi |G(x)|\,dx \ge 2\int_0^1 tg(t)\,dt + (1/6)\int_{5/2}^\infty t^{-1}g(t)\,dt.$$

Next, using again (2.6) with $k = 0$, we obtain

$$\int_0^1 tg(t)\,dt \ge \int_{1/2}^1 t^2(t^{-1}g(t))\,dt \ge (1/12)\int_1^{5/2} t^{-1}g(t)\,dt.$$

This is the lower bound in (2.2), and the proof is complete. □

3. General monotone functions

Theorem 2.1 is only an initial point for further study. Our aim is to relax the monotonicity condition. A natural attempt is to generalize Leindler's *rest of bounded variation* condition for sequences [7]. For an admissible function, say h, such a generalization to functions reads as

$$\int_x^\infty |h'(t)|\,dt \ll |h(x)| \tag{3.1}$$

for all $x \in (0,\infty)$, written $h \in RBV$.

Clearly, any admissible monotone function h satisfies (3.1). The reverse is not true. However, any admissible function h satisfying (3.1) can be represented as the difference of two monotone decreasing admissible functions. Indeed, let $h_1(x) = \int_x^\infty |h'(t)|\,dt$, then taking $h_2(x) = h_1(x) - h(x)$, we easily derive that both h_1 and h_2 are monotone decreasing and admissible. This means that all upper estimates from (2.1) and (2.2) are valid for such functions as well.

Thus, we are going to relax not only the monotonicity but (3.1) as well.

The condition we will study is an extension of the *general monotonicity* introduced in [12] for sequences. Such an extension to functions is given by

$$||dh||_{L^p(x,2x)} \ll \beta(x) \tag{3.2}$$

for all $x \in (0,\infty)$, some p, $1 \le p \le \infty$, and any h locally of bounded variation. A majorant $\beta(x)$ is a fixed non-negative function on $(0,\infty)$. Of course, we shall use $dh(t) = h'(t)\,dt$ when possible. We denote the introduced class by $GM_p(\beta)$. If $\beta(x) = |h(x)|$, we will write GM_p; and GM for GM_1.

It is obvious that any monotone function belongs to GM, and that GM is less restrictive than (3.1). Let us now investigate the properties of the introduced classes. Our study will be minimal in the sense that we give only the basic properties we use in this paper; more detailed research will appear in [9] devoted to the weighted integrability of Fourier transforms.

First, the Hölder inequality yields for $1 \le p_1 \le p_2 \le \infty$

$$GM_{p_2}(\beta^{(2)}) \subset GM_{p_1}(\beta^{(1)}), \qquad \text{where} \quad \beta^{(1)}(x) = x^{1/p_1 - 1/p_2}\beta^{(2)}(x).$$

To get "proper" embedding, that is, with the same β, one has to consider slightly modified classes of general monotone functions $\widetilde{GM}_p(\beta)$ defined as

$$\left(x^{-1}\int_x^{2x} |h'(t)|^p dt\right)^{1/p} \ll \beta(x)$$

for $1 \le p < \infty$ and the limit case as $p \to \infty$ for $\widetilde{GM}_\infty(\beta)$ (when both classes coincide). Since we do not use embedding properties, the initial definition suits us well.

Let us figure out when a general monotone function is of bounded variation. It is important in various respects, for example, the Fourier transform of a function of bounded variation is well defined as improper integral.

Lemma 3.1. *Let $h \in GM_p(\beta)$. If h' is integrable near the origin, then it is integrable over $\mathbb{R}_+$, that is, such h is of bounded variation on $\mathbb{R}_+$, provided that in addition*

$$\int_1^\infty t^{-1/p}\beta(t)\,dt < \infty. \tag{3.3}$$

Furthermore, we have

$$\int_0^\infty |h'(t)|\,dt \ll \int_0^2 |h'(t)|\,dt + \int_1^\infty t^{-1/p}\beta(t)\,dt.$$

Proof. We write

$$\begin{aligned}\int_0^\infty |h'(t)|\,dt &= (1/\ln 2)\Big(\int_0^1 + \int_1^\infty\Big)x^{-1}\int_x^{2x} |h'(t)|\,dt \\ &\le (1/\ln 2)\int_0^1 x^{-1}\int_x^{2x} |h'(t)|\,dt + C\int_1^\infty x^{-1/p}\beta(x)\,dx,\end{aligned}$$

where the last bound came from the Hölder inequality and $h \in GM_p(\beta)$.

We then obtain

$$\begin{aligned}\int_0^1 x^{-1}\int_x^{2x} |h'(t)|\,dt &= \int_0^1 |h'(t)|\int_{t/2}^t x^{-1}dx\,dt + \int_1^2 |h'(t)|\int_{t/2}^1 x^{-1}dx\,dt \\ &\le \ln 2\int_0^2 |h'(t)|\,dt,\end{aligned}$$

which completes the proof. □

Remark 3.2. If we omit the assumption of integrability of h' near the origin, the condition, which insures that h to be of bounded variation is similar to (3.3), but with integration over the whole $\mathbb{R}_+$. This can be meaningless, for example, for a monotone h, $p = 1$, and $\beta(x) = |h(x)|$ when we arrive at infinite integral $\int_0^\infty t^{-1}|h(t)|\,dt$.

The next properties are of special interest and importance.

Lemma 3.3. *For any $N > x$ we have for $h \in GM_p(\beta)$ and $c > 1$*

$$\int_x^N |h'(t)|\,dt \ll x^{1-1/p}\beta(x) + \int_x^N t^{-1/p}\beta(t)\,dt, \tag{3.4}$$

$$\int_x^N |h'(t)|\,dt \ll \int_{x/c}^N t^{-1/p}\beta(t)\,dt. \tag{3.5}$$

Proof. For $N \le 2x$, the estimate (3.4) trivially follows from Hölder's inequality. Suppose that $N > 2x$. With

$$\begin{aligned}\int_x^N u^{-1}\int_u^{2u} |h'(t)|\,dt\,du &= \int_x^{2x} |h'(t)|\ln(t/x)\,dt + \ln 2\int_{2x}^N |h'(t)|\,dt \\ &\quad + \int_N^{2N} |h'(t)|\ln(2N/t)\,dt\end{aligned}$$

in hand, we obtain

$$\begin{aligned}\ln 2\int_x^N |h'(t)|\,dt &\le \ln 2\left(\int_x^{2x}+\int_{2x}^N\right)|h'(t)|\,dt+\int_N^{2N}|h'(t)|\ln(2N/t)\,dt\\ &= \ln 2\int_x^{2x}|h'(t)|\,dt+\int_N^{2N}|h'(t)|\ln(2N/t)\,dt+\int_x^N u^{-1}\int_u^{2u}|h'(t)|\,dt\,du\\ &\qquad -\int_x^{2x}|h'(t)|\ln(t/x)\,dt-\int_N^{2N}|h'(t)|\ln(2N/t)\,dt\\ &= \int_x^N u^{-1}\int_u^{2u}|h'(t)|\,dt\,du+\int_x^{2x}|h'(t)|\ln(2x/t)\,dt.\end{aligned}$$

The right-hand side does not exceed

$$\ln 2\int_x^{2x}|dh(t)|+\int_x^N u^{-1}\int_u^{2u}|dh(t)|\,du$$

that, in turn, by Hölder's inequality and (3.2), is dominated, up to a constant multiplier, by

$$x^{1-1/p}\beta(x)+\int_x^N u^{-1/p}\beta(u)\,du.$$

To prove (3.5), considering (3.4) as $K(x)\ll L(x)$ and integrating it over $(z/c,z)$, we obtain

$$z(1-1/c)\int_z^N|h'(t)|\,dt\ll z\int_{z/c}^z t^{-1/p}\beta(t)\,dt+z(1-1/c)\int_{z/c}^N t^{-1/p}\beta(t)\,dt.$$

This inequality immediately reduces to (3.5), which completes the proof. □

4. Integrability of the Fourier transform

4.1. Integrability of the Fourier transform on $\mathbb{R}_+$

Many such conditions are known; see, e.g., [4, 14, 16, 8]. However, the classes we are going to study are different. A model case, in many respects, is that of monotone functions. First, G cannot be integrable on $\mathbb{R}_+$, since g being extended to the whole axis has discontinuity at the origin. This is not the case for F. The following result of Pólya type is well known (see, e.g., [14, 6.10, Theorem 124]).

Theorem 4.1. *Let f be a bounded non-negative convex function vanishing at infinity. Then F is positive and integrable on* $(0,\infty)$.

This fits our study since such f is necessarily monotone (see [4, 6.3.1]). We observe that convexity controls the behavior of the Fourier transform on the whole half-axis and not specifically near infinity. This is the case for much wider classes from [8] as well.

To get a flavor of functions with integrable Fourier transform, we note that such a function necessarily possesses certain smoothness. More precisely, if $F(x)$ or $G(x)$ is integrable on $\mathbb{R}_+$, then the integrals

$$\int_{\varepsilon}^{x/2} \frac{h(x+t)-h(x-t)}{t}\,dt,$$

where h is either f or g, are uniformly bounded; for the well-known prototype for Fourier series see [6, Ch. II, 10]. Indeed, expressing, say, $f(x+t)$ and $f(x-t)$ via the Fourier inversion, we obtain

$$\begin{aligned}
&\left|\int_{\varepsilon}^{x/2} \frac{f(x+t)-f(x-t)}{t}\,dt\right| \\
&\quad = \pi^{-1}\left|\int_{\varepsilon}^{x/2} t^{-1}\int_0^\infty F(u)[\cos u(x+t)-\cos u(x-t)]\,du\,dt\right| \\
&\quad \le 2\pi^{-1}\int_0^\infty |F(u)|\left|\int_{\varepsilon}^{x/2}\frac{\sin t}{t}\,dt\right|\,du.
\end{aligned}$$

The last integral on the right is uniformly bounded; the proof for g is the same.

In Theorem 4.1 we actually assume the monotonicity of both the function and its derivative.

Let us discuss how the assumption on the derivative may be relaxed. First, let us replace the monotonicity of f' with the RBV condition. As above, no chance to get something new under assumption (3.1) since every such (admissible) function is represented as the difference of two monotone decreasing (admissible) functions. In our case this leads to the difference of two convex functions, with common integral, which is a partial case of quasi-convex functions, that is, locally absolutely continuous with derivative locally of bounded variation and such that

$$\int_0^\infty t|df'(t)| < \infty. \tag{4.1}$$

Every such f is the difference of two convex functions $f_1(t) = \int_t^\infty \int_s^\infty |df'(u)|\,ds$ and $f_2(t) = f_1(t) - f(t)$; for f twice differentiable f_1'' and f_2'' are just non-negative. We obtain this case in full by assuming f not monotone but of bounded variation with the derivative satisfying (3.1). The cosine Fourier transform of such function is known to be integrable on the whole half-axis, see [4, Sect. 6.3].

Let us continue with assuming the derivative to be general monotone.

Theorem 4.2. *Let f, g be admissible and of bounded variation and let $\lim\limits_{t\to\infty} f'(t)$, $g'(t) = 0$ and $f', g' \in GM_p(\beta)$. Then*

$$\int_0^\infty |F(x)|\,dx \ll \int_0^\infty t^{1-1/p}\beta(t)\,dt, \tag{4.2}$$

and for $x > 0$

$$G(x) = x^{-1}g(\pi/(2x)) + \theta\gamma(x), \tag{4.3}$$

where $|\theta| \le C$ *and*

$$\int_0^\infty |\gamma(x)|\,dx \le \int_0^\infty t^{1-1/p}\beta(t)\,dt.$$

Proof. First, if $f' \in GM_p(\beta)$ and $\lim\limits_{t\to\infty} f'(t) = 0$, one has for any $q \in [1,\infty]$

$$f \in GM_q(\overline{\beta}), \qquad \text{where} \qquad \overline{\beta}(x) = x^{1/q} \int_{x/2}^\infty t^{-1/p}\beta(t)\,dt.$$

Indeed, by Lemma 3.3,

$$\left(\int_x^{2x} |f'(t)|^q\,dt\right)^{1/q} \le \left(\int_x^{2x} \left|\int_t^\infty df'(z)\right|^q dt\right)^{1/q} \ll x^{1/q}\int_{x/2}^\infty t^{-1/p}\beta(t)\,dt.$$

Integrating now by parts in the Fourier transform formula for F, we obtain

$$F(x) = -x^{-1}\left(\int_0^{\pi/(2x)} + \int_{\pi/(2x)}^\infty\right) f'(t)\sin xt\,dt.$$

We now have

$$\int_0^\infty x^{-1}\left|\int_0^{\pi/(2x)} f'(t)\sin xt\,dt\right| dx \le (\pi/2)\int_0^\infty |f'(t)|\,dt, \tag{4.4}$$

and, using the fact that $f \in GM_1(\overline{\beta})$,

$$\int_0^\infty |f'(t)|\,dt \ll \int_0^\infty x^{-1}\int_x^{2x} |f'(t)|\,dt\,dx$$
$$\ll \int_0^\infty x^{-1}\overline{\beta}(x)\,dx = \int_0^\infty \int_{x/2}^\infty t^{-1/p}\beta(t)\,dt\,dx \ll \int_0^\infty t^{1-1/p}\beta(t)\,dt.$$

Integrating again by parts and using assumptions of the theorem, we obtain

$$x^{-1}\int_{\pi/(2x)}^\infty f'(t)\sin xt\,dt = x^{-2}\int_{\pi/(2x)}^\infty \cos xt\,df'(t). \tag{4.5}$$

Then, using (3.5) with $c = \pi/2$, we get

$$\int_0^\infty x^{-2}\int_{\pi/(2x)}^\infty |df'(t)|\,dx \ll \int_0^\infty x^{-2}\int_{1/x}^\infty t^{-1/p}\beta(t)\,dt\,dx$$
$$\ll \int_0^\infty t^{1-1/p}\beta(t)\,dt. \tag{4.6}$$

Estimates of the sine transform go along the same lines with certain changes. We have

$$G(x) = \left(\int_0^{\pi/(2x)} + \int_{\pi/(2x)}^\infty\right) g(t)\sin xt\,dt = I_1 + I_2. \tag{4.7}$$

After twice integrating by parts, I_2 is estimated similarly to (4.5) and (4.6). Indeed,

$$\begin{aligned}\int_{\pi/(2x)}^{\infty} g(t)\sin xt\,dt &= x^{-1}\int_{\pi/(2x)}^{\infty} g'(t)\cos xt\,dt\\ &= x^{-2}g'(t)\sin xt\Big|_{\pi/(2x)}^{\infty} - x^{-2}\int_{\pi/(2x)}^{\infty}\sin xt\,dg'(t)\\ &= -x^{-2}g'(\pi/(2x)) - x^{-2}\int_{\pi/(2x)}^{\infty}\sin xt\,dg'(t),\end{aligned}$$

and $\int_0^\infty \left|x^{-2}g'(\pi/(2x))\right| dx$ is dominated by

$$\int_0^\infty |g'(t)|\,dt \ll \int_0^\infty t^{1-1/p}\beta(t)\,dt,$$

while the second term is estimated exactly as in (4.6). For I_1, we have

$$\begin{aligned}&\int_0^{\pi/(2x)} g(t)\sin xt\,dt\\ &\quad= \int_0^{\pi/(2x)} [g(t)-g(\pi/(2x))]\sin xt\,dt + \int_0^{\pi/(2x)} g(\pi/(2x))\sin xt\,dt\\ &\quad= x^{-1}g(\pi/(2x)) - \int_0^{\pi/(2x)} g'(u)\int_0^u \sin xt\,dt\,du.\end{aligned}$$

The integral on the right is estimated like in (4.4), which completes the proof. □

Let us now consider several examples of majorants β in assumptions of this theorem. One of the most important is the GM-condition, that is, $||dh||_{L^1(x,2x)} \ll |h(x)|$. Then we can rewrite estimate (4.2) as

$$\int_0^\infty |F(x)|\,dx \ll \int_0^\infty |f'(t)|\,dt,$$

with a similar estimate for $\int_0^\infty |\gamma(x)|\,dx$ in (4.3).

We note that the same estimates hold if we replace the GM-condition with the more general one

$$||dh||_{L^1(x,2x)} \ll \int_{x/c}^{cx} \frac{|h(t)|}{t}\,dt \quad \text{for some} \quad c>1 \tag{4.8}$$

(see for the discrete case [13]). Moreover, the result still holds if we suppose the limit version of (4.8), i.e.,

$$||dh||_{L^1(x,\infty)} \ll \int_{x/c}^{\infty} \frac{|h(t)|}{t}\,dt \quad \text{for some} \quad c>1, \tag{4.9}$$

which is more general than (4.8) (for the discrete case, see [5]).

The conditions considered, like those in [8], control the integrability of the Fourier transform on the whole $\mathbb{R}_+$. This is not surprising since they imply the quasi-convexity. Indeed, when $p = 1$ for the most general case, (4.9), we write

$$\int_0^\infty t|dh'(t)| = \int_0^\infty \int_u^\infty |dh'(t)|\,du \ll \int_0^\infty \int_{u/c}^\infty \frac{|h'(t)|}{t} dt\,du \ll \int_0^\infty |h'(u)|\,du.$$

where h is either f or g. What is specific is that the upper bound in the theorem is $\int_0^\infty |h'(u)|\,du$ rather than $\int_0^\infty t|dh'(t)|$. This is provided by the general monotonicity of the derivative.

Considering now functions from $GM_p(\beta)$, with $p > 1$, rather than the derivatives to be such, we obtain similar results with different bounds.

Theorem 4.3. *Let $h := f$ and $h := g$ be admissible functions satisfying $GM_p(\beta)$ for some $1 < p \le \infty$. Then*

$$\int_0^\infty |F(x)|\,dx \ll \int_0^\infty x^{-1/p}\beta(x)\,dx$$

and

$$G(x) = x^{-1}g(\pi/(2x)) + \theta\gamma(x),$$

where $\theta \le C$ and

$$\int_0^\infty |\gamma(x)|\,dx \le \int_0^\infty x^{-1/p}\beta(x)\,dx.$$

Proof. Multiplying both sides of (3.2) by $x^{-1/p}$ and integrating then over $(0,\infty)$, we obtain

$$\int_0^\infty x^{-1/p}||dh||_{L^p(x,2x)}\,dx \ll \int_0^\infty x^{-1/p}\beta(x)\,dx. \tag{4.10}$$

Assuming the right-hand side to be finite, we get for $1 < p < \infty$ exactly the Fomin type condition

$$\int_0^\infty \left(x^{-1}\int_x^{2x} |h'(t)|^p\,dt\right)^{1/p} dx < \infty,$$

and for $p = \infty$, the Sidon-Telyakovskiĭ type condition

$$\int_0^\infty \sup_{x\le t\le 2x} |h'(t)|\,dx < \infty;$$

all studied in detail in [8]. Now, Theorem 3 (or Theorem 4 for $p = \infty$) from [8] yields the result. The above argument (4.10) lets $\int_0^\infty x^{-1/p}\beta(x)\,dx$ to be the bound. □

Remark 4.4. For $p = 1$, the same argument as that in the beginning of the proof implies h to be only of bounded variation provided $\int_0^\infty x^{-1}\beta(x)\,dx < \infty$; unlike in the case $p > 1$ where we obtain belonging to specific integrability subspaces of the space of functions of bounded variation. Assuming bounded variation alone is by no means enough for integrability of the Fourier transform.

The last two theorems give convenient conditions for integrability of the Fourier transform, and even asymptotic formulas for the sine transform, but again on the whole half-axis.

Let us now study integrability of the Fourier transform separately near the origin and near infinity.

4.2. Integrability of the Fourier transform near the origin

We will first try to check which conditions are enough for the integrability of the Fourier transform of a general monotone function near the origin.

Theorem 4.5. *For f and g admissible and from $GM_1(\beta)$*

$$\int_0^\pi |F(x)|\,dx \ll \int_0^1 |f(t)|\,dt + \int_1^\infty t^{-1}\beta(t)\Big(1+\ln t\Big)\,dt \tag{4.11}$$

and

$$\int_0^\pi |G(x)|\,dx \ll \int_0^1 t|g(t)|\,dt + \int_1^\infty t^{-1}\beta(t)\Big(1+\ln t\Big)\,dt. \tag{4.12}$$

Proof. We proceed exactly like in Theorem 2.1. Using then (3.4) leads to (4.11) and (4.12) (more restrictive than the above conditions (2.1) and (2.2) for monotone functions). Indeed, let us see how it works for the last integral in (2.3). We have

$$\begin{aligned}\int_0^\pi x^{-1}\Big|\int_{\pi/x}^\infty f'(t)\sin xt\,dt\Big|\,dx &\le \int_0^1 x^{-1}\int_{1/x}^\infty |f'(t)|\,dt\,dx\\ &\ll \int_0^1 x^{-1}\Big[\beta(1/x)+\int_{1/x}^\infty u^{-1}\beta(u)\,du\Big]\,dx\\ &\le \int_1^\infty x^{-1}\beta(x)\,dx + \int_1^\infty u^{-1}\beta(u)\Big(1+\ln t\Big)\,du,\end{aligned}$$

and this is exactly the last integral in (4.11). The last integral in (2.4) as well as the corresponding integrals for $G(x)$ are treated similarly.

The proof is complete. □

Remark 4.6. Assuming in Theorem 4.5 that $f, g \in GM_p(\beta)$ with $p > 1$, we will obtain similar estimates but with $t^{-1/p}$ instead of t^{-1} in the corresponding integrals. However, Theorem 4.3 gives less restrictive condition for β at infinity.

Assuming general monotonicity of the derivative rather than that of the function, we obtain the following result.

Theorem 4.7. *Let f, g be admissible and let $\lim_{t\to\infty} f'(t), g'(t) = 0$ and $f', g' \in GM_p(\beta)$. Then*

$$\int_0^\pi |F(x)|\,dx \ll \int_0^1 t|f'(t)|\,dt + \int_{1/2}^\infty t^{1-1/p}\beta(t)\,dt, \tag{4.13}$$

and for $0 < x \le \pi$

$$G(x) = x^{-1}g(\pi/(2x)) + \theta\gamma(x), \tag{4.14}$$

where $|\theta| \le C$ *and*

$$\int_0^\pi |\gamma(x)|\,dx \le \int_0^1 t|g'(t)|\,dt + \int_{1/2}^\infty t^{1-1/p}\beta(t)\,dt.$$

Proof. The proof is just repeating the proof of Theorem 4.2 with integrating over $[0,\pi]$ instead of over $\mathbb{R}_+$. Let us give only two key estimates.

First, as in (4.4) we have

$$\int_0^\pi x^{-1}\Big|\int_0^{\pi/(2x)} f'(t)\sin xt\,dt\Big|\,dx \le \int_0^{1/2} t|f'(t)|\,dt \int_0^\pi dx \\ + \int_{1/2}^\infty t|f'(t)| \int_0^{\pi/(2t)} dx\,dt \ll \int_0^1 t|f'(t)|\,dt + \int_1^\infty |f'(t)|\,dt.$$

By (3.5),

$$\int_1^\infty |f'(t)|\,dt \le \int_1^\infty \int_t^\infty |df'(u)|\,dt \\ \ll \int_1^\infty \int_{t/2}^\infty u^{-1/p}\beta(u)\,du \ll \int_{1/2}^\infty t^{1-1/p}\beta(t)\,dt.$$

In the same way, instead of (4.6) we have (by (3.4))

$$\int_0^\pi x^{-2}\int_{\pi/(2x)}^\infty |df'(t)|\,dx \ll \int_0^\pi x^{-2}\Big[x^{-1+1/p}\beta(\pi/2x) + \int_{\pi/(2x)}^\infty t^{-1/p}\beta(t)\,dt\Big]\,dx \\ \ll \int_{1/2}^\infty t^{1-1/p}\beta(t)\,dt.$$

Estimates of G go along the same lines. The proof is complete. □

4.3. Integrability of the Fourier transform near infinity

To investigate integrability near infinity, let us apply the same approach as in Theorems 4.2 and 4.3. To show that it really makes sense, we shall specify β and p in the sequel.

Theorem 4.8. *Let f and g be admissible functions, with f of bounded variation on $[0,1]$, such that $\lim_{t\to\infty} f'(t), g'(t) = 0$ and $f', g' \in GM_p(\beta)$. Then*

$$\int_\pi^\infty |F(x)|\,dx \ll \int_0^1 |f'(t)|\,dt + \int_0^1 t^{-1}|f(0)-f(t)|\,dt \\ + \int_0^1 t^{1-1/p}\beta(t)\,dt + \int_1^\infty t^{-1/p}\beta(t)\,dt \tag{4.15}$$

and for $x > \pi$

$$G(x) = x^{-1}g(2\pi/x) + \theta\gamma(x), \tag{4.16}$$

where $|\theta| \le C$ *and*

$$\int_\pi^\infty |\gamma(x)|\,dx \le \int_0^2 t^{-1}|g(0)-g(t)|\,dt + \int_0^2 t^{1-1/p}\beta(t)\,dt + \int_2^\infty t^{-1/p}\beta(t)\,dt. \tag{4.17}$$

Proof. Let again

$$F(x) = \left(\int_0^{\pi/x} + \int_{\pi/x}^{\infty}\right) f(t)\cos xt\,dt.$$

First,

$$\int_0^{\pi/x} f(t)\cos xt\,dt = \int_0^{\pi/x} [f(t) - f(0)]\cos xt\,dt.$$

Next,

$$\int_\pi^\infty \int_0^{\pi/x} |f(t) - f(0)|\,dt\,dx \le \pi \int_0^1 t^{-1}|f(t) - f(0)|\,dt.$$

Further, integrating by parts twice and using that both f and f' vanish at infinity, we obtain

$$\int_{\pi/x}^\infty f(t)\cos xt\,dt = x^{-2}f'(\pi/x) - x^{-2}\int_{\pi/x}^\infty \cos xt\,df'(t). \tag{4.18}$$

Integrating the first term on the right over $[\pi,\infty)$ gives $\int_0^1 |f'(t)|\,dt$.

Applying Lemma 3.3 with $c = \pi$ to the last term on the right-hand side of (4.18), we have

$$\int_\pi^\infty x^{-2}\Big|\int_{\pi/x}^\infty \cos xt\,df'(t)\Big|\,dx \le \int_\pi^\infty x^{-2}\int_{\pi/x}^\infty |df'(t)|\,dx$$
$$\ll \int_\pi^\infty x^{-2}\int_{1/x}^\infty t^{-1/p}\beta(t)\,dt\,dx \ll \int_0^1 t^{1-1/p}\beta(t)\,dt + \int_1^\infty t^{-1/p}\beta(t)\,dt,$$

which gives the required estimate.

For G, we proceed along the same lines with slight difference. First, we split the integral as

$$G(x) = \left(\int_0^{2\pi/x} + \int_{2\pi/x}^{\infty}\right) g(t)\sin xt\,dt.$$

We estimate the first integral exactly as that for F using the fact that $\int_0^{2\pi/x}\sin tx\,dt = 0$. Integrating the second integral by parts twice, we obtain

$$\int_{2\pi/x}^\infty g(t)\sin xt\,dt = x^{-1}g(2\pi/x) - x^{-2}\int_{2\pi/x}^\infty \sin xt\,dg'(t).$$

The first term on the right gives the leading term in the asymptotic formula (4.16), while the second one is treated exactly in the same way as that in (4.18). The proof is complete. □

Let us discuss various particular cases of this theorem. First, for monotone g from the class in question the leading term in (4.16) cannot be integrable near infinity. Indeed, integrating it over (π,∞) leads to the integral $\int_0^2 t^{-1}|g(t)|\,dt$. But $g(0) \ne 0$ for monotone functions.

Second, our attempts rest on proper choices of β in (3.2). It is natural to substitute the Hardy transform of $|h|$ for β, i.e.,

$$\|dh(t)\|_{L^p(x,2x)} \ll x^{-1}\int_0^x |h(t)|\,dt. \tag{4.19}$$

In the case of $h = f'$ say, we have in place of (4.15)

$$\int_\pi^\infty |F(x)|\,dx \ll \int_0^1 t^{-1}|f(0)-f(t)|\,dt + \int_1^\infty t^{-1/p}|f'(t)|\,dt + \int_0^1 |f'(t)|\left(\left\{\begin{array}{lll} 1, & \text{if} & 1<p<\infty \\ \ln(2/t), & \text{if} & p=1 \end{array}\right.\right)dt$$

with a similar right-hand side in (4.17); here $p=\infty$ does not make sense at all.

Trying a weaker version with the absolute value of the Hardy transform of h itself, that is,

$$\|dh(t)\|_{L^p(x,2x)} \ll \left|\frac{1}{x}\int_0^x h(t)\,dt\right|, \tag{4.20}$$

we get a smaller class. When applied to $h = f'$, the bounds in (4.19) and (4.20) coincide for *monotone* f. Indeed, in this case

$$\int_0^x |f'(t)|\,dt = \left|\int_0^x f'(t)\,dt\right| = |f(0)-f(x)|.$$

Back to (4.20) with $h = f'$, we obtain for $1 \le p < \infty$ in (4.15)

$$\begin{aligned}\int_\pi^\infty |F(x)|\,dx &\ll \int_0^1 |f'(t)| + \int_0^1 t^{-1}|f(0)-f(t)|\,dt + \int_1^\infty t^{-1-1/p}|f(0)-f(t)|\,dt \\ &\ll \int_0^1 |f'(t)| + \int_0^1 \frac{\omega(f;t)}{t}\,dt + \sup_{t\in[0,\infty)}|f(t)|, \end{aligned}\tag{4.21}$$

where $\omega(f;\cdot)$ is the modulus of continuity of f. The integral $\int_\pi^\infty |\gamma(x)|\,dx$ in (4.17) has a similar bound.

Finally, applying Theorem 4.7, with $p = 1$ in (4.20), for integrability over $[0,\pi]$ we arrive at the boundedness of $\int_{1/2}^\infty t^{-1}|f(0)-f(t)|\,dt$. This is of sense only when $f(0) = 0$ which is of course not the case for monotone functions. This means that (4.20) with $h = f'$ controls integrability just near infinity.

Remark 4.9. We know (see the beginning of the proof of Theorem 4.2) that $f' \in GM_p(\beta)$ implies $f \in GM_q(\overline{\beta})$ for any $q \in [1,\infty]$ with $\overline{\beta}(x) = x^{1/q}\int_{x/2}^\infty t^{-1/p}\beta(t)\,dt$.

Using $f \in GM_1(\overline{\beta})$, we obtain

$$\int_0^1 |f'(t)|\,dt + \int_0^1 t^{-1}|f(0)-f(t)|\,dt \ll \int_0^1 t^{1-1/p}\Big(1+|\ln t|\Big)\beta(t)\,dt + \int_1^\infty t^{-1/p}\beta(t)\,dt.$$

Therefore, we can write in place of both (4.15) and (4.17)

$$\int_\pi^\infty |F(x)|\,dx, \int_\pi^\infty |\gamma(x)|\,dx \ll \int_0^1 t^{1-1/p}\Big(1+|\ln t|\Big)\beta(t)\,dt \\ + \int_1^\infty t^{-1/p}\beta(t)\,dt. \tag{4.22}$$

Analogously, estimating in the same way the first term on the right-hand side of (4.13), we obtain

$$\int_0^1 t|f'(t)|\,dt \ll \int_0^{1/2} t^{2-1/p}\beta(t)\,dt + \int_{1/2}^\infty t^{-1/p}\beta(t)\,dt,$$

with a similar estimate for g. Hence,

$$\int_0^\pi |F(x)|\,dx, \int_0^\pi |\gamma(x)|\,dx \ll \int_0^1 t^{2-1/p}\beta(t)\,dt + \int_1^\infty t^{1-1/p}\beta(t)\,dt. \tag{4.23}$$

Comparing now (4.22) and (4.23) with (4.2) in terms of the majorant β, we see that conditions providing integrability of the Fourier transform near the origin are less restrictive for β near the origin, while those near infinity allow us to relax restrictions for β just near infinity.

5. Applications to trigonometric series

We wish to apply the above-proved results to trigonometric series. Observe that we use either those where conditions are for integrability of the Fourier transform near the origin or finite part of those where integrability is over the whole half-axis.

The almost one hundred years old problem of the integrability of trigonometric series reads as follows. Given a trigonometric series

$$a_0/2 + \sum_{k=1}^\infty (a_k \cos kx + b_k \sin kx), \tag{5.1}$$

find assumptions on the sequences of coefficients $\{a_k\}, \{b_k\}$ under which the series is the Fourier series of an integrable function. We will say in this case that the trigonometric series is integrable. Frequently, the series

$$a_0/2 + \sum_{k=1}^\infty a_k \cos kx \tag{5.2}$$

and

$$\sum_{k=1}^\infty b_k \sin kx \tag{5.3}$$

are investigated separately, since there is a difference in their behavior.

Given series (5.2) or (5.3) with the null sequences of coefficients in an appropriate space (subspace of the space of sequences of bounded variation). We apply the following interpolation procedure. Set for $t \in [k-1, k]$

$$A(t) = a_k + (k-t)\Delta a_{k-1}, \qquad a_0 = 0,$$
$$B(t) = b_k + (k-t)\Delta b_{k-1},$$

where $\Delta d_k = d_k - d_{k+1}$. These two functions $A(t)$ and $B(t)$ take at integer points the values $\{a_k\}$ and $\{b_k\}$, respectively, and are linear in between.

The following result due to Trigub [16, Theorem 4.1.2] is a "bridge" between sequences of Fourier coefficients and Fourier transforms (for an extension, see the recent paper [15]; an earlier version, for functions with compact support, is due to Belinsky [3]):

$$\sup_{0<|x|\le\pi}\left|\int_{-\infty}^{+\infty}\varphi(t)e^{-ixt}\,dt - \sum_{-\infty}^{+\infty}\varphi(k)e^{-ikx}\right| \ll ||\varphi||_{BV}. \tag{5.4}$$

This is, in a sense, equiconvergence of the Fourier integral and trigonometric series, both generated by a function of bounded variation. Relation (5.4) allows us to pass from estimating trigonometric series (5.2) and (5.3) to estimating the Fourier transform of $A(t)$ and $B(t)$, respectively, and vice versa. More precisely, $A(t)$ and $B(t)$ satisfy assumptions of one of the theorems from the previous section they inherit from the corresponding assumptions on the sequences a_k and b_k. Then (5.4) delivers the claimed result for trigonometric series. This approach was suggested in [8], where the reader can find numerous references to important results on the integrability of trigonometric series, first of all to those by Kolmogorov, Sidon, Boas, Telyakovskiĭ, Fomin, etc. Many results of early period can be found in [1].

The above scheme along with corresponding routine calculations goes through smoothly in each of the next results. We thus omit the details.

First, let us write down a known result (see, e.g., [1]) that follows immediately from Theorem 2.1.

Corollary 5.1. *If a_k and b_k are monotone null-sequences, then*

$$\int_0^\pi |a_0/2 + \sum_{k=1}^\infty a_k \cos kx|\,dx \ll \sum_{k=1}^\infty k^{-1}|a_k|$$

and

$$\sum_{k=1}^\infty k^{-1}|b_k| \ll \int_0^\pi |\sum_{k=1}^\infty b_k \sin kx|\,dx \ll \sum_{k=1}^\infty k^{-1}|b_k|.$$

The same upper estimates are also true for RBV sequences as is discussed above.

Let us obtain, in the same way, integrability results for general monotone sequences of the coefficients of trigonometric series.

First of all, let us indicate how the general monotonicity for the sequence $d = \{d_k\}$ looks like, written $d \in GMS_p(\beta)$:

$$\left(\sum_{k=n}^{2n-1} |\triangle d_k|^p\right)^{1/p} \ll \beta_n, \tag{5.5}$$

with the usual modification for $p = \infty$.

It is clear that for $1 \le p_1 \le p_2 \le \infty$, one has the following embeddings

$$GMS_{p_1}(\beta^{(1)}) \subset GMS_{p_2}(\beta^{(1)}) \subset GMS_{p_1}(\beta^{(2)}),$$

where $\left\{\beta_n^{(2)} = n^{1/p_1 - 1/p_2}\beta_n^{(1)}\right\}_{n\in\mathbb{N}}$. In the above approach to get a function β we just take $\beta(x) = \beta_k$ for $k \le x < k+1$.

Corollary 5.2. *If a_k and b_k are null-sequences both from $GMS_1(\beta)$, then*

$$\int_0^\pi |a_0/2 + \sum_{k=1}^\infty (a_k \cos kx + b_k \sin kx)|\,dx \ll \sum_{k=1}^\infty k^{-1}\beta_k \ln k.$$

Remark 5.3. We took both sequences to be from $GMS_1(\beta)$ with the same β just for brevity. If each is from a class with different β we simply write similar assertions separately for cosine and sine series.

And finally using Theorems 4.7 and 4.3, we obtain integrability results for the classes of sequences of the coefficients that, to the best of our knowledge, have never been considered before (for corresponding earlier results, see [1, Ch. X, §7], [10]).

Theorem 5.4. *Let a_k and b_k be null-sequences of bounded variation, and let $\triangle a_k$ and $\triangle b_k$ be null-sequences from $GMS_p(\beta)$. Then for each x, $0 < x \le \pi$,*

$$\sum_{k=1}^\infty a_k \cos kx = \theta_1\gamma_1(x), \tag{5.6}$$

and

$$\sum_{k=1}^\infty b_k \sin kx = x^{-1}B(\pi/(2x)) + \theta_2\gamma_2(x), \tag{5.7}$$

where $|\theta_i| \le C$ and $\int_0^\pi |\gamma_j(x)|\,dx \le \sum\limits_{k=1}^\infty k^{1-1/p}\beta_k$, $j = 1, 2$.

For a partial result of this type, with first differences in RBVS, see the recent paper [11].

Theorem 5.5. *Let a_k and b_k be null-sequences satisfying* (5.5) *for some $1 < p \le \infty$. Then for each x, $0 < x \le \pi$,* (5.6) *and* (5.7) *hold but with $\int_0^\pi |\gamma_j(x)|\,dx \le \sum\limits_{k=1}^\infty k^{-1/p}\beta_k$, $j = 1, 2$.*

If one wishes to have integrability results in a more traditional form, the only thing to be done is to integrate (5.6) and (5.7).

Corollary 5.6. *Let* a_k *and* b_k *satisfy either assumptions of Theorem* 5.4 *or Theorem* 5.5 *provided the corresponding series in the bounds converge. Then the series* (5.2) *is the Fourier series of an integrable function, while the series* (5.3) *is such if and only if* $\sum_{k=1}^{\infty} k^{-1}|b_k| < \infty$.

Acknowledgements

A substantial part of the paper was written during the visit of E.L. to Scuola Normale Superiore, Pisa, Italy. The authors are indebted to F. Ricci for hospitality, stimulating discussions and perfect conditions for fruitful work. The research was supported by the RFFI (Grant N 08-01-00302), MTM 2008-05561-C02-02/MT, and the Leading Scientific Schools (Grant NSH-2787.2008.1).

References

[1] N. Bary, *A Treatise on Trigonometric Series*. Pergamon, Oxford, 1964.

[2] H. Bateman, A. Erdélyi, *Tables of integral transforms, Vol. I.* McGraw Hill Book Company, New York, 1954.

[3] E.S. Belinsky, *On asymptotic behavior of integral norms of trigonometric polynomials*. Metric Questions of the Theory of Functions and Mappings, Nauk. Dumka, Kiev **6** (1975), 15–24 (Russian).

[4] P.L. Butzer, R.J. Nessel, *Fourier Analysis and Approximation*. Birkhäuser Verlag, Basel, 1971.

[5] M. Dyachenko, S. Tikhonov, *Convergence of trigonometric series with general monotone coefficients*. Comptes Rendus Mathématique, Acad. Sci. Paris, Vol. 345, Is. 3, 1, (2007) 123–126.

[6] J.-P. Kahane, *Séries de Fourier absolument convergentes*. Springer-Verlag, Berlin-Heidelberg-New York, 1970.

[7] L. Leindler, *On the uniform convergence and boundedness of a certain class of sine series*. Anal. Math. **27** (2001), 279–285.

[8] E. Liflyand, *Fourier transforms of functions from certain classes*. Anal. Math. **19** (1993), 151–168.

[9] E. Liflyand, S. Tikhonov, *A concept of general monotonicity and applications*. Submitted for publication.

[10] S.A. Telyakovskiĭ, *On the behavior near the origin of sine series with convex coefficients*. Publ. l'Ins. Math., 58(72) (1995), 43–50.

[11] S.A. Telyakovskiĭ, *On the behavior of sine series near zero*. Makedon. Akad. Nauk. Umet. Oddel. Mat-Tehn. Nauk. Prolozi 21(2000), no. 1–2, 47–54(2002) (Russian).

[12] S. Tikhonov, *Trigonometric series with general monotone coefficients*. J. Math. Anal. Appl. **326** (2007), 721–735.

[13] S. Tikhonov, *Best approximation and moduli of smoothness: computation and equivalence theorems*. J. Approx. Theory **153** (2008), 19–39.

[14] E.C. Titchmarsh, *Introduction to the theory of Fourier integrals*. Oxford, 1937.

[15] R.M. Trigub, *A Generalization of the Euler-Maclaurin formula*. Mat. Zametki **61** (1997), 312–316 (Russian). – English translation in Math. Notes **61** (1997), 253–257.

[16] R.M. Trigub, E.S. Belinsky, *Fourier Analysis and Approximation of Functions*. Kluwer, 2004.

E. Liflyand
Department of Mathematics,
Bar-Ilan University,
52900 Ramat-Gan, Israel
e-mail: liflyand@math.biu.ac.il

S. Tikhonov
Scuola Normale Superiore
Piazza dei Cavalieri 7
56126 Pisa, Italy

and

ICREA and Centre de Recerca Matemàtica,
Apartat 50
E-08193 Barcelona, Spain
e-mail: s.tikhonov@sns.it

Analysis and Mathematical Physics
Trends in Mathematics, 397–408
© 2009 Birkhäuser Verlag Basel/Switzerland

Traces of Hörmander Algebras on Discrete Sequences

Xavier Massaneda, Joaquim Ortega-Cerdà and Myriam Ounaïes

Abstract. We show that a discrete sequence Λ of the complex plane is the union of n interpolating sequences for the Hörmander algebras A_p if and only if the trace of A_p on Λ coincides with the space of functions on Λ for which the divided differences of order $n-1$ are uniformly bounded. The analogous result holds in the unit disk for Korenblum-type algebras.

Mathematics Subject Classification (2000). 30E05, 42A85.

Keywords. Interpolating sequences, divided differences.

1. Definitions and statement

A function $p : \mathbb{C} \longrightarrow \mathbb{R}_+$, is called a *weight* if

(w1) There is a constant $K > 0$ such that $p(z) \geq K \ln(1 + |z|^2)$.

(w2) There are constants $D_0 > 0$ and $E_0 > 0$ such that whenever $|z - w| \leq 1$ then
$$p(z) \leq D_0 p(w) + E_0.$$

Let $H(\mathbb{C})$ denote the space of all entire functions. We consider the algebra
$$A_p = \Big\{ f \in H(\mathbb{C}),\quad \forall z \in \mathbb{C},\ |f(z)| \leq A\, e^{Bp(z)} \text{ for some } A > 0, B > 0 \Big\}.$$

Condition (w1) implies that A_p contains the polynomials and (w2) that it is closed under differentiation.

Definition 1.1. Given a discrete subset $\Lambda \subset \mathbb{C}$ we denote by $A_p(\Lambda)$ the space of sequences $\omega(\Lambda) = \{\omega(\lambda)\}_{\lambda \in \Lambda}$ of complex numbers such that there are constants $A, B > 0$ for which
$$|\omega(\lambda)| \leq A e^{Bp(\lambda)}, \quad \lambda \in \Lambda.$$

Partially supported by the Picasso programme (Action Intégrée) HF2006-0211. First and second authors supported by MEC grant MTM2008-05561-C02-01 and CIRIT grant 2005-SGR 00611.

We say that Λ is an *interpolating sequence for* A_p when for every sequence $\omega(\Lambda)\in A_p(\Lambda)$ there exists $f\in A_p$ such that $f(\lambda)=\omega(\lambda)$, $\lambda\in\Lambda$. In terms of the restriction operator

$$\begin{aligned}\mathcal{R}_\Lambda : A_p &\longrightarrow A_p(\Lambda)\\ f &\mapsto \{f(\lambda)\}_{\lambda\in\Lambda},\end{aligned}$$

Λ is interpolating when $\mathcal{R}_\Lambda(A_p)=A_p(\Lambda)$.

Definition 1.2. Let Λ be a discrete sequence in $\mathbb{C}$ and ω a function given on Λ. The *divided differences of* ω are defined by induction as follows

$$\begin{aligned}\Delta^0\omega(\lambda_1)&=\omega(\lambda_1)\ ,\\ \Delta^j\omega(\lambda_1,\dots,\lambda_{j+1})&=\frac{\Delta^{j-1}\omega(\lambda_2,\dots,\lambda_{j+1})-\Delta^{j-1}\omega(\lambda_1,\dots,\lambda_j)}{\lambda_{j+1}-\lambda_1}\qquad j\geq 1.\end{aligned}$$

For any $n\in\mathbb{N}$, denote

$$\Lambda^n=\{(\lambda_1,\dots,\lambda_n)\in\Lambda\times\overset{n}{\cdots}\times\Lambda\ :\ \lambda_j\neq\lambda_k \text{ if } j\neq k\},$$

and consider the set $X_p^{n-1}(\Lambda)$ consisting of the functions in $\omega(\Lambda)$ with divided differences of order n uniformly bounded with respect to the weight p, i.e., such that for some $B>0$

$$\sup_{(\lambda_1,\dots,\lambda_n)\in\Lambda^n}|\Delta^{n-1}\omega(\lambda_1,\dots,\lambda_n)|e^{-B[p(\lambda_1)+\cdots+p(\lambda_n)]}<+\infty\ .$$

Remark 1.3. It is clear that $X_p^n(\Lambda)\subset X_p^{n-1}(\Lambda)\subset\cdots\subset X_p^0(\Lambda)=A_p(\Lambda)$.

To see this assume that $\omega(\Lambda)\in X_p^n(\Lambda)$, i.e., there exists $B>0$ such that

$$C:=\sup_{(\lambda_1,\dots,\lambda_{n+1})\in\Lambda^{n+1}}\left|\frac{\Delta^{n-1}\omega(\lambda_2,\dots,\lambda_{n+1})-\Delta^{n-1}\omega(\lambda_1,\dots,\lambda_n)}{\lambda_{n+1}-\lambda_1}\right| \times e^{-B[p(\lambda_1)+\cdots+p(\lambda_{n+1})]}<\infty\ .$$

Then, given $(\lambda_1,\dots,\lambda_n)\in\Lambda^n$ and taking $\lambda_1^0,\dots,\lambda_n^0$ from a finite set (for instance the n first $\lambda_j^0\in\Lambda$ different of all λ_j) we have

$$\begin{aligned}\Delta^{n-1}\omega(\lambda_1,\dots,\lambda_n)&=\frac{\Delta^{n-1}\omega(\lambda_1,\dots,\lambda_n)-\Delta^{n-1}\omega(\lambda_1^0,\lambda_1,\dots,\lambda_{n-1})}{\lambda_n-\lambda_1^0}(\lambda_n-\lambda_0^1)\\ &+\frac{\Delta^{n-1}\omega(\lambda_1^0,\lambda_1,\dots,\lambda_{n-1})-\Delta^{n-1}\omega(\lambda_2^0,\lambda_1^0,\dots,\lambda_{n-2})}{\lambda_{n-1}-\lambda_2^0}(\lambda_{n-1}-\lambda_2^0)+\cdots\\ &+\frac{\Delta^{n-1}\omega(\lambda_{n-1}^0,\dots,\lambda_1^0,\lambda_1)-\Delta^{n-1}\omega(\lambda_n^0,\dots,\lambda_1^0)}{\lambda_1-\lambda_n^0}(\lambda_1-\lambda_n^0)+\Delta^{n-1}\omega(\lambda_n^0,\dots,\lambda_1^0)\end{aligned}$$

Then a direct estimate and (w1) show that for some $B > 0$ there is a constant $K(\lambda_1^0, \dots, \lambda_n^0)$ such that

$$\begin{aligned}\left|\Delta^{n-1}\omega(\lambda_1,\dots,\lambda_n)\right| &\le C\left(e^{B[p(\lambda_1^0)+\cdots+p(\lambda_n)]} + \cdots + e^{B[p(\lambda_{n-1}^0)+\cdots+p(\lambda_1)]}\right)\\ &\le K(\lambda_1^0,\dots,\lambda_n^0)e^{B[p(\lambda_1)+\cdots+p(\lambda_n)]},\end{aligned}$$

and the statement follows.

The main result of this note is modelled after Vasyunin's description of the sequences Λ in the unit disk such that the trace of the algebra of bounded holomorphic functions H^∞ on Λ equals the space of (hyperbolic) divided differences of order n (see [7], [8]). The analogue in our context is the following.

Theorem 1.4 (Main Theorem). *The identity $\mathcal{R}_\Lambda(A_p) = X_p^{n-1}(\Lambda)$ holds if and only if Λ is the union of n interpolating sequences for A_p.*

For the most usual of these weights there exists a complete description of the A_p-interpolating sequences, both in analytic and geometric terms. This is the case for doubling and radial weights (see [2, Corollary 4.8]), or for non-isotropic weights of the form $p(z) = |\operatorname{Im} z| + \log(1 + |z|)$ (see [5, Theorem 1]).

With similar techniques it should be possible to extend this result to an Hermite-type interpolation problem with multiplicities, along the lines of [6].

2. General properties

We begin by showing that one of the inclusions of Theorem 1.4 is immediate.

Proposition 2.1. *For all $n \in \mathbb{N}$, the inclusion $\mathcal{R}_\Lambda(A_p) \subset X_p^{n-1}(\Lambda)$ holds.*

Proof. Let $f \in A_p$. Let us show by induction on $j \ge 1$ that, for certain constants $A, B > 0$

$$|\Delta^{j-1} f(z_1,\dots,z_j)| \le Ae^{B[p(z_1)+\cdots+p(z_j)]} \qquad \text{for all } (z_1,\dots,z_j) \in \mathbb{C}^j.$$

As $f \in A_p$, we have $|\Delta^0 f(z_1)| = |f(z_1)| \le Ae^{Bp(z_1)}$.

Assume that the property is true for j and let $(z_1,\dots,z_{j+1}) \in \mathbb{C}^{j+1}$. Fix $z_1,\dots,z_j$ and consider z_{j+1} as the variable in the function

$$\Delta^j f(z_1,\dots,z_{j+1}) = \frac{\Delta^{j-1} f(z_2,\dots,z_{j+1}) - \Delta^{j-1} f(z_1,\dots,z_j)}{z_{j+1} - z_1}.$$

By the induction hypothesis,

$$\begin{aligned}&|\Delta^{j-1} f(z_2,\dots,z_{j+1}) - \Delta^{j-1} f(z_1,\dots,z_j)|\\ &\qquad \le A\big(e^{B[p(z_2)+\cdots+p(z_{j+1})]} + e^{B[p(z_1)+\cdots+p(z_j)]}\big) \le 2Ae^{B[p(z_1)+\cdots+p(z_{j+1})]}.\end{aligned}$$

Thus, if $|z_{j+1}-z_1|\geq 1$, we easily deduce the desired estimate. For $|z_{j+1}-z_1|\leq 1$, by the maximum principle and (w2):

$$\begin{aligned}|\Delta^j f(z_1,\dots,z_{j+1})| &\leq 2A \sup_{|\xi-z_1|=1} e^{B[p(z_1)+\cdots+p(z_j)+p(\xi)]}\\ &\leq Ae^{(B+D_0)[p(z_1)+\cdots+p(z_j)+p(z_{j+1})]}.\end{aligned}$$

□

Definition 2.2. A sequence Λ is *weakly separated* if there exist constants $\varepsilon>0$ and $C>0$ such that the disks $D(\lambda,\varepsilon e^{-Cp(\lambda)})$, $\lambda\in\Lambda$, are pairwise disjoint.

Remark 2.3. If Λ is weakly separated then $X_p^0(V)=X_p^n(V)$, for all $n\in\mathbb{N}$.

To see this it is enough to prove (by induction) that $X_p^0(\Lambda)\subset X_p^n(\Lambda)$ for all $n\in\mathbb{N}$. For $n=0$ this is trivial. Assume now that $X_p^0(\Lambda)\subset X_p^{n-1}(\Lambda)$. Given $\omega(\Lambda)\in X_p^0(\Lambda)$ we have

$$\begin{aligned}|\Delta^n\omega(\lambda_1,\dots,\lambda_{n+1})| &= \left|\frac{\Delta^{n-1}(\lambda_2,\dots,\lambda_{n+1})-\Delta^{n-1}(\lambda_1,\dots,\lambda_n)}{\lambda_{n+1}-\lambda_1}\right|\\ &\leq \frac{2A}{\varepsilon}e^{(B+C)[p(\lambda_1)+\cdots+p(\lambda_{n+1})]}\ .\end{aligned}$$

Lemma 2.4. *Let $n\geq 1$. The following assertions are equivalent:*

(a) *Λ is the union of n weakly separated sequences,*

(b) *There exist constants $\varepsilon>0$ and $C>0$ such that*

$$\sup_{\lambda\in\Lambda}\#[\Lambda\cap D(\lambda,\varepsilon e^{-Cp(\lambda)})]\leq n\ .$$

(c) $X_p^{n-1}(\Lambda)=X_p^n(\Lambda)$.

Proof. (a) ⇒(b). This is clear, by the weak separation.

(b) ⇒(a). We proceed by induction on $j=1,\dots,n$. For $j=1$, it is again clear by the definition of weak separation. Assume the property true for $j-1$. Let $1\geq\varepsilon>0$ and $C>0$ be such that $\sup_{\lambda\in\Lambda}\#[\Lambda\cap D(\lambda,\varepsilon e^{-Cp(\lambda)})]\leq j$. Put $\varepsilon'=e^{-E_0C}\varepsilon/2$ and $C'=D_0C$. By Zorn's Lemma, there is a maximal subsequence $\Lambda_1\subset\Lambda$ such that the disks $D(\lambda,\varepsilon' e^{-C'p(\lambda)})$, $\lambda\in\Lambda_1$, are pairwise disjoint. In particular Λ_1 is weakly separated. For any $\alpha\in\Lambda\setminus\Lambda_1$, there exists $\lambda\in\Lambda_1$ such that

$$D(\lambda,\varepsilon' e^{-C'p(\lambda)})\cap D(\alpha,\varepsilon' e^{-C'p(\alpha)})\neq\emptyset,$$

otherwise Λ_1 would not be maximal. Then $\lambda\in D(\alpha,\varepsilon e^{-Cp(\alpha)})$, since

$$|\lambda-\alpha|<\varepsilon' e^{-C'p(\lambda)}+\varepsilon' e^{-C'p(\alpha)}<\varepsilon e^{-Cp(\alpha)},$$

by (w2). Thus $D(\alpha,\varepsilon e^{-Cp(\alpha)})$ contains at most $j-1$ points of $\Lambda\setminus\Lambda_1$. We use the induction hypothesis to conclude that $\Lambda\setminus\Lambda_1$ is the union of $j-1$ weakly separated sequences and, by consequence, Λ is the union of j weakly separated sequences.

(b)⇒(c). It remains to see that $X_p^{n-1}(\Lambda)\subset X_p^n(\Lambda)$. Given $\omega(\Lambda)\in X_p^{n-1}(\Lambda)$ and points $(\lambda_1,\dots,\lambda_{n+1})\in\Lambda^{n+1}$, we have to estimate $\Delta^n\omega(\lambda_1,\dots,\lambda_{n+1})$. Under the assumption (b), at least one of these $n+1$ points is not in the disk $D(\lambda_1,\varepsilon e^{-Cp(\lambda_1)})$. Note that Λ^n is invariant by permutation of the $n+1$ points, thus we may assume that $|\lambda_1-\lambda_{n+1}|\geq \varepsilon e^{-Cp(\lambda_1)}$. Using the fact that $\omega(\Lambda)\in X_p^{n-1}(\Lambda)$, there are constants $A,B>0$ such that

$$\begin{aligned}|\Delta^n\omega(\lambda_1,\dots,\lambda_{n+1})| &\leq \frac{|\Delta^{n-1}\omega(\lambda_2,\dots,\lambda_{n+1})|+|\Delta^{n-1}\omega(\lambda_1,\dots,\lambda_n)|}{|\lambda_1-\lambda_{n+1}|}\\ &\leq Ae^{B[p(\lambda_1)+\cdots+p(\lambda_{n+1})]}.\end{aligned}$$

(c)⇒(b). We prove this by contraposition. Assume that for all $C,\varepsilon>0$, there exists $\lambda\in\Lambda$ such that $\#[\Lambda\cap D(\lambda,\varepsilon e^{-Cp(\lambda)})]>n$. Since Λ has no accumulation points, for any fixed $C>0$, we can extract from Λ a weakly separated subsequence $\mathcal{L}=\{\alpha^l\}_{l\in\mathbb{N}}$ such that $\#[(\Lambda\setminus\mathcal{L})\cap D(\alpha^l,1/l\,e^{-Cp(\alpha^l)})]\geq n$ for all l. Let us call $\lambda_1^l,\dots,\lambda_n^l$ the points of $\Lambda\setminus\mathcal{L}$ closest to α^l, arranged by increasing distance. In order to construct a sequence $\omega(\Lambda)\in X_p^{n-1}(\Lambda)\setminus X_p^n(\Lambda)$, put

$$\begin{aligned}\omega(\alpha^l)&=\prod_{j=1}^{n-1}(\alpha^l-\lambda_j^l),\text{ for all }\alpha^l\in\mathcal{L}\\ \omega(\lambda)&=0\text{ if }\lambda\in\Lambda\setminus\mathcal{L}.\end{aligned}$$

To see that $\omega(\Lambda)\in X^{n-1}(\Lambda)$ let us estimate $\Delta^{n-1}\omega(\lambda_1,\dots,\lambda_n)$ for any given vector $(\lambda_1,\dots,\lambda_n)\in\Lambda^n$. We don't need to consider the case where the points are distant, thus, as $\mathcal{L}$ is weakly separated, we may assume that at most one of the points is in $\mathcal{L}$. On the other hand, it is clear that $\Delta^{n-1}\omega(\lambda_1,\dots,\lambda_n)=0$ if all the points are in $\Lambda\setminus\mathcal{L}$. Then, taking into account that Δ^{n-1} is invariant by permutation, we will only consider the case where λ_n is some $\alpha^l\in\mathcal{L}$ and $\lambda_1,\dots,\lambda_{n-1}$ are in $\Lambda\setminus\mathcal{L}$. In that case,

$$|\Delta^{n-1}\omega(\lambda_1,\dots,\lambda_{n-1},\alpha^l)|=|\omega(\alpha^l)|\prod_{k=1}^{n-1}|\alpha^l-\lambda_k^l|^{-1}\leq 1,$$

as desired.

On the other hand, a similar computation yields

$$|\Delta^n\omega(\lambda_1^l,\dots,\lambda_n^l,\alpha^l)|=|\omega(\alpha^l)|\prod_{k=1}^{n}|\alpha^l-\lambda_k^l|^{-1}=|\alpha^l-\lambda_n^l|^{-1}\geq le^{Cp(\alpha^l)}.$$

Using (w2), for any constant $B>0$, and choosing $C=B(nD_0+1)$, we have

$$|\Delta^n\omega(\lambda_1^l,\dots,\lambda_n^l,\alpha^l)|e^{-B(p(\lambda_1^l)+\cdots+p(\lambda_n^l)+p(\alpha^l))}\geq le^{-BnE_0}\to+\infty.$$

We finally conclude that $\omega(\Lambda)\notin X_p^n(\Lambda)$. □

Corollary 2.5. *If Λ is an interpolating sequence, then it is weakly separated.*

Proof. If Λ is an interpolating sequence, then $\mathcal{R}_\Lambda(A_p)=X_p^0(\Lambda)$. On the other hand, by Proposition 2.1, $\mathcal{R}_\Lambda(A_p)\subset X_p^1(\Lambda)$. Thus $X_p^0(\Lambda)=X_p^1(\Lambda)$. We conclude by the preceding lemma applied to the particular case $n=1$. □

Lemma 2.6. *Let $\Lambda_1,\dots,\Lambda_n$ be weakly separated sequences. There exist positive constants a,b, B_1,B_2 and $\varepsilon>0$, a subsequence $\mathcal{L}\subset\Lambda_1\cup\dots\cup\Lambda_n$ and disks $D_\lambda=D(\lambda,r_\lambda)$, $\lambda\in\mathcal{L}$, such that*
(i) $\Lambda_1\cup\dots\cup\Lambda_n\subset\cup_{\lambda\in\mathcal{L}}D_\lambda$
(ii) $a\varepsilon e^{-B_1p(\lambda)}\le r_\lambda\le b\varepsilon e^{-B_2p(\lambda)}$ *for all* $\lambda\in\mathcal{L}$
(iii) $\operatorname{dist}(D_\lambda,D_{\lambda'})\ge a\varepsilon e^{-B_1p(\lambda)}$ *for all* $\lambda,\lambda'\in\mathcal{L}$, $\lambda\neq\lambda'$.
(iv) $\#(\Lambda_j\cap D_\lambda)\le 1$ *for all* $j=1,\dots,n$ *and* $\lambda\in\mathcal{L}$.

Proof. Let $0<\varepsilon<1$ and $C>0$ be constants such that

$$|\lambda-\lambda'|\ge\varepsilon e^{-C/D_0(p(\lambda)-E_0)},\qquad\forall\lambda,\lambda'\in\Lambda_j,\ \lambda\neq\lambda',\qquad\forall j=1,\dots,n\ ,\quad(2.1)$$

where $D_0\ge1$ and $E_0\ge0$ are given by (w2).

We will proceed by induction on $k=1,\dots,n$ to show the existence of a subsequence $\mathcal{L}_k\subset\Lambda_1\cup\dots\cup\Lambda_k$ and constants $C_k\ge C$, $B_k\ge0$ such that:

$(\mathrm{i})_k$ $\quad\Lambda_1\cup\dots\cup\Lambda_k\subset\cup_{\lambda\in\mathcal{L}_k}D(\lambda,R^k_\lambda)$,

$(\mathrm{ii})_k$ $\quad 2^{-3k}e^{-C_kp(\lambda)-B_k}\varepsilon\le R^k_\lambda\le\varepsilon e^{-Cp(\lambda)}\sum_{j=0}^{k-1}2^{-(3j+2)}\le 2/7e^{-Cp(\lambda)}\varepsilon$,

$(\mathrm{iii})_k$ $\quad\operatorname{dist}(D(\lambda,R^k_\lambda),D(\lambda',R^k_{\lambda'}))\ge2^{-3k}\varepsilon e^{-C_kp(\lambda)-B_k}$ for any $\lambda,\lambda'\in\mathcal{L}_k$, $\lambda\neq\lambda'$.

The constants C_k and B_k are chosen, in view of (w2), so that $C_kp(\lambda)+B_k\le C_{k+1}p(\lambda')+B_{k+1}$ whenever $|\lambda-\lambda'|\le1$.

Then it suffices to chose $\mathcal{L}=\mathcal{L}_n$, $r_\lambda=R^n_\lambda$, $a=e^{-B_n}2^{-3n}$, $b=2/7$, $B_1=C_n$ and $B_2=C$. As $r_\lambda<e^{-Cp(\lambda)}\varepsilon$, it is clear that $D(\lambda,r_\lambda)$ contains at most one point of each Λ_j, hence the lemma follows.

For $k=1$, the property is clearly verified with $\mathcal{L}_1=\Lambda_1$ and $R^1_\lambda=e^{-Cp(\lambda)}\varepsilon/4$.

Assume the property true for k and split $\mathcal{L}_k=\mathcal{M}_1\cup\mathcal{M}_2$ and $\Lambda_{k+1}=\mathcal{N}_1\cup\mathcal{N}_2$, where

$$\begin{aligned}\mathcal{M}_1&=\{\lambda\in\mathcal{L}_k\ :\ D(\lambda,R^k_\lambda+2^{-3k-2}\varepsilon e^{-C_kp(\lambda)-B_k})\cap\Lambda_{k+1}\neq\emptyset\},\\ \mathcal{N}_1&=\Lambda_{k+1}\cap\bigcup_{\lambda\in\mathcal{L}_k}D(\lambda,R^k_\lambda+2^{-3k-2}\varepsilon e^{-C_kp(\lambda)-B_k}),\\ \mathcal{M}_2&=\mathcal{L}_k\setminus\mathcal{M}_1,\\ \mathcal{N}_2&=\Lambda_{k+1}\setminus\mathcal{N}_1.\end{aligned}$$

Now, we put $\mathcal{L}_{k+1}=\mathcal{L}_k\cup\mathcal{N}_2$ and define the radii R^{k+1}_λ as follows:

$$R^{k+1}_\lambda=\begin{cases}R^k_\lambda+2^{-3k-2}\varepsilon e^{-C_kp(\lambda)-B_k} & \text{if }\lambda\in\mathcal{M}_1,\\ R^k_\lambda & \text{if }\lambda\in\mathcal{M}_2,\\ 2^{-3k-3}\varepsilon e^{-C_{k+1}p(\lambda)-B_{k+1}} & \text{if }\lambda\in\mathcal{N}_2.\end{cases}$$

It is clear that

$$\Lambda_1\cup\dots\cup\Lambda_{k+1}\subset\bigcup_{\lambda\in\mathcal{L}_{k+1}}D(\lambda,R^{k+1}_\lambda)$$

and, by the induction hypothesis,

$$2^{-3k-3}\varepsilon e^{-C_{k+1}p(\lambda)+B_{k+1})} \leq R_\lambda^{k+1} \leq \varepsilon e^{-Cp(\lambda)} \sum_{j=0}^{k} 2^{-3j-2} \leq 2/7\varepsilon e^{-Cp(\lambda)}.$$

In order to prove (iii)$_k$ take now $\lambda, \lambda' \in \mathcal{L}_{k+1}$, $\lambda \neq \lambda'$. We will verify that $\operatorname{dist}(D(\lambda, R_\lambda^{k+1}), D(\lambda', R_{\lambda'}^{k+1})) = |\lambda - \lambda'| - R_\lambda^{k+1} - R_{\lambda'}^{k+1} \geq 2^{-3k-3}\varepsilon e^{-C_{k+1}p(\lambda)-B_{k+1}}$ by considering different cases.

If $\lambda, \lambda' \in \mathcal{L}_k$ and $p(\lambda) \leq p(\lambda')$, then

$$\begin{aligned}\operatorname{dist}(D(\lambda, R_\lambda^{k+1}), D(\lambda', R_{\lambda'}^{k+1})) &\geq |\lambda - \lambda'| - R_\lambda^k - R_{\lambda'}^k - 2^{-3k-1}\varepsilon e^{-C_k p(\lambda)-B_k} \\ &\geq 2^{-3k-1}\varepsilon e^{-C_k p(\lambda)-B_k}.\end{aligned}$$

Assume now $\lambda, \lambda' \in \mathcal{N}_2$ and $p(\lambda) \leq p(\lambda')$. Condition (2.1) implies $|\lambda - \lambda'| \geq \varepsilon e^{-Cp(\lambda)}$, hence

$$\operatorname{dist}(D(\lambda, R_\lambda^{k+1}), D(\lambda', R_{\lambda'}^{k+1})) \geq (1 - 2^{-3k-2})\varepsilon e^{-Cp(\lambda)}.$$

If $\lambda \in \mathcal{M}_1$ and $\lambda' \in \mathcal{N}_2$ there exists $\beta \in \mathcal{N}_1$ such that $|\lambda - \beta| \leq R_\lambda^{k+1}$. There is no restriction in assuming that $|\lambda - \lambda'| \leq 1$. Then, using (2.1) on $\beta, \lambda' \in \Lambda_{k+1}$, we have

$$|\lambda - \lambda'| \geq |\beta - \lambda'| - |\lambda - \beta| \geq \varepsilon e^{-C/D_0(p(\beta)-E_0)} - R_\lambda^{k+1} \geq \varepsilon e^{-Cp(\lambda)} - R_\lambda^{k+1}\ .$$

The definition of $R_{\lambda'}^{k+1}$ together with the estimate $R_\lambda^{k+1} \leq 2/7\varepsilon e^{-Cp(\lambda)}$ yield

$$\begin{aligned}\operatorname{dist}(D(\lambda, R_\lambda^{k+1}), D(\lambda', R_{\lambda'}^{k+1})) &\geq \varepsilon e^{-Cp(\lambda)} - 2R_\lambda^{k+1} - R_{\lambda'}^{k+1} \\ &\geq \varepsilon e^{-Cp(\lambda)} - 2R_\lambda^k - 2^{-3k-1}\varepsilon e^{-C_k p(\lambda)-B_k} - 2^{-3k-3}\varepsilon e^{-C_{k+1}p(\lambda')-B_{k+1}} \\ &\geq \varepsilon e^{-Cp(\lambda)} - \frac{4}{7}\varepsilon e^{-Cp(\lambda)} - 2^{-3k}\varepsilon e^{-C_k p(\lambda)-B_k} \geq \varepsilon e^{-Cp(\lambda)}(3/4 - 2^{-3k}),\end{aligned}$$

as required.

Finally, if $\lambda \in \mathcal{M}_2$ and $\lambda' \in \mathcal{N}_2$, again, assuming that $|\lambda - \lambda'| \leq 1$, we have

$$\begin{aligned}\operatorname{dist}(D(\lambda, R_\lambda^{k+1}), D(\lambda', R_{\lambda'}^{k+1})) &= |\lambda - \lambda'| - R_\lambda^k - 2^{-3k-3}\varepsilon e^{-C_{k+1}p(\lambda')-B_{k+1}} \\ &\geq 2^{-3k-2}\varepsilon e^{-C_k p(\lambda)-B_k} - 2^{-3k-3}\varepsilon e^{-C_k p(\lambda)-B_k} \\ &\geq 2^{-3k-3}\varepsilon e^{-Cp(\lambda)}.\end{aligned}$$

□

3. Proof of Theorem 1.4. Necessity

Assume $\mathcal{R}_\Lambda(A_p) = X_p^{n-1}(\Lambda)$, $n \geq 2$. Using Proposition 2.1, we have $X_p^{n-1}(V) = X_p^n(V)$, and by Lemma 2.4 we deduce that $\Lambda = \Lambda_1 \cup \cdots \cup \Lambda_n$, where $\Lambda_1, \ldots, \Lambda_n$ are weakly separated sequences. We want to show that each Λ_j is an interpolating sequence.

Let $\omega(\Lambda_j) \in A_p(\Lambda_j) = X_p^0(\Lambda_j)$. Let $\cup_{\lambda\in\mathcal{L}} D_\lambda$ be the covering of Λ given by Lemma 2.6. We extend $\omega(\Lambda_j)$ to a sequence $\omega(\Lambda)$ which is constant on each $D_\lambda \cap \Lambda_j$

in the following way:

$$\omega_{|D_\lambda\cap\Lambda} = \begin{cases} 0 & \text{if } D_\lambda\cap\Lambda_j=\emptyset \\ \omega(\alpha) & \text{if } D_\lambda\cap\Lambda_j=\{\alpha\}\ . \end{cases}$$

We verify by induction that the extended sequence is in $X_p^{k-1}(\Lambda)$ for all k. It is clear that it belongs to $X_p^0(\Lambda)$. Assume that $\omega\in X_p^{k-2}(\Lambda)$ and consider $(\alpha_1,\dots,\alpha_k)\in\Lambda^k$. If all the points are in the same D_λ then $\Delta^{k-1}\omega(\alpha_1,\dots,\alpha_k)=0$, so we may assume that $\alpha_1\in D_\lambda$ and $\alpha_k\in D_{\lambda'}$ with $\lambda\neq\lambda'$. Then we have

$$|\alpha_1-\alpha_k|\geq a\varepsilon e^{-B_1p(\lambda)},$$

by Lemma 2.6 (iii). With this and the induction hypothesis it is clear that for certain constants $A,B>0$

$$\begin{aligned}|\Delta^{k-1}\omega(\alpha_1,\dots,\alpha_k)| &= \left|\frac{\Delta^{k-2}\omega(\alpha_2,\dots,\alpha_k)-\Delta^{k-2}\omega(\alpha_1,\dots,\alpha_k)}{\alpha_1-\alpha_k}\right| \\ &\leq Ae^{B[p(\alpha_1)+\cdots+p(\alpha_k)]}.\end{aligned}$$

In particular $\omega(\Lambda)\in X_p^{n-1}(\Lambda)$, and by assumption, there exist $f\in A_p$ interpolating the values $\omega(\Lambda)$. In particular f interpolates $\omega(\Lambda_j)$.

4. Proof of Theorem 1.4. Sufficiency

According to Proposition 2.1 we only need to see that $X_p^{n-1}(\Lambda)\subset\mathcal{R}_\Lambda(A_p)$.

Before going further, let us recall the following facts about interpolation in the spaces A_p.

Lemma 4.1. [1, Lemma 2.2.6] *Let Γ be an A_p-interpolating sequence. Then:*

(i) *For all $A,B>0$, there exist constants $A',B'>0$ such that for all sequences $\omega\in A_p(\Gamma)$ with $\sup_{\gamma\in\Gamma}|\omega(\gamma)|e^{-Bp(\gamma)}\leq A$ there exists $f\in A_p$ with $\sup_z|f(z)|e^{-B'p(z)}\leq A'$ and $f(\gamma)=\omega(\gamma)$ for all $\gamma\in\Gamma$..*

(ii) *There exists a constant $C>0$ such that $\sum_{\gamma\in\Gamma}e^{-Cp(\gamma)}<\infty$.*

Applying (i) to the sequences $\omega_\gamma=\{\delta_{\gamma,\gamma'}\}_{\gamma'\in\Gamma}$ it is easy to deduce that Γ is weakly separated. Property (ii) is just a consequence of the weak separation and properties (w1) and (w2).

Assume thus that $\Lambda=\Lambda_1\cup\cdots\cup\Lambda_n$ where $\Lambda_1,\dots,\Lambda_n$ are interpolating sequences. Recall that each Λ_j is weakly separated (Corollary 2.5). Consider also the covering of Λ given by Lemma 2.6.

Lemma 4.2. *There exist constants $A,B>0$ and a sequence $\{F_\lambda\}_{\lambda\in\mathcal{L}}\subset A_p$ such that:*

$$F_\lambda(\alpha)=\begin{cases}1 & \textit{if } \alpha\in\Lambda\cap D_\lambda \\ 0 & \textit{if } \alpha\in\Lambda\cap D_{\lambda'},\ \lambda'\neq\lambda\end{cases}$$

$$|F_\lambda(z)|\leq Ae^{B(p(\lambda)+p(z))}\quad \textit{for all } z\in\mathbb{C}.$$

Proof. Fix $\lambda \in \mathcal{L}$ and define $\omega(\Lambda)$ by

$$\omega(\alpha) = \begin{cases} \prod_{\beta \in \Lambda \cap D_\lambda} (\alpha - \beta)^{-1} & \text{if } \alpha \notin \Lambda \cap D_\lambda \\ 0 & \text{if } \alpha \in \Lambda \cap D_\lambda. \end{cases}$$

By Lemma 2.6 (iii), we have $|\alpha - \beta| \geq c\varepsilon e^{-Cp(\alpha)}$ whenever $\alpha \notin \Lambda \cap D_\lambda$, $\beta \in \Lambda \cap D_\lambda$. Since $\#(\Lambda \cap D_\lambda) \leq n$ we deduce that

$$|\omega(\alpha)| \leq (c\varepsilon)^{-n} e^{nCp(\alpha)}$$

Recall that Λ_j is an interpolating sequence for all $j = 1, \ldots, n$, thus there exist a n-indexed sequence $\{f_{\lambda,j}\}_{\lambda \in \mathcal{L}, j \in [[1,n]]} \subset A_p$ such that for all $z \in \mathbb{C}$,

$$|f_{\lambda,j}(z)| \leq Ae^{Bp(z)}, \qquad f_{\lambda,j}(\alpha) = \prod_{\beta \in \Lambda \cap D_\lambda} (\alpha - \beta)^{-1} \text{ if } \alpha \notin \Lambda_j \cap D_\lambda,$$

with the constants A and B independent of λ (see Lemma 4.1(i)).

The sequence of functions $\{F_\lambda\}_{\lambda \in \mathcal{L}}$ defined by

$$F_\lambda(z) = \prod_{j=1}^{n} \left[1 - \prod_{\beta \in \Lambda \cap D_\lambda} (z - \beta) f_{\lambda,j}(z) \right]$$

has the desired properties. □

Lemma 4.3. *For all $D > 0$, there exist $D' > 0$ and a sequence $\{G_\lambda\}_{\lambda \in \mathcal{L}} \subset A_p$ such that:*

$$G_\lambda(\alpha) = e^{Dp(\lambda)} \quad \textit{if} \quad \alpha \in \Lambda \cap D_\lambda.$$
$$|G_\lambda(z)| \leq Ae^{Bp(\lambda)} e^{D'p(z)} \quad \textit{for all } z \in \mathbb{C},$$

where $A, B > 0$ do not depend on D.

Proof. In this proof D' denotes a constant depending on D but not on λ, and its actual value may change from one occurrence to the other.

Let $\lambda \in \mathcal{L}$. Assume, without loss of generality, that $D_\lambda \cap \Lambda_j = \{\alpha_{\lambda,j}\}$ for all j. As Λ_1 is an interpolating sequence and $e^{Dp(\lambda)} \leq Ae^{D'p(\alpha_{\lambda,1})}$, by Lemma 4.1(i) there exists a sequence $\{h_{\lambda,1}\}_\lambda \subset A_p$ such that

$$h_{\lambda,1}(\alpha_{\lambda,1}) = e^{Dp(\lambda)}, \qquad |h_{\lambda,1}(z)| \leq Ae^{D'p(z)} \quad \text{for all } z \in \mathbb{C}.$$

Setting $H_{\lambda,1}(z) = h_{\lambda,1}(z)$, we have $H_{\lambda,1}(\alpha_{\lambda,1}) = e^{Dp(\lambda)}$. Now, as Λ_2 is A_p-interpolating and

$$\frac{|e^{Dp(\lambda)} - H_{\lambda,1}(\alpha_{\lambda,2})|}{|\alpha_{\lambda,2} - \alpha_{\lambda,1}|} = \frac{|H_{\lambda,1}(\alpha_{\lambda,1}) - H_{\lambda,1}(\alpha_{\lambda,2})|}{|\alpha_{\lambda,2} - \alpha_{\lambda,1}|} \leq Ae^{D'p(\alpha_{\lambda,2})},$$

there exists a sequence $\{h_{\lambda,2}\}_\lambda \subset A_p$ such that

$$h_{\lambda,2}(\alpha_{\lambda,2}) = \frac{e^{Dp(\lambda)} - H_{\lambda,1}(\alpha_{\lambda,2})}{\alpha_{\lambda,2} - \alpha_{\lambda,1}}, \qquad |h_{\lambda,2}(z)| \le Ae^{D'p(z)} \quad \text{for all } z \in \mathbb{C}.$$

Setting $H_{\lambda,2}(z) = h_{\lambda,1}(z) + h_{\lambda,2}(z)(z - \alpha_{\lambda,1})$. We have

$$H_{\lambda,2}(\alpha_{\lambda,1}) = H_{\lambda,2}(\alpha_{\lambda,2}) = e^{Dp(\lambda)}.$$

We proceed by induction to construct a sequence of functions $\{h_{\lambda,k}\}_\lambda \subset A_p$ such that

$$h_{\lambda,k}(\alpha_{\lambda,k}) = \frac{e^{Dp(\lambda)} - H_{\lambda,k-1}(\alpha_{\lambda,k})}{(\alpha_{\lambda,k} - \alpha_{\lambda,1}) \cdots (\alpha_{\lambda,k} - \alpha_{\lambda,k-1})}$$
$$|h_{\lambda,k}(z)| \le Ae^{D'p(z)} \quad \text{for all } z \in \mathbb{C}.$$

Then the function defined by $H_{\lambda,k}(z) = H_{\lambda,k-1}(z) + h_{\lambda,k}(z)(z-\alpha_{\lambda,1}) \cdots (z-\alpha_{\lambda,k-1})$ verifies

$$H_{\lambda,k}(\alpha_{\lambda,1}) = \cdots = H_{\lambda,k}(\alpha_{\lambda,k}) = e^{Dp(\lambda)}.$$

Finally, we set $G_\lambda = H_{\lambda,n}$. □

To proceed with the proof of the inclusion $X_p^{n-1}(\Lambda) \subset \mathcal{R}_\Lambda(A_p)$, let $\omega(\Lambda) \in X_p^{n-1}(\Lambda)$.

Fix $\lambda \in \mathcal{L}$ and let $\Lambda \cap D_\lambda = \{\alpha_1, \ldots, \alpha_k\}$, $k \le n$. We first consider a polynomial interpolating the values $\omega(\alpha_1), \ldots, \omega(\alpha_k)$:

$$P_\lambda(z) = \Delta^0\omega(\alpha_1) + \Delta^1\omega(\alpha_1, \alpha_2)(z - \alpha_1) + \cdots + \Delta^{k-1}\omega(\alpha_1, \ldots, \alpha_k) \prod_{j=1}^{k-1}(z - \alpha_j).$$

Notice that $P_\lambda \in A_p$, since $\omega(\Lambda) \in X_p^{n-1}(\Lambda)$ and by properties (w1) and (w2) we have

$$|P_\lambda(z)| \le A|z|^k e^{B[p(\alpha_1)+\cdots+p(\alpha_k)]} \le Ae^{B'[p(z)+p(\lambda)]} .$$

Now, define

$$f = \sum_{\lambda \in \mathcal{L}} F_\lambda G_\lambda P_\lambda e^{-Dp(\lambda)},$$

where D is a large constant to be chosen later on.

By the preceding estimates on G_λ and P_λ, there exist constants $A, B > 0$ not depending on D and a constant $D'' > 0$ such that, for all $z \in \mathbb{C}$, we have

$$|f(z)| \le Ae^{D''p(z)} \sum_{\lambda \in \mathcal{L}} e^{(B-D)p(\lambda)}.$$

In view of Lemma 4.1 (ii), taking $D = B + C$, the latter sum converges and $f \in A_p$.

To verify that f interpolates $\omega(\Lambda)$, let $\alpha \in \Lambda$ and let λ be the (unique) point of $\mathcal{L}$ such that $\alpha \in D_\lambda$. Then, $f(\alpha) = G_\lambda(\alpha)P_\lambda(\alpha)e^{-Dp(\alpha)} = P_\lambda(\alpha) = \omega(\alpha)$, as desired.

5. Similar results in the disk

The previous definitions and proofs can be adapted to produce analogous results in the disk. To do so one just needs to replace the Euclidean distance used in $\mathbb{C}$ by the pseudo-hyperbolic distance

$$\rho(z,\zeta) = \left|\frac{z-\zeta}{1-\bar{\zeta}z}\right| \qquad z,\zeta \in \mathbb{D},$$

and the Euclidean divided differences by their hyperbolic version

$$\delta^0\omega(\lambda_1) = \omega(\lambda_1) \ ,$$
$$\delta^j\omega(\lambda_1,\dots,\lambda_{j+1}) = \frac{\Delta^{j-1}\omega(\lambda_2,\dots,\lambda_{j+1}) - \Delta^{j-1}\omega(\lambda_1,\dots,\lambda_j)}{\frac{\lambda_{j+1}-\lambda_1}{1-\bar{\lambda}_1\lambda_{j+1}}} \qquad j \geq 1.$$

In this context a function $\phi : \mathbb{D} \longrightarrow \mathbb{R}_+$ is a *weight* if

(wd1) There is a constant $K > 0$ such that $\phi(z) \geq K \ln\left(\frac{1}{1-|z|}\right)$.

(wd2) There are constants $D_0 > 0$ and $E_0 > 0$ such that whenever $\rho(z,\zeta) \leq 1/2$ then

$$\phi(z) \leq D_0\phi(\zeta) + E_0.$$

The model for the associated spaces

$$A_\phi = \{f \in H(\mathbb{D}) : \sup_{z\in\mathbb{D}} |f(z)|e^{-B\phi(z)} < \infty \text{ for some } B > 0\},$$

is the Korenblum algebra $A^{-\infty}$, which corresponds to the choice $e^{-\phi(z)} = 1 - |z|$. The interpolating sequences for this and similar algebras have been characterised in [3] and [4].

With these elements, and replacing the factors $z-\alpha$ by $\frac{z-\alpha}{1-\bar{\alpha}z}$ when necessary, we can follow the proofs above and, mutatis mutandis, show that Theorem 1.4 also holds in this situation.

The only point that requires further justification is the validity of Lemma 4.1 for the weights ϕ. Condition (i) is a standard consequence of the open mapping theorem for (LF)-spaces applied to the restriction map $\mathcal{R}_\Lambda$, and the same proof as in [1, Lemma 2.2.6] holds. Applying (i) to the sequences $\omega_\lambda(\Lambda)$ defined by

$$\omega_\lambda(\lambda') = \begin{cases} 1 & \text{if } \lambda' = \lambda \\ 0 & \text{if } \lambda' \neq \lambda \end{cases}$$

we have functions $f_\lambda \in A_\phi$ interpolating these values and with growth control independent of λ. Since $1 = |f_\lambda(\lambda) - f_\lambda(\lambda')|$, an estimate on the derivative of f_λ shows that for some $C > 0$ and $\varepsilon > 0$ the pseudohyperbolic disks $D_H(\lambda, \varepsilon e^{-C\phi(\lambda)}) = \{z \in \mathbb{D} : \rho(z,\lambda) < e^{-C\phi(\lambda)}\}$ are pairwise disjoint. In particular the sum of their areas is finite, hence

$$\sum_{\lambda\in\Lambda} (1-|\lambda|)^2 e^{-2C\phi(\lambda)} < +\infty \ .$$

From this and condition (wd1) we finally obtain (ii).

References

[1] C.A. Berenstein and R. Gay, *Complex Analysis and Special Topics in Harmonic Analysis*, Springer Verlag, 1995.

[2] C.A. Berenstein and B.Q. Li, *Interpolating varieties for spaces of meromorphic functions*, J. Geom. Anal. **5** (1995), 1–48.

[3] J. Bruna and D. Pascuas, *Interpolation in* $A^{-\infty}$, J. London Math. Soc. (2) **40** (1989), no. 3, 452–466.

[4] X. Massaneda, *Density conditions for interpolation in* $A^{-\infty}$, J. Anal. Math. **79** (1999), 299–314.

[5] X. Massaneda, J. Ortega-Cerdà and M. Ounaïes, *A geometric characterization of interpolation in* $\hat{\mathcal{E}}'(\mathbb{R})$, Trans. Amer. Math. Soc. **358** (2006), 3459–3472.

[6] M. Ounaïes, *Interpolation by entire functions with growth conditions*, Mich. Math. J. **56** (2008), no. 1, 155–171.

[7] V.I. Vasyunin, *Traces of bounded analytic functions on finite unions of Carleson sets* (*Russian*). Investigations on linear operators and the theory of functions, XII. Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **126** (1983), 31–34.

[8] V.I. Vasyunin, *Characterization of finite unions of Carleson sets in terms of solvability of interpolation problems* (*Russian*). Investigations on linear operators and the theory of functions, XIII. Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **135** (1984), 31–35.

Xavier Massaneda and Joaquim Ortega-Cerdà
Departament de Matemàtica Aplicada i Anàlisi
Universitat de Barcelona
Gran Via 585
08071-Barcelona, Spain
e-mail: `xavier.massaneda@ub.edu`
`jortega@ub.edu`

Myriam Ounaïes
Institut de Recherche Mathématique Avancée
Université de Strasbourg
7 Rue René Descartes
67084 Strasbourg CEDEX, France
e-mail: `ounaies@math.u-strasbg.fr`

Analysis and Mathematical Physics
Trends in Mathematics, 409–426
© 2009 Birkhäuser Verlag Basel/Switzerland

Resonance Dynamics and Decoherence

Marco Merkli

Abstract. We present a rigorous analysis of the phenomenon of decoherence for general N-level quantum systems coupled a reservoir modelled by a thermal bosonic quantum environment. We present an explicit form of the dominant reduced dynamics of open systems. We give explicit results for a spin 1/2 (qubit), including decoherence and thermalization times. Our approach is based on a dynamical theory of quantum resonances. It yields the exact reduced dynamics of the small system and does not involve master equation or van Hove limit approximations. This approach is suitable for a wide variety of systems which are not explicitly solvable, including systems of interacting spins (registers of interacting qubits), for which the coupling between the system and the environment is fixed but small.

Mathematics Subject Classification (2000). Primary 82C10; Secondary 37L05.

Keywords. Decoherence, open quantum systems, reduced density matrix, thermal environment, spins, qubits, classical-quantum transition, non-equilibrium dynamics, quantum resonances, spectral deformation.

1. Introduction

We consider an open quantum system S + R, where S is a "system of interest" and R is an "environment" or "reservoir". Typically, S is a system under examination in a laboratory, like an atom, a molecule or an aggregate of spins. It is not possible in reality to isolate any physical system entirely from its surroundings R, and only if we take these surroundings into account can we consider the total system as being closed and evolving according to a Hamiltonian dynamics. The reservoir R is supposed to be a very *large* quantum system compared to S. An immediate question is how the reservoir influences the dynamics of the small system. Prime examples of effects S shows are *thermalization* and *decoherence*. The former means that, due to its interaction with R, the system S is driven to the equilibrium state at the temperature of R. The effect of decoherence is the subject of this contribution, and we will describe it in detail in the next section.

Typically, we may assume that we know the (relatively simple) *microscopic* structure of S, while our knowledge of R is limited to its *macroscopic* characterization. In other words, we shall assume the energy levels and corresponding states of S to be given to us, as well as the thermodynamic parameters of R (temperature, pressure, chemical potential. . .). We will always assume that the state space of S is finite-dimensional. Of course, in a true concrete analysis, we will also have to specify the reservoir R on a microscopic level. However, when studying the reduced dynamics of S, only the macroscopic properties of R will be left in the description. The details of the environment do not play any role. In order to avoid introducing the microscopic structure of the environment, often an effective dynamics of the small system S is introduced – however, any trustworthy effective dynamics has to be derived from a full microscopic model, to which certain reduction and/or approximation schemes are applied (e.g., Born- or Born-Markov approximations). Our approach is to start off with a fully microscopic model of S + R, to eliminate the degrees of freedom of R and to analyze the remaining reduced description of S. In this process, we do not employ any approximation, however, our results are perturbative in the strength of the coupling between S and R.

We will take reservoirs to be spatially infinitely extended quantum systems. This is not merely a mathematical convenience, but rather a physical necessity that is linked to the very phenomena we want to describe. Indeed, if we try to keep the reservoir very large but finite, then irreversible physical processes will not take place. One can understand this easily heuristically, since for finite systems, Hamiltonians have pure point spectrum, and so the dynamics will not drive the system to a final state. On the other hand, we may want to consider reservoirs which are just large, but maybe not infinitely extended, say an oven in a laboratory. The temporal behaviour of such systems is approximated by that of systems with infinitely extended reservoirs on time-scales which are large, but not too large, see, e.g., [5].

We focus in this paper on the phenomenon of decoherence. A definition of decoherence is the vanishing of off-diagonal matrix elements of the reduced density matrix of S. A state given by a diagonal density matrix is characterized by classical probabilities, in the sense that averages of observables are obtained by weighing averages in specific states with given probabilities. The quantum nature is contained in the off-diagonal reduced density matrix elements, which are responsible for interference effects typical for quantum mechanics [16]. In this sense, a decohering system undergoes a transition from quantum to classical behaviour.

2. Description of decoherence

The *pure states* of S + R are described by normalized vectors ψ in the Hilbert space $\mathfrak{H} = \mathfrak{H}_{\rm S} \otimes \mathfrak{H}_{\rm R}$. An observable A is a (self-adjoint) operator on $\mathfrak{H}$, its expectation value in the state ψ is $\langle A\rangle = \langle \psi, A\psi\rangle$. The dynamics is determined by the Hamiltonian (energy operator)

$$H = H_{\rm S} + H_{\rm R} + \lambda v,$$

where $H_{\rm S}$ and $H_{\rm R}$ are the Hamiltonians of S and R, $\lambda \in \mathbb{R}$ is a coupling constant and where v represents the interaction between S and R. The dynamical equation is the Schrödinger equation,

$$\mathrm{i}\hbar\partial_t\psi = H\psi.$$

We will set for convenience $\hbar = 1$, so that the state vector evolves as $\psi_t = \mathrm{e}^{-\mathrm{i}tH}\psi_0$. Not all states can be represented by a single vector ψ. *Mixed states* are determined by *density matrices* ρ on $\mathfrak{H}$. These are non-negative (self-adjoint) trace-class operators which are normalized as $\mathrm{Tr}\rho = 1$. The average of an observable A in the mixed state ρ is given by $\langle A\rangle = \mathrm{Tr}(\rho A)$. To any density matrix ρ, we can associate normalized vectors ψ_n and probabilities p_n, $n = 1, 2, \ldots$, s.t.

$$\rho = \sum_{n=1}^{\infty} p_n |\psi_n\rangle\langle\psi_n|, \tag{2.1}$$

where $|\psi_n\rangle\langle\psi_n|$ is the rank-one orthonormal projection onto $\mathbb{C}\psi_n$ (spectral decomposition of ρ). Since the evolution of ψ_n is given by $\mathrm{e}^{-\mathrm{i}tH}\psi_n$, it follows from (2.1) that the density matrix ρ evolves according to $\rho_t = e^{-\mathrm{i}tH}\rho_0\mathrm{e}^{\mathrm{i}tH}$.

Consider the system S to be finite-dimensional, $\mathfrak{H}_{\rm S} = \mathbb{C}^N$. An example of a mixed state of S is its equilibrium state at temperature $T = 1/\beta$, given by the density matrix $\rho_{{\rm S},\beta} \propto \mathrm{e}^{-\beta H_{\rm S}}$. It is readily seen that one cannot find any vector $\psi \in \mathbb{C}^N$ representing this state, i.e., having the property that $\langle\psi, A\psi\rangle$ equals $\mathrm{Tr}(\rho_{{\rm S},\beta}A)$ for all $A \in B(\mathfrak{H}_{\rm S})$. However, by enlarging the Hilbert space, such a vector can be found: view $\sqrt{\rho_{{\rm S},\beta}}$ as an element of the Hilbert space of Hilbert-Schmidt operators on $\mathfrak{H}_{\rm S}$ (this space is naturally isomorphic to $\mathfrak{H}_{\rm S} \otimes \mathfrak{H}_{\rm S}$). Then clearly $\langle A\rangle = \mathrm{Tr}(\rho_{{\rm S},\beta}A) = \big\langle\sqrt{\rho_{{\rm S},\beta}}, A\sqrt{\rho_{{\rm S},\beta}}\big\rangle_{\rm HS}$, where $\langle\kappa,\sigma\rangle_{\rm HS} = \mathrm{Tr}(\kappa^*\sigma)$ is the inner product of Hilbert-Schmidt operators.[1]

As mentioned in the introduction, the system R is infinitely extended in space. Consequently, even if it has a finite energy density, the total energy $H_{\rm R}$ is not well defined (is infinite). In fact, it is not even clear which Hilbert space can describe states of the infinitely extended system R. One constructs the system R via the thermodynamic limit. First, one takes a state $\rho_{{\rm R},\Lambda}$ of the reservoir constrained to a box $\Lambda \subset \mathbb{R}^3$, with fixed thermodynamic properties (such as temperature, density etc.). For each finite Λ, one knows the Hilbert space and the state. (For instance, a quantum gas in a box Λ is described by the Hilbert space $\oplus_{n\geq 0}L^2(\Lambda^n, \mathrm{d}^{3n}x)$ (Fock space), and since Λ is finite, the energy operator has discrete spectrum, so the Gibbs-state density matrix is well defined.) Then the size of the box is made larger and larger, $\Lambda \uparrow \mathbb{R}^3$. This defines averages of (localized) observables A in the infinitely extended state, $E(A) = \lim_{\Lambda\uparrow\mathbb{R}^3}\mathrm{Tr}(\rho_{{\rm R},\Lambda}A)$. One can now try to find a Hilbert space $\mathfrak{H}_{\rm R}$ and a normalized vector $\psi_{\rm R} \in \mathfrak{H}_{\rm R}$ such that $E(A) = \langle\psi_{\rm R}, A\psi_{\rm R}\rangle$. This is a difficult task in general, but explicit expressions for Hilbert spaces and vectors have been found in the important cases of infinitely extended ideal quantum

[1] This is a manifestation of a general fact: a state over a C^*-algebra can be represented by a vector state in a Hilbert space. This is the so-called Gelfand-Naimark-Segal representation [3].

gases in thermal equilibrium.[2] We understand that this construction has been carried out, and that the state of R is represented on the Hilbert space $\mathfrak{H}_{\rm R}$ by a vector (or a density matrix). We give a more detailed explanation of this procedure at the end of this section. For now we carry on with a more qualitative discussion.

Given the density matrix of the total system, ρ_t, how can we extract the dynamics of S? Define the *reduced density matrix* of S by

$$\overline{\rho}_t := {\rm Tr}_{\rm R}(\rho_t),$$

where the trace is taken over $\mathfrak{H}_{\rm R}$ only (partial trace). This is a density matrix on $\mathfrak{H}_{\rm S}$, and it satisfies

$${\rm Tr}_{\rm S}(\overline{\rho}_t A_{\rm S}) = {\rm Tr}_{\rm S+R}\big(\rho_t(A_{\rm S}\otimes \mathbb{1}_{\rm R})\big)$$

for all obvservables $A_{\rm S}\in B(\mathfrak{H}_{\rm S})$. The reduced density matrix contains all information to describe the evolution of expectation values of observables of S alone. The degrees of freedom of R and the effects of the interaction between S and R are encoded in $\overline{\rho}_t$, which acts on the Hilbert space of the system S only.

Let $\{\varphi_j\}_{j=1}^N$ be a fixed basis of $\mathfrak{H}_{\rm S}$ and denote the matrix elements of $\overline{\rho}_t$ as $[\overline{\rho}_t]_{m,n} := \langle\varphi_m, \overline{\rho}_t\varphi_n\rangle$. A definition of decoherence is the vanishing of off-diagonal reduced density matrix elements in the limit of large times,

$$\lim_{t\to\infty}[\overline{\rho}_t]_{m,n} = 0, \quad \forall m\neq n. \tag{2.2}$$

This is a *basis dependent* notion of disappearance of correlations,

$$\overline{\rho}_t = \sum_{m,n} c_{m,n}(t)|\varphi_m\rangle\langle\varphi_n| \longrightarrow \sum_m p_m(t)|\varphi_m\rangle\langle\varphi_m|, \tag{2.3}$$

as $t\to\infty$. Most often, the basis considered is the energy basis, consisting of eigenvectors of $H_{\rm S}$. A mixture of states φ_j of the form $\sum_{m,n} c_{m,n}|\varphi_m\rangle\langle\varphi_n|$ is called an incoherent mixture if all "off-diagonals" vanish, $c_{m,n}=0$ for $m\neq n$. Else it is called a coherent mixture of the φ_j. The process (2.3) is thus a transition of a coherent to an incoherent mixture. Hence the name decoherence.

2.1. An explicitly solvable model of decoherence

Consider S to be an N-level system, coupled to a reservoir R of thermal bosons at temperature $T=1/\beta$ through an *energy-conserving interaction* (see [15] for the qubit case, $N=2$, and [11] for general N).

The Hilbert space and Hamiltonian of S are given by $\mathfrak{H}_{\rm S}=\mathbb{C}^N$ and $H_{\rm S} = {\rm diag}(E_1,\ldots,E_N)$, respectively, and the interaction operator is $v = G\otimes\varphi(g)$, where $G={\rm diag}(\gamma_1,\ldots,\gamma_N)$. Here,

$$\varphi(g) = \frac{1}{\sqrt{2}}[a^*(g)+a(g)], \tag{2.4}$$

[2]For bosons, this is known as the Araki-Woods construction, for fermions it is the Araki-Wyss construction, [1, 2, 14].

where the $a^\#(g)$ are the usual bosonic creation and annihilation operators, smeared out with a *form factor* $g \in L^2(\mathbb{R}^3, \mathrm{d}^3k)$ (momentum space representation): $a(g) = \int_{\mathbb{R}^3} \overline{g(k)} a(k)\mathrm{d}^3k$, $a^*(g) = \int_{\mathbb{R}^3} g(k)a^*(k)\mathrm{d}^3k$, $[a(k), a^*(l)] = \delta(k-l)$.

Since $[H_\mathrm{S}, H] = [H_\mathrm{S}, H_\mathrm{S} + H_\mathrm{R} + v] = 0$ the energy of the small system is conserved. This model is exactly solvable. The solution is given by

$$[\overline{\rho}_t]_{m,n} = [\overline{\rho}_0]_{m,n} \mathrm{e}^{-\mathrm{i}t(E_m - E_n) + \mathrm{i}\alpha_{m,n}(t)}, \tag{2.5}$$

where

$$\alpha_{m,n}(t) = (\gamma_m^2 - \gamma_n^2)S(t) + \mathrm{i}(\gamma_m - \gamma_n)^2\Gamma(t) \tag{2.6}$$

$$\Gamma(t) = \int_{\mathbb{R}^3} |g(k)|^2 \coth(\beta|k|/2) \frac{\sin^2(|k|t/2)}{|k|^2} \mathrm{d}^3k \tag{2.7}$$

$$S(t) = \frac{1}{2}\int_{\mathbb{R}^3} |g(k)|^2 \frac{|k|t - \sin(|k|t)}{|k|^2} \mathrm{d}^3k. \tag{2.8}$$

The parameter β in the above expression for $\Gamma(t)$ is the inverse temperature of the reservoir. We immediately see that

1. The populations are constant, $[\overline{\rho}_t]_{m,m} = [\overline{\rho}_0]_{m,m}$ for all m and all t.
2. If $\gamma_m = \gamma_n$ for some $m \neq n$, then the corresponding off-diagonal matrix element does not decay (decoherence-free subspaces).
3. Full decoherence (2.2) occurs if and only if $\Gamma(t) \to \infty$ as $t \to \infty$. Whether this happens or not depends on the *infrared behaviour* (small k) of the form factor, as well as on the *space dimension*. Let the infrared behaviour be characterized by $g(k) \sim |k|^p$ as $|k| \sim 0$. We obtain in three space-dimensions

$$\lim_{t\to\infty} \frac{\alpha_{m,n}(t)}{t} = \frac{1}{2}(\gamma_m^2 - \gamma_n^2)\left\langle g, |k|^{-1}g\right\rangle + \mathrm{i}(\gamma_m - \gamma_n)^2 \begin{cases} 0 & \text{if } p > 0 \\ \text{const.} & \text{if } p = -1/2 \\ +\infty & \text{if } p < -1/2. \end{cases}$$

For $p = -1/2$ the off-diagonal matrix elements decay exponentially quickly, $|[\overline{\rho}_t]_{m,n}| \sim \mathrm{e}^{-\mathrm{const.}t(\gamma_m - \gamma_n)^2}$ and for $p < -1/2$ the decay is quicker. For $p > 0$ the function $h(k) := |g(k)|^2 \coth(\beta|k|/2)\,|k|^{-2}$ is integrable on $\mathbb{R}^3$. We write $\sin^2(|k|t/2) = \frac{1}{2}(1 - \cos(|k|t))$ and obtain from (2.7)

$$\Gamma(t) = \frac{1}{2}\int_{\mathbb{R}^3} \frac{|g(k)|^2}{|k|^2} \coth(\beta|k|/2)\mathrm{d}^3k - \mathrm{Re}\big(\widehat{F}(t)\big),$$

where $\widehat{F}$ is the Fourier transform of the function

$$F(r) = \frac{1}{4}\coth(\beta|r|/2) \int_{S^2} |g(|r|, \sigma)|^2 \mathrm{d}\sigma,$$

defined for and integrable on $r \in \mathbb{R}$ ($\mathrm{d}\sigma$ is the uniform measure on the sphere S^2). From the Riemann-Lebesgue lemma we know that $\lim_{t\to\infty} \widehat{F}(t) = 0$, so

$$\lim_{t\to\infty} \Gamma(t) = \frac{1}{2}\int_{\mathbb{R}^3} \frac{|g(k)|^2}{|k|^2} \coth(\beta|k|/2)\mathrm{d}^3k \neq 0, \quad \text{for } p > 0.$$

That is, for $p > 0$ we do not have full decoherence.

This is a non-demolition model (H_{S} conserved), in which processes of absorption and emission of quanta of the reservoir by the system S are suppressed. To enable such processes one needs interactions v which do not commute with H_{S}. In the latter case, one typically expects that *thermalization* takes place. The phenomenon of thermalization can be described as follows.

Let $\rho(\beta,\lambda)$ be the equilibrium state of the total system at temperature $T = 1/\beta$ (where λ is the coupling constant measuring the strength of interaction between S and R) and let $\rho_{t=0}$ be any initial density matrix (on $\mathfrak{H}$). Thermalization means that

$$\mathrm{Tr}_{\mathrm{S+R}}(\rho_t A) \longrightarrow \mathrm{Tr}_{\mathrm{S+R}}(\rho(\beta,\lambda)A), \quad \text{as } t\to\infty, \tag{2.9}$$

where A is any observable of the total system S + R. The convergence (2.9) implies that

$$\overline{\rho}_t \longrightarrow \overline{\rho}_\infty(\beta,\lambda) := \mathrm{Tr}_{\mathrm{R}}(\overline{\rho}(\beta,\lambda)),$$

as $t\to\infty$. An expansion of $\overline{\rho}_\infty(\beta,\lambda)$ in the coupling constant λ gives

$$\overline{\rho}_\infty(\beta,\lambda) = \overline{\rho}_\infty(\beta,0) + O(\lambda),$$

where $\overline{\rho}_\infty(\beta,0)$ is the *Gibbs state* of the system S. The Gibbs state (density matrix) is diagonal in the energy basis (diagonalizing H_{S}), however, the correction term $O(\lambda)$ is not, in general (see, e.g., [11] for explicit calculations for the qubit). This shows the following effect.

> *Even if* S *is initially in an incoherent superposition of energy eigenstates it will acquire some* "residual coherence" *of order* $O(\lambda)$ *in the process of thermalization. This leads us to defining decoherence in thermalizing systems as being the decay of off-diagonals of* $\overline{\rho}_t$ *to their* (*non-zero*) *limit values, i.e., to the corresponding off-diagonals of* $\overline{\rho}_\infty(\beta,\lambda)$.

In examining the vast literature on this topic (some references are [7, 15, 16, 17]) we have only encountered either models with energy-conserving interactions (which are explicitly solvable), or models with Markovian approximations with uncontrolled errors (master equations, Lindblad dynamics). The goal of our work is to describe decoherence for systems which may also exhibit thermalization, in a rigorous fashion (controlled perturbation expansion).

2.2. Description of the infinitely extended reservoir R

Before taking the thermodynamic limit, as outlined above, the reservoir confined to a box Λ is described by the bosonic Fock space

$$\mathfrak{H}_{\mathrm{R},\Lambda} = \bigoplus_{n\geq 0} L^2_{\mathrm{sym}}(\Lambda^n, \mathrm{d}^{3n}x), \tag{2.10}$$

where the subindex "sym" means that we take symmetric square-integrable functions only (indistinguishable Bose particles). The Hamiltonian is that of non-interacting particles, given by $H_{\mathrm{R},\Lambda} = \oplus_{n\geq 0} H^{(n)}_{\mathrm{R},\Lambda}$, with $H^{(n)}_{\mathrm{R},\Lambda} = \sum_{j=1}^n \sqrt{-\partial^2_{x_j}}$ (with periodic boundary conditions). The density matrix $\rho_{\mathrm{R},\beta,\Lambda} = Z^{-1}_{\mathrm{R},\beta,\Lambda}\mathrm{e}^{-\beta H_{\mathrm{R},\Lambda}}$ is a well-defined trace-class operator on the space (2.10), and the normalization factor $Z_{\mathrm{R},\beta,\Lambda}$ is chosen so that $\mathrm{Tr}(\rho_{\mathrm{R},\beta,\Lambda}) = 1$. One calculates [1, 14] $E_\beta(a^\#(f)) := \lim_{\Lambda\uparrow\mathbb{R}^3} \mathrm{Tr}\big(\rho_{\mathrm{R},\beta,\Lambda} a^\#(f)\big) = 0$, where $a^\#$ stands for either a or a^*, and

$$E_\beta(a^*(f)a(g)) := \lim_{\Lambda\uparrow\mathbb{R}^3} \mathrm{Tr}\big(\rho_{\mathrm{R},\beta,\Lambda} a^*(f)a(g)\big) = \left\langle g, \frac{1}{\mathrm{e}^{\beta|k|}-1} f \right\rangle, \tag{2.11}$$

where the square-integrable f, g are represented in Fourier transform in the inner product on the right-hand side.[3] All products of creation and annihilation operators can be calculated using the Wick theorem [3], so (2.11) (plus the vanishing of averages of $a(f)$ and $a^*(f)$) determines the infinitely extended thermal state E_β of R completely. We consider here only reservoir equilibrium states below critical density, i.e., in absence of Bose-Einstein condensate.

The Araki-Woods Hilbert space representation is given by

$$\mathfrak{H}_\mathrm{R} = \mathcal{F} \otimes \mathcal{F}, \tag{2.12}$$

where $\mathcal{F} = \oplus_{n\geq 0} L^2_{\mathrm{sym}}(\mathbb{R}^{3n}, \mathrm{d}^{3n}x)$,

$$\psi_\mathrm{R} = \Omega \otimes \Omega, \tag{2.13}$$

the product of the Fock vacua in $\mathcal{F}$, and

$$a^*_\beta(g) = a^*\left(\sqrt{\frac{e^{\beta|k|}}{\mathrm{e}^{\beta|k|}-1}}\, g\right) \otimes \mathbb{1} + \mathbb{1} \otimes a\left(\sqrt{\frac{1}{\mathrm{e}^{\beta|k|}-1}}\, \overline{g}\right),$$

where $\overline{g}$ is the complex conjugate of g, and where the $a^\#$ are the ordinary Fock creation and annihilation operators on $\mathcal{F}$. We also set $a_\beta(g) := [a^*_\beta(g)]^*$. It is easy to check that

$$E_\beta(a^*(f)a(g)) = \langle \psi_\mathrm{R}, a^*_\beta(f) a_\beta(g) \psi_\mathrm{R} \rangle .$$

This last equation shows us that we have successfully represented the thermal state of the infinitely extended R as a vector state on a concrete Hilbert space.

3. Dynamical resonance theory: Results

Let S be an N-level system, $\mathfrak{H}_\mathrm{S} = \mathbb{C}^N$, with energies $\{E_j\}_{j=1}^N$, and let R be the free massless Bose field, spatially infinitely extended in $\mathbb{R}^3$ in equilibrium at temperature $T = 1/\beta$, as described at the end of Section 2.

The interaction operator is obtained by taking the thermodynamic limit of $\lambda v_\Lambda = \lambda G \otimes \varphi(g_\Lambda)$. Here, $G = G^* \in B(\mathfrak{H}_\mathrm{S})$ is a self-adjoint $N \times N$ matrix and $\varphi(g_\Lambda)$ is the smoothed-out field operator (2.4) acting on $\mathfrak{H}_{\mathrm{R},\Lambda}$, (2.10), and $g_\Lambda(x) = \chi_\Lambda(x) g(x)$ is the function g, cut off by being set equal to zero outside Λ. It is customary to abbreviate this description by simply writing

$$v = \lambda G \otimes \varphi(g), \tag{3.1}$$

and the thermodynamic limit is understood to be taken automatically.

[3] Of course, one has to restrict this to functions for which the r.h.s. is well defined.

We denote the average of observables $A \in B(\mathfrak{H}_{\rm S})$ at time t by

$$\langle A\rangle_t := \mathrm{Tr}_{\rm S}(\overline{\rho}_t A), \tag{3.2}$$

and the ergodic average is denoted by

$$\langle\!\langle A\rangle\!\rangle_\infty := \lim_{T\to\infty}\frac{1}{T}\int_0^T \langle A\rangle_t \mathrm{d}t.$$

Our approach is based on a dynamical resonance theory, where resonances are treated in a setting of *spectral deformation* (see Section 4). This leads to the following regularity requirement which we assume to be fulfilled throughout this paper.

(A) The function

$$g_\beta(u,\sigma) := \sqrt{\frac{u}{1-\mathrm{e}^{-\beta u}}}\,|u|^{1/2}\begin{cases} g(u,\sigma) & \text{if } u\geq 0\\ \mathrm{e}^{\mathrm{i}\phi}\overline{g}(-u,\sigma) & \text{if } u<0\end{cases}$$

is such that $\vartheta\mapsto g_\beta(u+\vartheta,\sigma)$ has an analytic continuation, as a map $\mathbb{C}\to L^2(\mathbb{R}\times S^2, \mathrm{d}u\times\mathrm{d}\sigma)$, into $\{|\vartheta|<\tau\}$, for some $\tau>0$. Here, ϕ is an arbitrary fixed phase. (See [6] for the usefulness and physical interpretation of this phase.)

Examples of admissible g are $g(k) = g_1(\sigma)|k|^p\mathrm{e}^{-|k|^2}$, where $p=-1/2+n$, $n = 0,1,2,\ldots$, and $g_1(\sigma)=\mathrm{e}^{\mathrm{i}\phi}\overline{g}_1(\sigma)$. They include the physically most important cases, see also [15]. We point out that it is possible to weaken condition (A) considerably, at the expense of a mathematically more involved treatment, as mentioned in [11, 13]. The following result is the main result of [11]. We give an outline of the proof in Section 4.

Theorem 3.1 (Evolution of observables [11]**).** *There is a $\lambda_0>0$ s.t. the following statements hold for $|\lambda|<\lambda_0$, $t\geq 0$, and $A\in B(\mathfrak{H}_{\rm S})$.*

1. $\langle\!\langle A\rangle\!\rangle_\infty$ *exists.*
2. *We have*

$$\langle A\rangle_t - \langle\!\langle A\rangle\!\rangle_\infty = \sum_{\varepsilon\neq 0}\mathrm{e}^{\mathrm{i}t\varepsilon}R_\varepsilon(A) + O(\lambda^2\mathrm{e}^{-\tau t}), \tag{3.3}$$

where the ε are "resonance energies", $0\leq \mathrm{Im}\varepsilon<\tau/2$, *and $R_\varepsilon(A)$ are linear functionals of A which depend on the initial state $\rho_{t=0}$.*

3. *Let e be an eigenvalue of the operator $H_{\rm S}\otimes\mathbb{1}_{\rm S}-\mathbb{1}_{\rm S}\otimes H_{\rm S}$ (acting on $\mathfrak{H}_{\rm S}\otimes\mathfrak{H}_{\rm S}$). For $\lambda=0$ each ε coincides with one of the e and we have the following expansion for small λ*

$$\varepsilon\equiv\varepsilon_e^{(s)} = e-\lambda^2\delta_e^{(s)}+O(\lambda^4).$$

The $\delta_e^{(s)}$ satisfy $\mathrm{Im}\,\delta_e^{(s)}\leq 0$. They are eigenvalues of so-called level shift operators Λ_e, *and $s=1,\ldots,\nu(s)\leq$* mult (e) *labels the eigenvalue splitting. Furthermore, we have*

$$R_\varepsilon(A) = \sum_{(m,n)\in I_e}\kappa_{m,n}A_{m,n}+O(\lambda^2), \tag{3.4}$$

with $I_e = \{(m,n) \mid E_m - E_n = e\}$*, and where* $A_{m,n}$ *is the* (m,n)*-matrix element of* A *and the numbers* $\kappa_{m,n}$ *depend on the initial state.*

Discussion. Relation (3.3) gives a detailed picture of the dynamics of averages of observables. The resonance energies ε and the functionals R_ε can be calculated for concrete models, to arbitrary precision (in the sense of rigorous perturbation theory in λ). See Section 3.1 for explicit expressions for the qubit. In the absence of interaction ($\lambda = 0$) we have $\varepsilon = e \in \mathbb{R}$. Depending on the interaction each resonance energy ε may migrate into the upper complex plane, or it may stay on the real axis, as $\lambda \neq 0$. The averages $\langle A\rangle_t$ approach their ergodic means $\langle\!\langle A\rangle\!\rangle_\infty$ if and only if $\mathrm{Im}\varepsilon > 0$ for all $\varepsilon \neq 0$. In this case the convergence takes place on the time scale $[\mathrm{Im}\varepsilon]^{-1}$. Otherwise $\langle A\rangle_t$ oscillates. A sufficient condition for decay is that $\mathrm{Im}\delta_e^{(s)} < 0$ (and λ small).

There are two kinds of processes which drive the decay: energy-exchange processes and energy preserving ones. The former are induced by interactions enabling processes of absorption and emission of field quanta with energies corresponding to the Bohr frequencies of S (this is the "Fermi Golden Rule Condition"). Energy preserving interactions suppress such processes, allowing only for a phase change of the system during the evolution ("phase damping").

Even if the initial density matrix, $\rho_{t=0}$, is a product of the system and reservoir density matrices, the density matrix ρ_t at any subsequent moment of time $t > 0$ is *not* of product form. The evolution creates entanglement between the system and reservoir. Our technique does not require $\rho_{t=0}$ to be a product state [11].

Our next goal is to use Theorem 3.1 to describe in detail the decay of reduced density matrix elements. According to Theorem 3.1 the dynamics is governed by the resonance energies $\varepsilon_e^{(s)}$ whose lowest-order contributions $\delta_e^{(s)}$ are eigenvalues of level shift operators Λ_e. In what follows we assume that all eigenvalues $\delta_e^{(s)}$ are simple. We denote the corresponding eigenvector by $\eta_e^{(s)}$, and the eigenvector associated to the adjoint operator Λ_e^* with eigenvalue $\overline{\delta_e^{(s)}}$ is denoted by $\widetilde{\eta}_e^{(s)}$. They are normalized as $\left\langle \eta_e^{(s)}, \widetilde{\eta}_e^{(s)} \right\rangle = 1$. The assumption of simplicity of the spectrum of Λ_e is not necessary at all for our method, it is simply made to make the exposition somewhat simpler. Let $\{\varphi_n\}$ be an orthornomal basis of $\mathbb{C}^N$ diagonalizing the Hamiltonian of S, $H_{\mathrm{S}}\varphi_n = E_n\varphi_n$. The matrix element $[\overline{\rho}_t]_{m,n}$ is obtained by choosing the observable $A = |\varphi_n\rangle\langle\varphi_m|$ in (3.2). We denote the difference of two eigenvalues of H_{S} by $E_{m,n} = E_m - E_n$. A closer analysis of the functionals R_ε yields the following result, the proof of which we give in Section 5.

Theorem 3.2 (Dominant dynamics). *There is a constant* λ_1 *s.t. if* $0 < |\lambda| < \lambda_1$*, then for all* m, n *and all* $t \geq 0$

$$
\begin{aligned}
&[\overline{\rho}_t]_{m,n} - \langle\!\langle |\varphi_n\rangle\langle\varphi_m| \rangle\!\rangle_\infty \\
&= \sum_{\{s:\varepsilon_{E_{n,m}}^{(s)} \neq 0\}} \mathrm{e}^{\mathrm{i}t\varepsilon_{E_{n,m}}^{(s)}} \sum_{\{k,l:E_{l,k}=E_{n,m}\}} \sigma_{m,n;k,l}^{(s)}\ [\overline{\rho}_0]_{k,l} + O(\lambda^2 \mathrm{e}^{-t\gamma_{m,n}}),
\end{aligned}
\tag{3.5}
$$

where

$$\gamma_{m,n} = \min\{\mathrm{Im}\varepsilon_e^{(s)} : \varepsilon_e^{(s)} \neq 0 \text{ and } e \neq E_{n,m}\}.$$

The mixing constants $\sigma^{(s)}_{m,n;k,l}$ are given by

$$\sigma^{(s)}_{m,n;k,l} = \left\langle \widetilde{\eta}^{(s)}_{E_{n,m}}, \varphi_n \otimes \varphi_m \right\rangle \left\langle \varphi_l \otimes \varphi_k, \eta^{(s)}_{E_{n,m}} \right\rangle,$$

$\eta^{(s)}_{E_{n,m}}$ *and* $\widetilde{\eta}^{(s)}_{E_{n,m}}$ *being the resonance eigenvectors introduced above.*

Discussion. The group of matrix elements $[\overline{\rho}_t]_{m,n}$ associated to the same energy difference $e = E_n - E_m$ evolve in a coupled way, while groups belonging to different e evolve independently, in the regime of Theorem 3.2. It is clear that the eigenvalue $e = 0$ is always degenerate ($\varphi_k \otimes \varphi_k$ is always an associated eigenvector, for all k). One easily sees that if $e = E_n - E_m$ is simple then $\sigma^{(s)}_{m,n;k,l}$ vanishes unless $(k,l) = (m,n)$, in which case $\sigma^{(s)}_{m,n;k,l} = 1$ equals one (this follows simply from the fact that η_e and $\widetilde{\eta}_e$ belong to the spectral subspace associated to e). The main term of the r.h.s. of (3.5) is then simply $\mathrm{e}^{\mathrm{i}t\varepsilon_{E_{n,m}}}$, so

$$[\overline{\rho}_t]_{m,n} - \langle\!\langle |\varphi_n\rangle\langle\varphi_m| \rangle\!\rangle_\infty = \mathrm{e}^{\mathrm{i}t\varepsilon_{E_{n,m}}} [\overline{\rho}_0]_{m,n} + O(\lambda^2 \mathrm{e}^{-t\gamma_{m,n}}).$$

The usefulness of Theorem 3.2 is that it relates $[\overline{\rho}_t]_{m,n}$ to the initial conditions $[\overline{\rho}_0]_{k,l}$. We can understand how to arrive at Theorem 3.2 from Theorem 3.1 in the following way. The expansion (3.4) implies that the main term of $R_\varepsilon(|\varphi_n\rangle\langle\varphi_m|)$ is non-vanishing only if ε bifurcates out of $e = E_n - E_m$. This means that all contributions to the sum in (3.3) with ε not bifurcating out of $E_n - E_m$ are of order λ^2, and decaying according to $\mathrm{e}^{\mathrm{i}t\varepsilon}$. These terms, plus the $O(\lambda^2\mathrm{e}^{-\tau t})$ term in (3.3), constitute the remainder term in (3.5).

The constants $\gamma_{m,n}$ are typically of order λ^2 (they may be of higher order if the so-called Fermi Golden Rule condition for efficient coupling is not satisfied [9]). Expansion (3.5) is thus useful in the regime

$$\lambda^2 \mathrm{e}^{-t\lambda^2 \min\{\mathrm{Im}\,\delta_e^{(s)} : e \neq E_{n,m}\}} \ll e^{-t\lambda^2 \max\{\mathrm{Im}\,\delta^{(s)}_{E_{n,m}} : s=1,\ldots,\nu(E_{n,m})\}}.$$

In other words, given any finite maximal time of interest $t_{\max}$, there is a λ_1 s.t. if $0 < |\lambda| < \lambda_1$, expansion (3.5) is valid, and the remainder term is negligible for all $0 \leq t \leq t_{\max}$. The expansion (3.5) thus isolates the dominant dynamics.

3.1. Application: thermalization versus decoherence time for a qubit

A qubit, or spin 1/2, is described by the Hilbert-space of pure states $\mathbb{C}^2$. The Hamiltonian is $H_{\mathrm{S}} = \mathrm{diag}(E_1, E_2)$ (in the canonical basis of $\mathbb{C}^2$). We set $\Delta = E_2 - E_1 > 0$. The coupling operator is given by the self-adjoint operator

$$v = \begin{bmatrix} a & c \\ \overline{c} & b \end{bmatrix} \otimes \varphi(g),$$

where $\varphi(g)$ is given in (2.4). The operator $H_{\rm S}\otimes\mathbb{1}-\mathbb{1}\otimes H_{\rm S}$ has four eigenvalues, $e\in\{-\Delta,0,0,\Delta\}$. One calculates the resonance energies associated to these e to be (see also Theorem 3.1 and the next subsection)

$$\begin{aligned}\varepsilon_0^{(1)}(\lambda)&=0\\ \varepsilon_0^{(2)}(\lambda)&=\mathrm{i}\lambda^2|c|^2\xi(\Delta)+O(\lambda^4)\\ \varepsilon_\Delta(\lambda)&=\Delta+\lambda^2R+\frac{\mathrm{i}}{2}\lambda^2\left[|c|^2\xi(\Delta)+(b-a)^2\xi(0)\right]+O(\lambda^4)\\ \varepsilon_{-\Delta}(\lambda)&=-\overline{\varepsilon_\Delta(\lambda)}\end{aligned}$$

where we have set

$$\xi(\eta):=\pi\int_{\mathbb{R}^3}\coth\left(\frac{\beta|k|}{2}\right)|g(k)|^2\delta(\eta-|k|)\mathrm{d}^3k$$

and (P.V. denoting the principal value)

$$R=\frac{b^2-a^2}{2}\left\langle g,|k|^{-1}g\right\rangle+\frac{|c|^2}{2}\mathrm{P.V.}\int_{\mathbb{R}\times S^2}u^2\coth\left(\frac{\beta|k|}{2}\right)\frac{|g(|u|,\sigma)|^2}{u-\Delta}\mathrm{d}u\,\mathrm{d}\sigma. \quad (3.6)$$

The corresponding resonance eigenvectors (defined before Theorem 3.2) are as follows, where $\{\varphi_1,\varphi_2\}$ is the canonical orthonormal basis of $\mathbb{C}^2$, and where $\varphi_{i,j}=\varphi_i\otimes\varphi_j$:

$$\begin{aligned}\eta_0^{(1)}&=\varphi_{1,1}+\varphi_{2,2}, &\widetilde{\eta}_0^{(1)}&=\frac{1}{1+\mathrm{e}^{-\beta\Delta}}[\varphi_{1,1}+\mathrm{e}^{-\beta\Delta}\varphi_{2,2}],\\ \eta_0^{(2)}&=\varphi_{1,1}-\mathrm{e}^{\beta\Delta}\varphi_{2,2}, &\widetilde{\eta}_0^{(2)}&=\frac{1}{1+\mathrm{e}^{\beta\Delta}}[\varphi_{1,1}-\varphi_{2,2}],\end{aligned}$$

and $\eta_\Delta=\widetilde{\eta}_\Delta=\varphi_{2,1}$, $\eta_{-\Delta}=\widetilde{\eta}_\Delta=\varphi_{1,2}$. Note that $\eta_0^{(1)}$ is just the (not normalized) trace state on S. The mixing constants $\sigma^{(s)}_{m,n;k,l}$ (see Theorem 3.2) are thus

$$\begin{aligned}\sigma^{(1)}_{1,2;1,2}=\sigma^{(1)}_{2,1;2,1}&=1,\\ \sigma^{(2)}_{1,1;1,1}&=\left\langle\eta_0^{(2)},\varphi_{1,1}\right\rangle\left\langle\varphi_{1,1},\eta_0^{(2)}\right\rangle=\frac{1}{1+\mathrm{e}^{\beta\Delta}},\\ \sigma^{(2)}_{1,1;2,2}=\sigma^{(2)}_{2,2;1,1}&=\left\langle\eta_0^{(2)},\varphi_{1,1}\right\rangle\left\langle\varphi_{2,2},\eta_0^{(2)}\right\rangle=\frac{-1}{1+\mathrm{e}^{-\beta\Delta}},\\ \sigma^{(2)}_{2,2;2,2}&=\left\langle\eta_0^{(2)},\varphi_{2,2}\right\rangle\left\langle\varphi_{2,2},\eta_0^{(2)}\right\rangle=\frac{1}{1+\mathrm{e}^{-\beta\Delta}}.\end{aligned}$$

We shall assume that the Fermi Golden Rule is satisfied: $\xi(\Delta)\neq 0$. Then zero is a simple resonance eigenvalue, $\varepsilon_0^{(1)}=0$, and consequently, for $e=0$ the term $s=1$ is not present in the sum (3.5). Theorem 3.1 thus gives the following dominant dynamics:

$$[\overline{\rho}_t]_{1,1}-\langle\!\langle|\varphi_1\rangle\langle\varphi_1|\rangle\!\rangle_\infty\sim\mathrm{e}^{\mathrm{i}t\varepsilon_0^{(2)}(\lambda)}\left\{\frac{[\overline{\rho}_0]_{1,1}}{1+\mathrm{e}^{\beta\Delta}}-\frac{[\overline{\rho}_0]_{2,2}}{1+\mathrm{e}^{-\beta\Delta}}\right\}, \quad (3.7)$$

$$[\overline{\rho}_t]_{1,2}-\langle\!\langle|\varphi_2\rangle\langle\varphi_1|\rangle\!\rangle_\infty\sim\mathrm{e}^{\mathrm{i}t\varepsilon_\Delta(\lambda)}[\overline{\rho}_0]_{1,2}. \quad (3.8)$$

The dynamics for $[\overline{\rho}_t]_{2,2}$ and $[\overline{\rho}_t]_{2,1}$ are easily obtained also directly from Theorem 3.1, or by using that $[\overline{\rho}_t]_{1,1}+[\overline{\rho}_t]_{2,2}=1$ (since $\mathrm{Tr}\,\overline{\rho}_t=1$) and the fact that $\overline{\rho}_t$ is self-adjoint. We point out that since the system S+R approaches its (joint) equilibrium as $t\to\infty$, we have $\langle\!\langle|\varphi_1\rangle\langle\varphi_1|\rangle\!\rangle_\infty=\frac{\mathrm{e}^{\beta\Delta}}{1+\mathrm{e}^{\beta\Delta}}+O(\lambda^2)$ and $\langle\!\langle|\varphi_2\rangle\langle\varphi_1|\rangle\!\rangle_\infty=O(\lambda^2)$ (Gibbs distribution). This law can also be recovered by setting $t=0$ in (3.7), (3.8) and using that $[\overline{\rho}_0]_{2,2}=1-[\overline{\rho}_0]_{1,1}$.

The *thermalization time* (decay of diagonals) is $\tau_{\rm th}:=[\mathrm{Im}\varepsilon_0^{(2)}(\lambda)]^{-1}$, and the *decoherence time* (decay of off-diagonals) is $\tau_{\rm dec}:=[\mathrm{Im}\varepsilon_\Delta(\lambda)]^{-1}$. Their ratio is

$$\frac{\tau_{\rm th}}{\tau_{\rm dec}}=\frac{1}{2}\left[1+\frac{(b-a)^2}{|c|^2}\frac{\xi(0)}{\xi(\Delta)}\right]+O(\lambda^2).$$

Note that we have $\xi(0)>0$ for infra-red behaviour $g(k)\sim|k|^{-1/2}$ as $|k|\sim 0$ and $\xi(0)=0$ for more regular infra-red behaviour. Moreover, $\xi(0)\sim T$ and $\xi(\Delta)\sim$ const. >0, as the temperature $T\sim 0$.

Spin-Boson model. The Hamiltonian of S is given by [4, 8]

$$H_{\rm S}=-\frac{1}{2}\hbar\Delta_0\sigma_x+\frac{1}{2}\epsilon\sigma_z,$$

where the σ are Pauli matrices, Δ_0 is the bare tunnelling matrix element, and ϵ is the bias. The coupling operator is

$$v=\sigma_z\otimes\varphi(g).$$

This determines the matrix elements a,b,c in the general formulation, and we obtain

$$\frac{(b-a)^2}{|c|^2}=16\frac{\epsilon^2}{\hbar^2\Delta_0^2}\ .$$

This shows for instance that the thermalization time will become smaller relative to the decoherence time if the bias ϵ is decreased, or if the tunnelling parameter Δ_0 is increased.

Explicit form of the level shift operators. For the sake of completeness, we include the explicit form of the level shift operators Λ_e, $e=0,\pm\Delta$. By definition,

$$\Lambda_e=P_eI\overline{P}_e(\overline{L}_0-e+\mathrm{i}0)^{-1}\overline{P}_eIP_e,$$

where P_e is the spectral projection onto the eigenspace of $H_{\rm S}\otimes\mathbb{1}-\mathbb{1}\otimes H_{\rm S}$ associated to the eigenvalue e, $\overline{P}_e=\mathbb{1}-P_e$, $\overline{L}_0$ is the operator L_0 restricted to $\mathrm{Ran}\,\overline{P}_e$, and where I is the interaction operator, see [11] and Section 4. The explicit form of Λ_e has been calculated in [11] for a general N-level system coupled to the thermal Bose environment (Proposition 5.1 of [11]).[4] The explicit form of Λ_0, expressed in

[4]In the present work, we take the generator of dynamics to be the Liouville operator associated to the reference state $\psi_0=\psi_{{\rm S},\infty}\otimes\psi_{\rm R}$, see Section 4. In [11] the Liouville operator is taken with respect to the reference vector $\psi_{{\rm S},\beta}\otimes\psi_{\rm R}$. Those two choices are related by a simple transformation, and all physical results are independent of the particular choice of reference state.

the basis $\{\varphi_1 \otimes \varphi_1, \varphi_2 \otimes \varphi_2\}$ of Ran P_0 is

$$\Lambda_0 = \frac{\mathrm{i}}{2} \frac{|c|^2 \xi(\Delta)}{\mathrm{e}^{\beta\Delta} - 1} \begin{bmatrix} 1 & -1 \\ -e^{\beta\Delta} & \mathrm{e}^{\beta\Delta} \end{bmatrix}.$$

The dimension of Ran $P_{\pm\Delta}$ is one, so $\Lambda_{\pm\Delta}$ reduces simply to a number,

$$\Lambda_{\pm\Delta} = R \pm \frac{\mathrm{i}}{2}[|c|^2 \xi(\Delta) + (b-a)^2 \xi(0)],$$

where R is given in (3.6). Knowing the explicit form of the level shift operators, the expansions of the resonance energies and resonance eigenvectors are now easy to obtain.

4. Outline of resonance approach

Consider any observable $A \in B(\mathfrak{H}_{\mathrm{S}})$. We have

$$\begin{aligned} \langle A \rangle_t &= \mathrm{Tr}_{\mathrm{S}}\,[\overline{\rho}_t\ A] = \mathrm{Tr}_{\mathrm{S+R}}\,[\rho_t\ A \otimes \mathbb{1}_{\mathrm{R}}] \\ &= \left\langle \psi_0, \mathrm{e}^{\mathrm{i}tL_\lambda}\,[A \otimes \mathbb{1}_{\mathrm{S}} \otimes \mathbb{1}_{\mathrm{R}}]\,\mathrm{e}^{-\mathrm{i}tL_\lambda}\psi_0 \right\rangle . \end{aligned} \tag{4.1}$$

In the last step, we pass to the *representation Hilbert space* of the system (the GNS Hilbert space), where the initial density matrix is represented by the vector ψ_0 (in particular, the Hilbert space of the small system becomes $\mathfrak{H}_{\mathrm{S}} \otimes \mathfrak{H}_{\mathrm{S}}$), see also after equation (2.1) and Section 2.2. For this outline we take the initial state to be one represented by the product vector $\psi_0 = \psi_{\mathrm{S},\infty} \otimes \psi_{\mathrm{R}}$, where $\psi_{\mathrm{S},\infty}$ is the trace state of S, $\langle \psi_{\mathrm{S},\infty}, (A_{\mathrm{S}} \otimes \mathbb{1}_{\mathrm{S}})\psi_{\mathrm{S},\infty} \rangle = \frac{1}{N}\mathrm{Tr}\,(A_{\mathrm{S}})$, and where ψ_{R} is the equilibrium state of R at a fixed inverse temperature $0 < \beta < \infty$, (2.13). (This form of the initial state is is not necessary for our method to work, see [11].) The dynamics is implemented by the group of automorphisms $\mathrm{e}^{\mathrm{i}tL_\lambda} \cdot \mathrm{e}^{-\mathrm{i}tL_\lambda}$. The self-adjoint generator L_λ is called the *Liouville operator*. It is of the form $L_\lambda = L_0 + \lambda W$, where $L_0 = L_{\mathrm{S}} + L_{\mathrm{R}}$ represents the uncoupled Liouville operator, and λW is the interaction (represented in the GNS Hilbert space).

We borrow a trick from the analysis of open systems far from equilibrium: there is a (non-self-adjoint) generator K_λ s.t.

$$\begin{aligned} \mathrm{e}^{\mathrm{i}tL_\lambda} A \mathrm{e}^{-\mathrm{i}tL_\lambda} &= \mathrm{e}^{\mathrm{i}tK_\lambda} A \mathrm{e}^{-\mathrm{i}tK_\lambda} \quad \text{for all observables } A,\ t \geq 0, \text{ and} \\ K_\lambda \psi_0 &= 0. \end{aligned}$$

There is a standard way of constructing K_λ given L_λ and the reference vector ψ_0. K_λ is of the form $K_\lambda = L_0 + \lambda I$, where the interaction term undergoes a certain modification ($W \to I$), cf. [11]. As a consequence, formally, we may replace the propagators in (4.1) by those involving K. The resulting propagator which is directly applied to ψ_0 will then just disappear due to the invariance of ψ_0. One can carry out this procedure in a rigorous manner, obtaining the following resolvent representation [11]

$$\langle A \rangle_t = -\frac{1}{2\pi \mathrm{i}} \int_{\mathbb{R}-\mathrm{i}} \left\langle \psi_0, (K_\lambda(\omega) - z)^{-1}\,[A \otimes \mathbb{1}_{\mathrm{S}} \otimes \mathbb{1}_{\mathrm{R}}]\,\psi_0 \right\rangle \mathrm{e}^{\mathrm{i}tz} \mathrm{d}z, \tag{4.2}$$

where $K_\lambda(\omega) = L_0(\omega) + \lambda I(\omega)$, I is representing the interaction, and $\omega \mapsto K_\lambda(\omega)$ is a spectral deformation (translation) of K_λ. The latter is constructed as follows. There is a deformation transformation $U(\omega) = \mathrm{e}^{-\mathrm{i}\omega D}$, where D is the (explicit) self-adjoint generator of translations [11, 10] transforming the operator K_λ as

$$K_\lambda(\omega) = U(\omega) K_\lambda U(\omega)^{-1} = L_0 + \omega N + \lambda I(\omega). \tag{4.3}$$

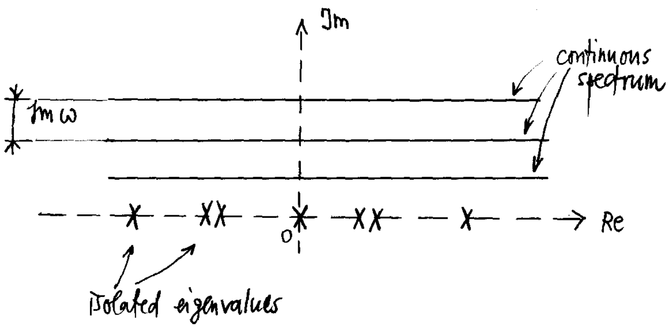


FIGURE 1. Spectrum of $K_0(\omega)$

Here, $N = N_1 \otimes \mathbb{1} + \mathbb{1} \otimes N_1$ is the total number operator of $\mathfrak{H}_{\mathrm{R}}$, (2.12), and where N_1 is the usual number operator on $\mathcal{F}$. N has spectrum $\mathbb{N} \cup \{0\}$, where 0 is a simple eigenvalue (vacuum eigenvector ψ_{R}). For real values of ω, $U(\omega)$ is a group of unitaries. The spectrum of $K_\lambda(\omega)$ depends on $\operatorname{Im}\omega$ and moves according to the value of $\operatorname{Im}\omega$, whence the name "spectral deformation". Even though $U(\omega)$ becomes unbounded for complex ω, the r.h.s. of (4.3) is a well-defined closed operator on a dense domain, analytic in ω at zero. Analyticity is used in the derivation of (4.2) and this is where the analyticity condition (A) before Theorem 3.1 comes into play. The operator $I(\omega)$ is infinitesimally small with respect to the number operator N. Hence we use perturbation theory in λ to examine the spectrum of $K_\lambda(\omega)$.

The point of the spectral deformation is that the (important part of the) spectrum of $K_\lambda(\omega)$ is much easier to analyze than that of K_λ, because the deformation uncovers the resonances of K_λ. We have (see Figure 1)

$$\operatorname{spec}\big(K_0(\omega)\big) = \{E_i - E_j\}_{i,j=1,\dots,N} \bigcup_{n\geq 1} \{\omega n + \mathbb{R}\},$$

because $K_0(\omega) = L_0 + \omega N$, L_0 and N commute, and the eigenvectors of $L_0 = L_{\mathrm{S}} + L_{\mathrm{R}}$ are $\varphi_i \otimes \varphi_j \otimes \psi_{\mathrm{R}}$. The continuous spectrum is bounded away from the isolated eigenvalues by a gap of size $\operatorname{Im}\omega$. For values of the coupling parameter λ small compared to $\operatorname{Im}\omega$, we can follow the displacements of the eigenvalues by using analytic perturbation theory. (Note that for $\operatorname{Im}\omega = 0$, the eigenvalues are imbedded into the continuous spectrum, and analytic perturbation theory is not valid! The spectral deformation is indeed very useful!)

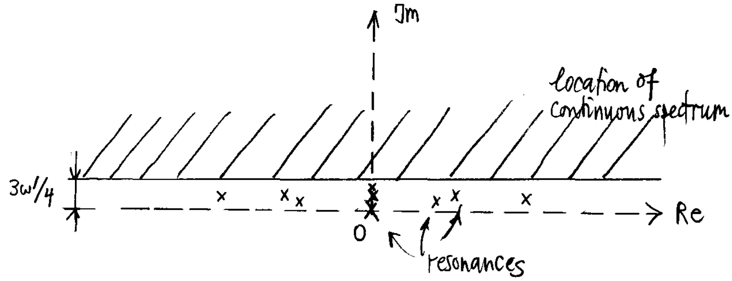


FIGURE 2. Spectrum of $K_\lambda(\omega)$. Resonances $\varepsilon_e^{(s)}$ are uncovered.

Theorem 4.1. *(See Fig. 2.) Fix* $\operatorname{Im}\omega$ *s.t.* $0 < \operatorname{Im}\omega < \tau$ *(where* τ *is as in Condition* (A) *given after* (3.2)*). There is a constant* $c_0 > 0$ *s.t. if* $|\lambda| \le c_0/\beta$ *then, for all* ω *with* $\operatorname{Im}\omega > \omega'$*, the spectrum of* $K_\lambda(\omega)$ *in the complex half-plane* $\{\operatorname{Im} z < \omega'/2\}$ *is independent of* ω *and consists purely of the distinct eigenvalues*

$$\{\varepsilon_e^{(s)} \ : \ e \in \operatorname{spec}(L_{\mathrm{S}}), s = 1, \ldots, \nu(e)\},$$

where $1 \le \nu(e) \le \operatorname{mult}(e)$ *counts the splitting of the eigenvalue* e*. Moreover,* $\lim_{\lambda\to 0} |\varepsilon_e^{(s)}(\lambda) - e| = 0$ *for all* s*, and we have* $\operatorname{Im}\varepsilon_e^{(s)} \ge 0$*. Also, the continuous spectrum of* $K_\lambda(\omega)$ *lies in the region* $\{\operatorname{Im} z \ge 3\omega'/4\}$*.*

Next we separate the contributions to the path integral in (4.2) coming from the singularities at the resonance energies and from the continuous spectrum. We deform the path of integration $z = \mathbb{R} - \mathrm{i}$ into the line $z = \mathbb{R} + \mathrm{i}\omega'/2$, thereby picking up the residues of poles of the integrand at $\varepsilon_e^{(s)}$ (all e, s). Let $\mathcal{C}_e^{(s)}$ be a small circle around $\varepsilon_e^{(s)}$, not enclosing or touching any other spectrum of $K_\lambda(\omega)$. We introduce the generally non-orthogonal Riesz spectral projections

$$Q_e^{(s)} = Q_e^{(s)}(\omega,\lambda) = -\frac{1}{2\pi\mathrm{i}}\int_{\mathcal{C}_e^{(s)}} (K_\lambda(\omega) - z)^{-1}\mathrm{d}z. \tag{4.4}$$

It follows from (4.2) that

$$\langle A\rangle_t = \sum_e \sum_{s=1}^{\nu(e)} \mathrm{e}^{\mathrm{i}t\varepsilon_e^{(s)}} \left\langle \psi_0, Q_e^{(s)}[A\otimes \mathbb{1}_{\mathrm{S}} \otimes \mathbb{1}_{\mathrm{R}}]\psi_0\right\rangle + O(\lambda^2 \mathrm{e}^{-\omega' t/2}). \tag{4.5}$$

Note that the imaginary parts of all resonance energies $\varepsilon_e^{(s)}$ are smaller than $\omega'/2$, so that the remainder term in (4.5) is not only small in λ, but it also decays faster than all of the terms in the sum! (See also Figure 3.)

Finally, we notice that all terms in (4.5) with $\varepsilon_e^{(s)} \ne 0$ will vanish in the ergodic mean limit, so

$$\langle\!\langle A\rangle\!\rangle_\infty = \lim_{T\to\infty}\frac{1}{T}\int_0^T \langle A\rangle_t \ \mathrm{d}t = \sum_{s:\varepsilon_0^{(s)}=0} \left\langle \psi_0, Q_0^{(s)}[A\otimes \mathbb{1}_{\mathrm{R}} \otimes \mathbb{1}_{\mathrm{R}}]\psi_0\right\rangle.$$

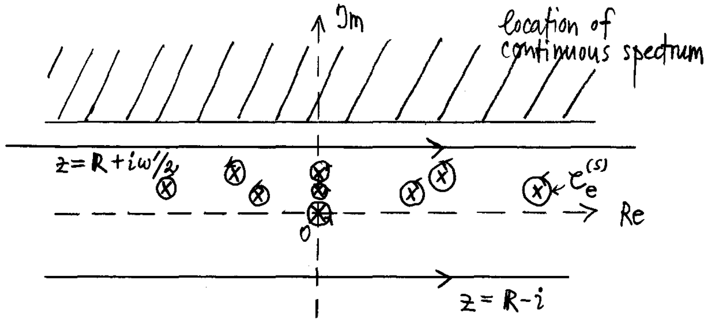


FIGURE 3. Contour deformation: $\int_{\mathbb{R}-\mathrm{i}} \mathrm{d}z = \sum_{e,s} \int_{\mathcal{C}_e^{(s)}} \mathrm{d}z + \int_{\mathbb{R}+\mathrm{i}\omega'/2} \mathrm{d}z$

The identification of the linear functionals

$$R_{\varepsilon_e^{(s)}}(A) = \left\langle \psi_0, Q_e^{(s)}[A \otimes \mathbb{1}_\mathrm{S} \otimes \mathbb{1}_\mathrm{R}]\psi_0 \right\rangle \tag{4.6}$$

(cf. (3.3)) is useful for concrete calculations, as well as in the proof of Theorem 3.2. This concludes the outline of the proof of Theorem 3.1.

5. Proof of Theorem 3.2

The proof is based on expansion (3.3) together with formula (4.6). We have $[\overline{\rho}_t]_{m,n} = \mathrm{Tr}\,(\overline{\rho}_t|\varphi_n\rangle\langle\varphi_m|)$, and it follows that

$$\begin{aligned} &[\overline{\rho}_t]_{m,n} - \langle\!\langle |\varphi_n\rangle\langle\varphi_m| \rangle\!\rangle_\infty \\ &\quad = \sum_{\{e,s:\varepsilon_e^{(s)}\neq 0\}} \mathrm{e}^{\mathrm{i}t\varepsilon_e^{(s)}} \left\langle \psi_0, Q_e^{(s)}[|\varphi_n\rangle\langle\varphi_m| \otimes \mathbb{1}_\mathrm{S}]\psi_0 \right\rangle + O(\lambda^2 \mathrm{e}^{-\omega' t/2}). \end{aligned} \tag{5.1}$$

We leave out the trivial $\mathbb{1}_\mathrm{R}$. Remembering that $\psi_0 = \psi_{\mathrm{S},\infty} \otimes \psi_\mathrm{R}$, where $\psi_{\mathrm{S},\infty}$ is the trace state of S, represented by the vector $\frac{1}{\sqrt{N}} \sum_{j=1}^N \varphi_j \otimes \varphi_j$, we see that

$$[|\varphi_n\rangle\langle\varphi_m| \otimes \mathbb{1}_\mathrm{S}]\psi_0 = \frac{1}{\sqrt{N}} \varphi_n \otimes \varphi_m \otimes \psi_\mathrm{R}. \tag{5.2}$$

We shall treat in here the case where all resonance eigenvalues $\varepsilon_e^{(s)}$ are simple (the general case is dealt with in a similar fashion). Thus $Q_e^{(s)} = |\chi_e^{(s)}\rangle\langle\widetilde{\chi}_e^{(s)}|$ is a rank-one projection, with $K_\lambda(\omega)\chi_e^{(s)} = \varepsilon_e^{(s)}\chi_e^{(s)}$, $K_\lambda(\omega)^*\widetilde{\chi}_e^{(s)} = \overline{\varepsilon_e^{(s)}}\widetilde{\chi}_e^{(s)}$ and with the normalization $\left\langle \chi_e^{(s)}, \widetilde{\chi}_e^{(s)} \right\rangle = 1$. We expand the resonance eigenvectors in powers of λ,

$$\chi_e^{(s)} = \eta_e^{(s)} \otimes \psi_\mathrm{R} + O(\lambda), \quad \widetilde{\chi}_e^{(s)} = \widetilde{\eta}_e^{(s)} \otimes \psi_\mathrm{R} + O(\lambda), \tag{5.3}$$

where $\eta_e^{(s)}$, $\widetilde{\eta}_e^{(s)}$ are eigenvectors of the level shift operator Λ_e associated to the eigenvalue $\delta_e^{(s)}$ and its complex conjugate, respectively (see also before Theorem

3.2 and [11]). Λ_e acts on the eigenspace $P(L_{\rm S}=e)$, and $\eta_e^{(s)}, \widetilde{\eta}_e^{(s)} \in \operatorname{Ran} P(L_{\rm S}=e)$. We obtain

$$Q_e^{(s)} = |\eta_e^{(s)}\rangle\langle\widetilde{\eta}_e^{(s)}| \otimes |\psi_{\rm R}\rangle\langle\psi_{\rm R}| + R_1(\lambda),$$

where R_1 satisfies $\langle\psi_{\rm R}|R_1(\lambda)|\psi_{\rm R}\rangle = O(\lambda^2)$. (This term is of order λ^2 and not only λ since the average of the interaction (3.1) vanishes in the vacuum state.) Combining (5.2) and (5.3) and setting $\varphi_{m,n} = \varphi_m \otimes \varphi_n$, we arrive at

$$\begin{aligned}
&\left\langle \psi_0, Q_e^{(s)}[|\varphi_n\rangle\langle\varphi_m| \otimes \mathbb{1}_{\rm S}]\psi_0 \right\rangle \\
&= \frac{\delta_{e=E_{n,m}}}{\sqrt{N}} \left\langle \psi_0, \eta_e^{(s)} \otimes \psi_{\rm R} \right\rangle \left\langle \widetilde{\eta}_e^{(s)}, \varphi_{n,m} \right\rangle + O(\lambda^2) \\
&= \frac{\delta_{e=E_{n,m}}}{\sqrt{N}} \sum_{\{l,k:E_{l,k}=e\}} \langle \psi_0, \varphi_{l,k} \otimes \psi_{\rm R}\rangle \left\langle \varphi_{l,k}, \eta_e^{(s)} \right\rangle \left\langle \widetilde{\eta}_e^{(s)}, \varphi_{n,m} \right\rangle + O(\lambda^2).
\end{aligned}$$

(The δ is the Kronecker delta here.) The initial values are recovered from the first scalar product on the r.h.s.,

$$\frac{1}{\sqrt{N}} \langle \psi_0, \varphi_{l,k} \otimes \psi_{\rm R}\rangle = \langle \psi_0, [|\varphi_l\rangle\langle\varphi_k| \otimes \mathbb{1}_{\rm S}]\psi_0\rangle = [\overline{\rho}_0]_{k,l}.$$

This shows that

$$\left\langle \psi_0, Q_e^{(s)}[|\varphi_n\rangle\langle\varphi_m| \otimes \mathbb{1}_{\rm S}]\psi_0 \right\rangle = \delta_{e,E_{n,m}} \sum_{\{l,k:E_{l,k}=E_{n,m}\}} \sigma_{m,n;k,l}[\overline{\rho}_0]_{k,l} + O(\lambda^2), \quad (5.4)$$

where the "mixing coefficients" $\sigma_{m,n;k,l}^{(s)}$ are defined in Theorem 3.2. We use expression (5.4) in (5.1),

$$\begin{aligned}
&[\overline{\rho}_t]_{m,n} - \langle\!\langle |\varphi_n\rangle\langle\varphi_m| \rangle\!\rangle_\infty \qquad (5.5) \\
&= \sum_{\{e,s:\varepsilon_e^{(s)}\neq 0\}} \mathrm{e}^{\mathrm{i}t\varepsilon_e^{(s)}} \left[\delta_{e,E_{n,m}} \sum_{\{l,k:E_{l,k}=E_{n,m}\}} \sigma_{m,n;k,l}^{(s)}[\overline{\rho}_0]_{k,l} + O(\lambda^2) \right] + O(\lambda^2 \mathrm{e}^{-\omega' t/2}).
\end{aligned}$$

The main term in the sum selects $e = E_{n,m}$ and only the summation over $s : \varepsilon_{E_{n,m}}^{(s)} \neq 0$ remains. This yields the dominant part in the r.h.s. of formula (3.5). The remainder is

$$\sum_{\{e:e\neq E_{n,m},\ s:\varepsilon_e^{(s)}\neq 0\}} \mathrm{e}^{\mathrm{i}t\varepsilon_e^{(s)}} O(\lambda^2) + O(\lambda^2 \mathrm{e}^{-\omega' t/2}),$$

which is $O(\lambda^2 \mathrm{e}^{-t\gamma_{m,n}})$, as indicated in Theorem 3.2. This concludes the proof of Theorem 3.2.

References

[1] Araki, H., Woods, E.J.: *Representations of the canonical commutation relations describing a nonrelativistic infinite free Bose gas* J. Math. Phys., **4** 637–662 (1963)

[2] Araki, H., Wyss, W.: *Representations for canonical anticommutation relations.* Helv. Phys. Acta **37** 136–159 (1964)

[3] Bratteli, O., Robinson, D.W., Operator Algebras and Quantum Statistical Mechanics I, II. Texts and Monographs in Physics, Springer-Verlag, 1987

[4] Caldeira, A.O., Leggett, A.J.: *Quantum tunnelling in a dissipative system.* Ann. Phys. **149**, 374–456 (1983)

[5] Fröhlich, J., Merkli, M., Ueltschi, D.: *Dissipative transport: thermal contacts and tunneling junctions.* Ann. Henri Poincaré **4**, no. 5, 897–945 (2003)

[6] Fröhlich, J., Merkli, M.: *Another return of "Return to Equilibrium".* Commun. Math. Phys. **251**, 235–262 (2004)

[7] Joos, E., Zeh, H.D., Kiefer, C., Giulini, D., Kupsch, J. Stamatescu, I.O.: Decoherence and the appearance of a classical world in quantum theory. Second edition. *Springer Verlag, Berlin,* 2003

[8] Leggett, A.J., Chakravarty, S., Dorsey, A.T., Fisher, M.P.A., Garg, A., Zwerger, W.: *Dynamics of the dissipative two-state system.* Rev. Mod. Phys. **59**, no. 1, 1–85 (1987)

[9] Merkli, M.: *Level shift operators for open quantum systems.* J. Math. Anal. Appl. **327** no. 1, 376–399 (2007)

[10] Merkli, M., Mück, M., Sigal, I.M.: *Instability of equilibrium states for coupled heat reservoirs at different temperatures*, J. Funct. Anal. **243**, Issue 1, 310–344 (2006)

[11] Merkli, M., Sigal, I.M., Berman, G.P.: *Resonance theory of decoherence and thermalization*, Ann. Phys. **323** (2008), no. 2, 373–412.

[12] Merkli, M., Sigal, I.M., Berman, G.P.: *Dynamics of collective decoherence and thermalization.* Ann. Phys. **323** (2008), no. 12, 3091–3112.

[13] Merkli, M., Sigal, I.M., Berman, G.P.: *Decoherence and thermalization*, Phys. Rev. Lett. **98**, no. 13, 130401, 4pp. (2007)

[14] Merkli, M.: *The Ideal Quantum Gas.* Open quantum systems. I, 183–233, Lecture Notes in Math., 1880, Springer, Berlin, 2006

[15] Palma, M.G., Suominen, K.-A., Ekert, A.: *Quantum computers and dissipation*, Proc. R. Soc. Lond. A **452**, 567–584 (1996)

[16] Schlosshauer, M., Decoherence and the quantum-to-classical transition, The frontiers collection, Springer Verlag, 2007

[17] Zurek, W.H.: *Decoherence, einselection, and the quantum origins of the classical.* Rev. Mod. Phys. **75**, 715–775 (2003)

Marco Merkli
Department of Mathematics and Statistics
Memorial University of Newfoundland
St. John's, NL
Canada, A1C 5S7
e-mail: `merkli@math.mun.ca`

Analysis and Mathematical Physics
Trends in Mathematics, 427–439
© 2009 Birkhäuser Verlag Basel/Switzerland

Ramified Integrals, Casselman Phenomenon, and Holomorphic Continuations of Group Representations

Yuri A. Neretin

To Mark Iosifovich Graev on his 85^{th} birthday

Abstract. Let G be a real semisimple Lie group, K its maximal compact subgroup, and $G_{\mathbb{C}}$ its complexification. It is known that all K-finite matrix elements on G admit holomorphic continuations to branching functions on $G_{\mathbb{C}}$ having singularities at a prescribed divisor. We propose a geometric explanation of this phenomenon.

1. Introduction

1.1. Casselman's theorem. Let G be a real semisimple Lie group, let K be a maximal compact subgroup. Let $G_{\mathbb{C}}$ be the complexification of G.

Let ρ be an infinite-dimensional irreducible representation of G in a complete separable locally convex space W[1]. Recall that a vector $w \in W$ is K-finite if the orbit $\rho(G)v$ spans a finite-dimensional subspace in W.[2]

A K-finite matrix element is a function on G of the form

$$f(g) = \ell(\rho(g)v),$$

where v is a K-finite vector in W and ℓ is a K-finite linear functional, i.e., a K-finite element of the dual representation.

Supported by the grant FWF, project P19064, Russian Federal Agency for Nuclear Energy, Dutch grant NWO.047.017.015, and grant JSPS-RFBR-07.01.91209 .

[1]the case of unitary representations in Hilbert spaces is sufficiently non-trivial.

[2]Let us rephrase the definition. We restrict ρ to the subgroup K and decompose the restriction into a direct sum $\sum V_i$ of finite-dimensional representations of K. Finite sums of the form $\sum_{v_j \in V_j} v_j$ are precisely all K-finite vectors.

Theorem 1.1. [3] *There is an (explicit) complex submanifold $\Delta \subset G_{\mathbb{C}}$ of codimension 1 such that each K-finite matrix element of G admits a continuation to an analytic multi-valued ramified function on $G_{\mathbb{C}} \setminus \Delta$.*

Example. Let $G = \mathrm{SL}(2,\mathbb{R})$ be the group of real matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, whose determinant $= 1$. Then $K = \mathrm{SO}(2)$ consists of matrices $\begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$, where $\varphi \in \mathbb{R}$; the group $G_{\mathbb{C}}$ is the group of complex 2×2 matrices with determinant $= 1$. The submanifold $\Delta \subset \mathrm{SL}(2,\mathbb{C})$ is a union of the following four manifolds

$$a = 0, \quad b = 0, \quad c = 0, \quad d = 0. \tag{1.1}$$

Indeed, in this case, there exists a canonical K-eigenbasis. All the matrix elements in this basis are Gauss hypergeometric functions of the form

$${}_2F_1(\alpha,\beta;\gamma;\theta), \quad \text{where } \theta = \frac{ad}{bc},$$

where the indices α, β, γ depend on parameters of a representation and of a pair of basis elements (see [6]).

Points of ramification of ${}_2F_1$ are $\theta = 0, 1, \infty$. Since $ad - bc = 1$, only $\theta = 0$ and $\theta = \infty$ are admissible; this implies (1.1). □

Thus a representation ρ of a real semisimple group admits a continuation to an analytic matrix-valued function on $G_{\mathbb{C}}$ having singularities at Δ. This fact seems to be strange if we look to explicit constructions of representations.

Our purpose is to clarify this phenomenon and to find a direct geometric construction of the analytic continuation. We achieve this aim for a certain special case (namely, for principal maximally degenerate series of $\mathrm{SL}(n,\mathbb{R})$, see Section 2) and formulate a general conjecture (Section 3). It seems that our explanation (a reduction to the 'Thom isotopy Theorem'), see [4], [5]) is trivial. However, as far as I know it is not known for experts in representation theory.

Addendum contains a general discussion of holomorphic continuations of representations.

2. Isotopy of cycles

2.1. Principal degenerate series for the groups $\mathrm{SL}(n,\mathbb{R})$**.** Let $G = \mathrm{SL}(n,\mathbb{R})$ be the group of all real matrices with determinant $= 1$. The maximal compact subgroup $K = \mathrm{SO}(n)$ is the group of all real orthogonal matrices.

[3] This theorem was obtained in famous preprints of W. Casselman on the Subrepresentation Theorem. Unfortunately, these preprints are unavailable for the author; however these results were included to the paper of W. Casselman and Dr. Milicic [1]. There are (at least) two known proofs; the original proof is based on properties of system of partial differential equations for matrix elements [1], also by a simple trick [3] the theorem can be reduced to properties of Heckman–Opdam hypergeometric functions [2].

Denote by $\mathbb{RP}^{n-1} \subset \mathbb{CP}^{n-1}$ the real and complex projective spaces; recall that the manifold $\mathbb{RP}^{n-1}$ is orientable iff n is even.

Denote by $d\omega$ the $\mathrm{SO}(n)$-invariant Lebesgue measure on $\mathbb{RP}^{n-1}$, let $d(\omega g)$ be its pushforward under the map $\mathrm{SL}(n,\mathbb{R})$, denote by

$$J(g,x) := \frac{d\omega g}{d\omega}$$

the Jacobian of a transformation g at a point x.

Fix $\alpha \in \mathbb{C}$. Define a representation $T_\alpha(g)$ of the group $\mathrm{SL}(n,\mathbb{R})$ in the space $C^\infty(\mathbb{RP}^{n-1})$ by the formula

$$T_\alpha(g)f(x) = f(xg)J(g,x)^\alpha.$$

The representations T_α are called *representations of principal degenerate series.* If $\alpha \in \frac{1}{2} + i\mathbb{R}$, then this representation is unitary in $L^2(\mathbb{RP}^{n-1})$.

2.2. Discriminant submanifold Δ. Denote by g^t the transpose of a matrix g. Denote by Δ the submanifold in $\mathrm{SL}(n,\mathbb{C})$ consisting of matrices g such that the equation

$$\det(gg^t - \lambda) = 0$$

has a multiple root.

> *We wish to construct a continuation of the function $g \mapsto T_\alpha(g)$ to a multi-valued function on* $\mathrm{SL}(n,\mathbb{C}) \setminus \Delta$.

For simplicity, *we assume n is even.*[4]

2.3. Invariant measure. Denote by $x_1 : x_2 : \cdots : x_n$ the homogeneous coordinates in the projective space. The $\mathrm{SO}(n)$-invariant $(n-1)$-form on $\mathbb{RP}^{n-1}$ is given by

$$d\omega(x) = \Big(\sum_j x_j^2\Big)^{-n/2} \sum_j (-1)^j x_j \, dx_1 \dots \widehat{dx_j} \dots dx_n.$$

This expression can be regarded as a meromorphic $(n-1)$-form on $\mathbb{CP}^{n-1}$ having a pole on the quadric

$$Q(x) := \sum x_j^2 = 0.$$

Now we can treat the Jacobian $J(g,x)$ as a meromorphic function on $\mathbb{CP}^{n-1}$ having a zero at the quadric $Q(x) = 0$ and a pole on the shifted quadric $Q(gx) = 0$.

2.4. K-finite functions. The following functions span the space of K-finite functions on $\mathbb{RP}^{n-1}$:

$$f(x) = \frac{\prod x_j^{k_j}}{(\sum x_j^2)^{\sum k_j/2}}, \qquad \text{where } \sum k_j \text{ is even.}$$

Evidently, they have singularities at the quadric $Q(x) = 0$ mentioned above.

2.5. K-finite matrix elements. K-finite matrix elements are given by the formula

$$\{f_1, f_2\} = \int_{\mathbb{RP}^{n-1}} f_1(x) f_2(xg) J(g,x)^\alpha d\omega(x). \tag{2.1}$$

[4] If n is odd, then we must replace the integrand in (2.1) by an $(n-1)$-form on two sheet covering of $\mathbb{CP}^{n-1} \setminus \mathbb{RP}^{n-1}$. Also we must replace the cycle $\mathbb{RP}^{n-1}$ by its two-sheet covering.

The integrand is a holomorphic form on $\mathbb{CP}^{n-1}$ of maximal degree ramified over the quadrics $Q(x) = 0$, $Q(xg) = 0$. Denote by $\mathfrak{U} = \mathfrak{U}[g]$ the complement to these quadrics. Therefore locally in $\mathfrak{U}$ the integrand is a closed $(n-1)$-form. Hence we can replace $\mathbb{RP}^{n-1}$ by an arbitrary isotopic cycle C in $\mathfrak{U}$.

2.6. Reduction to the Pham Theorem. Now let $g(s)$ be a path in $\mathrm{SL}(n, \mathbb{C})$ starting in $\mathrm{SL}(n, \mathbb{R})$. For each s one has a pair $Q(x) = 0$, $Q(x \cdot g(s)) = 0$ of quadrics and the corresponding complement $\mathfrak{U}[g(s)]$.

> *Is it possible to construct an isotopy $C(s)$ of the cycle $\mathbb{RP}^{n-1}$ such that $C(s) \subset \mathfrak{U}[g(s)]$ for all s?*

Now recall the following theorem of F. Pham [4] (see, also, V.A. Vasiliev [5]).

Theorem 2.1. *Let $R_1(s), \ldots, R_l(s)$ be nonsingular complex hypersurfaces in $\mathbb{CP}^k$ depending on a parameter. Assume that R_j are transversal (at all points for all values of the parameter s). Then each cycle $Q(s_0)$ in the complement of $\cup R_j(s_0)$ admits an isotopy $Q(s)$ such that for each s a cycle $Q(s)$ is contained in the complement of $\cup R_j(s)$.*

2.7. Transversality of quadrics

Lemma 2.2. *Let A, B be non-degenerate symmetric matrices. Assume that all the roots of the characteristic equation*

$$\det(A - \lambda B) = 0$$

are pairwise distinct. Then the quadrics $\sum a_{ij} x_i x_j = 0$ and $\sum b_{ij} x_i x_j = 0$ are transversal.

By the Weierstrass theorem such pair of quadrics can be reduced to

$$\sum \lambda_j x_j^2 = 0, \qquad \sum x_j^2 = 0, \tag{2.2}$$

where λ_j are the roots of the characteristic equation. If they are not transversal at a point x, then the rank of the Jacobi matrix

$$\begin{pmatrix} \lambda_1 x_1 & \ldots & \lambda_n x_n \\ x_1 & \ldots & x_n \end{pmatrix}$$

is 1. Therefore

$$(\lambda_i - \lambda_j) x_i x_j = 0 \quad \text{for all } i,\, j. \tag{2.3}$$

The system (2.3), (2.2) is inconsistent. □

2.8. Last step. In our case, the matrices of quadratic forms are gg^t and 1. Therefore, by the virtue of the Pham Theorem a desired isotopy of the cycle $\mathbb{RP}^{n-1}$ exists.

3. General case

By the Subrepresentation Theorem, all the irreducible representations of a semisimple group G are subrepresentations of the principal (generally, non-unitary) series. Therefore, it suffices to construct analytic continuations for representations of the principal series.

For definiteness, we discuss the spherical principal series of the group $G = \mathrm{SL}(n, \mathbb{R})$.

3.1. Spherical principal series for $G = \mathrm{SL}(n, \mathbb{R})$. Denote by $\mathrm{Fl}(\mathbb{R}^n)$ the space of all complete flags of subspaces

$$\mathcal{W} : 0 \subset W_1 \subset \cdots \subset W_{n-1} \subset \mathbb{R}^n$$

in $\mathbb{R}^n$; here $\dim W_k = k$. By $\mathrm{Gr}_k(\mathbb{R}^n)$ we denote the Grassmannian of all k-dimensional subspaces in $\mathbb{R}^n$. By γ_k we denote the natural projection $\mathrm{Fl}(\mathbb{R}^n) \to \mathrm{Gr}_k(\mathbb{R}^n)$.

By ω_k we denote the $\mathrm{SO}(n)$-invariant measure on $\mathrm{Gr}_k(\mathbb{R}^n)$. For $g \in \mathrm{GL}(n, \mathbb{R})$ we denote by $J_k(g, V)$ the Jacobian of the transformation $V \mapsto Vg$ of $\mathrm{Gr}_k(\mathbb{R}^n)$,

$$J_k(g, V) = \frac{d\omega_k(Vg)}{d\omega_k(V)}.$$

Fix $\alpha_1, \ldots, \alpha_{n-1} \in \mathbb{C}$. The representation T_α of the spherical principal series of the group $\mathrm{SL}(n, \mathbb{R})$ acts in the space $C^\infty(\mathrm{Fl}(\mathbb{R}^n))$ by the formula

$$T_\alpha(g) f(\mathcal{W}) = f(\mathcal{W} \cdot g) \prod_{k=1}^{n-1} J_k(g, \gamma_k(\mathcal{W}))^{\alpha_k}.$$

3.2. Singularities. Consider the symmetric bilinear form in $\mathbb{C}^n$ given by

$$B(x, y) = \sum x_j y_j.$$

By $L_k \subset \mathrm{Gr}_k(\mathbb{C}^n)$ we denote the set of all the k-dimensional subspaces, where the form B is degenerate[5]. By $\mathcal{L} \subset \mathrm{Fl}(\mathbb{C}^n)$ we denote the set of all the flags $W_1 \subset \cdots \subset W_{n-1}$, where $W_k \in L_k$ for some k.

In fact, all K-finite functions on $\mathrm{Fl}(\mathbb{R}^n)$ admit analytic continuations to $\mathrm{Fl}(\mathbb{C}^n) \setminus \mathcal{L}$ (a singularity on $\mathcal{L}$ is a pole or a two-sheet ramification).

3.3. A conjecture

Conjecture 3.1. *Let $\gamma(t)$ be a path on $\mathrm{GL}(n, \mathbb{C})$ avoiding the discriminant submanifold Δ, let $\gamma(0) \in \mathrm{SL}(n, \mathbb{R})$. Then there is an isotopy $C(t)$ of the cycle $C(t_0) := \mathrm{Fl}(\mathbb{R}^n)$ in the space $\mathrm{Fl}(\mathbb{C}^n)$ avoiding the submanifolds $\mathcal{L}$ and $\mathcal{L} \cdot g(s)$*

Such isotopy produces an analytic continuation of representations of principal series of $\mathrm{SL}(n, \mathbb{R})$.

---

[5] Equivalently, we can consider all the $(k-1)$-dimensional subspaces in $\mathbb{CP}^{n-1}$ tangent to the quadric $\sum x_j^2 = 0$.

References

[1] Casselman, W., Miličić, Dr. *Asymptotic behavior of matrix coefficients of admissible representations.* Duke Math. J. 49 (1982), no. 4, 869–930.

[2] Heckman, G.I., Opdam, E.M., *Root systems and hypergeometric functions. I.* Compositio Math, 64(1987), 329–352;

[3] Neretin, Yu.A. *K-finite matrix elements of irreducible Harish–Chandra modules are hypergeometric.* Funct. Anal. Appl., 41 2007, 295–302.

[4] Pham, F., *Introduction l'étude topologique des singularités de Landau.*, Paris, Gauthier-Villars, 1967

[5] Vasiliev, V.A., *Ramified integrals*, Moscow, Independent University, 2000; Engl. transl., Kluwer, 1995

[6] Vilenkin N.Ya. *Special functions and the theory of group representations.* Amer. Math. Soc., 1968 (translated from Russian 1965 edition).

Addendum. Survey of holomorphic continuations of representations

Let G be a connected linear Lie group. Denote by $G_{\mathbb{C}}$ its complexification. Let ρ be an irreducible representation of G (in a Fréchet space). We are interested in the following problems:

- Is it possible to extend ρ holomorphically to $G_{\mathbb{C}}$?
- Is it possible to extend ρ holomorphically to an open domain $U \subset G_{\mathbb{C}}$.

See, also, [16], Section 1.5.

A.1. Weyl trick. Let ρ be a finite-dimensional representation of a semisimple Lie group G. Then ρ admits a holomorphic continuation to the group $G_{\mathbb{C}}$.

A.2. Why the Weyl trick does not survive for infinite-dimensional unitary representations? Let G be a noncompact Lie group, let ρ be its irreducible faithful unitary representation. Let X be a noncentral element of the Lie algebra $\mathfrak{g}$. It is more-or-less obvious that the operator $\rho(X)$ is unbounded.

Then, for t, $s \in \mathbb{R}$,

$$\rho(\exp(t + is)X) = \exp(is\rho(X)) \exp(t\rho(X)).$$

Since $i\rho(X)$ is self-adjoint, then $\exp(t\rho(X))$ is unitary; on the other hand $\exp(is\rho(X))$ have to be unbounded for all positive s or for all negative s (and usually it is unbounded for all s.

However, this argument does not remove completely an idea of holomorphic continuation, since it remain two following logical possibilities

- a holomorphic extension exists in spite of the unboundedness of operators.
- If a spectrum of X is contained on the positive half-line, then $\exp(tX)$ is defined for negative t. We can hope to construct something from elements of this kind.

The second variant is realized for Olshanski semigroups, see below, the first variant is general, this follows from the Nelson Theorem.

A.3. Nelson's paper. In 1959 E. Nelson [13] proved that each unitary irreducible representation ρ of a real Lie group G has a dense set of analytic vectors. This implies that ρ can be extended analytically to a sufficiently small neighborhood of G in $G_{\mathbb{C}}$.

Usually this continuation can be done in a constructive way as it is explained below (see also [16], Section 1.5.).

A.4. Induced representations. First, we recall the definition of induced representations.

Consider a Lie group G and its closed connected subgroup H. Let ρ be a representation of H in a *finite-dimensional* complex space V.

These data allow to construct canonically a vector bundle (*skew product*) over G/H with a fiber V. Recall a construction (see, for instance, [8]). Consider the direct product $G \times V$ and the equivalence relation

$$(g, v) \sim (gh^{-1}, \rho(h)v), \qquad \text{where } g \in G,\ v \in V,\ h \in H.$$

Denote by $R = G \times_V H$ the quotient space. The standard map $G \to G/H$ determines a map $R \to G/H$ (we simply forget v). A fiber can be (noncanonically) identified with V.

Next, the group G acts on $G \times V$ by transformations

$$(g, v) \mapsto (rg, v), \qquad \text{where } g \in G,\ v \in V,\ r \in G.$$

This action induces an action of G on $G \times_H V$, the last action commutes with the projection $G \times_H V \to G/H$.

Therefore G acts in the space of sections of the bundle $G \times_H V \to G/H$ (because the graph of a section is a subset in the total space; the group G simply moves subsets). The *induced representation* $\pi = \mathrm{Ind}_H^G(\rho)$ *is the representation of* G *in a space of sections of the bundle* $G \times_H V \to G/H$.

The most important example are principal series, which were partially discussed above.

Our definition is not satisfactory since rather often it is necessary to specify the space of sections (for instance, smooth functions, L^2-functions, distributions, etc.). This discussion is far beyond our purpose, for the moment let us consider the space $C^\infty[G/H; \rho]$ of smooth sections.

A.5. Analytic continuation of induced representations. Here we discuss some heuristic arguments (see [16], Section 1.5). Their actual usage depends on the explicit situation.

Denote by $H_{[\mathbb{C}]}$ the complexification of the group H inside $G_{\mathbb{C}}$.

The (finite-dimensional) representation ρ admits a holomorphic extension to a representation of the universal covering group of $\widetilde{H}_{[\mathbb{C}]}$ of $H_{[\mathbb{C}]}$ in the space V. For a moment, let us require two assumptions[6]

– $H_{[\mathbb{C}]}$ *is closed in* $G_{\mathbb{C}}$.

– ρ is a linear representation of $H_{[\mathbb{C}]}$.

Under these assumptions, the same construction of a skew-product produces the bundle $(G_{\mathbb{C}}) \times_{H_{\mathbb{C}}} V \to G_{\mathbb{C}}/H_{\mathbb{C}}$. Moreover,

$$G \times_H V \;\subset\; (G_{\mathbb{C}}) \times_{H_{\mathbb{C}}} V.$$

Now let us agree on the next assumption[7]. *Let the space of holomorphic sections of* $(G_{\mathbb{C}}) \times_{H_{\mathbb{C}}} V$ *be dense in* C^∞ *on* G/H. Then we get a holomorphic continuation of the induced representation π to the whole complex group $G_{\mathbb{C}}$. More precisely, we slightly reduce the space of representation, but the 'formulae' determining a representation are the same.

This variant is realized for all nilpotent Lie groups.

A.6. Nilpotent Lie groups

Example. Let a, b, c range in $\mathbb{R}$, Consider operators $T(a,b,c)$ in $L^2(\mathbb{R})$ given by

$$T(a,b,c)f(x) = f(x+a)e^{ibx+c}.$$

They form a 3-dimensional group, namely the Heisenberg group. Now let a, b, c range in $\mathbb{C}$ and f ranges in the space $Hol(\mathbb{C})$ of entire functions. Then the same formula determines a representation of the complex Heisenberg group in $Hol(\mathbb{C})$. After this operation, the space of representation completely changes. However it is easy to find a dense subspace in $L^2(\mathbb{R})$ consisting of holomorphic functions and invariant with respect to all the operators $T(a,b,c)$. □

Now, let G be a simply connected nilpotent Lie group. By the Kirillov Theorem [7], each unitary representation of a nilpotent Lie group G is induced from a one-dimensional representation of a subgroup H; the manifolds G/H are equivalent to standard spaces $\mathbb{R}^m$. Therefore, $G_{\mathbb{C}}/H_{\mathbb{C}}$ is the standard complex space $\mathbb{C}^m$, and we get a representation of $G_{\mathbb{C}}$ in the space of entire functions.

However, the following Goodman–Litvinov Theorem (R. Goodman [4], G.L. Litvinov [11], [12]) is more delicate.

Theorem. *Let ρ be an irreducible unitary representation of a nilpotent group G in a Hilbert space W. There exists a* (*noncanonical*) *dense subspace Y with its own Fréchet topology and holomorphic representation $\widetilde{\rho}$ of $G_{\mathbb{C}}$ in the space W coinciding with ρ on G.*

[6]The second assumption is very restrictive. It does not hold for the parabolic induction.
[7]If it does not hold, then we go to Subsection A.7, where all the assumptions are omitted.

Let us explain how to produce a subspace Y. Let ρ be a unitary representation of a nilpotent group G in a space W. Let f be an entire function on $G_{\mathbb{C}}$ (it is specified below). Consider the operator

$$\rho(f) = \int_G f(g)\rho(g)\, dg.$$

Let $r \in G_{\mathbb{C}}$. We write formally

$$\rho(r)\rho(f) = \int_G f(g)\rho(rg)\, dg = \int_G f(gr^{-1})\rho(g).$$

Assume that for each $r \in G_{\mathbb{C}}$ the function $\gamma_r(g) := f(gr^{-1})$ is integrable on G. Under this condition we can define operators

$$\rho(r) : \Big\{\text{Image of } \rho(f)\Big\} \to W$$

as just now.

In fact, we need a subspace Z in $L^1(G)$ consisting of entire functions and invariant with respect to complex shifts. To be sure that

$$Y := \cup_{f\in Z}\mathrm{Im}(\rho(f)) \subset W$$

is dense, we need a sequence of positive $f_j \in Z$ converging to δ-functions; then $\rho(f_j)v$ converges to v for all $v \in W$. In what follows we describe a simple trick that allows to construct many functions f and a subspace Z.

A construction of functions f. First, let $G = T_n$ be the unipotent upper triangular subgroup of order n. Let t_{ij}, $i < j$, be the natural coordinates on T_n. Write them in the order

$$t_{12}, t_{23}, t_{34}, \dots, t_{(n-1)n}, t_{13}, t_{24}, \dots, t_{(n-2)n}, t_{14}, \dots$$

and re-denote these coordinates by x_1, x_2, x_3, ... In this notation, the right shift $g \mapsto gr^{-1}$ is an affine transformation of the form

$$(x_1, x_2, x_3, \dots) \mapsto (x_1 + a_1, x_2 + a_2 + b_{21}x_1, x_3 + a_3 + b_{31}x_1 + b_{32}x_2, \dots).$$

Now we can choose a desired function f in the form

$$f(x) = \exp\Big\{-\sum p_j x_j^{2k_j}\Big\}, \quad \text{where } p_j > 0 \text{ and } k_1 > k_2 > \dots > 0 \text{ are integers.}$$

For an arbitrary nilpotent G we apply the Ado theorem (in fact, the standard proof, see [17], produces a polynomial embedding of G to some T_n with very large n).

A.7. Local holomorphic continuations. Now let $G \supset H$ be arbitrary.

The construction of a skew product survives locally. It determines a holomorphic bundle on a (noncanonical) neighborhood $U \subset G_{\mathbb{C}}/H_{\mathbb{C}}$ of G/H. Denote by $\mathcal{A}(U)$ the space of holomorphic sections of this bundle. Let $r \in G_{\mathbb{C}}$ satisfies $r \cdot G/H \subset U$. Then r determines an operator $\pi(r) : \mathcal{A}(U) \to C^\infty$ and we get a local analytic continuation of the induced representation.

A.8. Local analytic continuations for semisimple groups. For definiteness, consider $G = \mathrm{SL}(n, \mathbb{R})$. Denote by P be the minimal parabolic (i.e., P is the group of upper triangular matrices), Then G/P is the flag space mentioned above. Next, $G_\mathbb{C} = \mathrm{SL}(n, \mathbb{C})$, and $P_{[\mathbb{C}]}$ is the group of complex upper-triangular matrices of the form

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots \\ 0 & b_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}, \qquad \text{where } b_{ii} \neq 0.$$

Evidently, the group $P_\mathbb{C}$ is not simply connected.

Fix $s_j \in \mathbb{C}$ and consider the one-dimensional character

$$\chi_s(B) := \prod_{j=1}^{n} b_{jj}^{s_j}$$

of P. The function χ_s is defined on $P_\mathbb{C}$ only locally, however this is sufficient for the arguments of the previous subsection.

This kind of arguments can be easily applied to an arbitrary representation of a nondegenerate principal series. Keeping in the mind the Subquotient Theorem, we easily get the following statement.

Observation. *Let G be a semisimple Lie group. Then there are (noncanonical) open sets $U_1 \subset U_2 \subset G_\mathbb{C}$ containing G such that $U_2 \supset U_1 \cdot U_1$ and the following property holds. Let ρ be an irreducible representation of G in a Fréchet space W. Then there is a (noncanonical) dense subspace $Y \subset W$ (equipped with its own Fréchet topology) and an operator-valued holomorphic function $\widetilde{\rho} : U_2 \to \mathrm{Hom}(Y, W)$ such that $\widetilde{\rho} = \rho$ on G and*

$$\widetilde{\rho}(g_1)\widetilde{\rho}(g_2)y = \widetilde{\rho}(g_1 g_2)y \qquad \text{for } g_1,\ g_2 \in U_1,\ y \in Y$$

Certainly, the operators $\widetilde{\rho}(g)$ are unbounded in the topology of $\mathrm{Hom}(W, W)$.

A.9. Crown. D.N. Akhiezer and S.G. Gindikin (see [1]) constructed a certain explicit domain $\mathcal{A} \subset G_\mathbb{C}$ ('crown') to which all the spherical functions of a real semisimple group G can be extended. Also the crown is a domain of holomorphy of all irreducible representations of G, see B. Krotz, R. Stanton, [9], [10].

The relation of their constructions with our previous considerations are not completely clear.

A.10. Olshanski semigroups. In all the previous examples, the operators of holomorphic continuation are unbounded in the initial topology. There is an important exception.

Unitary highest weight representations of a semisimple Lie group G admit holomorphic continuations to a certain subsemigroup $\Gamma \subset G_\mathbb{C}$ (M.I. Graev [6], G.I. Olshanski [19]). Since this situation is well understood, we omit further discussion, see also [18].

A.11. Infinite-dimensional groups. Induction (in different variants) is the main tool of construction of representations of Lie groups.

For infinite-dimensional groups the induction exists[8] but it is a secondary tool (however, the algebraic variant of induction is important for infinite-dimensional Lie algebras). A more effective instrument are symplectic and orthogonal spinors.

Let us realize the standard real orthogonal group $O(2n)$ as a group of $(n+n)\times(n+n)$ complex matrices g having the structure

$$g = \begin{pmatrix} \Phi & \Psi \\ \overline{\Psi} & \overline{\Phi} \end{pmatrix}$$

that are orthogonal in the following sense

$$g \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} g^t = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Actually we do the following. Consider the space $\mathbb{R}^{2n}$ equipped with a standard basis $e_1, \dots, e_n, e_{n+1}, \dots, e_{2n}$. Then we pass to the space $\mathbb{C}^{2n}$ and write matrices of real orthogonal operators in the basis

$$e_1 + ie_{n+1},\, e_2 + ie_{n+2},\, \dots,\, e_n + ie_{2n}, e_1 - ie_{n+1},\, e_2 - ie_{n+2},\, \dots,\, e_n - ie_{2n}$$

Next, we set $n = \infty$. Denote by $OU(2\infty)$ the group of all bounded matrices of the same structure satisfying an additional condition: Ψ is a Hilbert–Schmidt matrix (i.e., the sum $\sum |\psi_{kl}|^2 < \infty$).

By the well-known theorem of F.A. Berezin [2], [3] and D. Shale-W. Stinespring [20], the spinor representation is well defined on the group $OU(2\infty)$.

Numerous infinite-dimensional groups G can be embedded in a natural way to $OU(2\infty)$, after this we can restrict the spinor representation to G.

For instance, for the loop group $C^\infty(S^1, SO(2n))$ we consider the natural action in the space $L^2(S^1, \mathbb{R}^{2n})$ and define the operator of the complex structure in this space via Hilbert transform (see, for instance, [16]). Applying the spinor representation, we get the so-called basic representation of the loop group. To obtain other highest weight representations we apply restrictions and tensoring.

The group $OU(2\infty)$ admits a complexification $OGL(2\infty, \mathbb{C})$ consisting of complex matrices $g = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ that are orthogonal in the same sense with Hilbert–Schmidt blocks B and C. The spinor representation of $OU(2\infty)$ admits a holomorphic continuation to the complex group $OGL(2\infty)$, see [14], [16], the operators of the holomorphic continuation are unbounded in the initial topology, but are bounded on a certain dense Fréchet subspace equipped with its own topology.

This produces a highest weight representation of complex loop groups as free byproducts (see another approach in R. Goodman, N. Wallach [5]).

More interesting phenomenon arises for the group Diff of diffeomorphisms of circle, in this case the analytic continuation exists in spite of the nonexistence $\mathrm{Diff}_{\mathbb{C}}$, see [15].

[8]If a group acts on the space with measure, then it acts in L^2.

Acknowledgments

The author is grateful to Alexei Rosly and Victor Vasiliev for discussion of this subject.

References

[1] Akhiezer, D.N.; Gindikin, S.G. *On Stein extensions of real symmetric spaces.* Math. Ann. 286 (1990), no. 1–3, 1–12.

[2] Berezin, F.A. *Canonical transformations in the second quantization representation.* (Russian) Dokl. Akad. Nauk SSSR 150 1963 959–962.

[3] Berezin F.A. *The method of second quantization.* Nauka, Moscow, 1965; English transl.: Academic Press, 1966.

[4] Goodman, R., *Holomorphic representations of nilpotent Lie groups,* J. Funct. Anal., 31 (1979), 115–137.

[5] Goodman, R.; Wallach, N.R. *Structure and unitary cocycle representations of loop groups and the group of diffeomorphisms of the circle.* J. Reine Angew. Math. 347 (1984), 69–133.

[6] Graev, M.I., *Unitary representations of real semisimple Lie groups.* Trans. Moscow Math. Soc, v. 7., 1958, 335–389. English transl. in Amer. Math. Soc. Translations. Series 2, Vol. 66: Thirteen papers on group theory, algebraic geometry and algebraic topology.

[7] Kirillov, A.A. *Unitary representations of nilpotent Lie groups.* (Russian) Russian Math. Surveys, 17 (1962), 53–104.

[8] Kirillov, A.A. *Elements o f the theory of representations.* Springer-Verlag, Berlin-New York, 1976.

[9] Krötz, B.; Stanton, R.J. *Holomorphic extensions of representations. I. Automorphic functions.* Ann. of Math. (2) 159 (2004), no. 2, 641–724.

[10] Krötz, B.; Stanton, R.J. *Holomorphic extensions of representations. II. Geometry and harmonic analysis.* Geom. Funct. Anal. 15 (2005), no. 1, 190– 245.

[11] Litvinov, G.L. *On completely reducible representations of complex and real Lie groups.* Funct. Anal. Appl., v. 3 (1969), 332–334.

[12] Litvinov, G.L. *Group representations in locally convex spaces, and topological group algebras.* (Russian) Trudy Sem. Vektor. Tenzor. Anal. 16 (1972), 267–349; English transl. in Selecta Math. Soviet. 7 (1988), 101–182.

[13] Nelson, E., *Analytic vectors.* Ann. Math., 70 (1959), 572–615.

[14] Neretin, Yu.A. *On the spinor representation of* $\mathrm{O}(\infty, C)$. Dokl. Akad. Nauk SSSR 289 (1986), no. 2, 282–285. English transl.: Soviet Math. Dokl. 34 (1987), no. 1, 71–74.

[15] Neretin, Yu.A. *Holomorphic continuations of representations of the group of diffeomorphisms of the circle.* Mat. Sbornik 180 (1989), no. 5, 635–657, 720; translation in Russ. Acad. Sci. Sbornik. Math., v. 67 (1990); available via `www.mat.univie.ac.at/~neretin`

[16] Neretin, Yu.A. *Categories of symmetries and infinite-dimensional groups.* London Mathematical Society Monographs, 16, Oxford University Press, 1996.

[17] Neretin, Yu.A. *A construction of finite-dimensional faithful representation of Lie algebra.* Proceedings of the 22nd Winter School "Geometry and Physics" (Srni, 2002). Rend. Circ. Mat. Palermo (2) Suppl. No. 71 (2003), 159–161.

[18] Neretin Yu.A. *Lectures on Gaussian integral operators and classical groups.* Available via `http://www.mat.univie.ac.at/ neretin/lectures/lectures.htm`

[19] Olshanskii, G.I. *Invariant cones in Lie algebras, Lie semigroups and holomorphic discrete series.* Funct. Anal. Appl. 15, 275–285 (1982).

[20] Shale, D., Stinespring, W.F., *Spinor representations of infinite orthogonal groups.* J. Math. Mech. 14 1965 315–322.

Yuri A. Neretin
Math. Dept., University of Vienna
Nordbergstrasse, 15
Vienna, Austria

and

Institute for Theoretical and Experimental Physics
Bolshaya Cheremushkinskaya, 25
Moscow 117259, Russia

e-mail: `neretin@mccme.ru`

URL: `www.mat.univie.ac.at/~neretin, wwwth.itep.ru/~neretin`

Analysis and Mathematical Physics
Trends in Mathematics, 441–453
© 2009 Birkhäuser Verlag Basel/Switzerland

The Stability of Solitary Waves of Depression

Nguyet Thanh Nguyen and Henrik Kalisch

Abstract. We provide a sufficient condition for the orbital stability of negative solitary-wave solutions of the regularized long-wave equation. In particular, it is found that solitary waves with speed $c < -\frac{1}{6}$ are orbitally stable.

Mathematics Subject Classification (2000). Primary 35Q53; Secondary 35B35.

Keywords. Solitary waves, orbital stability, long wave equations.

1. Introduction

In this article, the dynamic stability of negative solitary-wave solutions of the regularized long-wave equation

$$u_t + u_x + \left(u^2\right)_x - u_{xxt} = 0, \tag{1.1}$$

is investigated. This equation which is also called the BBM equation, is used to model the propagation of small-amplitude surface waves on a fluid running in a long narrow channel. For an account of modeling properties of (1.1), the reader may consult the work of Benjamin et al. [6], Peregrine [14] and Whitham [17]. As is well known, equation (1.1) admits solitary-wave solutions of the form $u(x,t) = \Phi(x-ct)$. Indeed, when this ansatz is substituted into (1.1), there appears the ordinary differential equation

$$-c\Phi + \Phi + c\Phi'' + \Phi^2 = 0, \tag{1.2}$$

where $\Phi' = \frac{d\Phi}{d\xi}$, for $\xi = x - ct$. It is elementary to check that a solution of (1.2) is given by

$$\Phi(\xi) = \tfrac{3}{2}(c-1)\ \mathrm{sech}^2\Big(\tfrac{1}{2}\sqrt{\tfrac{c-1}{c}}\,\xi\Big). \tag{1.3}$$

These solutions are strictly positive progressive waves which propagate to the right without changing their profile over time. As can be seen from the expression (1.3), solitary waves with positive propagation velocity are defined only when $c > 1$. It is

This work was supported in part by the Research Council of Norway and the ESF Research Networking Programme HCAA.

well known that these positive solitary waves are dynamically stable with respect to small perturbations. As observed by one of the authors in [12], the formula (1.3) is still valid when $c < 0$, resulting in a solitary wave of depression which propagates to the left. Surprisingly, the stability of these solitary-waves of depression depends on the speed c. In fact, it was shown in [12] that for negative values of c close to zero, the solitary waves are unstable.

The original proof of stability of positive solitary waves was given by Benjamin [5] and Bona [7], using previous ideas of Boussinesq concerning the characterization of solitary waves as extremals of a constrained minimization problem [1, 9]. While the main thrust of their work was in the direction of the Korteweg-deVries equation

$$u_t + u_x + \left(u^2\right)_x + u_{xxx} = 0,$$

their proof is also applicable to the regularized long-wave equation (1.1). In fact, the proof of stability of positive solitary waves appears in the appendix of [5]. The method of Benjamin has subsequently been refined and extended, and a general theory has been developed [2, 4, 8, 10, 11, 16]. It appears however that almost all previous work has exclusively focused on positive solitary waves. In order to treat negative solitary waves, the general theory developed in [10, 16] cannot be applied straightforwardly, and it is our purpose here to indicate a complete proof of stability of negative solitary waves. Thus the main contribution of the present article is the proof of the following theorem.

Theorem. *The solitary wave* Φ *with velocity* c *is stable if* $c < -\frac{1}{6}$.

Observe that this theorem provides a sufficient condition for the stability of negative solitary waves. We must hasten to mention however that numerical computations in [12, 13] suggest that our result is not sharp. For the sake of clarity, we closely follow the original proof of Benjamin without paying much heed to the more general theory.

Figure 1 is depicting a stable solitary wave of depression, with velocity $c = -1.2$ and amplitude $\max_x |\Phi| = 3.3$, propagating to the left.

2. Preliminaries

As already observed by Benjamin and others [5, 6], a solitary wave cannot be stable in the strictest sense of the word. To understand this, consider two solitary-waves of different height, centered initially at the same point. Since the two waves have different amplitudes, they have different velocities according to the formula (1.3). As time passes the two waves will drift apart, no matter how small the initial difference was. However, in the situation just described, it is evident that two solitary waves with slightly differing height will stay similar in shape during the time evolution. Measuring the difference in shape therefore will give an acceptable notion of stability. This sense of orbital stability was introduced by Benjamin [5]. We say the solitary wave is orbital stable, if for a solution u of the equation (1.1)

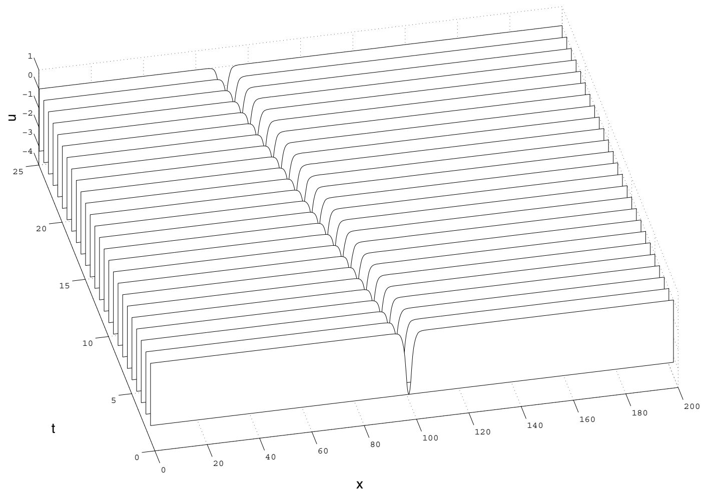


FIGURE 1. Stable solitary wave of depression with velocity $c = -1.2$.

that is initially sufficiently close to a solitary-wave will always stay close to a translation of the solitary-wave during the time evolution. A more mathematically precise definition is as follows. For any $\epsilon > 0$, consider the tube

$$U_\epsilon = \{u \in H^1 : \inf_s \|u - \tau_s \Phi\|_{H^1} < \epsilon\}, \tag{2.1}$$

where $\tau_s \Phi(x) = \Phi(x - s)$ is a translation of Φ. The set U_ϵ is an ϵ-neighborhood of the collection of all translates of Φ.

Definition 2.1. The solitary wave is **stable** if for any $\epsilon > 0$, there exists $\delta > 0$ such that if $u_0 = u(\cdot, 0) \in U_\delta$, then $u(\cdot, t) \in U_\epsilon, \forall t \in \mathbb{R}$. The solitary wave Φ is **unstable** if Φ is not stable.

The proof of stability is based on the conservation of certain integral quantities under the action of the evolution equation. Equation (1.1) has four invariant integrals. In particular, the functionals

$$V(u) = \tfrac{1}{2} \int_{-\infty}^{\infty} (u^2 + u_x^2)\, dx, \tag{2.2}$$

and

$$E(u) = \int_{-\infty}^{\infty} (\tfrac{1}{2} u^2 + \tfrac{1}{3} u^3)\, dx, \tag{2.3}$$

are critically important to the proof of stability of Φ. Note that $V(g) = \frac{1}{2}\|g\|_{H^1}^2$. The properties of these functionals are summarized in the following proposition.

Proposition 2.2. *Suppose u is a smooth solution of* (1.1) *with sufficient spatial decay. Then V and E are constant as functions of t, invariant with respect to spatial translations and continuous with respect to the $H^1(\mathbb{R})$-norm.*

Proof. The proof is standard. To see conservation in time for V and E, multiply the equation by u and $(-u-u^2+u_{xt})$, respectively. Translation invariance means

$$V(u)=V(\tau_s(u));\ \text{and } E(u)=E(\tau_s(u)),\ \forall s\in\mathbb{R}.$$

This follows immediately from the definition. Finally, we prove continuity of E with respect to the $H^1(\mathbb{R})$-norm. Let $\{w_n\}$ be any sequence in $H^1(\mathbb{R})$ such that $\lim_{n\to\infty}\|w_n-w\|_{H^1}=0$. Then

$$\begin{aligned}\big|E(w_n)-E(w)\big| &= \Big|\tfrac{1}{2}\int_{-\infty}^{\infty}[w_n^2-w^2]dx+\tfrac{1}{3}\int_{-\infty}^{\infty}[w_n^3-w^3]dx\Big|\\ &\le \tfrac{1}{2}\Big|\int_{-\infty}^{\infty}(w_n-w)(w_n+w)dx\Big|+\tfrac{1}{3}\Big|\int_{-\infty}^{\infty}(w_n-w)(w_n^2+w_nw+w^2)dx\Big|.\end{aligned}$$

Using the Cauchy-Schwarz inequality, this can be dominated by

$$\tfrac{1}{2}\|w_n-w\|_{L^2}\|w_n+w\|_{L^2}+\tfrac{1}{3}\|w_n-w\|_{L^2}\|w_n^2+w_nw+w^2\|_{L^2}.$$

Thus there appears the estimate

$$\begin{aligned}&|E(w_n)-E(w)|\\ &\le \|w_n-w\|_{H^1}\Big\{\tfrac{1}{2}\big(\|w_n\|_{H^1}+\|w\|_{H^1}\big)+\tfrac{1}{3}\big(\|w_n^2\|_{H^1}+\|w_nw\|_{H^1}+\|w^2\|_{H^1}\big)\Big\}.\end{aligned}$$

This expression approaches 0 as $n\to\infty$, because $w_n$and $w\in H^1(\mathbb{R})$ imply $\|w^rw_n^s\|_{H^1}<\infty$ for r, $s=0,1,2$; Thus, $\lim_{n\to\infty}|E(w_n)-E(w)|=0$. □

It is well known that the initial value problem for (1.1) is globally well posed. In fact, as soon as local existence is established, the conservation of the H^1-norm can be exploited to obtain a global solution. For the exact proof, the reader may consult the articles of Benjamin et al. [6] and Albert and Bona [3].

The notation used in this article is the standard notation in the theory of partial differential equations. Since all functions considered here are real-valued, we take the L^2-inner product to be $\langle f,g\rangle=\int_{-\infty}^{\infty}f(x)\,g(x)\,dx$. We will also have occasion to consider the L^2-inner product on the half-line, and this will be denoted by $\langle f,g\rangle_{L^2[0,\infty)}=\int_0^{\infty}f(x)\,g(x)\,dx$.

3. Orbital stability

In this section, orbital stability of the solitary waves of depression will be proved. Consider for a moment the difference in $L^2(\mathbb{R})$ of a solitary wave and a general solution of (1.1). Intuitively, for each u in $H^1(\mathbb{R})$, there is an $\alpha\in\mathbb{R}$, such that

$$\int_{-\infty}^{\infty}\Big\{u(\xi+\alpha(u))-\Phi(\xi)\Big\}^2d\xi=\inf_{a\in\mathbb{R}}\int_{-\infty}^{\infty}\Big\{u(\xi+a)-\Phi(\xi)\Big\}^2d\xi. \tag{3.1}$$

If the integral on the right is a differentiable function of a, and $\|u\|_{L^2} = \|\Phi\|_{L^2}$, then $\alpha(u)$ could be determined by solving the equation

$$\big\langle u(\cdot + \alpha(u)), \Phi' \big\rangle = 0. \tag{3.2}$$

A formal proof of the existence of a $\alpha(u)$ proceeds with the use the implicit function theorem as follows.

Proposition 3.1. *There is $\epsilon > 0$, such that there exists a C^1-mapping $\alpha : U_\epsilon \longrightarrow \mathbb{R}$, with the property that $\big\langle u(\cdot + \alpha(u)), \Phi' \big\rangle = 0$ for every $u \in U_\epsilon$.*

Proof. For a given $u \in U_\epsilon$, consider the functional

$$F : (u, \alpha) \longmapsto \int_{-\infty}^{\infty} u(\xi + \alpha(u))\Phi'(\xi)d\xi.$$

Observe that

$$\frac{dF(\Phi, 0)}{d\alpha} = \int_{-\infty}^{\infty} (\Phi'(\xi))^2 d\xi > 0,$$

and

$$F(\Phi, 0) = \int_{-\infty}^{\infty} \Phi(\xi)\Phi'(\xi)d\xi. = \frac{1}{2}\int_{-\infty}^{\infty} \frac{d}{d\xi}\Phi^2(\xi)d\xi = \frac{1}{2}[\Phi^2(\infty) - \Phi^2(-\infty)] = 0$$

Therefore, by the implicit function theorem, there exists a C^1-map $\alpha(u)$ near Φ such that

$$\big\langle u(\cdot + \alpha(u)), \Phi' \big\rangle = 0.$$

By translation invariance, the size of the neighborhood is the same everywhere. □

A crucial ingredient in the proof of stability is the fact that for all c less than some critical speed c^*, the functional $E(u)$ attains its minimum value when restricted to functions for which $V(u) = V(\Phi)$. In fact, we have the following explicit estimate.

Proposition 3.2. *If $c < -\frac{1}{6}$, there are $\beta > 0$, and $\epsilon > 0$ such that*

$$E(u) - E(\Phi) \geq \frac{\beta}{2}\,\|u(\cdot + \alpha(u)) - \Phi\|_{H^1}^2,$$

for all $u \in U_\epsilon$, satisfying $V(u) = V(\Phi)$.

Proof. The demonstration of this theorem follows the ideas outlined in the work of Benjamin [5]. In that work, however, the focus was on positive solitary waves. To accommodate negative solitary waves, the proof has to be modified accordingly.

For each u in U_ϵ such that $V(u) = V(\Phi)$, let $v = u(\cdot + \alpha(u)) - \Phi$, where α is defined in Proposition 3.1. Let $\Delta V = V(\Phi + v) - V(\Phi)$, and note that $\Delta V = 0$. However, according to the definition of V, we also have

$$\Delta V = \frac{1}{2}\int_{-\infty}^{\infty} \Big\{v^2 + v'^2 + 2\Phi v + 2\Phi' v'\Big\}d\xi. \tag{3.3}$$

Defining ΔE in a similar way, we see that

$$\Delta E = E(\Phi + v) - E(\Phi) = \int_{-\infty}^{\infty} \left\{ \tfrac{1}{2}v^2 + \Phi v + \Phi^2 v + \Phi v^2 + \tfrac{1}{3}v^3 \right\} d\xi. \tag{3.4}$$

On the other hand, since $\Delta V = 0$, we may also write

$$\Delta E = \Delta E - c\Delta V.$$

Therefore, in view of equations (3.3) and (3.4) there appears the expression

$$\Delta E = \int_{-\infty}^{\infty} \left\{ \tfrac{1}{2}v^2 + \Phi v^2 + \tfrac{1}{3}v^3 - \tfrac{1}{2}cv^2 - \tfrac{1}{2}cv'^2 \right\} d\xi + \int_{-\infty}^{\infty} \left\{ \Phi - c\Phi + \Phi^2 + c\Phi'' \right\} v \, d\xi.$$

Since Φ satisfies (1.2), the integrand of the second integral vanishes identically. Therefore,

$$\Delta E = \frac{1}{2}\int_{-\infty}^{\infty} \left\{ -cv'^2 + (-c + 1 + 2\Phi)v^2 \right\} d\xi + \frac{1}{3}\int_{-\infty}^{\infty} v^3 d\xi \, .$$

Thus we have $\Delta E = \delta^2 E + \delta^3 E$, where $\delta^2 E$ and $\delta^3 E$ are the second and third variation of E, respectively. In order to obtain a lower bound for $\delta^2 E$, v is written as the sum of an even function f and an odd function g. Since Φ itself is even, it can be shown directly that the even and odd parts of v contribute independently to $\delta^2 E$. In other words, we have

$$\delta^2 E = \delta^2 E(f) + \delta^2 E(g). \tag{3.5}$$

The contribution to $\delta^2 E$ from even functions is obtained as follows.

Lemma 3.3. *If $c < -\frac{1}{6}$, there are positive constants κ_1 and κ_2 such that*

$$\delta^2 E(f) \geq \kappa_1 \|f\|_{H^1}^2 - \kappa_2 \|v\|_{H^1}^3 .$$

Proof. The estimate for the lower bound of the contribution of $\delta^2 E(f)$, where f is an even function will be obtained by comparison with the integral

$$J = \int_0^{\infty} \left\{ 2\sqrt{\tfrac{c}{c-1}}\; f'^2 + \left[\tfrac{\mu}{2}\sqrt{\tfrac{c-1}{c}} + \tfrac{20}{3}\sqrt{\tfrac{1}{c(c-1)}}\; \Phi \right] f^2 \right\} d\,\xi,$$

where μ is a constant to be specified later. The substitution $z = \frac{1}{2}\sqrt{\frac{c-1}{c}}\;\xi$ yields $\Phi = \frac{3}{2}(c-1)\,\mathrm{sech}^2 z$, and puts the integral in the simpler form

$$J = \int_0^{\infty} \left\{ \left(\tfrac{\partial f}{\partial z} \right)^2 + \left(\mu - 20\,\mathrm{sech}^2 z \right) f^2 \right\} dz.$$

From here one, derivatives with respect to z will be denoted by $\frac{\partial}{\partial z}$, while derivatives with respect to ξ will be indicated by a prime. Moreover, the integrands will be interpreted as functions of z or ξ as indicated by the variable of integration. Next, we will make use of the spectral theory for a certain linear operator on the Hilbert space $L^2(0,\infty)$. Let the operator $\mathcal{L}_e$ be defined by

$$\mathcal{L}_e = -\frac{d^2}{dz^2} - 20\,\mathrm{sech}^2 z, \tag{3.6}$$

with domain defined as those functions $\phi \in H^2(0,\infty)$ that satisfy the boundary condition $\phi'(0) = 0$. As it turns out, $\mathcal{L}_e$ has only two negative eigenvalues:

$$\lambda_1 = -16 \text{ and } \lambda_2 = -4,$$

with corresponding normalized eigenfunctions

$$\psi_1 = \sqrt{\tfrac{35}{16}}\,\mathrm{sech}^4 z, \text{ and } \psi_2 = \sqrt{\tfrac{5}{8}}\,\big(6\,\mathrm{sech}^2 z - 7\,\mathrm{sech}^4 z\big). \tag{3.7}$$

FIGURE 2. The spectrum of $\mathcal{L}_e$.

The rest of the singular set consists of positive continuous spectrum. Now it can be seen from the expressions in (3.7) that ψ_1 and ψ_2 have $\mathrm{sech}^4 z$ as a common term. Thus we may write $\mathrm{sech}^2 z$ as a linear combination of ψ_1 and ψ_2. In particular,

$$\mathrm{sech}^2 z = \tfrac{7}{6}\sqrt{\tfrac{16}{35}}\,\psi_1 + \tfrac{1}{6}\sqrt{\tfrac{8}{5}}\,\psi_2. \tag{3.8}$$

The connection between J and $\mathcal{L}_e$ becomes apparent as follows.

$$\big\langle \mathcal{L}_e f, f\big\rangle_{L^2[0,\infty)} = \int_0^\infty \{f'^2 - 20\,\mathrm{sech}^2 z\; f^2\}dz.$$

One should recognize the right-hand side of this equation as the first and last term in the integral J. By the spectral theorem, the left-hand side of this equation is equal to

$$-16\;F_1^2 - 4\;F_2^2 + \int_0^\infty F^2(\lambda)\;\lambda d\rho(\lambda), \tag{3.9}$$

where $\rho(\lambda)$ is the spectral-function on $\mathbb{R}$. The coefficients F_1, F_2, and $F(\lambda)$ are defined by

$$F_{1,2} = \int_0^\infty \psi_{1,2}\; f(z)dz \quad \text{and} \quad F(\lambda) = \int_0^\infty \psi(z;\lambda) f(z)dz, \tag{3.10}$$

where $\psi(z;\lambda)$ is the generalized eigenfunction corresponding to the continuous spectrum of $\mathcal{L}_e$. Moreover, the spectral decomposition

$$f(z) = F_1\psi_1 + F_2\psi_2 + \int_0^\infty \psi(z;\lambda)F(\lambda)d\rho(\lambda),$$

and Parseval's identity give

$$\int_0^\infty f^2 dz = F_1^2 + F_2^2 + \int_0^\infty F^2 d\rho(\lambda). \tag{3.11}$$

From equation (3.9) and (3.11), we obtain a new form for the integral J as follows:

$$J = (\mu - 16)F_1^2 + (\mu - 4)F_2^2 + \int_0^\infty (\mu + \lambda)F^2 d\rho(\lambda). \tag{3.12}$$

Next, we introduce the notation $p = \|v\|_{H^1}$, and write F_2 as a linear combination of F_1 and p^2. Here, the constraint $\Delta V = 0$ will play a major role. According to equation (3.3), the constraint $\Delta V = 0$ is equivalent to

$$-p^2 = 2\int_{-\infty}^\infty (\Phi v + \Phi' v')\, d\xi.$$

On the right-hand side, use integration by parts in the last term yields

$$-p^2 = 2\int_{-\infty}^\infty (\Phi - \Phi'')\, v\, d\xi.$$

In light of equation (1.2) and the fact that Φ is an even function, this equation can be put in the form

$$-p^2 = \frac{4}{c}\int_0^\infty (\Phi + \Phi^2)\, f\, d\xi.$$

On the right-hand side of this equation, use the expression (1.3) for Φ, and make a change variable $z = \frac{1}{2}\sqrt{\frac{c-1}{c}}\ \xi$. Then (3.7) and (3.10) can be used to put the equation in the form

$$\int_0^\infty \operatorname{sech}^2 z\ f dz = -\tfrac{3}{2}\sqrt{\tfrac{16}{35}}\ (c-1)F_1 - \tfrac{1}{12}\sqrt{\tfrac{c}{c-1}}\ p^2. \tag{3.13}$$

Using the definition of F_2, (3.13), the definition of F_1, it appears that

$$F_2 = AF_1 + Ip^2,$$

where A and I are defined by

$$A = (-9c + 2)\sqrt{\tfrac{2}{7}}, \text{ and } I = -\tfrac{1}{2}\sqrt{\tfrac{5}{8}}\ \sqrt{\tfrac{c}{c-1}}. \tag{3.14}$$

Thus equation (3.12) becomes

$$J = \left[\mu - 16 + (\mu - 4)A^2\right]F_1^2 + (\mu - 4)(2AIF_1p^2 + I^2p^4) + \int_0^\infty (\mu + \lambda)\, F^2\, d\rho(\lambda).$$

Now for positive μ, the integral term is nonnegative. We choose μ is such a way that the coefficient of F_1^2 in the expression for J is nonnegative. Thus we need

$$\mu \geq \frac{4(4 + A^2)}{1 + A^2}. \tag{3.15}$$

Since $\frac{4(4+A^2)}{1+A^2} > 4$, the coefficient $(\mu - 4)$ is then automatically also positive. Therefore, with this choice of μ, J can be estimated below by

$$J \geq \frac{24}{1+A^2} AIF_1 p^2.$$

By using straightforward inequalities

$$F_1^2 \leq \int_0^\infty f^2 dz \leq \frac{1}{2}\int_{-\infty}^\infty v^2 dz \leq \frac{1}{2}\int_{-\infty}^\infty \left\{v^2 + \tfrac{1}{4}\tfrac{c-1}{c}\ v'^2\right\} dz$$
$$= \frac{1}{4}\sqrt{\tfrac{c-1}{c}} \int_{-\infty}^\infty \left\{v^2 + v'^2\right\} d\xi = \frac{1}{4}\sqrt{\tfrac{c-1}{c}}\ p^2,$$

we obtain the lower bound for the integral as

$$J \geq \frac{12AI}{1+A^2}\left(\tfrac{c-1}{c}\right)^{\frac{1}{4}} \|v\|_{H^1}^3. \tag{3.16}$$

Finally, a lower bound for the even contribution to $\delta^2 E$ is found as follows.

$$\delta^2 E(f) = \int_0^\infty \Big\{ -\tfrac{2c}{5} f'^2 + (-c+1)\big(1 - \tfrac{3}{20}\,\mu\big) f^2 \Big\} d\xi + \frac{3}{10}\sqrt{c(c-1)}\ J$$
$$\geq \frac{1}{2}\int_{-\infty}^\infty \Big\{ -\tfrac{2c}{5} f'^2 + (-c+1)\big(1 - \tfrac{3}{20}\mu\big) f^2 \Big\} d\xi$$
$$+ \frac{18AI}{5(1+A^2)}\sqrt{c(c-1)}\ \left(\tfrac{c-1}{c}\right)^{\frac{1}{4}} \|v\|_{H^1}^3.$$

Now we need the coefficient $1 - \frac{3}{20}\mu$ to be positive, and considering (3.15), this is possible only if $\frac{20}{3} > \mu \geq \frac{4(4+A^2)}{1+A^2}$. But by (3.14), this inequality can be satisfied only if $c < -\frac{1}{6}$. Thus as long as $c < -\frac{1}{6}$, we have the estimate

$$\delta^2 E(f) \geq \kappa_1 \int_{-\infty}^\infty \left(f'^2 + f^2\right) d\xi - \kappa_2 \|v\|_{H^1}^3, \tag{3.17}$$

where $\kappa_1 = \min\big(-\frac{c}{5}, \frac{1}{2}(-c+1)(1 - \frac{3}{20}\mu)\big)$, and $\kappa_2 = \frac{-18AI}{5(1+A^2)}\sqrt{c(c-1)}\ \left(\frac{c-1}{c}\right)^{\frac{1}{4}}$ are positive constants. □

Next, we turn to the contribution to $\delta^2 E$ from the odd part of v.

Lemma 3.4. *For all $c < 0$, there holds the estimate*

$$\delta^2 E(g) \geq \frac{-c}{8} \|g\|_{H^1}^2.$$

Proof. For the odd contribution to $\delta^2 E$, the result in Proposition 3.1 will play an important role. By virtue of this result, we have

$$\int_{-\infty}^\infty (\Phi + v)\Phi' d\xi = 0.$$

Since Φ is even, this is the same as

$$\int_{-\infty}^{\infty} g\Phi' d\xi = 0. \tag{3.18}$$

Let s be a positive constant, and consider the linear operator

$$\mathscr{L}_o = \frac{1}{s}\Big[-\frac{d^2}{d\xi^2} - 3s\,\mathrm{sech}^2\big(\tfrac{1}{2}\sqrt{s}\;\xi\big)\Big], \tag{3.19}$$

defined on those functions $\phi \in H^2(0,\infty)$ that satisfy the boundary condition $\phi(0) = 0$. As illustrated in Figure 3, $\mathscr{L}_o$ has only one negative eigenvalue $\lambda_1 = -1$ with corresponding eigenfunction

$$\theta_1 = \frac{d}{d\xi}\left\{3s\,\mathrm{sech}^2\big(\tfrac{1}{2}\sqrt{s}\;\xi\big)\right\}, \tag{3.20}$$

and the rest of the spectrum of $\mathscr{L}_o$ is positive continuous [5].

FIGURE 3. The spectrum of $\mathscr{L}_o$.

Now the spectral theorem asserts that

$$\langle \mathscr{L}_o g, g\rangle_{L^2[0,\infty)} = -\langle g, \theta_1\rangle^2_{L^2[0,\infty)} + \int_0^\infty \lambda G^2(\lambda) d\rho(\lambda),$$

where $G(\lambda) = \int_0^\infty \theta(\xi;\lambda) g(\xi) d\xi$, and $\theta(\xi;\lambda)$ is a generalized eigenfunction of $\mathscr{L}_o$. If we now choose $s = \frac{c-1}{c}$, where $c < 0$ then it follows from (3.18) that

$$\langle g, \theta_1\rangle_{L^2[0,\infty)} = 0.$$

Thus we obtain

$$\langle \mathscr{L}_o g, g\rangle_{L^2[0,\infty)} \geq 0. \tag{3.21}$$

On the other hand, after integration by parts, there appears

$$0 \leq \langle \mathscr{L}_o g, g\rangle_{L^2[0,\infty)} = \frac{c}{c-1}\int_0^\infty \Big\{ g'^2 - 3\tfrac{c-1}{c}\,\mathrm{sech}^2\big(\tfrac{1}{2}\sqrt{\tfrac{c-1}{c}}\;\xi\big)\; g^2\Big\} d\xi.$$

Thus it is immediate that the following integral is nonnegative.

$$K = \int_0^\infty \left\{ g'^2 - \tfrac{2}{c}\Phi\, g^2\right\} d\xi \geq 0. \tag{3.22}$$

The contribution to $\delta^2 E$ due to the odd part of v may now be estimated by comparison with the integral K.

$$\delta^2 E(g) = -\frac{3c}{4}K + \frac{1}{2}\int_0^\infty \left\{\Phi + \tfrac{3}{2}(-c+1)\right\} g^2 d\xi + \frac{1}{4}\int_0^\infty \left\{-cg'^2 + (-c+1)\, g^2\right\} d\xi.$$

In light of (3.22), $-c+1 \geq -c$, and the fact that $\Phi \geq \frac{3}{2}(c-1)$ [cf.(1.3) with $c<0$], there follows

$$\delta^2 E(g) \geq \frac{-c}{8} \int_{-\infty}^{\infty} (g'^2 + g^2) d\xi \ . \tag{3.23}$$

□

Proof of Proposition 3.2*:* Using (3.5), Lemma (3.3), and Lemma (3.4), it is plain that

$$\delta^2 E \geq \beta \int_{-\infty}^{\infty} (v'^2 + v^2) d\xi - \kappa_2 \|v\|^3_{H^1}, \tag{3.24}$$

where $\beta = \min(\frac{-c}{8}, \kappa_1)$ is a positive constant, and κ_1 and κ_2 are defined in (3.17).

Now using the inequality $\sup_{\xi\in\mathbb{R}} |v(\xi)| \leq \frac{1}{\sqrt{2}} \|v\|_{H^1}$ [1], there appears an estimate for $\delta^3 E$.

$$\delta^3 E = -\frac{1}{3} \int_{-\infty}^{\infty} (-v) v^2 d\xi \geq -\frac{1}{3} \sup |v| \int_{-\infty}^{\infty} v^2 d\xi \geq -\frac{1}{3\sqrt{2}} \|v\|^3_{H^1}. \tag{3.25}$$

Combining (3.24) and (3.25) yields the final estimate

$$\Delta E \geq \beta \int_{-\infty}^{\infty} (v'^2 + v^2) d\xi - \gamma \|v\|^3_{H^1} = \|v\|^2_{H^1} (\beta - \gamma \|v\|_{H^1}).$$

where $\gamma = \kappa_2 + \frac{1}{3\sqrt{2}}$. Therefore, if $\|v\|_{H^1}$ is sufficiently small, say $\|v\|_{H^1} < \frac{\beta}{2\gamma}$, we obtain

$$\Delta E \geq \frac{\beta}{2} \|v\|^2_{H^1}.$$

□

Finally, we will close this section by showing a necessary condition for stability of the solitary-wave.

Theorem. *The solitary wave* Φ *with velocity* c *is stable if* $c < -\frac{1}{6}$.

Proof. The proof is based on the techniques of of Bona, Grillakis, Souganidis Shatah, and Strauss in [8, 10]. In particular, the theorem will be proved by contradiction as follows. Suppose Φ is not stable, then there exists an $\epsilon > 0$, and a sequence of initial data $u_n^0 \in H^1(\mathbb{R})$ and corresponding solutions $u_n \in C(\mathbb{R}, H^1(\mathbb{R}))$ with $u_n(\cdot, 0) = u_0^n$, such that

$$\lim_{n\to\infty} \|u_n^0 - \Phi\|_{H^1} = 0, \tag{3.26}$$

but

$$\sup_{t>0} \inf_{s\in\mathbb{R}} \|u_n(\cdot, t) - \tau_s \Phi(\cdot)\|_{H^1} \geq \tfrac{1}{2}\epsilon,$$

for large enough n. By the continuity of u_n in t, we can pick the first time t_n so that

$$\inf_{s\in\mathbb{R}} \|u_n(\cdot, t_n) - \tau_s \Phi(\cdot)\|_{H^1} = \tfrac{1}{2}\,\epsilon. \tag{3.27}$$

In other words, $u_n(\cdot, t_n) \in \partial U_{\frac{1}{2}\epsilon}$.

[1]This is known as the Sobolev lemma. The reader may refer to [5] for a simple proof.

Since V is continuous in $H^1(\mathbb{R})$ and invariant under time evolution, we have $\lim_{n\to\infty} V(u_n^0) = V(\Phi)$, and consequently

$$\lim_{n\to\infty} V(u(\cdot,t_n)) = V(\Phi). \tag{3.28}$$

Choose a sequence $w_n \in H^1(\mathbb{R})$, such that $V(w_n) = V(\Phi)$ and $\lim_{n\to\infty} \|w_n - u_n(\cdot,t_n)\|_{H^1} = 0$. [2] Note that by H^1-continuity of E, and time invariance,

$$\lim_{n\to\infty} \big[E(w_n) - E(\Phi)\big] = 0,$$

and also note that $w_n \in U_\epsilon$ for large n. On the other hand, so long as ϵ is small enough, Proposition (3.2) shows that

$$E(w_n) - E(\Phi) \geq \frac{\beta}{2}\|w_n(\cdot + \alpha(w_n)) - \Phi\|_{H^1}^2,$$

where β is the constant defined in (3.24). Therefore, since $\alpha(u)$ is a continuous function, it appears that

$$\lim_{n\to\infty} \|u_n(\cdot,t_n) - \Phi(\cdot - \alpha(u_n(\cdot,t_n)))\|_{H^1} = 0.$$

Finally, this is a contradiction to (3.27) □

References

[1] J.P. Albert, *Concentration compactness and the stability of solitary-wave solutions to nonlocal equations.* Applied analysis (Baton Rouge, LA, 1996), 1–29, Contemp. Math. **221**, Amer. Math. Soc., Providence, RI, 1999.

[2] J.P. Albert and J.L. Bona, *Total positivity and the stability of internal waves in stratified fluids of finite depth.* The Brooke Benjamin special issue (University Park, PA, 1989). IMA J. Appl. Math. **46** (1991), 1–19.

[3] J.P. Albert and J.L. Bona, *Comparisons between model equations for long waves.* J. Nonlinear Sci. **1** (1991), 345–374.

[4] J.P. Albert and J.L. Bona and D.B. Henry, *Sufficient conditions for stability of solitary-wave solutions of model equations for long waves.* Phys. D **24** (1987), 343–366.

[5] T.B. Benjamin, *The stability of solitary waves.* Proc. Roy. Soc. London A **328** (1972), 153–183.

[6] T.B. Benjamin, J.B. Bona, and J.J. Mahony, *Model equations for long waves in nonlinear dispersive systems.* Philos. Trans. Roy. Soc. London A **272** (1972), 47–78.

[7] J.L. Bona, *On the stability theory of solitary waves.* Proc. Roy. Soc. London A **344** (1975), 363–374.

[8] J.L. Bona, P.E. Souganidis and W.A. Strauss, *Stability and instability of solitary waves of Korteweg-de Vries type.* Proc. Roy. Soc. London A **411** (1987), 395–412.

[2] The sequence defined by $w_n = (\|\Phi\|_{H^1}/\|u_n\|_{H^1})\, u_n(\cdot,t_n)$ will do the job.

[9] J. Boussinesq, *Théorie des ondes et des remous qui se propagent le long d'un canal rectangulaire horizontal, en communiquant au liquide contenu dans ce canal des vitesses sensiblement pareilles de la surface au fond.* J. Math. Pures Appl. **17** (1872), 55–108.

[10] M. Grillakis, J. Shatah and W.A. Strauss, *Stability theory of solitary waves in the presence of symmetry.* J. Funct. Anal. **74** (1987), 160–197.

[11] I.D. Iliev, E.Kh. Khristov and K.P. Kirchev, *Spectral methods in soliton equations.* Pitman Monographs and Surveys in Pure and Applied Mathematics **73**. Longman Scientific & Technical, Harlow; copublished in the United States with John Wiley & Sons, Inc., New York, 1994.

[12] H. Kalisch, *Solitary Waves of Depression.* J. Computational Analysis and Application **8** (2006), 5–24.

[13] N. Nguyen and H. Kalisch, *Orbital Stability of Negative Solitary Waves*, to appear in Math. Comput. Simulation.

[14] P.G. Peregrine, *Calculations of the development of an undular bore.* J. Fluid Mech. **25** (1966), 321–330.

[15] J. Shatah and W. Strauss, *Instability of nonlinear bound states.* Commun. Math. Phys. **100** (1985), 173–190.

[16] P.E. Souganidis and W.A. Strauss, *Instability of a class of dispersive solitary waves.* Proc. Roy. Soc. Edinburgh **114A** (1990), 195–212.

[17] G.B. Whitham, *Linear and Nonlinear Waves.* Wiley, New York, 1974.

Nguyet Thanh Nguyen
Department of Mathematics
University of Bergen
Johannes Brunsgate 12
5008 Bergen, Norway
e-mail: `Nguyet.Nguyen@math.uib.no`

Henrik Kalisch
Department of Mathematics
University of Bergen
Johannes Brunsgate 12
5008 Bergen, Norway
e-mail: `henrik.kalisch@math.uib.no`

Analysis and Mathematical Physics
Trends in Mathematics, 455–463
© 2009 Birkhäuser Verlag Basel/Switzerland

Singular and Tangent Slit Solutions to the Löwner Equation

Dmitri Prokhorov and Alexander Vasil'ev

Abstract. We consider the Löwner differential equation generating univalent maps of the unit disk (or of the upper half-plane) onto itself minus a single slit. We prove that the circular slits, tangent to the real axis are generated by Hölder continuous driving terms with exponent 1/3 in the Löwner equation. Singular solutions are described, and the critical value of the norm of driving terms generating quasisymmetric slits in the disk is obtained.

Mathematics Subject Classification (2000). Primary 30C35, 30C20; Secondary 30C62.

Keywords. Univalent function, Löwner equation, Slit map.

1. Introduction

Let $\mathbb{D} = \{z \in \mathbb{C} : \ |z| < 1\}$ be the unit disk and $\mathbb{T} := \partial\mathbb{D}$. The famous Löwner equation was introduced in 1923 [3] in order to represent a dense subclass of the whole class of univalent conformal maps $f(z) = z(1 + c_1 z + \cdots)$ in $\mathbb{D}$ by the limit

$$f(z) = \lim_{t\to\infty} e^t w(z,t), \quad z \in \mathbb{D},$$

where $w(z,t) = e^{-t} z(1 + c_1(t)z + \cdots)$ is a solution to the equation

$$\frac{dw}{dt} = -w\frac{e^{iu(t)} + w}{e^{iu(t)} - w}, \quad w(z,0) \equiv z, \tag{1}$$

with a continuous driving term $u(t)$ on $t \in [0,\infty)$, see [3, page 117]. All functions $w(z,t)$ map $\mathbb{D}$ onto $\Omega(t) \subset \mathbb{D}$. If $\Omega(t) = \mathbb{D} \setminus \gamma(t)$, where $\gamma(t)$ is a Jordan curve in $\mathbb{D}$ except one of its endpoints, then the driving term $u(t)$ is uniquely defined and we call the corresponding map w a *slit map*. However, from 1947 [5] it is known that

The first author was partially supported by the Russian Foundation for Basic Research (grant 07-01-00120) and the second by the grants of the Norwegian Research Council #177355/V30, the European Science Foundation RNP HCAA, and the Scandinavian Network Analysis and Applications (Nordforsk).

solutions to (1) with continuous $u(t)$ may give non-slit maps, in particular, $\Omega(t)$ can be a family of hyperbolically convex digons in $\mathbb{D}$.

Marshall and Rohde [4] addressed the following question: *Under which condition on the driving term* $u(t)$ *the solution to* (1) *is a slit map?* Their result states that if $u(t)$ is Lip(1/2) (Hölder continuous with exponent 1/2), and if for a certain constant $C_{\mathbb{D}} > 0$, the norm $\|u\|_{1/2}$ is bounded $\|u\|_{1/2} < C_{\mathbb{D}}$, then the solution w is a slit map, and moreover, the Jordan arc $\gamma(t)$ is s quasislit (a quasiconformal image of an interval within a Stolz angle). As they also proved, a converse statement without the norm restriction holds. The absence of the norm restriction in the latter result is essential. On one hand, Kufarev's example [5] contains $\|u\|_{1/2} = 3\sqrt{2}$, which means that $C_{\mathbb{D}} \leq 3\sqrt{2}$. On the other hand, Kager, Nienhuis, and Kadanoff [1] constructed exact slit solutions to the half-plane version of the Löwner equation with arbitrary norms of the driving term.

Let us give here the half-plane version of the Löwner equation. Let $\mathbb{H} = \{z : \operatorname{Im} z > 0\}$, $\mathbb{R} = \partial\mathbb{H}$. The functions $h(z,t)$, normalized near infinity by $h(z,t) = z - 2t/z + b_{-2}(t)/z^2 + \cdots$, solving the equation

$$\frac{dh}{dt} = \frac{-2}{h - \lambda(t)}, \quad h(z,0) \equiv z, \tag{2}$$

where $\lambda(t)$ is a real-valued continuous driving term, map $\mathbb{H}$ onto a subdomain of $\mathbb{H}$. The question about the slit mappings and the behaviour of the driving term $\lambda(t)$ in the case of the half-plane $\mathbb{H}$ was addressed by Lind [2]. The techniques used by Marshall and Rohde carry over to prove a similar result in the case of the equation (2), see [4, page 765]. Let us denote by $C_{\mathbb{H}}$ the corresponding bound for the norm $\|\lambda\|_{1/2}$. The main result by Lind is the sharp bound, namely $C_{\mathbb{H}} = 4$.

In some papers, e.g., [1, 2], the authors work with equations (1, 2) changing $(-)$ to $(+)$ in their right-hand sides, and with the mappings of slit domains onto $\mathbb{D}$ or $\mathbb{H}$. However, the results remain the same for both versions.

Marshall and Rohde [4] remarked that there exist many examples of driving terms $u(t)$ which are not Lip(1/2), but which generate slit solutions with simple arcs $\gamma(t)$. In particular, if $\gamma(t)$ is tangent to $\mathbb{T}$, then $u(t)$ is never Lip(1/2).

Our result states that if $\gamma(t)$ is a circular arc tangent to $\mathbb{R}$, then the driving term $\lambda(t) \in$Lip(1/3). Besides, we prove that $C_{\mathbb{D}} = C_{\mathbb{H}} = 4$, and consider properties of singular solutions to the one-slit Löwner equation.

The authors are grateful for the referee's remarks which improved the presentation.

2. Circular tangent slits

We shall work with the half-plane version of the Löwner equation and with the sign $(+)$ in the right-hand side, consequently with the maps of slit domains onto $\mathbb{H}$.

We construct a mapping of the half-plane $\mathbb{H}$ slit along a circular arc $\gamma(t)$ of radius 1 centered on i onto $\mathbb{H}$ starting at the origin directed, for example, positively. The inverse mapping we denote by $z = f(w,t) = w - 2t/w + \cdots$. Then

$\zeta = 1/f(w,t)$ maps $\mathbb{H}$ onto the lower half-plane slit along a ray co-directed with $\mathbb{R}^+$ and having the distance 1/2 between them. Let ζ_0 be the tip of this ray. Applying the Christoffel-Schwarz formula we find f in the form

$$\frac{1}{f(w,t)} = \int\limits_0^{1/w} \frac{(1-\gamma w)\,dw}{(1-\alpha w)^2(1-\beta w)} = \frac{\beta-\gamma}{(\alpha-\beta)^2}\log\frac{w-\alpha}{w-\beta} + \frac{\alpha-\gamma}{\alpha-\beta}\frac{1}{w-\alpha}, \quad (3)$$

where the branch of logarithm vanishes at infinity, and $f(w,t)$ is expanded near infinity as

$$f(w,t) = w - \frac{2t}{w} + \cdots$$

The latter expansion gives us two conditions: there is no constant term and the coefficient is $-2t$ at w, which implies $\gamma = 2\alpha + \beta$ and $\alpha(\alpha + 2\beta) = -6t$. The condition $\operatorname{Im}\zeta_0 = -1/2$ yields

$$\frac{-2\alpha}{(\alpha-\beta)^2} = \frac{1}{2\pi}.$$

Then, $\beta = \alpha + 2\sqrt{-\alpha\pi}$, and $\alpha(3\alpha + 4\sqrt{-\alpha\pi}) = -6t$. Considering the latter equation with respect to α we expand the solution $\alpha(t)$ in powers of $t^{1/3}$. Hence,

$$\alpha(t) = -\left(\frac{9}{4\pi}\right)^{1/3} t^{2/3} + A_2 t + A_3 t^{4/3} + \cdots$$

and

$$\beta(t) = (12\pi)^{1/3} t^{1/3} + B_2 t^{2/3} + \cdots$$

Formula (3) in the expansion form regarding to $1/w$ gives

$$\frac{\beta-\alpha}{2\pi}\,\frac{1}{w} + \frac{\beta^2-\alpha^2}{4\pi}\,\frac{1}{w^2} + \cdots + \left(1 + 2\frac{\alpha}{\beta} + 2\frac{\alpha^2}{\beta^2} + \cdots\right)\left(\frac{1}{w} + \frac{\alpha}{w^2} + \cdots\right) = \zeta. \quad (4)$$

Remember that this formula is obtained under the conditions $\gamma = 2\alpha + \beta$ and $(\alpha - \beta)^2 = 4\alpha\pi$. We substitute the expansions of $\alpha(t)$ and $\beta(t)$ in this formula and consider it as an equation for the implicit function $w = h(z,t)$. Calculating coefficients $B_2 \dots B_4$ in terms of $A_2, \dots, A_4$, and verifying $A_2 = -3/4\pi$ we come to the following expansion for $h(z,t)$:

$$w = h(z,t) = h(\frac{1}{\zeta}, t) = \frac{1}{\zeta} + 2\zeta t + \frac{3}{2}(12\pi)^{1/3} t^{4/3} + \cdots.$$

This version of the Löwner equation admits the form

$$\frac{dh}{dt} = \frac{2}{h - \lambda(t)}, \quad h(z,0) \equiv z. \quad (5)$$

Being extended onto $\mathbb{R} \setminus \lambda(0)$ the function $h(z,t)$ satisfies the same equation. Let us consider $h(z,t)$, $z \in \widehat{\mathbb{H}} \setminus \lambda(0)$ with a singular point at $\lambda(0)$, where $\widehat{\mathbb{H}}$ is the closure of $\mathbb{H}$. Then

$$\lambda(t) = h(z,t) - \frac{2}{dh(z,t)/dt} = \lambda(0) + (12\pi)^{1/3} t^{1/3} + \cdots$$

about the point $t = 0$. Thus, the driving term $\lambda(t)$ is Lip(1/3) about the point $t = 0$ and analytic for the rest of the points t.

Remark 2.1. *The radius of the circumference is not essential for the properties of* $\lambda(t)$*. Passing from* $h(z,t)$ *to the function* $\frac{1}{r}h(rz,t)$ *we recalculate the coefficients of the function* $h(z,t)$ *and the corresponding coefficients in the expansion of* $\lambda(t)$ *that depend continuously on* r*. Therefore, they stay within bounded intervals whenever* r *ranges within the bounded interval.*

Remark 2.2. *In particular, the expansion for* $h(z,t)$ *reflects the Marshall and Rohde's remark* [4, page 765] *that the tangent slits can not be generated by driving terms from* Lip(1/2)*.*

3. Singular solutions for slit images

Suppose that the Löwner equation (5) with driving term $\lambda(t)$ generates a map $h(z,t)$ from $\Omega(t) = \mathbb{H} \setminus \gamma(t)$ onto $\mathbb{H}$, where $\gamma(t)$ is a quasislit. Extending h to the boundary $\partial\Omega(t)$ we obtain a correspondence between $\gamma(t) \subset \partial\Omega(t)$ and a segment $I(t) \subset \mathbb{R}$, while the remaining boundary part $\mathbb{R} = \partial\Omega(t) \setminus \gamma(t)$ corresponds to $\mathbb{R} \setminus I(t)$. The latter mapping is described by solutions to the Cauchy problem for the differential equation (5) with the initial data $h(x,0) = x \in \mathbb{R} \setminus \lambda(0)$. The set $\{h(x,t) : x \in \mathbb{R} \setminus \lambda(0)\}$ gives $\mathbb{R} \setminus I(t)$, and $\lambda(t)$ does not catch $h(x,t)$ for all $t \geq 0$, see [2] for details.

The image $I(t)$ of $\gamma(t)$ can be also described by solutions $h(\lambda(0),t)$ to (5), but the initial data $h(\lambda(0),0) = \lambda(0)$ forces h to be singular at $t = 0$ and to possess the following properties.

(i) There are two singular solutions $h^-(\lambda(0),t)$ and $h^+(\lambda(0),t)$ such that $I(t) = [h^-(\lambda(0),t), h^+(\lambda(0),t)]$.
(ii) $h^{\pm}(\lambda(0),t)$ are continuous for $t \geq 0$ and have continuous derivatives for all $t > 0$.
(iii) $h^-(\lambda(0),t)$ is strictly decreasing and $h^+(\lambda(0),t)$ is strictly increasing, so that $h^-(\lambda(0),t) < \lambda(t) < h^+(\lambda(0),y)$.

We will focus on studying the singularity character of $h^{\pm}$ at $t = 0$.

Theorem 3.1. *Let the Löwner differential equation* (5) *with the driving term* $\lambda \in$ Lip(1/2)*,* $\|\lambda\|_{1/2} = c$*, generate slit maps* $h(z,t) : \mathbb{H} \setminus \gamma(t) \to \mathbb{H}$ *where* $\gamma(t)$ *is a quasislit. Then* $h^+(\lambda(0),t)$ *satisfies the condition*

$$\lim_{t\to 0+} \sup \frac{h^+(\lambda(0),t) - h^+(\lambda(0),0)}{\sqrt{t}} \leq \frac{c + \sqrt{c^2+16}}{2},$$

and this estimate is the best possible.

Proof. Assume without loss of generality that $h^+(\lambda(0),0) = \lambda(0) = 0$. Denote $\varphi(t) := h^+(\lambda(0),t)/\sqrt{t}$, $t > 0$. This function has a continuous derivative and satisfies the differential equation

$$t\varphi'(t) = \frac{2}{\varphi(t) - \lambda(t)/\sqrt{t}} - \frac{\varphi(t)}{2}.$$

This implies together with Property (iii) that $\varphi'(t) > 0$ iff

$$\frac{\lambda(t)}{\sqrt{t}} < \varphi(t) < \varphi_1(t) := \frac{\lambda(t)}{2\sqrt{t}} + \sqrt{\frac{\lambda^2(t)}{4t} + 4}.$$

Observe that $\varphi_1(t) \leq A := (c + \sqrt{c^2 + 16})/2$.

Suppose that $\lim_{t\to 0+} \sup \varphi(t) = B > A$, including the case $B = \infty$. Then there exists $t^* > 0$, such that $\varphi(t^*) > B - \epsilon > A$, for a certain $\epsilon > 0$. If $B = \infty$, then replace $B - \epsilon$ by $B' > A$. Therefore, $\varphi'(t^*) < 0$ and $\varphi(t)$ increases as t runs from t^* to 0. Thus, $\varphi(t) > B - \epsilon$ for all $t \in (0, t^*)$ and we obtain from (5) that

$$\frac{dh^+(\lambda(0), t)}{dt} \leq \frac{2}{\sqrt{t}(B - \epsilon - c)},$$

for such t. Integrating this inequality we get

$$h^+(\lambda(0), t) \leq \frac{4\sqrt{t}}{B - \epsilon - c} < \frac{4\sqrt{t}}{A - c},$$

that contradicts our supposition. This proves the estimate of Theorem 3.1.

In order to attain the equality sign in Theorem 3.1, one chooses $\lambda(t) = c\sqrt{t}$. Then $h^+(\lambda(0), t) = A\sqrt{t}$ solves equation (5) with singularity at $t = 0$. This completes the proof. □

Remark 3.1. *Estimates similar to Theorem* 3.1 *hold for the other singular solution* $h^-(\lambda(0).t)$.

Remark 3.2. *Let us compare Theorem* 3.1 *with the results from Section* 2. *The image of a circular arc* $\gamma(t) \subset \mathbb{H}$ *tangent to* $\mathbb{R}$ *is* $I(t) = [h^-(\lambda(0), t), h^+(\lambda(0), t)]$, *where* $h^-(\lambda(0), t) = \alpha(t) = -(9/4\pi)^{1/3} t^{2/3} + \cdots$, *and* $h^+(\lambda(0), t) = \beta(t) = (12\pi)^{1/3} t^{1/3} + \cdots$, *so that* $h^-(\lambda(0), t) \in \mathrm{Lip}(2/3)$ *and* $h^+(\lambda(0), t) \in \mathrm{Lip}(1/3)$.

Remark 3.3. *Singular solutions to the differential equation* (5) *appear not only at* $t = 0$ *but at any other moment* $\tau > 0$. *More precisely, there exist two families* $h^-(\gamma(\tau), t)$ *and* $h^+(\gamma(\tau), t)$, $\tau \geq 0$, $t \geq \tau$, *of singular solutions to* (5) *that describe the image of arcs* $\gamma(t)$, $t \geq \tau$ *under map* $h(z, t)$. *They correspond to the initial data* $h(\gamma(\tau), \tau) = \lambda(\tau)$ *in* (5) *and satisfy the inequalities* $h^-(\gamma(\tau), t) < \lambda(t) < h^+(\gamma(\tau), t)$, $t > \tau$. *These two families of singular solutions have no common inner points and fill in the set*

$$\{(x, t) : h^-(\lambda(0), t) \leq x \leq h^+(\lambda(0), t), 0 \leq t \leq t_0\},$$

for some t_0.

4. Critical norm values for driving terms

In this section we discuss the results and techniques of Marshall and Rohde [4] and Lind [2]. The authors of [4] proved the existence of $C_{\mathbb{D}}$ such that driving terms $u(t) \in \mathrm{Lip}(1/2)$ with $\|u\|_{1/2} < C_{\mathbb{D}}$ in (1) generate quasisymmetric slit maps. This result remains true for an absolute number $C_{\mathbb{H}}$ in the half-plane version of the Löwner differential equation (2), see, e.g., [2].

Lind [2] claimed that the disk version (1) of the Löwner differential equation is 'more challenging', than the half-plane version (2). Working with the half-plane version she showed that $C_{\mathbb{H}} = 4$. The key result is based on the fact that if $\lambda(t) \in \mathrm{Lip}(1/2)$ in (2), and $h(x,t) = \lambda(t)$, say at $t = 1$, then $\Omega(t) = h(\mathbb{H}, t)$ is not a slit domain and $\|\lambda\|_{1/2} \geq 4$. Moreover, there is an example of $\lambda(t) = 4 - 4\sqrt{1-t}$ that yields $h(2,1) = \lambda(1)$. Although there may be more obstacles for generating slit half-planes than that of the driving term λ catching up some solution h to (2), Lind showed that this is basically the only obstacle. The latter statement was proved by using techniques of [4].

We will modify here the main Lind's reasonings so that they could be applied to the disk version of the Löwner equation. After that it remains to refer to [4] and [2] to state that $C_{\mathbb{D}}$ also equals 4.

Suppose that slit disks $\Omega(t)$ correspond to $u \in \mathrm{Lip}(1/2)$ in (1) with the sign '+' in its right-hand side instead of '−'. Then the maps $w(z,t)$ are extended continuously to $\mathbb{T} \setminus \{e^{iu(0)}\}$. Let $z_0 \in \mathbb{T} \setminus \{e^{iu(0)}\}$, and let $\alpha(t, \alpha_0) := \arg w(z_0, t)$ be a solution to the following real-valued initial value problem

$$\frac{d\alpha(t)}{dt} = \cot\frac{\alpha - u}{2}, \quad \alpha(0) = \alpha_0. \tag{6}$$

Similarly, suppose that slit half-planes $\Omega(t)$ correspond to $\lambda \in \mathrm{Lip}(1/2)$ in (2) with the sign '+' in its right-hand side instead of '−'. Then the maps $h(z,t)$ are extended continuously to $\mathbb{R} \setminus \lambda(0)$. Let $x_0 \in \mathbb{R} \setminus \lambda(0)$ and let $x(t, x_0) := h(x_0, t)$ be a solution to the following real-valued initial value problem

$$\frac{dx(t)}{dt} = \frac{2}{x(t) - \lambda(t)}, \quad x(t_0) = x_0. \tag{7}$$

For all $t \geq 0$, $\tan((\alpha(t) - u(t))/2) \neq 0$ in (6), and $x(t) - \lambda(t) \neq 0$ in (7) (see [2] for the half-plane version). Let us show a connection between the solutions $\alpha(t)$ to (6), and $x(t)$ to (7), where the driving terms $u(t)$ and $\lambda(t)$ correspond to each other.

Lemma 4.1. *Given $\lambda(t) \in \mathrm{Lip}(1/2)$, there exists $u(t) \in \mathrm{Lip}(1/2)$, such that equations (6) and (7) have the same solutions. Conversely, given $u(t) \in \mathrm{Lip}(1/2)$ there exists $\lambda(t) \in \mathrm{Lip}(1/2)$, such that equations (6) and (7) have the same solutions.*

Proof. Given $\lambda(t) \in \mathrm{Lip}(1/2)$, denote by $x(t, x_0)$ a solution to the initial value problem (7). Then the solution $\alpha(t, \alpha_0)$ to the initial value problem (6) is equal to $x(t, \alpha_0)$ when

$$\tan\frac{\alpha - u}{2} = \frac{x - \lambda}{2}$$

and

$$x_0 = \lambda(0) + 2\tan\frac{\alpha_0 - u(0)}{2}.$$

The function $u(t)$ is normalized by choosing

$$u(0) = x_0 - \arctan\frac{x_0 - \lambda(0)}{2}.$$

This condition makes α_0 and x_0 equal. Hence, the first part of Lemma 1 is true if we put

$$u(t) = x(t, x_0) - 2\arctan\frac{x(t, x_0) - \lambda(t)}{2}. \tag{8}$$

Obviously, (8) preserves the Lip(1/2) property.

Conversely, given $u(t) \in \text{Lip}(1/2)$, a solution $x(t, x_0)$ is equal to $\alpha(t, \alpha_0)$ when

$$\lambda(t) = \alpha(t, \alpha_0) - 2\tan\frac{\alpha(t, \alpha_0) - u(t)}{2}. \tag{9}$$

Again (9) preserves the Lip(1/2) property. This ends the proof. □

Observe that in some extreme cases relations (8) or (9) preserve not only the Lipschitz class but also its norm. Lind [2] gave an example of the driving term $\lambda(t) = 4 - 4\sqrt{1-t}$ in (7). It is easily verified that $x(t, 2) = 4 - 2\sqrt{1-t}$. If $t = 1$, then $x(1, 2) = \lambda(1) = 4$, and λ cannot generate slit half-plane at $t = 1$. This implies that $C_{\mathbb{H}} \leq 4$. Going from (7) to (6) we use (8) to put

$$u(t) = x(t, 2) - 2\arctan\frac{x(t, 2) - \lambda(t)}{2} = 4 - 2\sqrt{1-t} - 2\arctan\sqrt{1-t}.$$

From Lemma 4.1 we deduce that $\alpha(1, 2) = u(1)$. Hence u cannot generate slit disk at $t = 1$, and $C_{\mathbb{D}} \leq \|u\|_{1/2}$. Since

$$\sup_{0\leq t<1}\frac{u(1) - u(t)}{\sqrt{1-t}} = \sup_{0\leq t<1}\left(2 + 2\frac{\arctan\sqrt{1-t}}{\sqrt{1-t}}\right) = 4,$$

we have that $\|u\|_{1/2} \leq 4$. It is now an easy exercise to show that $\|u\|_{1/2} = 4$. This implies that $C_{\mathbb{D}} \leq 4$.

Lemma 4.2. *Let $u \in \text{Lip}(1/2)$ in (6) with $u(0) = 0$ and $\alpha_0 \in (0, \pi)$. Suppose that $\alpha(t)$ is a solution to (6) and $\alpha(1) = u(1)$. Then $\|u\|_{1/2} \geq 4$.*

Proof. Observe that $\alpha(t)$ is increasing on $[0, 1]$, and $\alpha(t) - u(t) > 0$ on $(0, 1)$. Let $u \in \text{Lip}(1/2)$ in (3), and $\|u\|_{1/2} = c$. Then,

$$\alpha(t) - u(t) \leq \alpha(1) - u(1) + c\sqrt{1-t} = c\sqrt{1-t}. \tag{10}$$

Given $\epsilon > 0$, there exists $\delta > 0$, such that

$$\tan\frac{c\sqrt{1-t}}{2} < \frac{c\sqrt{1-t}}{2}(1 + \epsilon),$$

for $1-\delta < t < 1$ and all $0 < c \leq 4$. We apply this inequality to (6) and obtain that

$$\frac{d\alpha}{dt} \geq \cot\frac{c\sqrt{1-t}}{2} > \frac{2}{c\sqrt{1-t}(1+\epsilon)}.$$

Integrating gives that

$$\alpha(1) - \alpha(t) \geq \frac{4\sqrt{1-t}}{c(1+\epsilon)}.$$

This allows us to improve (10) to

$$\alpha(t) - u(t) \leq \alpha(1) - \frac{4\sqrt{1-t}}{c(1+\epsilon)} - u(1) + c\sqrt{1-t} = \left(c - \frac{4}{c(1+\epsilon)}\right)\sqrt{1-t}. \quad (11)$$

Repeating these iterations we get

$$\alpha(t) - u(t) \leq c_n\sqrt{1-t},$$

where $c_0 = c$, $c_{n+1} = c - 4/[(1+\varepsilon)c_n]$, and $c_n > 0$. Let g_n be recursively defined by (see Lind [2])

$$g_1(y) = y - \frac{4}{y}, \quad g_n(y) = y - \frac{4}{g_{n-1}(y)}, \quad n \geq 2.$$

It is easy to check that $c_n < g_n((1+\varepsilon)c) < (1+\varepsilon)c_n$

Lind [2] showed that $g_n(y_n) = 0$ for an increasing sequence $\{y_n\}$, and $g_{n+1}(y)$ is an increasing function from (y_n, ∞) to $\mathbb{R}$. So $c(1+\epsilon) > y_n$ for all n, and it remains to apply Lind's result [2] that $\lim_{n\to\infty} y_n = 4$. Hence, $c \geq 4/(1+\epsilon)$. The extremal estimate is obtained if $\epsilon \to 0$ which leads to $c \geq 4$. This completes the proof. □

Now Lind's reasonings in [2] based on the techniques from [4] give a proof of the following statement.

Proposition 4.1. *If* $u \in \mathrm{Lip}(1/2)$ *with* $\|u\|_{1/2} < 4$, *then the domains* $\Omega(t)$ *generated by the Löwner differential equation* (1) *are disks with quasislits.*

In other words, Proposition 4.1 states that $C_{\mathbb{D}} = C_{\mathbb{H}} = 4$.

References

[1] W. Kager, B. Nienhuis, L.P. Kadanoff, *Exact solutions for Löwner evolutions*, J. Statist. Phys. **115** (2004), no. 3–4, 805–822.

[2] J. Lind, *A sharp condition for the Löwner equation to generate slits*, Ann. Acad. Sci. Fenn. Math. **30** (2005), no. 1, 143–158.

[3] K. Löwner, *Untersuchungen über schlichte konforme Abbildungen des Einheitskreises. I*, Math. Ann. **89** (1923), no. 1–2, 103–121.

[4] D.E. Marshall, S. Rohde, *The Löwner differential equation and slit mappings*, J. Amer. Math. Soc. **18** (2005), no. 4, 763–778.

[5] P.P. Kufarev, *A remark on integrals of Löwner's equation*, Doklady Akad. Nauk SSSR (N.S.) **57**, (1947). 655–656 (in Russian).

Dmitri Prokhorov
Department of Mathematics and Mechanics,
Saratov State University,
Astrakhanskay Str. 83,
410012, Saratov, Russia
e-mail: prokhorov@sgu.ru

Alexander Vasil'ev
Department of Mathematics,
University of Bergen,
Johannes Brunsgate 12, Bergen 5008, Norway
e-mail: alexander.vasiliev@math.uib.no

Analysis and Mathematical Physics
Trends in Mathematics, 465–471
© 2009 Birkhäuser Verlag Basel/Switzerland

A Remark on Amoebas in Higher Codimensions

Alexander Rashkovskii

Abstract. It is shown that tube sets over amoebas of algebraic varieties of dimension q in $\mathbb{C}_*^n$ (and, more generally, of almost periodic holomorphic chains in $\mathbb{C}^n$) are q-pseudoconcave in the sense of Rothstein. This is a direct consequence of a representation of such sets as supports of positive closed currents.

Mathematics Subject Classification (2000). Primary 32A60; Secondary 32A07, 42A75.

Keywords. Amoeba, k-pseudoconvexity, almost periodic chain, closed positive current.

1. Introduction

Let V be an algebraic variety in $\mathbb{C}_*^n = (\mathbb{C} \setminus 0)^n$. Its image $\mathcal{A}_V = \operatorname{Log} V$ under the mapping $\operatorname{Log}(z_1, \dots, z_n) = (\log|z_1|, \dots, \log|z_n|)$ is called the *amoeba* of A. The notion was introduced in [9] and has found numerous applications in complex analysis and algebraic geometry, see the survey [11].

The amoeba of V is a closed set with non-empty complement $\mathcal{A}_V^c = \mathbb{R}^n \setminus \mathcal{A}_V$. If V is of codimension 1, then each component of $\mathcal{A}_V^c$ is convex because $\operatorname{Log}^{-1}(\mathcal{A}_V^c)$ is the intersection of a family of domains of holomorphy. This is no longer true for varieties of higher codimension; nevertheless, some rudiments of convexity do take place. As shown by Henriques [10], if $\operatorname{codim} V = k$, then $\mathcal{A}_V^c$ is $(k-1)$-convex, a notion defined in terms of homology groups for sections by k-dimensional affine subspaces. A local result, due to Mikhalkin [11], states that $\mathcal{A}_V$ has no *supporting k-cap*, i.e., a ball B in a k-dimensional plane such that $\mathcal{A}_V \cap B$ is nonempty and compact, while $\mathcal{A}_V \cap (B + \epsilon v) = \emptyset$ for some $v \in \mathbb{R}^n$ and all sufficiently small $\epsilon > 0$.

The notion of amoeba was adapted by Favorov [3] to zero sets of holomorphic almost periodic functions in a tube domain as "shadows" cast by the zero sets to the base of the domain; a precise definition is given in Section 4. In [2], Henriques' result was extended to amoebas of zero sets of so-called regular holomorphic almost periodic mappings. This was done by a reduction to the case considered in [10] where the proof was given by methods of algebraic geometry.

In this note, we propose a different approach to convexity properties of amoebas in higher codimensions. It is purely analytical and works equally well for both algebraic and almost periodic situations. Moreover, we get our (pseudo)convexity results as a by-product of a representation of an amoeba as the support of a certain natural measure determined by the "density" of the zero set.

Let us start with a hypersurface case. When $V = \{P(z) = 0\} \subset \mathbb{C}_*^n$ is defined by a Laurent polynomial P, the function

$$N_P(y) = \frac{1}{(2\pi)^n} \int_{[-\pi,\pi]^n} \log |P(e^{y_1+i\theta_1}, \ldots, e^{y_n+i\theta_n})| \, d\theta,$$

known in tropical mathematics community as the *Ronkin function*, is convex in $\mathbb{R}^n$ and linear precisely on each connected component of $\mathcal{A}_V^c$. This means that the support of the current $dd^c N_P(\operatorname{Im} z)$ equals $T_{\mathcal{A}_V} = \mathbb{R}^n + i\mathcal{A}_V$, the tube set in $\mathbb{C}^n$ with base $\mathcal{A}_V$. Since the complement to the support of a positive closed current of bidegree $(1,1)$ is pseudoconvex (being a domain of existence for a pluriharmonic function), this implies pseudoconvexity of $T_{\mathcal{A}_V^c}$ and thus convexity of every component of $\mathcal{A}_V^c$. Of course, the function N_P gives much more than simply generating the amoeba (see, for example, [18], [8], [12]).

The same reasoning applies to amoebas of holomorphic almost periodic functions f with Ronkin's function N_P replaced by the mean value $\mathcal{M}_f(y)$ of $\log |f|$ over the real planes $\{x + iy : \, x \in \mathbb{R}^n\}$, $y \in \mathbb{R}^n$.

What we will do in the case of codimension $k > 1$, is presenting $T_{\mathcal{A}_V}$ as the support of a closed positive current of bidegree (k,k) (namely, a mean value current for the variety or, more generally, for a holomorphic chain) and then using a theorem on $(n-k)$-pseudoconcavity, in the sense of Rothstein, of supports of such currents due to Fornaess and Sibony [7]. In addition, we show that for a closed set $\Gamma \subset \mathbb{R}^n$, Rothstein's $(n-k)$-pseudoconcavity of T_Γ implies the absence of k-supporting caps of Γ (Proposition 2.2).

We obtain our main result, Theorem 4.1, for arbitrary almost periodic holomorphic chains, which is a larger class than zero sets of regular almost periodic holomorphic mappings, and the situation with algebraic varieties (Corollary 4.2) is a direct consequence. The existence of the mean value currents was established in [4], so we just combine it together with the theorem on supports of positive closed currents. In this sense, this note is just a simple illustration of how useful the mean value currents are.

2. Rothstein's q-pseudoconvexity

We will use the following notion of q-pseudoconvexity, due to W. Rothstein [19], see also [14]. Given $0 < q < n$ and $\alpha, \beta \in (0,1)$, the set

$$H = \{(z,w) \in \mathbb{C}^{n-q} \times \mathbb{C}^q : \, \|z\|_\infty < 1, \; \|w\|_\infty < \alpha \text{ or } \beta < \|z\|_\infty < 1, \; \|w\|_\infty < 1\}$$

is called an $(n-q,q)$-Hartogs figure; here $\|z\|_\infty = \max_j |z_j|$. Note that its convex hull $\hat{H}$ is the unit polydisc in $\mathbb{C}^n$. An open subset Ω of a complex n-dimensional manifold M is said to be *q-pseudoconvex* in M if for any $(n-q,q)$-Hartogs figure H and a biholomorphic map $\Phi : \hat{H} \to M$, the condition $\Phi(H) \subset \Omega$ implies $\Phi(\hat{H}) \subset \Omega$. If this is the case, we will also say that $M \setminus \Omega$ is *q-pseudoconcave* in M.

Loosely speaking, the q-pseudoconvexity is the Kontinuitätssatz with respect to $(n-q)$-polydiscs; usual pseudoconvexity is equivalent to $(n-1)$-pseudoconvexity.

Theorem 2.1. ([7], Cor. 2.6) *The support of a positive closed current of bidimension (q,q) on a complex manifold M is q-pseudoconcave in M.*

It is easy to see that for tube sets, $(n-k)$-pseudoconcavity implies absence of k-caps in the sense of Mikhalkin.

Proposition 2.2. *Let Γ be a closed subset of a convex open set $D \subset \mathbb{R}^n$. If the tube set $T_\Gamma = \mathbb{R}^n + i\Gamma$ is $(n-k)$-pseudoconcave in the tube domain $T_D = \mathbb{R}^n + iD$, then Γ has no k-supporting caps.*

Proof. Assume Γ has a k-supporting cap B. We assume that the vector v in the definition of the cap is orthogonal to B (in the general case, one gets an image of a Hartogs figure under a non-degenerate linear transform). Choose coordinates in $\mathbb{R}^n$ such that

$$B = \{(y', y'') \in \mathbb{R}^k \times \mathbb{R}^{n-k} : \ \|y'\|_\infty < 1, \ y'' = 0\},$$

$$\{\beta < \|y'\|_\infty < 1, \ \|y''\|_\infty < 1\} \subset D \setminus \Gamma, \quad \beta \in (0,1),$$

and $B + \epsilon v \subset D \setminus \Gamma$ for all $\epsilon \in (0,1)$, where $v = (0, v'')$, $\|v''\|_\infty = 1$. Since D is open, $\{y : \ \|y'\|_\infty < 1, \ \|y'' - \frac{1}{2}v''\|_\infty < \alpha\} \subset D \setminus \Gamma$ for some $\alpha \in (0,1)$. Therefore, the $\frac{i}{2}v''$-shift of the corresponding $(k, n-k)$-Hartogs figure H is a subset of the tube set $T_D \setminus T_\Gamma$. Since B is a subset of the shifted polydisc $\hat{H} + \frac{i}{2}v''$ and $B \cap \Gamma \neq \emptyset$, the set T_Γ is not $(n-k)$-pseudoconcave. ☐

3. Almost periodic holomorphic chains

Here we recall some facts from Ronkin's theory of holomorphic almost periodic mappings and currents; for details, see [15], [17], [4], [5], and the survey [6].

Let $\mathcal{T}_t$ denote the translation operator on $\mathbb{R}^n$ by $t \in \mathbb{R}^n$, then for any function f on $\mathbb{R}^n$, $(\mathcal{T}_t^* f)(x) = f(\mathcal{T}_t x) = f(x+t)$.

A continuous mapping f from $\mathbb{R}^n$ to a metric space X is called *almost periodic* if the set $\{\mathcal{T}_t^* f\}_{t \in \mathbb{R}^n}$ is relatively compact in $C(\mathbb{R}^n, X)$ with respect to the topology of uniform convergence on $\mathbb{R}^n$. The collection of all almost periodic mappings from $\mathbb{R}^n$ to X will be denoted by $\mathrm{AP}(\mathbb{R}^n, X)$.

As is known from classical theory of almost periodic functions, any function $f \in \mathrm{AP}(\mathbb{R}^n, \mathbb{C})$ has its mean value $\mathcal{M}_f$ over $\mathbb{R}^n$,

$$\mathcal{M}_f = \lim_{s \to \infty} (2s)^{-n} \int_{\Pi_s} f \, dm_n,$$

where $\Pi_s = \{x \in \mathbb{R}^n : \ \|x\|_\infty < s\}$ and m_n is the Lebesgue measure in $\mathbb{R}^n$.

Let D be a convex domain in $\mathbb{R}^n$, $T_D = \mathbb{R}^n + iD$. A continuous mapping $f : T_D \to X$ is called *almost periodic on* T_D if $\{\mathcal{T}_t^* f\}_{t\in\mathbb{R}^n}$ is a relatively compact subset of $C(T_D, X)$ with respect to the topology of uniform convergence on each tube subdomain $T_{D'}$, $D' \Subset D$. The collection of all almost periodic mappings from T_D to X will be denoted by $\mathrm{AP}(T_D, X)$.

The set $\mathrm{AP}(T_D, \mathbb{C})$ can be defined equivalently as the closure (with respect to the topology of uniform convergence on each tube subdomain $T_{D'}$, $D' \Subset D$) of the collection of all exponential sums with complex coefficients and pure imaginary exponents (frequencies). The mean value of $f \in \mathrm{AP}(T_D, \mathbb{C})$ is a continuous function of $\operatorname{Im} z$. The collection of all holomorphic mappings $f \in \mathrm{AP}(T_D, \mathbb{C}^k)$ will be denoted by $\mathrm{HAP}(T_D, \mathbb{C}^k)$. In particular, any mapping from $\mathbb{C}^n$ to $\mathbb{C}^k$, whose components are exponential sums with pure imaginary frequencies, belongs to $\mathrm{HAP}(\mathbb{C}^n, \mathbb{C}^k)$.

The notion of almost periodicity can be extended to distributions. For example, a measure μ on T_D is called almost periodic if $\phi(t) = \int (\mathcal{T}_t)_* \phi \, d\mu \in \mathrm{AP}(\mathbb{R}^n, \mathbb{C})$ for every continuous function ϕ with compact support in T_D. Furthermore, it can be extended to holomorphic chains as follows.

Let $Z = \sum_j c_j V_j$ be a holomorphic chain on $\Omega \subset \mathbb{C}^n$ supported by an analytic variety $|Z| = \cup_j V_j$ of pure dimension q. Its integration current $[Z]$ acts on test forms ϕ of bidegree (q, q) with compact support in Ω (shortly, $\phi \in \mathcal{D}_{q,q}(\Omega)$) as

$$([Z], \phi) = \int_{\mathrm{Reg}\,|Z|} \gamma_Z \phi = \sum_j c_j \int_{\mathrm{Reg}\ V_j} \phi,$$

where the function γ_Z takes constant positive integer values on the connected components of $\mathrm{Reg}\,|Z|$. The q-dimensional volume of Z in a Borel set $\Omega_0 \subset \Omega$ is

$$\mathrm{Vol}_Z(\Omega_0) = \int_{\Omega_0 \cap \mathrm{Reg}\,|Z|} \gamma_Z \beta_q$$

(the mass of the trace measure of $[Z]$ in Ω_0). If f is a holomorphic mapping on Ω such that $|Z| = f^{-1}(0)$ and $\gamma_Z(z)$ equals the multiplicity of f at z, the chain will be denoted by Z_f.

A q-dimensional holomorphic chain Z on T_D is called an *almost periodic holomorphic chain* if $(\mathcal{T}_t^*[Z], \phi) \in AP(T_D, \mathbb{C})$ for any test form $\phi \in \mathcal{D}_{q,q}(T_D)$. Here $\mathcal{T}_t^* S = \sum \alpha_{IJ}(z+t)\, dz^I \wedge d\bar{z}^J$ is the pullback of the current $S = \sum \alpha_{IJ}(z)\, dz_I \wedge d\bar{z}_J$.

For any $f \in \mathrm{HAP}(T_D, \mathbb{C})$, the chain (divisor) Z_f is always almost periodic; on the other hand, there exist almost periodic divisors (starting already from dimension $n = 1$) that are not divisors of any holomorphic almost periodic function; when $n > 1$, even a periodic divisor need not be the divisor of a periodic holomorphic function [16]. The situation with higher-dimensional mappings is even worse, since the chain Z_f generated by $f \in \mathrm{HAP}(T_D, \mathbb{C}^k))$, $k > 1$, need not be almost periodic [4]. It is however so if the mapping f is *regular*, that is, if $\operatorname{codim} |Z_g| = k$ or $|Z_g| = \emptyset$ for every mapping g from the closure of the set $\{\mathcal{T}_t^* f\}_{t\in\mathbb{R}^m}$ [4], [5]. A sufficient regularity condition [15] shows that such mappings are generic.

Now we can turn to construction of the current that plays central role in our considerations, the details can be found in [5]. Let Z be an almost periodic holomorphic chain of dimension q. For any test form $\phi \in \mathcal{D}_{q,q}(T_D)$, the mean value $\mathcal{M}_{\phi_Z}$ of the function $\phi_Z(t) := (\mathcal{T}_t^*[Z], \phi) \in \mathrm{AP}(\mathbb{R}^n, \mathbb{C})$ defines the *mean value current* $\mathcal{M}_Z$ of Z by the relation

$$(\mathcal{M}_Z, \phi) = \mathcal{M}_{\phi_Z}.$$

The current is closed and positive. Since $\mathcal{M}_Z$ is translation invariant with respect to x, its coefficients have the form $\mathcal{M}_{IJ} = m_n \otimes \mathcal{M}'_{IJ}$, where $\mathcal{M}'_{IJ}$ are Borel measures in D. In addition, if $\psi = \sum \psi_{IJ} dz_I \wedge d\bar{z}_J$ is a form with coefficients $\psi_{IJ} \in \mathcal{D}(D)$ and χ_s is the characteristic function of the cube Π_s, then there exists the limit

$$\lim_{s\to\infty} (2s)^{-n}([Z], \chi_s \psi) = (\mathcal{M}'_Z, \psi'),$$

where $\mathcal{M}'_Z = \sum \mathcal{M}'_{IJ} dy_I \wedge dy_J$ and $\psi' = \sum \psi_{IJ} dy_I \wedge dy_J$.

The trace measure $\mu_Z = \mathcal{M}_Z \wedge \beta_q$ can also be written as $\mu_Z = m_n \otimes \mu'_Z$, where μ'_Z is a positive Borel measure on D. The following result shows that it can be viewed as a density of the chain Z along $\mathbb{R}^n$.

Theorem 3.1. ([4], [5]) *Let Z be an almost periodic holomorphic chain in a tube domain T_D. For any open set $G \Subset D$ such that $\mu'_Z(\partial G) = 0$, one has*

$$\lim_{s\to\infty} (2s)^{-n} \operatorname{Vol}_Z(\Pi_s + iG) = \mu'_Z(G);$$

in addition, $\mu'_Z(G) = 0$ if and only if $|Z| \cap T_G = \emptyset$.

Remark 3.2. For $Z = Z_f$ with regular $f \in \mathrm{HAP}(T_D, \mathbb{C}^k)$, Theorem 3.1 was proved in [15] (for $k = n$) and [13] ($k < n$), without using the notion of almost periodic chain. The current $\mathcal{M}_{Z_f}$ can be constructed as follows. The coefficients a_{IJ} of the current $\log|f|(dd^c \log|f|)^{k-1}$ are locally integrable functions on T_D, almost periodic in the sense of distributions: $(\mathcal{T}_t^* a_{IJ}, \phi) \in \mathrm{AP}(T_D, \mathbb{C})$ for any test function $\phi \in \mathcal{D}(T_D)$. Therefore, they possess their mean values $A_{IJ} = \mathcal{M}_{a_{IJ}}$, and the current $\mathcal{M}_{Z_f} = dd^c(\sum A_{IJ} dz_I \wedge d\bar{z}_J)$.

4. Amoebas

Following [3], if Z is an almost periodic holomorphic chain in T_D, then its *amoeba* $\mathcal{A}_Z$ is the closure of the projection of $|Z|$ to D:

$$\mathcal{A}_Z = \overline{\operatorname{Im}|Z|},$$

where the map $\operatorname{Im} : \mathbb{C}^n \to \mathbb{R}^n$ is defined by $\operatorname{Im}(z_1, \dots, z_n) = (\operatorname{Im} z_1, \dots, \operatorname{Im} z_n)$. When $Z = Z_f$ for a regular mapping $f \in \mathrm{HAP}(T_D, \mathbb{C}^p)$, we write simply $\mathcal{A}_f$.

Our convexity result is stated in terms of the tube set $T_{\mathcal{A}_Z} = \mathbb{R}^n + i\mathcal{A}_Z$.

Theorem 4.1. *If Z is an almost periodic holomorphic chain of dimension q in a tube domain $T_D \subseteq \mathbb{C}^n$, then $T_{\mathcal{A}_Z} = \operatorname{supp} \mathcal{M}_Z$, where $\mathcal{M}_Z$ is the mean value current of the chain Z. Therefore, $T_{\mathcal{A}_Z}$ is q-pseudoconcave in T_D. In particular, for any regular mapping $f \in \mathrm{HAP}(T_D, \mathbb{C}^k)$, the set $T_{\mathcal{A}_f}$ is $(n-k)$-pseudoconcave.*

Proof. By Theorem 3.1, $\mathcal{A}_Z = \operatorname{supp} \mu'_Z$, which can be rewritten as

$$T_{\mathcal{A}_Z} = \operatorname{supp} m_n \otimes \mu'_Z = \operatorname{supp} \mathcal{M}_Z.$$

Since the current $\mathcal{M}_Z$ is positive and closed, Theorem 2.1 implies the corresponding pseudoconcavity. □

This covers the algebraic case as well by means of the map $\mathrm{E} : \mathbb{C}^n \to \mathbb{C}^n_*$, $\mathrm{E}\,(z_1, \dots, z_n) = (e^{-iz_1}, \dots, e^{-iz_n})$. For a Laurent polynomial P, the exponential sum E^*P is periodic in $T_{\mathbb{R}^n}$, and its mean value $\mathcal{M}_{\log|E^*P|}$ coincides with the Ronkin function N_P. Furthermore, given an algebraic variety $V \subset \mathbb{C}^n_*$, its pullback E^*V is almost periodic (actually, periodic) in $\mathbb{C}^n$ and $\mathcal{A}_{\mathrm{E}^*V} = \mathcal{A}_V$, which gives

Corollary 4.2. *The set $T_{\mathcal{A}_V^c}$ for an algebraic variety $V \subset \mathbb{C}^n_*$ of pure codimension k is $(n-k)$-pseudoconvex.*

References

[1] J.-P. Demailly, *Monge-Ampère operators, Lelong numbers and intersection theory.* Complex Analysis and Geometry (Univ. Series in Math.), ed. by V. Ancona and A. Silva, Plenum Press, New York, 1993, 115–193.

[2] A. Fabiano, J. Guenot, J. Silipo, *Bochner transforms, perturbations and amoebae of holomorphic almost periodic mappings in tube domains.* Complex Var. Elliptic Equ. **52** (2007), no. 8, 709–739.

[3] S.Yu. Favorov, *Holomorphic almost periodic functions in tube domains and their amoebas.* Comput. Methods Funct. Theory **1** (2001), no. 2, 403–415.

[4] S.Yu. Favorov, A.Yu. Rashkovskii and L.I. Ronkin, *Almost periodic currents and holomorphic chains.* C. R. Acad. Sci. Paris **327**, Série I (1998), 302–307.

[5] S.Yu. Favorov, A.Yu. Rashkovskii and L.I. Ronkin, *Almost periodic currents, divisors and holomorphic chains.* Israel. Math. Conf. Proc. **15** (2001), 67–88.

[6] S.Yu. Favorov and A.Yu. Rashkovskii, *Holomorphic almost periodic functions.* Acta Appl. Math. **65** (2001), 217–235.

[7] J.E. Fornaess and N. Sibony, *Oka's inequality for currents and applications.* Math. Ann. **301** (1995), no. 3, 399–419.

[8] M. Forsberg, M. Passare, A. Tsikh, *Laurent determinants and arrangements of hyperplane amoebas.* Adv. Math. **151** (2000), no. 1, 45–70.

[9] I.M. Gelfand, M.M. Kapranov, A.V. Zelevinsky, *Discriminants, resultants, and multidimensional determinants.* Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1994.

[10] A. Henriques, *An analogue of convexity for complements of amoebas of varieties of higher codimension, an answer to a question asked by B. Sturmfels.* Adv. Geom. **4** (2004), no. 1, 61–73.

[11] G. Mikhalkin, *Amoebas of algebraic varieties and tropical geometry.* Different faces of geometry, 257–300, Int. Math. Ser. (N. Y.), 3, Kluwer/Plenum, New York, 2004.

[12] M. Passare and H. Rullgård, *Monge-Ampère measures, and triangulations of the Newton polytope.* Duke Math. J. **121** (2004), no. 3, 481–507.

[13] A. Rashkovskii, *Currents associated to holomorphic almost periodic mappings.* Mat. Fizika, Analiz, Ceometria **2** (1995), no. 2, 250–269.

[14] O. Riemenschneider, *Über den Flächeninhalt analytischer Mengen und die Erzeugung k-pseudokonvexer Gebiete.* Invent. Math. **2** (1967), 307–331.

[15] L.I. Ronkin, *Jessen's theorem for holomorphic almost periodic mappings.* Ukrainsk. Mat. Zh. **42** (1990), 1094–1107.

[16] L.I. Ronkin, *Holomorphic periodic functions and periodic divisors.* Mat. Fizika, Analiz i Geometria **2** (1995), no. 1, 108–122.

[17] L.I. Ronkin, *Almost periodic distributions and divisors in tube domains.* Zap. Nauchn. Sem. POMI **247** (1997), 210–236.

[18] L.I. Ronkin, *On zeros of almost periodic functions generated by functions holomorphic in a multicircular domain.* Complex analysis in modern mathematics, 239–251, FAZIS, Moscow, 2001.

[19] W. Rothstein, *Ein neuer Beweis des Hartogsschen Hauptsatzes und seine Ausdehnung auf meromorphe Funktionen.* Math. Z. **53** (1950). 84–95.

Alexander Rashkovskii
Tek/nat, University of Stavanger
4036 Stavanger, Norway
e-mail: alexander.rashkovskii@uis.no

Analysis and Mathematical Physics
Trends in Mathematics, 473–505
© 2009 Birkhäuser Verlag Basel/Switzerland

Quadratic Differentials and Weighted Graphs on Compact Surfaces

Alexander Yu. Solynin

Abstract. We prove that for every simply connected graph Γ embedded in a compact surface $\mathcal{R}$ of genus $g \geq 0$, whose edges e^i_{kj} carry positive weights w^i_{kj}, there exist a complex structure on $\mathcal{R}$ and a Jenkins-Strebel quadratic differential $Q(z)\,dz^2$, whose critical graph Φ_Q complemented, if necessary, by second degree vertices on its edges, is homeomorphic to Γ on $\mathcal{R}$ and carries the same set of weights. In other words, *every positive simply connected graph on $\mathcal{R}$ can be analytically embedded in $\mathcal{R}$*. We also discuss a problem on the extremal partition of $\mathcal{R}$ relative to such analytical embedding. As a consequence, we establish the existence of systems of disjoint simply connected domains on $\mathcal{R}$ with a prescribed combinatorics of their boundaries, which carry proportional harmonic measures on their boundary arcs.

Keywords. Quadratic differential, Riemann surface, embedded graph, boundary combinatorics, harmonic measure, extremal partition.

1. Introduction and Main Theorem

In this paper we consider two structures on a compact surface $\mathcal{R}$ of genus $g \geq 0$. The first one is a combinatorial structure defined by a simply connected weighted graph $\Gamma = \{V, E, F, W\}$ embedded in $\mathcal{R}$ with the set of vertices $V = \{v_k\}$, set of edges $E = \{e^i_{kj}\}$, set of faces $F = \{f_k\}$, and the set of non-negative weights $W = \{w^i_{kj}\}$. The graph Γ is said to be simply connected if each of its faces f_k is a simply connected domain. Throughout the paper, $m \geq 1$ and $n \geq 1$ will be reserved to denote the cardinalities of the sets V and F, respectively. We will always assume that edges of Γ are enumerated in such a way that e^i_{kj} separates faces f_k and f_j and carries the weight $w^i_{kj} > 0$. Then, of course, e^i_{kj} and e^i_{jk} denote the same edge of Γ and $w^i_{kj} = w^i_{jk}$.

This research was supported in part by NSF grant DMS-0525339.

The second structure is a conformal structure associated with a special type of quadratic differentials. Namely, we will consider meromorphic quadratic differentials with poles of order at most 2, whose domain configuration consists of circle domains or strip domains exclusively. We remind the reader that, in general, the domain configuration of a quadratic differential on a compact Riemann surface may include circle domains, ring domains, strip domains, end domains, as well as density structures; see [11, Chapter 3]. The so-called *Jenkins-Strebel quadratic differentials*, i.e., quadratic differentials whose domain configuration consists of circle domains and/or ring domains only, play an important role in the theory of conformal and quasiconformal mappings, where they appear naturally as solutions to numerous extremal problems [9, 11, 14, 19, 20, 22].

The primary goal of the present paper is to prove Theorem 1 below and Theorem 3 given in Section 4 and discuss some related questions. Theorem 1 shows that combinatorial and conformal structures may coexist on the same configuration. In its formulation, we use standard terminology from graph theory and the theory of quadratic differentials, which will be explained further in Section 2.

Theorem 1. *For every simply connected graph $\Gamma = \{V, E, F, W\}$ embedded in a compact surface $\mathcal{R}$ of genus $g \geq 0$ carrying positive weights $w^i_{kj} \in W$ on its edges $e^i_{kj} \in E$, there is a complex structure on $\mathcal{R}$, turning $\mathcal{R}$ into a compact Riemann surface of genus g, and a Jenkins-Strebel quadratic differential $Q(z)\,dz^2$, unique up to conformal automorphisms of $\mathcal{R}$, the critical weighted graph $\Phi_Q = \{V_Q, E_Q, F_Q, W_Q\}$ of which, complemented, if necessary, by second degree vertices on its edges, is homeomorphic to Γ on $\mathcal{R}$ and carries the same weights as Γ does.*

Each face $D_k \in F_Q$ is a circle domain of $Q(z)\,dz^2$ centered at some double pole of $Q(z)\,dz^2$.

In other words, Theorem 1 asserts that every simply connected weighted graph Γ embedded in $\mathcal{R}$ can be realized as a critical graph of some Jenkins-Strebel quadratic differential. Any such realization will be called an *analytic embedding* of Γ in $\mathcal{R}$. Since every edge of Φ_Q is a geodesic arc with respect to the metric $|Q(z)|^{1/2}\,|dz|$, any analytic embedding can be thought as a *geodesic embedding* in $\mathcal{R}$. It also worth mentioning that Theorem 1 can be easily reformulated in terms of the moduli space $\mathcal{M}_g$ of compact Riemann surfaces of genus g.

Let $\mathbb{S}^2$ denote the two-dimensional unit sphere, which will be identified with the Riemann sphere $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. In the case $\mathcal{R} = \mathbb{S}^2$, Theorem 1 can be stated in the following form, which is convenient for possible applications in complex analysis.

Theorem 2. *For every connected plane graph $\Gamma = \{V, E, F, W\}$ carrying positive weights $w^i_{kj} \in W$ on its edges $e^i_{kj} \in E$, there is a Jenkins-Strebel quadratic differential $Q(z)\,dz^2$, unique up to a Möbius self-map of $\overline{\mathbb{C}}$, the critical weighted graph $\Phi_Q = \{V_Q, E_Q, F_Q, W_Q\}$ of which, complemented, if necessary, by second degree vertices on its edges, is homeomorphic to Γ on $\overline{\mathbb{C}}$ and carries the same weights as Γ does.*

Each face $D_k \in F_Q$ is a circle domain of $Q(z)\,dz^2$ centered at some double pole, say b_k, of $Q(z)\,dz^2$.

Let α_k denote the length (total weight) of the boundary cycle in Γ consisting of all boundary edges of ∂f_k in which each cut edge of Γ is transversed twice. Let $A_Q = \{a_1, \dots, a_l\}$, $A_Q \subset V_Q$, be the set of all simple poles of $Q(z)\,dz^2$ and let $b_n = \infty$. Then $Q(z)\,dz^2$ has the form

$$Q(z)\,dz^2 = \frac{P(z)}{\prod_{k=1}^{l}(z-a_k)\prod_{k=1}^{n-1}(z-b_k)^2}\,dz^2, \tag{1.1}$$

where

$$P(z) = C(z-p_1)\cdots(z-p_{l+2n-4}) \tag{1.2}$$

is a polynomial of degree $l+2n-4$ such that $P(z)$ and $R(z) = \prod_{k=1}^{l}(z-a_k)\prod_{k=1}^{n-1}(z-b_k)^2$ are relatively prime and

$$P(b_k) = -\frac{\alpha_k^2}{4\pi^2}\prod_{j=1}^{l}(b_k-a_j)\prod_{j=1}^{\prime\, n-1}(b_k-b_j)^2, \quad k=1,\dots,n-1, \qquad C = -\frac{\alpha_n^2}{4\pi^2}. \tag{1.3}$$

Here $\prod'$ denotes the product taken over all nonzero terms.

Theorems 1 and 2 may be useful in both, the graph theory and complex analysis, and may have applications to the qualitative theory of differential equations in the complex plane. In complex analysis, they give a rather efficient approach to use combinatorics.

In Section 2, we introduce necessary notations and explain terminology related to plane graphs and quadratic differentials. In Section 3, we collect the necessary results from Jenkins's theory of extremal partitions. In particular, we will discuss two extremal problems related to quadratic differentials and give differentiation formulas for their solutions. This section also contains two technical lemmas, which will be used in Section 7.

In Section 4, we discuss another aspect of the problem under consideration that is related to the problem on proportional harmonic measures studied in [2, 3]. The main result of this section, stated in Theorem 3, shows that the domain structure F_Q of the quadratic differential (1.1) is an essentially unique cellular structure on $\mathcal{R}$ with boundary combinatorics defined by Γ such that for every pair of indices k and j, every boundary arc $\gamma \subset \partial D_k \cap \partial D_j$ carries harmonic measures $\omega(\gamma, D_k, b_k)$ and $\omega(\gamma, D_j, b_j)$ proportional with respect to the weights α_k and α_j.

In Sections 5 and 6, we prove uniqueness and existence parts of Theorems 1 and 3, respectively. In Section 7, we discuss a problem on the extremal partition of $\mathcal{R}$ that leads to the same quadratic differential as in Theorems 1 and 3. In the final section, we consider some interesting particular cases of Theorem 2. Our intention here is to demonstrate a variety of problems where Theorem 1 may have applications.

This paper represents an extended version of the author's talk given at the international conference on Analysis and Mathematical Physics: "New Trends in Complex and Harmonic Analysis" in May 7–12, 2007, Bergen, Norway.

2. Notations and terminology

1. Graphs embedded in $\mathcal{R}$. Throughout the paper, $\Gamma = \{V, E, F, W\}$ will denote a simply connected weighted graph described in the Introduction. We always assume that its edges e^i_{kj} are open Jordan arcs on $\mathcal{R}$ such that the closure $\bar{e}^i_{kj}$ of e^i_{kj} is a rectifiable closed Jordan arc or a rectifiable closed Jordan curve on $\mathcal{R}$. By $A = A_\Gamma$ we will denote the set of all vertices $v_k \in V$ such that $\deg v_k = 1$. Later on, this set will play a special role.

Since Γ is simply connected, each face $f_k \in F$ is a simply connected domain on $\mathcal{R}$, whose boundary ∂f_k represents a closed walk of Γ in which each cut edge is transversed twice. We will call this walk the *boundary cycle*[1] around f_k. By α_k we denote the length (total weight) of the boundary cycle around f_k, i.e.,

$$\alpha_k = \sum_{i,j} w^i_{kj}, \tag{2.1}$$

where the weight w^i_{kjk} of each cut edge $e^i_{kk} \in \partial f_k$ is counted twice. In the latter case, the cut edge e^i_{kk} represents two distinct boundary arcs of f_k.

A function $\tau : \mathcal{R} \times [0,1] \to \mathcal{R}$ is called a *deformation* of $\mathcal{R}$ if the following conditions are fulfilled:

1) τ is continuous on $\mathcal{R} \times [0,1]$,
2) $\tau(\cdot, 0)$ is the identity mapping,
3) for every fixed $t \in [0,1]$, $\tau(\cdot, t)$ is a homeomorphism from $\mathcal{R}$ onto itself.

Every deformation τ transforms a graph Γ embedded in $\mathcal{R}$ into a *homeomorphic* graph Γ_τ embedded in $\mathcal{R}$. More precisely, by $\Gamma_\tau(t')$, $0 \le t' \le 1$, we will denote the graph obtained from Γ via deformation $\tau(P, t)$, when the real parameter t varies from 0 to t'. Then $\Gamma_\tau = \Gamma_\tau(1)$. By $\mathcal{G}_\Gamma$ we will denote the set of all graphs embedded in $\mathcal{R}$, which are homeomorphic to Γ on $\mathcal{R}$.

A set $M \subset \mathcal{R}$ is called a fixed set of a deformation τ if $\tau(P, t) = P$ for all $P \in M$ and all $0 \le t \le 1$. In this case we also say that τ is a deformation on $\mathcal{R} \setminus M$.

By a cell (D, b) we mean a hyperbolic simply connected domain $D \subset \mathcal{R}$ with a distinguished center $b \in D$. Accordingly, F becomes a *cellular structure* $[F, B]$ of Γ if every $f_k \in F$ is considered as a cell centered at some point $b_k \in f_k$. Here the set $B = \{b_1, \ldots, b_n\}$ may be thought as a set of isolated vertices added to the graph Γ.

Let $\Gamma' = \{V', E', F', W\}$ be a graph on $\mathcal{R}$ dual of Γ, which has B as a set of its non-isolated vertices and the set A_Γ defined above as its set of isolated vertices. We will write $A_\Gamma = \{a_1, \ldots, a_l\}$.

Let $\mathcal{G}'_\Gamma = \mathcal{G}_{\Gamma'}$ be the collection of all graphs $\widetilde{\Gamma}$ on $\mathcal{R}$ homeomorphic to Γ' on $\mathcal{R}$. If $\Gamma' \in \mathcal{G}'_\Gamma$, then its edges are in a natural one-to-one correspondence with the edges of Γ. Namely, by l^i_{kj} we will denote the edge of $\Gamma' \in \mathcal{G}'_\Gamma$ transversal to the

[1]Boundary cycle is not, in general, a cycle as it is usually defined in the graph theory since its cut edges are transversed twice.

edge e^i_{kj} of Γ. We also assume that l^i_{kj} carries the weight w^i_{kj} as e^i_{kj} does. Then α_k defined by (2.1) can be interpreted as the total weight of the vertex $b_k \in V'$.

2. Quadratic differentials. We remind the reader that a quadratic differential on a Riemann surface, in particular on $\overline{\mathbb{C}}$, is locally defined by a form $Q(z)\,dz^2$, where Q is a meromorphic function of a local coordinate z. If $z = z(\zeta)$ is a conformal change of variables then

$$Q_1(\zeta)\,d\zeta^2 = Q(z(\zeta))(z'(\zeta))^2\,d\zeta^2 \tag{2.2}$$

represents $Q(z)\,dz^2$ in terms of ζ. We refer to [9, 11, 13, 14, 19, 20, 22] for basic properties and applications of quadratic differentials.

Now we introduce some objects related to a quadratic differential $Q(z)\,dz^2$ defined on $\mathcal{R}$. To avoid trivialities, we exclude from our consideration the case when $\mathcal{R} = \overline{\mathbb{C}}$ and $Q(z)\,dz^2$ is conformally equivalent to $Cz^{-2}\,dz^2$ with some $C \in \mathbb{C}$.

Every quadratic differential on $\mathcal{R}$ defines the so-called Q-metric $|Q(z)|^{1/2}\,|dz|$. Accordingly, the Q-length of a rectifiable arc $\gamma \subset \overline{\mathbb{C}}$, which does not depend on a local coordinate z, is given by

$$|\gamma|_Q = \int_\gamma |Q(z)|^{1/2}\,|dz|.$$

A maximal curve or arc γ such that $Q(z)\,dz^2 > 0$ (respectively, $Q(z)\,dz^2 < 0$) along γ is called a trajectory (respectively, orthogonal trajectory) of $Q(z)\,dz^2$. Zeros and poles of Q are critical points of $Q(z)\,dz^2$. Let V_Q be the set of all zeros and simple poles of $Q(z)\,dz^2$. Any trajectory/orthogonal trajectory having at least one of its terminal points in V_Q is called a critical trajectory/orthogonal trajectory, respectively. By Φ_Q we denote the *critical set* of $Q(z)\,dz^2$, i.e., Φ_Q is the union of all critical trajectories of $Q(z)\,dz^2$ and their end points. According to the Basic Structure Theorem [11, Theorem 3.5], the set $(\mathcal{R}\setminus\text{Closure}\,(\Phi_Q))\cup\text{Interior}\,(\text{Closure}\,(\Phi_Q))$ consists of a finite number of domains called the *domain configuration* of $Q(z)\,dz^2$. As we already mentioned in the Introduction the domain configuration may include circle domains, ring domains, strip domains, end domains, as well as density structures.

A circle domain D of $Q(z)\,dz^2$ is a maximal simply connected domain swept out by regular closed trajectories of $Q(z)\,dz^2$ surrounding a second-order pole a that is the only singularity of $Q(z)\,dz^2$ in D. A circle domain is necessarily bounded by a finite number of critical trajectories of $Q(z)\,dz^2$ and their terminal points in V_Q.

Similarly, a ring domain D of $Q(z)\,dz^2$ is a maximal non-degenerate doubly-connected domain swept out by regular closed trajectories of $Q(z)\,dz^2$. Again, each boundary component of D consists of a finite number of critical trajectories of $Q(z)\,dz^2$ and their terminal points in V_Q.

In modern literature, quadratic differentials whose domain configuration consists of circle domains and/or ring domains only are often called the *Jenkins-Strebel* *differentials*. They were first studied by J.A. Jenkins [10, 12] in connection with extremal problems in conformal mapping. Then Jenkins's theory was developed further and found many important applications in works of Strebel, Renelt, Kuz'mina, and others.

If $Q(z)\,dz^2$ is a Jenkins-Strebel quadratic differential on $\mathcal{R}$, then its critical set Φ_Q can be considered as a weighted graph $\Phi_Q = \{V_Q, E_Q, F_Q, W_Q\}$ called the *critical graph* of $Q(z)\,dz^2$. Namely, we define V_Q as above and $E_Q = \{\nu^i_{kj}\}$ as the collection of all critical trajectories of $Q(z)\,dz^2$. Then, F_Q is the domain configuration of $Q(z)\,dz^2$ consisting of all circle domains and/or ring domains of $Q(z)\,dz^2$. Finally, for every edge $\nu^i_{kj} \in E_Q$ we prescribe its Q-length as a weight:

$$w^i_{kj} = \int_{\nu^i_{kj}} |Q(z)|^{1/2}\,|dz|. \tag{2.3}$$

A strip domain (or *digon*) G of $Q(z)\,dz^2$ is a maximal simply connected domain swept out by regular trajectories γ of $Q(z)\,dz^2$, each of which has its initial and terminal points at some poles q_1 and q_2 (possibly coincident) each of order ≥ 2 such that q_1 and q_2 represent distinct boundary points of G. We want to emphasize here that ∂G consists of two boundary arcs (sides of the digon G) joining q_1 and q_2 and that each of these boundary arcs contains a zero or simple pole of $Q(z)\,dz^2$.

Quadratic differentials $Q(z)\,dz^2$, whose domain configuration consists of digons only, each of which has its vertices at the second-order poles of $Q(z)\,dz^2$ with the radial trajectory structure in their neighborhoods, also have found important applications in extremal problems. We will christen them as *Kuz'mina quadratic differentials* after Galina V. Kuz'mina, who inspired their study in early 1980's. She also obtained some important results related to quadratic differentials of this type, see [14, 15, 16, 17].

If $Q(z)\,dz^2$ is a Jenkins-Strebel quadratic differential without ring domains in its domain configuration, then the orthogonal quadratic differential $-Q(z)\,dz^2$ is a Kuz'mina quadratic differential. As one can easily see from examples, the converse is not true in general.

Let $A = \{a_1, \ldots, a_l\}$ and $B = \{b_1, \ldots, b_n\}$ be the sets of simple poles and double poles of a Kuz'mina quadratic differential $Q(z)\,dz^2$. Let $\{G^i_{kj}\}$ be the domain configuration of $Q(z)\,dz^2$. We assume here that a digon G^i_{kj} has its vertices at the poles b_k and b_j, possibly coincident. Let w^i_{kj} be the height of the digon G^i_{kj}, i.e., w^i_{kj} is the Q-length of an arc $\gamma \subset G^i_{kj}$ of any orthogonal trajectory of $Q(z)\,dz^2$ joining the sides of G^i_{kj}. Let γ^i_{kj} be the regular trajectory of $Q(z)\,dz^2$ selected in G^i_{kj} such that the set $G^i_{kj} \setminus \gamma^i_{kj}$ consists of two digons, each of which has height $w^i_{kj}/2$.

Now we can consider a weighted graph $\Gamma_Q = \{V, E, F, W\}$ on $\mathcal{R}$ with the set of vertices $V = A \cup B$, set of edges $E = \{\gamma^i_{kj}\}$, and set of positive weights

$W = \{w^i_{kj}\}$. The graph Γ_Q constructed in such way will be called *the trajectory graph* of a Kuz'mina quadratic differential $Q(z)\,dz^2$.

A Kuz'mina quadratic differential $Q(z)\,dz^2$ is called *associated* with a weighted graph Γ on $\mathcal{R}$ if its trajectory graph Γ_Q is homeomorphic to Γ on $\mathcal{R}$ and carries the same set of weights as Γ does.

3. Extremal partitions and quadratic differentials

In this section, we collect results of Jenkins's theory of extremal partitions, which will be used later. We refer to the original paper [10], monographs [9, 11, 14, 20], and surveys [13, 16, 19] for a general account and technical details of this theory.

Let $A = \{a_1, \ldots, a_l\}$ and $B = \{b_1, \ldots, b_n\}$ be two sets of distinct points on a Riemann surface $\mathcal{R}$ called *punctures* and *centers*, respectively. Below we always assume that each center $b_k \in B$ is supplied with a fixed local coordinate z_k such that $z_k(b_k) = 0$. Let $\mathcal{R}' = \mathcal{R} \setminus A$ and let $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ be a set of positive weights. Then for a given family $D = \{D_1, \ldots, D_n\}$ of non-overlapping cells D_k on $\mathcal{R}'$ centered at $b_k \in D_k$, we consider the following weighted sum:

$$M_1(D) = \sum_{k=1}^{n} \alpha_k^2 m_{z_k}(D_k, b_k).$$

Here $m_z(D, b)$ denotes the reduced module of a simply connected domain D at $b \in D$ with respect to the local coordinate z, which can be defined as

$$m_z(D, b) = \frac{1}{2\pi} \log\, R_z(D, b) = \frac{1}{2\pi} \log\, |f'(0)|,$$

were $R_z(D, b)$ is the conformal radius of D at b with respect to z, f is the Riemann mapping from the unit disc $\mathbb{D} = \{\zeta \in \mathbb{C} : |\zeta| < 1\}$ onto D normalized by $f(0) = b$, and $f'(0)$ is the derivative with respect to the local coordinate z. Below we also use the following notations. For $r > 0$ and $a \in \mathbb{C}$, let $\mathbb{D}_r(a) = \{\zeta \in \mathbb{C} : |\zeta - a| < r\}$, $C_r(a) = \partial\mathbb{D}_r(a)$. If $a = \infty$, then $\mathbb{D}_r(a) = \{\zeta \in \overline{\mathbb{C}} : |\zeta| > 1/r\}$ and $C_r(a) = \partial\mathbb{D}_r(a)$.

Problem 1. For given sets of punctures A and centers B with the corresponding set $Z = \{z_1, \ldots, z_n\}$ of local coordinates, and a given set of positive weights α, find the supremum

$$\mathcal{M}_1 = \sup\, M_1(D)$$

taken over all families D described above and find all families $D^* = \{D_1^*, \ldots, D_n^*\}$ such that $M_1(D^*) = \mathcal{M}_1$.

Of course, Problem 1 is just a particular case of the original Jenkins's problem studied in [10, 12, 14, 19, 20].

Now we define a similar problem for digons, which study began with the works of Emel'yanov [8] and Kuz'mina [15]. By a digon we will mean a hyperbolic simply connected domain G on $\mathcal{R}$ with two distinguished points, b_1 and b_2, called *vertices*, on its boundary. As before, we assume that vertices b_1 and b_2 are supplied with fixed local coordinates z_1 and z_2. Every digon G considered below will have well-defined non-zero interior angles, say φ_1 and φ_2, at its vertices b_1 and b_2, respectively. In addition, we always will assume that for every $\varepsilon > 0$ sufficiently small the intersection $G \cap C_\varepsilon(b_k)$ consists of a single open arc denoted by $\gamma_G(b_k, \varepsilon)$. Here and below, circles and disks on $\mathcal{R}$ are defined in terms of distinguished local coordinates. For example, $C_\varepsilon(b_k) = \{P \in \mathcal{R} : |z_k(P)| = \varepsilon\}$, $\mathbb{D}_\varepsilon(b_k) = \{P \in \mathcal{R} : |z_k(P)| < \varepsilon\}$, etc. Then for all $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ small enough, the set $G_{\varepsilon_1,\varepsilon_2} = G \setminus \left(\overline{\mathbb{D}}_{\varepsilon_1}(b_1) \cup \overline{\mathbb{D}}_{\varepsilon_2}(b_2)\right)$ can be considered as a quadrilateral having $\gamma_G(b_1, \varepsilon_1)$ and $\gamma_G(b_2, \varepsilon_2)$ as a pair of distinguished sides on its boundary. For any given quadrilateral Q, by $\mathrm{Mod}(Q)$ we will denote the module of Q with respect to the family of locally rectifiable arcs separating the distinguished sides of Q, see [11, Chapter 2]. Then the module of the digon G (with respect to local coordinates z_1, z_2) is defined by

$$m_{z_1,z_2}(G, b_1, b_2) = \lim_{\varepsilon_1,\varepsilon_2 \to 0} \left(\mathrm{Mod}(G_{\varepsilon_1,\varepsilon_2}) + \frac{1}{\varphi_1} \log \varepsilon_1 + \frac{1}{\varphi_2} \log \varepsilon_2 \right),$$

provided that a finite limit exists.

Let Γ be a simply connected weighted graph on $\mathcal{R}$ defined in Section 2. Let $\Gamma' \in \mathcal{G}'_\Gamma$ be a dual of Γ, which has isolated vertices $A = \{a_1, \ldots, a_l\}$ and non-isolated vertices $B = \{b_1, \ldots, b_n\}$ with the total weights $\alpha = \{\alpha_1, \ldots, \alpha_n\}$.

A system $G = \{G^i_{kj}\}$ of non-overlapping digons G^i_{kj} on $\mathcal{R} \setminus (A \cup B)$ such that $G^i_{kj} = G^i_{jk}$ for all possible k, j, and i, is called *admissible* for Γ' if the following conditions are satisfied:

(1) A digon G^i_{kj} has its vertices at the points b_k and b_j of B (possibly coincident).
(2) Non-zero angles φ^i_{kj} and φ^i_{jk} of G^i_{kj} at the vertices b_k and b_j and the module $m^i_{kj} = m_{z_k,z_j}(G^i_{kj}, b_k, b_j)$ are well defined.
(3) $\varphi^i_{kj} : \varphi^i_{jk} = \alpha_j : \alpha_k$ for all k and j and all possible i.
(4) Selecting in each digon G^i_{kj} a Jordan arc l^i_{kj}, which joins the vertices b_k and b_j, we obtain a graph Γ'' on $\mathcal{R}$ with the set of vertices $V = A \cup B$ and set of edges $\{l^i_{kj}\}$ carrying weights $\{w^i_{kj}\}$, which is homeomorphic to the graph Γ' on the punctured surface $\mathcal{R} \setminus V$.

Let $\mathbf{G}_{\Gamma'}$ be the set of all systems $G = \{G^i_{kj}\}$ admissible for Γ'. Then on $\mathbf{G}_{\Gamma'}$ we consider the weighted sum of the reduced moduli of digons:

$$M_2(G) = -\sum (w^i_{kj})^2 m_{z_k,z_j}(G^i_{kj}, b_k, b_j),$$

where each digon G^i_{kj} is counted precisely once.

Problem 2. For a given graph Γ' described above, find the supremum

$$\mathcal{M}_2 = \sup M_2(G)$$

taken over all systems G admissible for Γ' and find all systems $G^* = \{G^{i*}_{kj}\}$ in $\mathbf{G}_{\Gamma'}$ such that $M_2(G^*) = \mathcal{M}_2$.

It is well known (see [10, 14, 19, 20]) that Problem 1 has a unique extremal configuration $D^* = \{D^*_1, \ldots, D^*_n\}$ formed by circle domains of a Jenkins-Strebel quadratic differential $Q_1(z)\,dz^2$. If $\mathcal{R} = \overline{\mathbb{C}}$ and $b_n = \infty$, then $Q_1(z)\,dz^2$ has the form

$$Q_1(z)\,dz^2 = \frac{P_1(z)}{\prod_{k=1}^{l}(z-a_k)\prod_{k=1}^{n-1}(z-b_k)^2}\,dz^2, \tag{3.1}$$

where P_1 is a polynomial of degree $2n+l-4$ satisfying conditions (1.3).

It is less known (see [8, 13, 15, 16, 19, 22]) that Problem 2 also has a unique extremal configuration $G^* = \{G_{kj}^{i*}\}$ formed by digons of a Kuz'mina quadratic differential $Q_2(z)\,dz^2$. In the case $\mathcal{R} = \overline{\mathbb{C}}$ and $b_n = \infty$, $Q_2(z)\,dz^2$ has the form

$$Q_2(z)\,dz^2 = -\frac{P_2(z)}{\prod_{k=1}^{l}(z-a_k)\prod_{k=1}^{n-1}(z-b_k)^2}\,dz^2. \tag{3.2}$$

Here P_2 is a polynomial of degree $2n+l-4$ satisfying conditions (1.3).

In addition, we want to mention that the metric $(w_{kj}^i)^{-1}|Q_2(z)|^{1/2}\,|dz|$ is extremal for the module problem in the digon G_{kj}^i. In particular,

$$\int_\gamma |Q_2(z)|^{1/2}\,|dz| \ge w_{kj}^i \tag{3.3}$$

for every rectifiable arc $\gamma \subset G_{kj}^i$ joining the sides of G_{kj}^i.

In what follows, we always assume that the weights α_k and w_{kj}^i are fixed. Then $\mathcal{M}_1 = \mathcal{M}_1(a_1,\dots,a_l,b_1,\dots,b_n)$ becomes a function of the punctures $a_1,\dots,$ a_l and centers $b_1,\dots,b_n$, which also depends on the corresponding local coordinates $z_1,\dots,z_l,z_{l+1},\dots,z_{l+n}$. For each choice of local coordinates, $\mathcal{M}_1$ is a continuous function of the punctures and centers, which is single-valued on

$$\widetilde{\mathcal{R}}^{l+n} = \{(P_1,\dots,P_{l+n}) \in \mathcal{R}^{l+n} : P_j \ne P_k, \quad \text{for } j \ne k\}.$$

In contrast, the sum $\mathcal{M}_2 = \mathcal{M}_2(a_1,\dots,a_l,b_1,\dots,b_n)$ can be defined as a single-valued function only locally, i.e., on a sufficiently small neighborhood of every point of $\widetilde{\mathcal{R}}^{l+n}$. Considered globally on $\widetilde{\mathcal{R}}^{l+n}$, $\mathcal{M}_2(a_1,\dots,a_l,b_1,\dots,b_n)$ becomes, in general, a path dependent infinitely-valued function. In particular, if $\mathcal{R} = \overline{\mathbb{C}}$, then $\mathcal{M}_2$ is infinitely-valued if $l+n \ge 4$. To give a precise definition of this multi-valued function, we fix a finite atlas on $\mathcal{R}$, then we fix $A^0 \in \widetilde{\mathcal{R}}^l$ and $B^0 \in \widetilde{\mathcal{R}}^n$ such that $(A^0,B^0) \in \widetilde{\mathcal{R}}^{l+n}$, and then we fix a graph Γ_0' as in Problem 2. For these *initial conditions* we define $\mathcal{M}_2(a_1^0,\dots,a_l^0,b_1^0,\dots,b_n^0)$ to be the solution to Problem 2 for A^0, B^0, and Γ_0'. Then we deform $\mathcal{R}$ via deformation $\tau = \tau(P,t)$. Let Γ_τ' be the graph obtained from Γ_0' via this deformation. Let $a_k = \tau(a_k^0,1)$, $1 \le k \le l$, and let $b_k = \tau(b_k^0,1)$, $1 \le k \le n$.

By the branch of $\mathcal{M}_2(a_1,\dots,a_l,b_1,\dots,b_n)$ corresponding to the deformation τ, we mean the solution to Problem 2 for the graph $(\Gamma')_\tau$.

The gradient theorems for solutions of Problems 1 and 2, the study of which was initiated by this author in 1980's, see [7, 16, 18, 19], show that the function $\mathcal{M}_1$ and every single-valued branch of $\mathcal{M}_2$ are smooth. Below we give a precise statement for the case of the Riemann sphere, i.e., we assume till the end of this section that $\mathcal{R} = \overline{\mathbb{C}}$.

Theorem A (cf. Theorems 5.4–5.6, [19]). *Let $\mathcal{R} = \overline{\mathbb{C}}$. If all punctures $a_1^0, \dots, a_l^0$ and centers $b_1^0, \dots, b_n^0$ are distinct and the varying parameter a_k or b_k is finite, then $\mathcal{M}_1$ and $\mathcal{M}_2$ are locally single-valued continuously differentiable functions of this parameter at (A^0, B^0) and the corresponding partial derivatives are given by*

$$\frac{\partial \mathcal{M}_s}{\partial a_k} = \pi \text{Res}\,[Q_s, a_k^0], \qquad \frac{\partial \mathcal{M}_s}{\partial b_k} = \pi \text{Res}\,[Q_s, b_k^0], \qquad s = 1, 2, \tag{3.4}$$

where $Q_1(z)\,dz^2$ and $Q_2(z)\,dz^2$ are quadratic differentials extremal for Problem 1 *and Problem* 2*, respectively.*

Next, we prove some results about global behavior of the multi-valued function $\mathcal{M}_2$ for the case $\mathcal{R} = \overline{\mathbb{C}}$. Let z_1 and z_2 be distinct points on $\overline{\mathbb{C}}$ and let $\overline{\mathbb{C}}(z_1, z_2) = \overline{\mathbb{C}} \setminus \{z_1, z_2\}$. Let $\gamma : [0, 1] \to \overline{\mathbb{C}}$ be a simple arc in $\overline{\mathbb{C}}(z_1, z_2)$ from ζ_1 to ζ_2. By the index of γ with respect to z_1 and z_2 we mean the normalized change in the argument of $\frac{z - z_1}{z - z_2}$, when z varies from ζ_1 to ζ_2 along γ, i.e.,

$$\text{ind}_{z_1, z_2}(\gamma) = \frac{1}{2\pi} \Delta_\gamma \left(\arg \frac{z - z_1}{z - z_2} \right). \tag{3.5}$$

In (3.5) we assume, of course, that $z_1, z_2 \neq \infty$. If $z_k = \infty$ for $k = 1$ or $k = 2$, then we put $\text{ind}_{z_1,z_2}(\gamma) = (-1)^{k+1}(2\pi)^{-1}\Delta_\gamma\,(\arg(z - z_k))$.

Now let Γ be a connected plane graph and let $\Gamma' \in \mathcal{G}'_\Gamma$ be a graph dual of Γ as in Problem 2. Then the index of Γ' is defined by

$$\text{ind}(\Gamma') = \max_{l_{kj}^i} \max \left(\text{ind}_{z_1, z_2} l_{kj}^i \right), \tag{3.6}$$

where the inner maximum is taken over all pairs of distinct points z_1 and z_2 such that $z_1, z_2 \in (A \cup B) \setminus \{b_k, b_j\}$. We remind the reader that $A = \{a_1, \dots, a_l\}$ and $B = \{b_1, \dots, b_n\}$ denote the set of isolated vertices of Γ' and set of non-isolated vertices of Γ', respectively. Also, l_{kj}^i joins the vertices b_k and b_j. Thus, $\text{ind}\,(\Gamma')$ is a non-negative real number.

Let $G_{\Gamma'}(V')$, $V' = A \cup B$, denote the family of graphs $\tilde{\Gamma}$ with isolated vertices $A = \{a_1, \dots, a_l\}$ and non-isolated vertices $B = \{b_1, \dots, b_n\}$, for each of which there is a deformation $\tau = \tau(z, t)$ transforming Γ' into $\tilde{\Gamma}$ such that $\tau(a_k, 1) = a_k$ for all $1 \le k \le l$ and $\tau(b_k, 1) = b_k$ for all $1 \le k \le n$. If, in addition, τ is a deformation on $\overline{\mathbb{C}} \setminus V'$, then we say that the graphs Γ' and $\tilde{\Gamma}$ belong to the same homotopic class. Let $\mathcal{H}_{\Gamma'}(V') = \{\mathcal{H}\}$ denote the collection of all homotopic classes of the family $G_{\Gamma'}(V')$. Let $\mathcal{H} \in \mathcal{H}_{\Gamma'}(V')$. One can easily see that $\text{ind}\,(\Gamma_1) = \text{ind}\,(\Gamma_2)$ if $\Gamma_1, \Gamma_2 \in \mathcal{H}$. Thus, the index of $\mathcal{H}$, denoted by $\text{ind}\,(\mathcal{H})$, can be defined as a common index of all graphs in $\mathcal{H}$.

Since the number of edges and vertices of all graphs under consideration is fixed, the following elementary lemma is an easy consequence of our definition of the index, see (3.6).

Lemma 1. *For every plane graph $\Gamma' \in \mathcal{G}'_\Gamma$ with the set of isolated vertices A and set of non-isolated vertices B, and for every positive integer N, the number of distinct homotopy classes $\mathcal{H} \in \mathcal{H}_{\Gamma'}(V')$ such that* $\text{ind}\,(\mathcal{H}) \le \text{N}$ *is finite.*

Let z_1 and z_2 be distinct points in the closed annulus $\overline{A}(r_1, r_2) = \{z : r_1 \leq |z| \leq r_2\}$, $0 < r_1 < r_2$. Let G be a digon on $\overline{\mathbb{C}}(z_1, z_2)$ having its vertices at $\zeta_1 = 0$ and $\zeta_2 = \infty$. Let $\operatorname{ind}_{z_1,z_2}(G)$ denote the common index of all simple arcs $\gamma \subset G$ joining ζ_1 and ζ_2. Suppose that each side of G is a piecewise analytic arc (not necessarily Jordan). Then each of the sets $G \cap C_{r_1}(0)$ and $G \cap C_{r_2}(0)$ consists of a finite number of circular arcs. Let $\mathcal{L}$ be the collection of all such arcs, which join the sides of G. Let e_1 be the arc in $\mathcal{L}$, which separate inside G the vertex ζ_1 from all other arcs of $\mathcal{L}$. Similarly, let $e_2 \in \mathcal{L}$ be the arc in G separating ζ_2 from all other arcs of $\mathcal{L}$. The arcs e_1 and e_2 divide G into three simply connected components. By $G_{r_1 r_2}$, we will denote those of them, which has both arcs e_1 and e_2 on its boundary. Then $G_{r_1 r_2}$ can be considered as a quadrilateral with e_1 and e_2 as its pair of distinguished sides.

Lemma 2. *Let r_1 and r_2, $0 < r_1 < r_2 < \infty$, be fixed and let z_1 and z_2 be distinct points in $\overline{A}(r_1, r_2)$. There is a constant $C_0 > 0$, depending on r_1 and r_2, but not on z_1 and z_2, such that*

$$\operatorname{Mod}(G_{r_1,r_2}) \geq C_0 n \tag{3.7}$$

for every quadrilateral $G_{r_1 r_2}$, which corresponds to a digon G in $\overline{\mathbb{C}}(z_1, z_2)$ described above such that $\operatorname{ind}_{z_1,z_2}(G) \geq 2n$.

Proof. Since the module of a quadrilateral is invariant under conformal mappings and reflections, we may assume without loss of generality that $z_1 = 1$, $|z_2| \geq 1$, and $\Im z_2 \geq 0$.

Let G be a digon satisfying the assumptions of the lemma such that

$$\operatorname{ind}_{z_1,z_2}(G) \geq 2n, \quad n \geq 1. \tag{3.8}$$

Let L be a circular arc in the closed upper half-plane joining z_1 and z_2, which is orthogonal to the real axis. Since the boundary of G is piecewise analytic, the intersection $G \cap L$ consists of a finite number of open circular arcs. By $l_1, \dots, l_N$ we denote those of them, which join the opposite sides of G. We assume that these arcs are enumerated in a "natural order" in G. The latter means that for every $2 \leq k \leq N$, the arc l_{k-1} separates l_k from the vertex $\zeta_1 = 0$ inside G. We also put $l_0 = \zeta_1$, $l_{N+1} = \zeta_2$.

The set $G \setminus \cup_{k=1}^{N} l_k$ consists of $N+1$ simply connected components. Let R_k, $k = 1, \dots, N+1$, denote the component, which has l_{k-1} and l_k on its boundary. Let γ be a Jordan arc in G, which joins the vertices ζ_1 and ζ_2 and meets each arc l_k exactly once. Let $\gamma_k = \gamma \cap R_k$. Then

$$\operatorname{ind}_{z_1,z_2}(\gamma) = \sum_{k=1}^{N+1} \operatorname{ind}_{z_1,z_2}(\gamma_k). \tag{3.9}$$

Since γ_k does not intersect L, one can easily see that $|\operatorname{ind}_{z_1,z_2}(\gamma_k)| < 2$ for all $1 \leq k \leq N+1$. This together with (3.8) and (3.9) implies that among the arcs γ_k there are at least n arcs, each of which has index $\operatorname{ind}_{z_1,z_2}(\gamma_k) \geq 1$.

If $1 \leq \operatorname{ind}_{z_1,z_2}(\gamma_k) < 2$, then one can easily see that γ_k separates the point $\zeta_1 = 0$ from $\zeta_2 = \infty$ inside the domain $\overline{\mathbb{C}} \setminus L$. Therefore in this case the domain

R_k can be considered as a quadrilateral in $\overline{\mathbb{C}} \setminus L$ having its distinguished sides on L. Let

$$K(z_1, z_2) = \max\left((\mathrm{Mod}(\Omega)^{-1}\right), \tag{3.10}$$

where the maximum is taken over all quadrilaterals Ω in $\mathbb{C} \setminus (L \cup \{0\})$ having their distinguished sides on L and separating $\zeta_1 = 0$ from $\zeta_2 = \infty$ in $\overline{\mathbb{C}} \setminus L$.

Of course, problem (3.10) is just a particular case of Jenkins's problem on the extremal partitions for one homotopy class. It easily follows from general results of Jenkins's theory that $K(z_1, z_2)$ is positive and continuous on the set $\overline{A}(r_1, r_2) \times \overline{A}(r_1, r_2)$. Therefore there is a constant $C_0 > 0$ such that $\mathrm{Mod}(\Omega) \geq C_0$ for every quadrilateral Ω described above and every pair of points z_1 and z_2 in $\overline{A}(r_1, r_2)$.

Finally, applying Grötzsch's lemma (see [11, Theorems 2.6 and 2.7]) to the quadrilateral $G_{r_1 r_2}$ and system of quadrilaterals R_k in $G_{r_1 r_2}$ such that $\mathrm{ind}_{\mathrm{z_1,z_2}}(\gamma_\mathrm{k}) \geq 1$, we obtain

$$\mathrm{Mod}(G_{r_1,r_2}) \geq \sum \mathrm{Mod}(R_k) \geq C_0 n.$$

This sum, containing at least n terms, is taken over all quadrilaterals R_k such that $\mathrm{ind}_{\mathrm{z_1,z_2}}(\gamma_\mathrm{k}) \geq 1$. The proof is complete. □

Let $r > 0$ and let z_1 and z_2 be distinct points in $\mathbb{D}_r^* = \overline{\mathbb{C}} \setminus \mathbb{D}_r(0)$. Let G be a digon as in Lemma 2 but having both its vertices ζ_1 and ζ_2 at $z = 0$. Let $\mathcal{L}_0$ be the collection of all arcs in $G \cap C_r(0)$. For $k = 1, 2$, let e_k denote the arc in $\mathcal{L}_0$, which separates inside G the vertex ζ_k from all other arcs in $\mathcal{L}_0$. Finally, let G_r be a simply connected component of $G \setminus (e_1 \cup e_2)$, which has both arcs e_1 and e_2 on its boundary. Then G_r can be considered as a quadrilateral with e_1 and e_2 as a distinguished pair of its sides. Lemma 3 below shows that inequality (3.7) remains valid for digons G having both their vertices ζ_1 and ζ_2 at the same point, say at $z = 0$. Its proof is almost identical with the proof of Lemma 2 and therefore is left to the reader.

Lemma 3. *There is a constant $C_1 > 0$, depending on r but not on z_1 and z_2, such that $\mathrm{Mod}(G_r) \geq C_1 n$ for every quadrilateral G_r described above such that $\mathrm{ind}_{\mathrm{z_1,z_2}}(\mathrm{G}) \geq 2\mathrm{n}$.*

4. Proportional harmonic measures

Another aspect of the problem on extremal partitions is related to the question about proportional harmonic measures. This question was first discussed by J. Akeroyd [2], who considered two "crescent regions" in the unit disc $\mathbb{D}$. We remind the reader that the harmonic measure $\omega(\cdot, D, a)$ is a unique Borel probability measure on the boundary ∂D of a Dirichlet domain D such that

$$h(a) = \int_{\partial D} h(z)\, d\omega(z, D, a)$$

for all functions h harmonic on D and continuous on $\overline{D}$, see [6].

Let D_1 and D_2 be disjoint domains on a Riemann surface $\mathcal{R}$ having a Jordan arc L on their common boundary $\partial D_1 \cap \partial D_2$. We say that D_1 and D_2 carry proportional harmonic measures on L if there are points $b_1 \in D_1$ and $b_2 \in D_2$ and positive constants c_1, c_2 such that

$$c_1\omega(E, D_1, b_1) = c_2\omega(E, D_2, b_2)$$

for every Borel set $E \subset L$.

Now we modify this definition of proportionality for the case of a slit lying on the boundary of a simply connected domain D. Let L be an open Jordan arc on ∂D such that $G = D \cup L$ is a doubly-connected domain. Then L is a slit in G, which defines two boundary arcs, say L^+ and L^-, on ∂D. For $E \subset L$, let E^+ and E^- denote the set E considered as subsets of boundary arcs L^+ and L^-, respectively. We say that D carries proportional (equal) harmonic measures on a slit L if there is a point $a \in D$ such that $\omega(E^+, D, a) = \omega(E^-, D, a)$ for every Borel set $E \subset L$.

Definitions given above can be extended to systems of more than two domains. Let $[F, B]$ be the cellular structure of a simply connected weighted graph $\Gamma = \{V, E, F, W\}$ on $\mathcal{R}$, where $F = \{D_1, \dots, D_n\}$ and $B = \{b_1, \dots, b_n\}$. Thus, here D_k (not f_k!) denotes the face of Γ centered at b_k. Let $\mathcal{F}_\Gamma$ denote the set of all cellular structures $[\widetilde{F}, \widetilde{B}]$ associated with some graph $\widetilde{\Gamma}$ homeomorphic to Γ on $\mathcal{R}$. Let α_k be the length of the boundary cycle ∂D_k defined by (2.1) and let $\delta^i_{kj} = w^i_{kj}/\alpha_k$ be the *relative weight* of the edge e^i_{kj} with respect to the boundary cycle ∂D_k.

We say that $[F, B]$ carries harmonic measures proportional with respect to the set of weights $W = \{w^i_{kj}\}$ if

$$\omega(e^i_{kj}, D_k, b_k) = \delta^i_{kj} \quad \text{for all } k,\ j, \text{ and } i, \tag{4.1}$$

and

$$\alpha_k\omega(E, D_k, b_k) = \alpha_j\omega(E, D_j, b_j)$$

for every Borel set $E \subset \partial D_k \cap \partial D_j$.

The problem on proportional harmonic measures is to find, for a given graph $\Gamma = \{V, E, F, W\}$, all cellular structures $[\widetilde{F}, \widetilde{B}] \in \mathcal{F}_\Gamma$, which carry harmonic measures proportional with respect to the set of weights $W = \{w^i_{kj}\}$.

For planar linear graphs with central symmetry a solution to this problem follows from Theorem 1.3 in [3]. The following theorem solves this problem for any simply connected weighted graph Γ on $\mathcal{R}$.

Theorem 3. *For every simply connected weighted graph Γ on a compact surface $\mathcal{R}$ of genus $g \geq 0$ defined in Theorem 1 there is a complex structure on $\mathcal{R}$, turning $\mathcal{R}$ into a compact Riemann surface of genus g, and there is a cellular structure $[F^*, B^*] \in \mathcal{F}_\Gamma$ on $\mathcal{R}$, unique up to a conformal automorphism of $\mathcal{R}$, which carries harmonic measures proportional with respect to the set of weights $W = \{w^i_{kj}\}$.*

Let $F^* = \{D_1, \dots, D_n\}$, $B^* = \{b_1, \dots, b_n\}$, and let $A^* = \{a_1, \dots, a_l\}$ denote the set of all vertices of degree 1 of the graph Γ^* corresponding to $[F^*, B^*]$. The *configuration F^* coincides with the domain configuration F_Q of a Jenkins-Strebel quadratic differential $Q(z)\,dz^2$ defined by Theorem* 1.

In the case when $\mathcal{R} = \overline{\mathbb{C}}$ and $b_n = \infty$, F^* coincides with the domain configuration F_Q of a Jenkins-Strebel quadratic differential $Q(z)\,dz^2$ defined by formulas (1.1)–(1.3) *of Theorem* 2.

Now we remind some simple properties of quadratic differentials, which will be used in the proof of Theorem 3. A point a is a second-order pole of $Q(z)\,dz^2$ in its circle domain D if and only if for every local coordinate z with $z(a) = 0$ there exists $c > 0$ such that

$$Q(z) = -\frac{c^2}{4\pi^2}\frac{1}{z^2} + \frac{a_1}{z} + \cdots \tag{4.2}$$

near $z = 0$. If γ is a trajectory of $Q(z)\,dz^2$ in D, then

$$c = |\gamma|_Q = \int_\gamma |Q(z)|^{1/2}\,|dz| = \int_\gamma Q^{1/2}(z)\,dz$$

is the Q-length of γ. Here we assume that $Q^{1/2}\,dz > 0$ along the corresponding trajectory.

Let $\zeta = f(P)$ map D conformally onto the unit disk $\mathbb{D}$ such that $f(a) = 0$ and $f(b) = 1$ for some $b \in \partial D$. Then

$$f(P) = \exp\left\{\frac{2\pi i}{c}\int_b Q^{1/2}(z)\,dz\right\},$$

where $z = z(P)$ is a local coordinate such that $z((a) = 0$; see [11, Section 3.3]. The following simple lemma and its corollary, which in case of planar domains were used in [3], reveal a role played by quadratic differentials in problems on domains carrying proportional harmonic measures on their boundaries.

Lemma 4 (cf. [3]**).** *Let D be a circle domain of a quadratic differential $Q(z)\,dz^2$, which in terms of a local coordinate $z = z(P)$ such that $z(a) = 0$ has expansion* (4.2) *at $z = 0$. Then*

$$d\omega(z, D, a) = c^{-1}|Q(z)|^{1/2}\,|dz| \quad \textit{for all}\ \ z = z(P)\ \textit{such that}\ P \in \partial D. \tag{4.3}$$

Corollary 1 (cf. [3]**).** *Let D_1 and D_2 be circle domains of $Q(z)\,dz^2$ centered at a_1 and a_2, respectively. Let c_1 and c_2 be Q-lengths of trajectories of $Q(z)\,dz^2$ in the domains D_1 and D_2. Let L be an open Jordan arc on $\partial D_1 \cap \partial D_2$. If Q is meromorphic on L, then for every Borel set $E \subset L$,*

$$c_1\omega(E, D_1, a_1) = c_2\omega(E, D_2, a_2). \tag{4.4}$$

Lemma 5. *Let D_1 and D_2 be non-overlapping simply connected domains on $\mathcal{R}$ having a Jordan arc L on their common boundary $\partial D_1 \cap \partial D_2$. Let $\zeta = f(P)$ map D_1 conformally onto $\mathbb{D}$. If D_1 and D_2 carry proportional harmonic measures on L, then f can be analytically continued across L. In particular, L is an analytic arc.*

Proof. By assumptions of the lemma, there are points $a_1 \in D_1$ and $a_2 \in D_2$ and positive constants c_1 and c_2 such that (4.4) holds for every Borel set $E \subset L$. Postcomposing with a Möbius map, if necessary, we may assume without loss of generality that $f(a_1) = 0$ and $f(L) = \{e^{i\varphi} : 0 < \varphi < \varphi_1\}$ for some $0 < \varphi_1 \le 2\pi$.

Let $\zeta = g(P)$ map D_2 conformally onto $\mathbb{D}^* = \overline{\mathbb{C}} \setminus \overline{\mathbb{D}}$ such that $g(a_2) = \infty$ and $g(L) = \{e^{i\varphi} : 0 < \varphi < \varphi_2\}$. Here $\varphi_2 = (c_2/c_1)\varphi_1$ by (4.4).

Let $\widetilde{D}_2 = g^{-1}(\mathbb{D}')$, where $\mathbb{D}' = \mathbb{D}^* \setminus \{z = t : 1 \le t \le \infty\}$, and let $\tilde{g}(P)$ be a single-valued branch of $(g(P))^{c_1/c_2}$ defined on $\widetilde{D}_2$ by condition $1^{c_1/c_2} = 1$. Then $\tilde{g}$ is analytic on $\widetilde{D}_2$ and continuous on $\widetilde{D}_2 \cup L$.

Consider the function

$$\Phi(P) = \begin{cases} f(P) & \text{if } P \in D_1 \cup L \\ \tilde{g}(P) & \text{if } P \in \widetilde{D}_2. \end{cases}$$

Since D_1 and D_2 carry proportional harmonic measures on L, it follows from (4.4) that $f(\tau) = \tilde{g}(\tau)$ for every $\tau \in L$. Thus, Φ is continuous and one-to-one on $D_1 \cup \widetilde{D}_2 \cup L$. Since $|\Phi(\tau)| = 1$ for all $\tau \in L$, the latter implies that Φ is analytic in $D_1 \cup \widetilde{D}_2 \cup L$. Hence, L is an analytic arc. The proof is complete. □

The proof above shows also that the analytic continuation of f into $\widetilde{D}_2$ is given by a single-valued branch of the function g^{c_1/c_2}.

Lemma 6 below, which proof is left to the reader, shows that a similar result also holds for a simply connected domain carrying proportional harmonic measures on a boundary slit.

Lemma 6. *Let D carry proportional* (*equal*) *harmonic measures on its boundary slit L and let $\zeta = f(P)$ map D conformally onto $\mathbb{D}$ such that $f(L^+) = l^+$, $f(L^-) = l^-$. Then l^+ and l^- are disjoint arcs of equal length on $\mathbb{T}$ and the inverse mapping $g(\zeta) = f^{-1}(\zeta)$ can be continued analytically onto a doubly-connected domain $\overline{\mathbb{C}} \setminus (\mathbb{T} \setminus (l^+ \cup l^-))$.*

5. Proof of uniqueness

In this section, we prove the uniqueness assertion of Theorem 3, which in turn implies the uniqueness assertion of Theorem 1.

(1) *Proof of the uniqueness assertion of Theorem* 3. Assume that there are two cellular structures $[F_1, B_1]$ and $[F_2, B_2]$ in $\mathcal{F}_\Gamma$ corresponding to graphs Γ_1 and Γ_2, respectively, each of which carries harmonic measures proportional with respect to the weights $\{\alpha_k\}$ and $\{w^i_{kj}\}$. For $s = 1, 2$, let $F_s = \{D^s_1, \dots, D^s_n\}$, $B_s = \{b^s_1, \dots, b^s_n\}$, where $b^s_k \in D^s_k$, $1 \le k \le n$. Let $(e^i_{kj})_s$ denote the edge of Γ_s corresponding to the edge e^i_{kj} of Γ. Let $\zeta = f_{k,s}(P)$ map D^s_k conformally onto the unit disc $\mathbb{D}$ such that $f_{k,s}(b^s_k) = 0$. Let $(\tau^i_{kj})_s = f_{k,s}((e^i_{kj})_s)$ be the image of the boundary arc $(e^i_{kj})_s \subset \partial D^s_k$ under the mapping $f_{k,s}$. Let $\mathcal{A}_{k,s} = \{(\tau^i_{kj})_s\}_{i,j}$ be the collection of all arcs $(\tau^i_{kj})_s$ on $\mathbb{T}$ corresponding to the domain D^s_k and function $f_{k,s}$, $s = 1, 2$, $k = 1, \dots, n$. Since each of the structures $[F_1, A_1]$ and $[F_2, A_2]$ carries

harmonic measures proportional with respect to the same weights, it follows from (4.1) that the systems $\mathcal{A}_{k,1}$ and $\mathcal{A}_{k,2}$ coincide up to rotation about the origin. Rotating $\mathcal{A}_{k,2}$, if necessary, we may assume that $\mathcal{A}_{k,1}$ and $\mathcal{A}_{k,2}$ coincide.

For $1 \le k \le n$, let $\Phi_k(P) = f_{k,2}^{-1}(f_{k,1}(P))$. Then Φ_k maps D_k^1 conformally onto D_k^2 such that $\Phi_k(b_k^1) = b_k^2$. Since each of the domains D_k^1 and D_k^2 is bounded by a finite number of Jordan arcs, it follows from our argument above that Φ_k maps the boundary arc $(e_{kj}^i)_1$ of D_k^1 continuously and one-to-one in the sense of boundary correspondence onto the boundary arc $(e_{kj}^i)_2$ of D_k^2. In particular, the end points of $(e_{kj}^i)_1$ correspond to the end points of $(e_{kj}^i)_2$ under this mapping.

For $k \ne j$, let D_k^s and D_j^s have an edge $(e_{kj}^i)_s$ on their common boundary. By Lemma 5, $(e_{kj}^i)_s$ is an analytic arc. We claim that $\Phi_j(P)$ gives an analytic continuation of the function $\Phi_k(P)$ across $(e_{kj}^i)_1$.

Indeed, by our remark after Lemma 5, the analytic continuation of the function $f_{k,s}$, $s = 1, 2$, across $(e_{kj}^i)_s$ is given by some single-valued branch $g_s(P)$ of the function $e^{i\theta_{k,j}}/(f_{j,s}(P))^{\alpha_k/\alpha_j}$, where $\theta_{k,j}$ is a real constant independent of s. We have

$$g_s^{-1}(\zeta) = f_{j,s}^{-1}((e^{i\theta_{k,j}}\zeta^{-1})^{\alpha_j/\alpha_k}), \qquad s = 1, 2.$$

Therefore,

$$\Phi_j(z) = f_{j,2}^{-1}(f_{j,1}(P)) = g_2^{-1}(g_1(P))$$

is an analytic continuation of Φ_k across $(e_{kj}^i)_1$.

Similarly, using Lemma 6, one can show that Φ_k is single-valued and analytic on every slit $(e_{kk}^i)_1$ lying on the boundary of D_k^1.

Consider the function $\Phi : \mathcal{R} \to \mathcal{R}$ defined by

$$\Phi(P) = \Phi_k(P) \quad \text{if } P \text{ is in the closure of } D_k^1,\ 1 \le k \le n.$$

Our argument above shows that Φ is well defined (single-valued), continuous and one-to-one from $\mathcal{R}$ onto $\mathcal{R}$. Moreover, Φ is analytic on $\mathcal{R}$ except, possibly, the vertices of the graph Γ_1. Therefore using Riemann's theorem on the removable singularity, Φ is analytic and one-to-one from $\mathcal{R}$ onto $\mathcal{R}$. Hence, Φ is a conformal automorphism of $\mathcal{R}$. This proves the uniqueness assertion of Theorem 3. □

(2) *Proof of the uniqueness assertion of Theorem* 1. Assume that there are two complex structures on $\mathcal{R}$ turning $\mathcal{R}$ into compact Riemann surfaces $\mathcal{R}_1$ and $\mathcal{R}_2$ such that $\mathcal{R}_k$, $k = 1, 2$, admits a Jenkins-Strebel quadratic differential $Q_k(z)\,dz^2$, the critical weighted graph $\Phi_{Q_k} = \{V_{Q_k}, E_{Q_k}, F_{Q_k}, W_{Q_k}\}$ of which, complemented, if necessary, by second degree vertices on its edges, is homeomorphic to Γ on $\mathcal{R}$ and carries the same weights as Γ does.

For $k = 1, 2$, let $F_{Q_k} = \{f_{Q_k}^s\}_{s=1}^n$. Let $B_k = \{b_{k,s}\}_{s=1}^n$, where $b_{k,s}$ denotes the double pole of $Q_k(z)\,dz^2$ lying in the circle domain $f_{Q_k}^s$. It follows from Corollary 1 that $[F_{Q_1}, B_1]$ and $[F_{Q_2}, B_2]$ are cellular structures on $\mathcal{R}_1$ and $\mathcal{R}_2$, respectively, each of which carries harmonic measures proportional with respect to the same set of weights $W = \{w_{kj}^i\}$ defined in Theorem 1.

Therefore, it follows from the uniqueness assertion of Theorem 3 proved above that there is a conformal mapping $f : \mathcal{R}_1 \to \mathcal{R}_2$ which transplants the quadratic differential $Q_1(z)\,dz^2$ defined on $\mathcal{R}_1$ to the quadratic differential $Q_2(z)\,dz^2$ defined on $\mathcal{R}_2$. Thus, the required conformal structure and the corresponding Jenkins-Strebel quadratic differential are defined uniquely up to conformal autohomorphisms of $\mathcal{R}$. □

6. Proof of existence

Here we work in the reverse order compared to the previous section. First, we prove the existence assertion of Theorem 1. The latter easily implies the existence assertion of Theorem 3.

(1) *Proof of the existence assertion of Theorem* 1. Without loss of generality we may assume that Γ has no vertices of degree 2. Indeed, if v_l is a vertex of degree 2 and $e^{i_1}_{jk}$, $e^{i_2}_{jk}$ are edges having v_l as their end point, then we can replace $e^{i_1}_{jk}$ and $e^{i_2}_{jk}$ with a single edge $\tilde{e}_{jk} = e^{i_1}_{jk} \cup e^{i_2}_{jk} \cup v_l$ with the weight $w^{i_1}_{jk} + w^{i_2}_{jk}$. Continuing this gluing procedure, we reduce the proof to the case of graphs without vertices of degree 2.

Consider n copies, say $\mathbb{D}^1, \dots, \mathbb{D}^n$, of the unit disk. For $k = 1, \dots, n$, let $\mathcal{A}_k = \{\tau^i_{kj}\}_{i,j}$ be a collection of mutually disjoint open arcs on the boundary of $\mathbb{D}^k$ satisfying the following conditions:

(1) $\mathcal{A}_k$ contains an arc τ^i_{kj} with indexes j and i if and only if E contains an edge e^i_{kj}.
(2) $\cup_{i,j} \bar{\tau}^i_{kj} = \mathbb{T}$;
(3) The cyclic order in which the arcs τ^i_{kj} follow on the circle $\mathbb{T}$ coincides with the cyclic order in which the edges e^i_{kj} follow on the boundary cycle around f_k.
(4) length $(\tau^i_{kj}) = 2\pi\delta^i_{kj}$, where δ^i_{kj} is the relative weight of γ^i_{kj} defined in Section 4.

Next we construct a compact surface by gluing the boundary arcs of the disks $\mathbb{D}^k$, $k = 1, \dots, n$, as follows. Let $\mathcal{A}_k$ contain an arc τ^i_{kj} (and therefore $\mathcal{A}_j$ contains an arc τ^i_{jk}) and let $\zeta^i_{kj} = e^{i\varphi^i_{kj}}$ and $\eta^i_{kj} = e^{i\psi^i_{kj}}$ with $\varphi^i_{kj} < \psi^i_{kj} < \varphi^i_{kj} + 2\pi$ be the initial point of τ^i_{kj} and the terminal point of τ^i_{kj}, respectively, assuming positive orientation of $\mathbb{T}$. In this case, we glue $\mathbb{D}^k$ and $\mathbb{D}^j$ along the closure of τ^i_{kj} by identifying the points of $\bar{\tau}^i_{kj}$ with the points of $\bar{\tau}^i_{jk}$ according to the following rule:

$$\bar{\tau}^i_{kj} \ni e^{i\varphi^i_{kj}} \zeta^{2\delta^i_{kj}} \longleftrightarrow e^{i\psi^i_{jk}} \zeta^{-2\delta^i_{jk}} \in \bar{\tau}^i_{jk}. \tag{6.1}$$

Here $\zeta = e^{i\theta}$, $0 \le \theta \le \pi$.

Applying this gluing procedure for all arcs τ^i_{kj}, we obtain the resulting compact surface $\widetilde{\mathcal{R}}$ that is homeomorphic to $\mathcal{R}$. Below, by P and $z = z_P$ we denote a point on $\widetilde{\mathcal{R}}$ and its affix on $\mathbb{C}$, respectively. Thus, $z = z_P \in \overline{\mathbb{D}}$. Also, by $\mathbb{D}^k_P$ we denote the disk $\mathbb{D}^k$ considered as a subset of $\widetilde{\mathcal{R}}$. The gluing procedure described above generates a weighted graph $\widetilde{\Gamma} = \{\widetilde{V}, \widetilde{E}, \widetilde{F}, \widetilde{W}\}$ on $\widetilde{\mathcal{R}}$ with $\widetilde{V} = \{\zeta^i_{kj}\}$, $\widetilde{E} = \{\tau^i_{kj}\}$, $\widetilde{F} = \{\mathbb{D}^k\}_{k=1}^n$, and $\widetilde{W} = \{w^i_{kj}\}$. We emphasize once more that τ^i_{kj} and τ^i_{jk} denote the same edge of $\widetilde{\Gamma}$ since the points of these arcs are identified via formula (6.1). In a similar way this identification works for the end points ζ^i_{kj} and η^i_{jk} as well. Namely, each pair ζ^i_{kj}, η^i_{jk} defines the same point $P \in \widetilde{\mathcal{R}}$. In addition, several such pairs $\zeta^{i_s}_{k_s j_s}$, $\eta^{i_s}_{j_s k_s}$, being identified via (6.1), may define the same vertex of $\widetilde{\Gamma}$. It is not difficult to see that the surface $\widetilde{\mathcal{R}}$ with the graph $\widetilde{\Gamma}$ embedded in $\widetilde{\mathcal{R}}$ is homeomorphic to the surface $\mathcal{R}$ with the graph Γ embedded in $\mathcal{R}$.

To uniformize $\widetilde{\mathcal{R}}$, we introduce a complex atlas $\mathcal{U}$ as follows:

(1) Each disk $\mathbb{D}^k_P$ supplied with its "natural" local coordinate $z = z_P$ is considered as a chart $(\mathbb{D}^k_P, z_P)$ of $\mathcal{U}$.

(2) For each arc τ^i_{kj} we assign a chart (U^i_{kj}, q^i_{kj}) as follows. Let $\Delta^i_{kj} = \{P \in \mathbb{D}^k_P :\ \varphi^i_{kj} < \arg z_P < \psi^i_{kj}\}$ and let $U^i_{kj} = \{P \in \widetilde{\mathcal{R}} :\ P \in \Delta^i_{kj} \cup \Delta^i_{jk} \cup \tau^i_{kj}\}$. Then U^i_{kj} is a simply connected open subset of $\widetilde{\mathcal{R}}$. A corresponding local coordinate $q^i_{kj} : U^i_{kj} \to \mathbb{C}$ can be introduced by

$$q^i_{kj}(P) = \begin{cases} \left(e^{-i\varphi^i_{kj}} z_P\right)^{\frac{1}{2\delta^i_{kj}}} & \text{if } P \in \Delta^i_{kj} \cup \tau^i_{kj}, \\ \left(e^{-i\psi^i_{jk}} z_P\right)^{-\frac{1}{2\delta^i_{jk}}} & \text{if } P \in \Delta^i_{jk}. \end{cases} \tag{6.2}$$

It follows from (6.1) that q^i_{kj} is continuous and one-to-one from U^i_{kj} onto the upper half-plane $\{\zeta \in \mathbb{C} :\ \Im\,\zeta > 0\}$.

(3) Now we assign a chart for each vertex of $\widetilde{\Gamma}$. Assume that $P_l \in \widetilde{\mathcal{R}}$, $l = 1, \dots, m$, is a vertex of $\widetilde{\Gamma}$ of degree $d \ge 3$. In this case, we can find precisely d distinct initial points $\zeta^{i_1}_{k_1 j_1}, \dots, \zeta^{i_d}_{k_d j_d}$, corresponding to the edges $\tau^{i_1}_{k_1 k_2}, \dots, \tau^{i_d}_{k_d j_d}$, such that the points $\zeta^{i_s}_{k_s j_s} = e^{i\varphi^{i_s}_{k_s j_s}}$, $s = 1, \dots, d$, represent P_l on the boundary of $\mathbb{D}^{k_s}_P$. We should note here that the disks $\mathbb{D}^{k_s}_P$ and $\mathbb{D}^{k_p}_P$ are not necessarily distinct. We may assume that $\zeta^{i_s}_{k_s j_s}$, $s = 1, \dots, d$, are enumerated in such way that the corresponding faces $\mathbb{D}^{k_1}_P, \dots, \mathbb{D}^{k_{d+1}}_P$, where $\mathbb{D}^{k_{d+1}}_P := \mathbb{D}^{k_1}_P$, follow clockwise around P_l. Then, of course, $j_s = k_{s+1}$ for $s = 1, \dots, d$.

For $\varepsilon > 0$ sufficiently small and $s = 1, \dots, d$, define

$$V^{\varepsilon,s}_{P_l} = \{P \in \overline{\mathbb{D}^{k_s}_P} :\ |\arg z_P - \varphi^{i_s}_{k_s k_{s+1}}| < 2\varepsilon \delta^{i_1}_{k_1 k_2} \alpha_{k_1}/\alpha_{k_s}\}.$$

Let $V^\varepsilon_{P_l} = \cup_{s=1}^d V^{\varepsilon,s}_{P_l}$. Assuming that the points of the corresponding circular parts of the boundaries of $V^{\varepsilon,s}_{P_l}$ and $V^{\varepsilon,s+1}_{P_l}$ are identified according to the rule (6.1), the set $V^\varepsilon_{P_l}$ can be considered as an open neighborhood of $P_l \in \widetilde{\mathcal{R}}$. A corresponding local coordinate $\zeta = q_l(P)$ can be introduced as $\zeta = \left(\frac{w-1}{w+1}\right)^{2/d}$, where $w = w(P)$ is defined by

$$w = \begin{cases} \left(e^{-i\varphi^{i_1}_{k_1k_2}} z_P\right)^{\frac{1}{2\delta^{i_1}_{k_1k_2}}} & \text{if } P \in V^{\varepsilon,1}_{P_l}, \\ \left(e^{-i\psi^{i_{s-1}}_{k_sk_{s-1}}} z_P\right)^{\frac{(-1)^{s+1}}{2\delta^{i_1}_{k_1k_2}} \frac{\alpha_{k_s}}{\alpha_{k_1}}} & \text{if } P \in V^{\varepsilon,s}_{P_l},\ s = 2, \dots, d. \end{cases} \tag{6.3}$$

Here α_k denotes the length of the boundary cycle around f_k defined by (2.1). Using the identification rule (6.1), one can easily verify that $q_l(P)$ defined above maps $V^{\varepsilon}_{P_l}$ continuously and one-to-one onto some neighborhood of the point $0 \in \mathbb{C}$.

The explicit expressions in (6.2) and (6.3) show that the local coordinates introduced in (1)–(3) are conformally compatible. Combining all the charts introduced in (1)–(3), we obtain the desired complex atlas $\mathcal{U}$ on $\widetilde{\mathcal{R}}$.

Now for each chart introduced above, we define a quadratic differential $Q(\zeta)\, d\zeta^2$ satisfying the desired conditions. First, we put

$$Q(\zeta)\, d\zeta^2 = -\frac{\alpha_k^2}{4\pi^2} \frac{d\zeta^2}{\zeta^2} \quad \text{with } \zeta = z_P, \quad \text{if } P \in \mathbb{D}^k_P,\ k = 1, \dots, n. \tag{6.4}$$

This immediately implies that each disk $\mathbb{D}^k_P$ is a circle domain of $Q(\zeta)\, d\zeta^2$ and every closed trajectory of $Q(\zeta)\, d\zeta^2$ lying in $\mathbb{D}^k_P$ has the Q-length equal to α_k.

To define $Q(\zeta)\, d\zeta^2$ in terms of the local coordinate (6.2), we put

$$Q(\zeta)\, d\zeta^2 = -\frac{\left(w^i_{kj}\right)^2}{4\pi^2} \frac{d\zeta^2}{\zeta^2} \quad \text{with } \zeta = q^i_{kj}(z_P), \quad \text{if } P \in U^i_{kj}. \tag{6.5}$$

Equation (6.5) implies that the Q-length of τ^i_{kj} equals w^i_{kj}.

Finally, to define $Q(\zeta)\, d\zeta^2$ in a vicinity of the vertex P_l, we fix $\varepsilon > 0$ sufficiently small and then put

$$Q(\zeta)\, d\zeta^2 = -\frac{d^2 \left(w^{i_1}_{k_1k_2}\right)^2}{\pi^2} \frac{\zeta^{d-2}}{(\zeta^d - 1)^2}\, d\zeta^2 \quad \text{with } \zeta = q_l(z_P), \quad \text{if } P \in V^{\varepsilon}_{P_l}. \tag{6.6}$$

Changing variables in (6.4) via (6.2) or (6.3) and using the transformation rule (2.2), one can easily convert the quadratic differential (6.4) into the form given by (6.5) or (6.6), respectively. Thus, equations (6.4), (6.5), and (6.6) define the same quadratic differential on $\widetilde{\mathcal{R}}$.

By our construction, this quadratic differential has disks $\mathbb{D}^k_P$ as its circle domains and the weighted graph $\widetilde{\Gamma}$ as its critical graph. In addition, each open set corresponding to a local coordinate introduced in (1), (2), or (3) has a nonempty open intersection with at least one of the circle domains $\mathbb{D}^k_P$, $k = 1, \dots, n$. This implies that the domain configuration of the constructed quadratic differential $Q(\zeta)\, d\zeta^2$ consists precisely of the domains $\mathbb{D}^1_P, \dots, \mathbb{D}^n_P$. In particular, it does not contain density structures. Thus, $Q(\zeta)\, d\zeta^2$ is a Jenkins-Strebel quadratic differential on $\widetilde{\mathcal{R}}$. This completes the proof of the existence assertion of Theorem 1. □

(2) *Proof of the existence assertion of Theorem* 3. The domain configuration of the quadratic differential $Q(\zeta)\,d\zeta^2$ constructed in the proof above satisfies all topological requirements of Theorem 3. In addition, it follows from our construction in the proof above and Corollary 1 that this domain configuration, considered as a cellular structure on $\widetilde{\mathcal{R}}$, carries harmonic measures proportional with respect to the weights $\{w_{kj}^i\}$. □

Remark. In the first version of this paper, to prove the existence of quadratic differentials with required properties, we used a special extremal problem, which will be discussed in the next section. The referee of that version suggested that uniformization can be used to prove the existence assertion. A simple existence proof given above is based on his suggestion.

7. An extremal problem related to Theorem 1

The well-known heuristic Teichmüller principle, valid for a wide class of extremal problems on conformal and quasiconformal mappings, asserts that every such problem is intrinsically related to a certain quadratic differential. So, a question of interest is what extremal problem is hiding behind the quadratic differential appearing in Theorems 1 and 3? Of course, if a conformal structure on $\mathcal{R}$ is prescribed and the positions of double poles $b_1, \dots, b_n$ are already determined, then the quadratic differential of Theorem 1 is just a quadratic differential corresponding to the Jenkins problem on the extremal partition of $\mathcal{R}$. The main question here is how to deal with the case of varying Riemann surfaces and moving centers $b_1, \dots, b_n$. In this section, we first discuss this question for a general compact surface of genus $g \geq 0$. Then, we give a detailed treatment to this problem for the case $\mathcal{R} = \mathbb{S}^2$.

Let Γ be a simply connected graph embedded in a compact surface $\mathcal{R}$ as described in Theorem 1. Let Γ' be a graph dual of Γ as defined in Section 2. Let $V' = A \cup B$, $L = \{l_{kj}^i\}$, and $W = \{w_{kj}^i\}$ be the sets of vertices, edges, and weights of Γ', respectively. Here $A = \{a_1, \dots, a_l\}$ denotes the set of all first degree vertices of Γ and $B = \{b_1, \dots, b_n\}$ denotes the set of centers $b_k \in f_k$. Equivalently, A is the set of all isolated vertices of Γ'.

Let $\mathcal{U}$ be a complex atlas turning $\mathcal{R}$ into a Riemann surface $\mathcal{R}_{\mathcal{U}}$ in the corresponding moduli space. Let $\mathcal{M}_{\mathcal{U},1}$ be the solution to Problem 1 on $\mathcal{R}_{\mathcal{U}}$ for the punctures $A = \{a_1, \dots, a_l\}$, centers $B = \{b_1, \dots, b_n\}$, and weights $\alpha = \{\alpha_1, \dots, \alpha_n\}$. Let $\mathcal{M}_{\mathcal{U},2}$ be the solution to Problem 2 on $\mathcal{R}_{\mathcal{U}}$ for the graph Γ'.

To unify notations, we will write $C = (A, B) = (c_1, \dots, c_{l+n})$, where $c_k = a_k$ for $1 \leq k \leq l$ and $c_k = b_{k-l}$ for $l < k \leq l + n$. Keeping the weights w_{kj}^i, and therefore the total weights α_k, to be fixed, we consider the multi-valued function

$$\mathcal{M}_{\mathcal{U}} = \mathcal{M}_{\mathcal{U}}(A, B) = \mathcal{M}_{\mathcal{U}}(C) = \mathcal{M}_{\mathcal{U}}(c_1, \dots, c_{l+n}) = \mathcal{M}_{\mathcal{U},1} + \mathcal{M}_{\mathcal{U},2},$$

which is locally single-valued and continuous on the set $\widetilde{\mathcal{R}}_{\mathcal{U}}^{l+n}$.

Now we are ready to state the main problem of this section.

Problem 3. For a given weighted graph Γ embedded into a compact surface $\mathcal{R}$ of genus $g \geq 0$, find the supremum

$$\mathcal{M} = \mathcal{M}(\mathcal{R}, \Gamma) = \sup \mathcal{M}_{\mathcal{U}}(c_1, \ldots, c_{l+n}) \tag{7.1}$$

taken over all complex structures $\mathcal{U}$ on $\mathcal{R}$, all vectors $(c_1, \ldots, c_{l+n}) \in \widetilde{\mathcal{R}}_{\mathcal{U}}^{l+n}$, and all branches of $\mathcal{M}_{\mathcal{U}}$.

By Emel'yanov's comparison theorem for dual problems on extremal partitions (see [19, Theorem 3.3] for a version of this theorem for Riemann surfaces),

$$\mathcal{M}_{\mathcal{U}} = \mathcal{M}_{\mathcal{U},1} + \mathcal{M}_{\mathcal{U},2} \leq 0 \quad \text{on } \widetilde{\mathcal{R}}_{\mathcal{U}}^{l+n} \tag{7.2}$$

for every branch of $\mathcal{M}_{\mathcal{U}}$.

Furthermore, equality holds in (7.2) if and only if $Q_{\mathcal{U},1}(z)\,dz^2 = -Q_{\mathcal{U},2}(z)\,dz^2$. Here $Q_{\mathcal{U},1}(z)\,dz^2$ and $Q_{\mathcal{U},2}(z)\,dz^2$ denote quadratic differentials on $\mathcal{R}_{\mathcal{U}}$ extremal for the problems on $\mathcal{M}_{\mathcal{U},1}$ and $\mathcal{M}_{\mathcal{U},2}$, respectively. In other words, equality holds in (7.2) if and only if the Kuz'mina quadratic differential for the problem on $\mathcal{M}_{\mathcal{U},2}$ coincides up to a sign with the Jenkins-Strebel quadratic differential for the problem on $\mathcal{M}_{\mathcal{U},1}$. The quadratic differential defined by Theorems 1 and 3 is precisely of this kind. This gives the following solution to Problem 3.

Theorem 4. *Let $\mathcal{R}$ and Γ be a compact surface and a graph as in Problem 3. Then there is a conformal structure $\mathcal{U}_0$, a vector $(c_1^0, \ldots, c_{l+n}^0) \in \widetilde{\mathcal{R}}_{\mathcal{R}_0}^{l+n}$, and a branch $\widehat{\mathcal{M}}_{\mathcal{U}_0}$ of $\mathcal{M}_{\mathcal{U}_0}$ such that*

$$\widehat{\mathcal{M}}_{\mathcal{U}_0}(c_1^0, \ldots, c_{l+n}^0) = \mathcal{M}(\mathcal{R}, \Gamma) = 0.$$

Furthermore, the complex structure $\mathcal{U}_0$, the vector $(c_1^0, \ldots, c_{l+n}^0)$, and the branch $\widehat{\mathcal{M}}_{\mathcal{U}_0}$ are determined uniquely up to conformal automorphisms.

Problem 3 suggests another way (although rather technical) to establish the existence of a quadratic differential in Theorems 1 and 3. Namely, one can start with a maximizing sequence of complex structures $\{\mathcal{U}^{(k)}\}_{k=1}^{\infty}$, vectors $\{C^{(k)}\}_{k=1}^{\infty}$, and single-valued branches $\{\widehat{\mathcal{M}}_{\mathcal{U}^{(k)}}\}_{k=1}^{\infty}$ such that $\widehat{\mathcal{M}}_{\mathcal{U}^{(k)}}(C^{(k)}) \to \mathcal{M}(\mathcal{R}, \Gamma)$ and then show that, for some subsequence of indices $k_s \to \infty$, the corresponding subsequence of sets of faces converges to a domain configuration of an appropriate quadratic differential. Below in this section, we show how this can be done in the case of the Riemann sphere, i.e., we assume that $\mathcal{R} = \overline{\mathbb{C}}$. A similar, but even more technical proof works for a general compact surface as well.

Let us consider the maximization problem 7.1 for the multi-valued function $\mathcal{M}$ on the set $\widetilde{\mathbb{C}}^{l+n}$. The known results on the change of the reduced module under conformal mapping, cf. [19, Lemma 1.3], show that $\mathcal{M}$ is Möbius invariant, i.e.,

$$\mathcal{M}(\varphi(c_1), \ldots, \varphi(c_{l+n})) = \mathcal{M}(c_1, \ldots, c_{l+n})$$

for an appropriate choice of branches of $\mathcal{M}$ and every Möbius map φ.

Without loss of generality we may assume that Γ has no vertices of degree 2.

1. In the case $l+n=3$, the required quadratic differentials can be found explicitly. Assume, for example, that $n=3$, $l=0$. In two other cases, when $n=2$, $l=1$ and $n=1$, $l=2$, verification is similar and is left to the reader. Let Γ' be dual of the graph Γ. We may assume that Γ' has its vertices at $b_1=0$, $b_2=1$, and $b_3=\infty$. Since $l=0$, Γ' has no loops. Since $\deg v_k\neq 2$ for $v_k\in V$, Γ' has no parallel edges belonging to the same homotopic class on the punctured sphere $\mathbb{C}\setminus\{0,1\}$. Thus, we are left with the following three subcases:

(1) Γ' has edges $l_{13}=(-\infty,0)$ and $l_{23}=(1,+\infty)$ with positive weights w_{13} and w_{23}, respectively.
(2) Γ' has edges l_{13} and l_{23} as in (1) and the edge $l_{12}=(0,1)$ with weight $w_{13}>0$.
(3) Γ' has edges w_{13}, w_{23} as above and the edge $l_{33}=\{z=(1/2)+it:-\infty<t<\infty\}$ with weight $w_{33}>0$.

In each of these cases the corresponding quadratic differential

$$Q_i(z)\,dz^2=-\frac{1}{4\pi^2}\frac{P_i(z)}{z^2(z-1)^2}\,dz^2,\quad i=1,2,3,$$

can be found explicitly. Namely we have:

Case 1: $P_1(z)=\alpha_3^2(z-p)^2$ with $\alpha_3=w_{13}+w_{23}$, $\alpha_1=w_{13}$, and $p=\alpha_1\alpha_3^{-1}$;

Case 2: $P_2(z)=\alpha_3^2(z-a)(z-\bar a)$ with $\alpha_3=w_{13}+w_{23}$, $\alpha_1=w_{12}+w_{13}$, $\alpha_2=w_{12}+w_{23}$, and a defined by conditions $|a|=\alpha_1\alpha_3^{-1}$, $|1-a|=\alpha_2\alpha_3^{-1}$, $\Im a>0$;

Case 3: $P_3(z)=\alpha_3^2(z-p_1)(z-p_2)$ with $\alpha_3=w_{13}+w_{23}+w_{33}$, $\alpha_1=w_{13}$, $\alpha_2=w_{23}$, and $0<p_1<p_2<1$ defined by equations $p_1p_2=\alpha_1^2\alpha_3^{-2}$, $p_1+p_2=1+\alpha_1^2\alpha_3^{-2}-\alpha_2^2\alpha_3^{-2}$.

2. Let $l+n\geq 4$. Our first goal is to show that $\mathcal{M}$ achieves its maximal value

$$M=\sup\mathcal{M}(c_1,\dots,c_{l+n}),\quad -\infty<M\leq 0,$$

at some point $C^*=(A^*,B^*)\in\widetilde{\mathbb{C}}^{l+n}$, where $A^*=(a_1^*,\dots,a_l^*)\in\widetilde{\mathbb{C}}^l$, $B^*=(b_1^*,\dots,b_n^*)\in\widetilde{\mathbb{C}}^n$, and for some single-valued branch of $\mathcal{M}$.

(a) Let Γ_s', $s=1,2,\dots$, be a maximizing sequence of graphs and let $C_s=(A_s,B_s)\in\widetilde{\mathbb{C}}^{l+n}$ be the sequence of vertex sets of Γ_s'. Without loss of generality we may assume that $C_s\to C^*=(A^*,B^*)=(c_1^*,\dots,c_{l+n}^*\}\in\overline{\mathbb{C}}^{l+n}$ as $s\to\infty$. To emphasize that objects under consideration depend on s, we will use an additional lower index s. Accordingly, we will write Γ_s, $C_s=(A_s,B_s)$, $\mathcal{M}_s$, $\mathcal{M}_{1,s}$, $\mathcal{M}_{2,s}$, etc. Then

$$\mathcal{M}_s=\mathcal{M}(c_{1,s},\dots,c_{l+n,s})\to M\quad\text{as } s\to\infty. \tag{7.3}$$

We claim that using auxiliary Möbius maps, selecting subsequences, and changing numeration, if necessary, we may either achieve a situation when all the coordinates of C^* are distinct or, alternatively, we can find indices j_1, j_2, j_3, and j_4 such that conditions (a), (b), and (c) below are fulfilled and, in addition, at least one of the conditions (d_1) or (d_2) is satisfied:

(a) $c_{j_1,s}=0$, $c_{j_2,s}=\infty$, $c_{j_3,s}=1$ for all s.
(b) $|c_{k,s}|\leq 1$ for all $k\neq j_2$ and all s.

(c) $c_{j_4,s} \to 0$ as $s \to \infty$.

(d_1) The graph Γ'_s contains a loop, called $(l^1_{j_1,j_1})_s$, which has its ends at $c_{j_1,s} = 0$ and separates $c_{j_2,s} = \infty$ from $c_{j_3,s} = 1$ on $\overline{\mathbb{C}} \setminus \{0\}$.

(d_2) The graph Γ'_s contains an edge, called $(l^1_{j_1,j_2})_s$, which joins $c_{j_1,s} = 0$ and $c_{j_2,s} = \infty$.

Indeed, using a suitable Möbius map we may first satisfy the following three conditions: $c_{j_1,s} = b_{1,s} = 0$, $c_{j_2,s} = \infty$, where $c_{j_2,s} = b_{2,s}$ if $n \geq 2$ and $b_{j_2,s} = a_{1,s}$ otherwise, and $\max_{k\geq 3} |c_{k,s}| = 1$ for all $s = 1, 2, \ldots$ For every given s, the number of vertices $c_{k,s}$ is finite. Therefore selecting subsequences and rotating about the origin, if necessary, we may assume in addition that $c_{j_3,s} = 1$ for some index $j_3 \neq j_1, j_2$ and all $s = 1, 2, \ldots$ We still assume that $c_{k,s} \to c^*_k$ for $1 \leq k \leq l + n$ as $s \to \infty$.

If $c^*_1, \ldots, c^*_{l+n}$ are not all distinct, then we have to consider the following cases:

(i) Let $\#B^* = 1$. Since $c_{j_1,s} = b_{1,s} = 0$ and $c_{j_2,s} = \infty$, this assumption implies that $c_{j_2,s} = a_{1,s}$ is an isolated vertex of Γ'_s. Therefore $n = 1$ and B_s contains the only one vertex $b_{1,s} = 0$. Hence every edge of Γ'_s is a loop having its ends at 0. In particular, L_s contains a loop, which separates $c_{j_2,s} = \infty$ from $c_{j_3,s} = 1$ on $\mathbb{C} \setminus \{0\}$. Since $c^*_1, \ldots, c^*_{l+n}$ are not all distinct, there are sequences $c_{i_1,s}, \ldots, c_{i_p,s}$, $2 \leq p \leq l + n - 2$, each of which converges to the same limit $a \in \overline{\mathbb{D}}$. If $a = 0$, then conditions (a), (b), (c), and (d_1) are satisfied.

If $a \neq 0$, then we can choose two sequences, call them $c_{i_1,s}$ and $c_{i_2,s}$, such that $c_{i_1,s} \to a$, $c_{i_2,s} \to a$. Since $c_{i_1,s}$ is an isolated vertex of Γ'_s, there is a loop, called $(l^1_{1,1})_s$, which separates $c_{i_1,s}$ from all other vertices of Γ'_s on $\overline{\mathbb{C}} \setminus \{0\}$. Let φ_s be a Möbius map such that $\varphi_s(0) = 0$, $\varphi_s(c_{i_1,s}) = \infty$, $\varphi_s(c_{i_2,s}) = 1$. Then one can easily see that $\varphi_s(\infty) \to 0$ as $s \to \infty$. If $|\varphi_s(c_{k,s})| \leq 1$ for all $c_{k,s} \neq \infty$, then conditions (a), (b), (c), and (d_1) are satisfied. If $|\varphi_s(c_{k,s})| > 1$ for some $c_{k,s} \neq \infty$, then using suitable dilations, selecting subsequences, and changing numeration, if necessary, we will get the same conditions.

(ii) Let $\#B^* \geq 2$. Assume that there are sequences $c_{i_1,s}, \ldots, c_{i_p,s}$, $2 \leq p \leq l+n-2$, each of which converges to the same limit $a \in \overline{\mathbb{D}}$. Assume, in addition, that at least one of these sequences, let $c_{i_1,s}$, is a sequence of centers, i.e., $c_{i_1,s} \in B_s$.

In this case we can find sequences $c_{k',s} \in B_s$ and $c_{m',s} \in B_s$ such that $c_{k',s} \to a$, $c_{m',s} \to b \neq a$ and such that for every s the graph Γ'_s has an edge, called $(\tilde{\gamma}_{k',m'})_s$, which joins the vertices $c_{k',s}$ and $c_{m',s}$.

Now, using Möbius maps φ_s such that $\varphi_s(c_{k',s}) = 0$, $\varphi_s(c_{m',s}) = \infty$ for $s = 1, 2, \ldots$, and then selecting subsequences, changing numeration, and using dilations, if necessary, we will satisfy requirements (a), (b), (c), and (d_2).

(iii) Let $\#B^* \geq 2$. Assume as in case **(ii)** that there are sequences $c_{i_1,s}, \ldots, c_{i_p,s}$, $2 \leq p \leq l + n - 2$, each of which converges to the same limit $a \in \overline{\mathbb{D}}$. In contrast to case **(ii)**, we assume now that there are no sequences of centers, which converge to a. Then, of course, $a \neq 0$ since $c_{j_1,s} = b_{1,s} = 0$.

Since $c_{i_1,s}$ is an isolated vertex of Γ'_s, there is a loop $\tilde{l}_s$ of Γ'_s having its ends at some vertex $\tilde{b}_s$, which separates $c_{i_1,s}$ from all other vertices of Γ'_s on $\overline{\mathbb{C}} \setminus \{\tilde{b}_s\}$. Let φ_s be a Möbius map such that $\varphi_s(\tilde{b}_s) = 0$, $\varphi_s(c_{i_1,s}) = \infty$, $\varphi_s(c_{i_2,s}) = 1$. As in the case **(i)**, one can easily see that $\varphi_s(\infty) \to 0$ and $\varphi_s(0) \to 0$ as $s \to \infty$. Therefore conditions (a), (c), and (d$_1$) are satisfied. To satisfy condition (b), we apply, as above, suitable dilations, selections of subsequences, and changing numeration if necessary.

Our next claim is that if conditions (a), (b), and (c) are fulfilled and, in addition, at least one of conditions (d$_1$) or (d$_2$) is satisfied, then

$$\mathcal{M}_s = \mathcal{M}(c_{1,s}, \ldots, c_{l+n,s}) \to -\infty \quad \text{as } s \to \infty. \tag{7.4}$$

To prove this claim, we first introduce notations, which also will be used in part **(b)** of this proof. If $b_{k,s} \neq \infty$, then let $d_{k,s} = \min_{c_{j,s} \neq b_{k,s}} |b_{k,s} - c_{j,s}|$ be the distance from $b_{k,s}$ to the closest vertex of Γ'_s different from $b_{k,s}$. If $b_{k,s} = \infty$, then we put $d_{k,s} = \max_{c_{j,s} \neq b_{k,s}} |c_{j,s}|$. Of course, under conditions (a) and (b) above, $d_{j_2,s} = 1$.

Let $D_{1,s}, \ldots, D_{n,s}$ be the extremal partition of Problem 1 for the set of punctures A_s, set of centers B_s, and set of weights α. Then Koebe's 1/4-theorem, see [11, Theorem 2.9], implies that

$$m(D_{k,s}, b_{k,s}) \leq \frac{1}{2\pi} \log(4d_{k,s}), \quad 1 \leq k \leq n, \quad s = 1, 2, \ldots. \tag{7.5}$$

Therefore,

$$\mathcal{M}_{1,s} = \sum_{k=1}^{n} \alpha_k^2 m(D_{k,s}, b_{k,s}) \leq \frac{1}{2\pi} \sum_{k=1}^{n} \alpha_k^2 \log d_{k,s} + C \tag{7.6}$$

with some real C independent of s.

If $b_{k,s} \neq \infty$, then let $\Delta_{k,s} = \{z : |z - b_{k,s}| < d_{k,s}/4\}$. If $b_{k,s} = \infty$, then we put $\Delta_{k,s} = \{z \in \overline{\mathbb{C}} : |z| > 4d_{k,s}\}$. Then for every given s, $\Delta_{1,s}$, $\Delta_{2,s}$, $\ldots$, $\Delta_{n,s}$ are disjoint discs on $\overline{\mathbb{C}}$ centered at the vertices $b_{1,s}, \ldots, b_{n,s}$ of the graph Γ'_s. Let $Q_2(z,s)\,dz^2$ be the extremal quadratic differential corresponding to Problem 2 defined for the graph Γ'_s with the set of vertices $C_s = (A_s, B_s)$. Let $\{(G^i_{kj})_s\}$ be the domain structure of $Q_2(z,s)\,dz^2$. Here for every i and s, $(G^i_{kj})_s$ is a digon having its vertices at $b_{k,s}$ and $b_{j,s}$. Let $(\widetilde{G}^i_{kj})_s$ denote the connected component of the intersection $((G^i_{kj})_s) \bigcap \Delta_{k,s}$, which contains the vertex $b_{k,s}$ on its boundary. We note that in case $k = j$ the digon $(G^i_{kk})_s$ has both its vertices at the same point and therefore we have two such connected components. In this case, $(\widetilde{G}^i_{kj})_s$ will denote the union of these two components. Let $\widetilde{\Delta}_{k,s} = \bigcup_{j,i}((\widetilde{G}^i_{kj})_s)$.

If γ is a closed Jordan curve in $\Delta_{k,s}$ separating $b_{k,s}$ from $\partial\Delta_{k,s}$, then (3.3) and (2.1) yield

$$\alpha_k^{-1} \int_{\gamma \cap \widetilde{\Delta}_{k,s}} |Q_2(z,s)|^{1/2}\, |dz| \geq 1.$$

Therefore the restriction of the metric $\alpha_k^{-1}|Q_2(z,s)|^{1/2}|dz|$ onto the set $\widetilde{\Delta}_{k,s}$ is an admissible metric for the reduced module problem in the punctured disc $\Delta_{k,s} \setminus \{b_{k,s}\}$. Hence for $1 \le k \le n$ and $s = 1, 2, \ldots$, we have

$$m(\Delta_{k,s}, b_{k,s}) = \frac{1}{2\pi}\log(d_{k,s}/4) \le \lim_{\varepsilon\to 0^+}\left(\alpha_k^{-2}\iint_{\widetilde{\Delta}^{\varepsilon}_{k,s}} |Q_2(z,s)|\,dA_z + \frac{1}{2\pi}\log\varepsilon\right), \tag{7.7}$$

where $\widetilde{\Delta}^{\varepsilon}_{k,s} = \widetilde{\Delta}_{k,s} \setminus \{z : |z - b_{k,s}| < \varepsilon\}$ if $k \ne 2$ and $\widetilde{\Delta}^{\varepsilon}_{2,s} = \{z : |z| > 1/\varepsilon\} \setminus \widetilde{\Delta}_{2,s}$.

Suppose that conditions (a), (b), (c), and condition (d_2) are satisfied. Then the extremal decomposition of Problem 2 for the graph Γ'_s contains a digon, called $(G^1_{12})_s$, which has its vertices at $b_{1,s} = 0$ and $b_{2,s} = \infty$. Let R_s be the quadrilateral defined, as in Lemma 2, for the digon $\left(G^1_{12}\right)_s$ and the annulus $A(s) = A(r_1, r_2)$ with $r_1 = d_{1,s}/4$, $r_2 = 4d_{2,s}$. Let l'_s and l''_s denote the sides of R_s joining the circles $\partial\Delta_{1,s}$ and $\partial\Delta_{2,s}$. By (3.3), the metric $(w^1_{12})^{-1}\,|Q_2(z,s)|\,|dz|$ is admissible for the module problem in R_s. Therefore,

$$\operatorname{Mod}(R_s) \le (w^1_{12})^{-2}\iint_{R_s} |Q_2(z,s)|\,dA_z, \qquad s = 1, 2, \ldots \tag{7.8}$$

Since every closed Jordan curve separating the boundary circles of the annulus $A(s)$ contains an arc joining the sides l'_s and l''_s of R_s, the well-known comparison principle for module/extremal length, see [1], implies that

$$\operatorname{Mod}(R_s) \ge \operatorname{Mod}(A(s)) = \frac{1}{2\pi}\log(16/d_{1,s}) \to +\infty \quad \text{as } s \to \infty \tag{7.9}$$

since $d_{1,s} \to 0$ by condition (c).

Finally, combining relations (7.5)–(7.8), we obtain

$$\begin{aligned}
\mathcal{M}_s = \mathcal{M}_{1,s} + \mathcal{M}_{2,s} &= \sum_{k=1}^{n}\alpha_k^2 m(D_{k,s}, b_{k,s}) - \sum_{k,j,i}(w^i_{kj})^2 m((G^i_{kj})_s, b_{k,s}, b_{j,s}) \\
&\le C - \iint_{R_s} |Q_2(z,s)|\,dA_z \qquad\qquad (7.10)\\
&\quad + \frac{1}{2\pi}\sum_{k=1}^{n}\alpha_k^2\log d_{k,s} - \lim_{\varepsilon\to 0}\sum_{k=1}^{n}\left(\iint_{\widetilde{\Delta}^{\varepsilon}_{k,s}} |Q_2(z,s)|\,dA_z + \frac{\alpha_k^2}{2\pi}\log\varepsilon\right) \\
&\le C - \iint_{R_s} |Q_2(z,s)|\,dA_z \le C - \operatorname{Mod}(R_s),
\end{aligned}$$

which together with (7.9) implies (7.4). The latter, of course, contradicts (7.3).

The proof given above remains valid if we replace condition (d_2) by condition (d_1). We still assume that conditions (a), (b), and (c) are satisfied. In this case, R_s will denote the quadrilateral in $\overline{\mathbb{C}} \setminus \overline{\Delta}_{1,s}$ defined for the digon $(G^1_{11})_s$ as in Lemma 3. Then R_s has its pair of distinguished sides on $\partial\Delta_{1,s}$ and separates the point $c_{j_2,s} = \infty$ from $c_{j_3,s} = 1$ on $\overline{\mathbb{C}} \setminus \overline{\Delta}_{1,s}$. All inequalities in (7.10) remain valid in this case. The well-known estimates for the module of quadrilaterals, related to

Grötzsch's ring, see [1], show that $\operatorname{Mod}(R_s) \to \infty$ as $s \to \infty$. This together with (7.10) again contradicts (7.3).

Therefore in the rest of this proof we may assume that all coordinates of the limit point $C^* = (A^*, B^*)$ are distinct.

(b) Next we show that all graphs Γ'_s of the maximizing sequence defined in part **(a)** can be chosen to be in the same homotopy class defined by some deformation of $\overline{\mathbb{C}}$ with the initial point $C^* = (A^*, B^*)$. This will allow us to apply differentiation formulas (3.4) to locate the maxima of $\mathcal{M}$.

Arguing by contradiction, we assume that all graphs Γ'_s are homotopically distinct in a vicinity of the point C^*. Then Lemma 1 implies that the sequence of indices $\{\operatorname{ind}(\Gamma'_s)\}_{s=1}^{\infty}$ can not be bounded. Therefore there is a subsequence of graphs, still denoted by Γ'_s, such that

$$\operatorname{ind}_{c_{1,s}c_{2,s}}(l_s) \to +\infty \tag{7.11}$$

for some vertices $c_{1,s}$ and $c_{2,s}$ of Γ'_s and some edge l_s of Γ'_s. Let $c_{3,s}$ and $c_{4,s}$ be the ends of l_s. Here we assume that each of the vertices $c_{1,s}$ and $c_{2,s}$ is distinct from the other three vertices but the vertices $c_{3,s}$ and $c_{4,s}$ may coincide.

Assume first that $c_{3,s} \neq c_{4,s}$. Then, changing numeration if necessary, we may assume that $c_{3,s} = a_{1,s} = 0$, $c_{4,s} = a_{2,s} = \infty$, $c_{1,s} = 1$. We claim, as in part **(a)**, that $\mathcal{M}_s = \mathcal{M}_s(c_{1,s}, \ldots, c_{l+n,s}) \to -\infty$ as $s \to \infty$. To prove this, we notice that all inequalities (7.5)–(7.8) remain valid for the considered sequence of graphs Γ'_s and for the corresponding sequence of extremal partitions. All notations related to these inequalities are the same as in part **(a)**. Therefore inequalities (7.10) also remain valid in the case under consideration. Taking into account (7.11), we conclude from Lemma 2 that $\operatorname{Mod}(R_s) \to \infty$ as $s \to \infty$. The latter together with (7.10) implies that $\mathcal{M}_s \to -\infty$ as $s \to \infty$ contradicting our assumption that $\mathcal{M}_s$ is a maximizing sequence.

Taking into account Lemma 3, one can easily see that similar argument works also in the case $c_{3,s} = c_{4,s}$.

(c) Since $\mathcal{M}_1$ and $\mathcal{M}_2$ are locally single-valued and continuous on $\widetilde{\mathbb{C}}^{l+n}$, our arguments in parts **(a)** and **(b)** show that $\mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2$ achieves its maximal value for some single-valued branch of $\mathcal{M}$ at some point $(c_1^*, \ldots, c_{l+n}^*) \in \widetilde{\mathbb{C}}^{l+n}$. Since $\mathcal{M}$ is invariant under Möbius maps, we may assume without loss of generality that $c_{l+n}^* = b_n^* = \infty$, $c_{l+n-1}^* = 0$, $c_{l+n-2}^* = 1$.

The point $(c_1^*, \ldots, c_{l+n}^*)$ is a critical point of $\mathcal{M}$. Hence all partial derivatives $\partial\mathcal{M}/\partial c_k$, $1 \le k \le l+n-3$, must vanish at $(c_1^*, \ldots, c_{l+n}^*)$. Using differentiation formulas (3.4), we find

$$\frac{\partial}{\partial b_k}\mathcal{M} = \pi\, \frac{d}{dz}\left((z-b_k)^2 Q_1(z)\right)\Big|_{z=b_k} + \pi\, \frac{d}{dz}\left((z-b_k)^2 Q_2(z)\right)\Big|_{z=b_k} \tag{7.12}$$

for all varying centers b_k and

$$\frac{\partial}{\partial a_k}\mathcal{M} = \pi \lim_{z\to a_k}((z-a_k)Q_1(z)) + \pi \lim_{z\to a_k}((z-a_k)Q_2(z)) \tag{7.13}$$

for all varying punctures a_k.

From (7.12) and (7.13) taking into account (3.1) and (3.2), we obtain

$$\frac{\partial}{\partial b_k}\mathcal{M}(C^*) = \pi\,(P_1'(b_k^*) + P_2'(b_k^*)) \prod\nolimits_{j=1}^{l}(b_k^* - a_j^*)^{-1} \prod\nolimits_{j=1}^{\prime\, n-1}(b_k^* - b_j^*)^{-2} = 0 \tag{7.14}$$

and

$$\frac{\partial}{\partial a_k}\mathcal{M}(C^*) = \pi\,(P_1(a_k^*) + P_2(a_k^*)) \prod\nolimits_{j=1}^{\prime\, l}(a_k^* - a_j^*)^{-1} \prod\nolimits_{j=1}^{n-1}(a_k^* - b_j^*)^{-2} = 0. \tag{7.15}$$

In (7.14) we assume, of course, that $n \geq 4$ and $1 \leq k \leq n-3$. In addition, if $n = 4$, then the second product in (7.14) is taken over empty set of terms. In this case, we assume that the value of this product is 1. Similarly, the first product in (7.15) is 1 if $l = 1$. In (7.15) we assume that $1 \leq k \leq k_0$, where $k_0 = \min\{l, l+n-3\}$.

Therefore if $n \geq 4$, then (7.14) gives the following $n-3$ necessary conditions for critical points of $\mathcal{M}$:

$$P_1'(b_k^*) = -P_2'(b_k^*), \qquad 1 \leq k \leq n-3. \tag{7.16}$$

If $l \geq 1$, then (7.15) in its turn gives k_0 necessary conditions:

$$P_1(a_k^*) = -P_2(a_k^*), \qquad 1 \leq k \leq k_0. \tag{7.17}$$

It follows from formulas (3.1) and (3.2) in Section 3 that each of the polynomials P_1 and $-P_2$ satisfies equations (1.3). For $n \geq 2$, this gives another $n-1$ conditions:

$$P_1(b_k^*) = -P_2(b_k^*), \quad \text{for } 1 \leq k \leq n-1. \tag{7.18}$$

Let $P(z) = P_1(z) + P_2(z)$. Since P_1 and $-P_2$ are polynomials of degree $l+2n-4$, each of which has the highest coefficient $-\alpha_n^2/4\pi^2$, the polynomial P has degree $\leq l+2n-5$. Since all points $a_1^*, \ldots, a_l^*, b_1^*, \ldots, b_n^*$ are distinct, equations (7.16)–(7.18) imply that the polynomial P has at least $l+2n-4$ zeros counting multiplicity. Therefore, P must vanish identically. Then, of course, $P_1(z) = -P_2(z)$ and hence $Q_1(z)\,dz^2 = -Q_2(z)\,dz^2$.

Summing up, if $\mathcal{M}$ achieves its maximum at $(c_1^*, \ldots, c_{l+n}^*) \in \overline{\mathbb{C}}^{l+n}$ with $c_{l+n}^* = a_1^* = \infty$, $c_{l+n-1}^* = 0$, and $c_{l+n-2}^* = 1$, then for $1 \leq k \leq l+n-1$ all coordinates c_k^* are finite and distinct. Let $Q_1(z)\,dz^2$ be the Jenkins-Strebel quadratic differential associated with Problem 1 for punctures $a_1^*, \ldots, a_l^*$, centers $b_1^*, \ldots, b_n^*$, and weights $\alpha_1, \ldots, \alpha_n$. Then the orthogonal differential $Q_2(z)\,dz^2 = -Q_1(z)\,dz^2$ is a Kuz'mina quadratic differential associated with Problem 2 such that the trajectory graph Γ'_{Q_2} of $Q_2(z)\,dz^2$ is homeomorphic to Γ' on $\overline{\mathbb{C}}$. In addition, the side γ_{kj}^i of Γ'_{Q_2} corresponding to the side l_{kj}^i of Γ' carries the same weight w_{kj}^i as l_{kj}^i does. This implies that the critical graph Γ_{Q_1}, which is dual of Γ'_{Q_2}, is homeomorphic to Γ on $\overline{\mathbb{C}}$ and the side ν_{kj}^i of Γ_{Q_1}, which is transversal to γ_{kj}^i, has the

Q_1-length w^i_{kj}. Therefore the quadratic differential $Q_1(z)\,dz^2$ satisfies all conditions required by Theorem 2. This establishes the existence of a Jenkins-Strebel quadratic differential with the desired properties.

8. Examples and remarks

In this section we consider some important particular cases of Theorem 2.

(a) *Trees and continua of the minimal logarithmic capacity.* Let $\Gamma = \{V, E, F, W\}$ be a free tree with l leaves and positive lengths of its edges. Then $n = 1$, F contains only one face $D = \overline{\mathbb{C}} \setminus (V \cup E)$, and we may assume that the length of ∂D is 1. By Theorem 2, there is a quadratic differential of the form

$$Q(z)\,dz^2 = -\frac{1}{4\pi^2}\frac{P(z)}{\prod_{k=1}^{l}(z-a_k)}\,dz^2,$$

where $P(z) = z^{l-2} + \cdots + c_0$ and $R(z) = \prod_{k=1}^{l}(z-a_k)$ are relatively prime, whose critical graph Γ_Q, complemented, if necessary, by second degree vertices on its edges, is homeomorphic to Γ and carries the same weights.

It is well known, see [14, Chapter 1], that the set $K_Q = V_Q \cup E_Q$ is extremal for Chebotarev's problem on continua of the minimal logarithmic capacity containing the points $a_1, \ldots, a_l$. Thus, Theorem 2 shows in this particular case that *every positive free tree can be realized uniquely up to a linear mapping as a continuum of minimal logarithmic capacity on* $\mathbb{C}$. A detailed treatment of Chebotarev's problem and related trees was given in a recent paper of P.M. Tamrazov [21].

(b) *Triangulation.* Assume that every face $f_k \in F$ is bounded by three distinct edges. Then Γ induces a finite triangulation on $\mathbb{S}^2$. Now, Theorem 2 says that every triangulation with prescribed lengths of sides of all triangles can be constructed as a conformal triangulation induced by some Jenkins-Strebel quadratic differential. In addition, such a conformal triangulation is unique if we fix the vertices of its initial triangle.

(c) *Cells with a fixed perimeter.* Let $\Gamma = \{V, E, F, W\}$ be a plane weighted graph such that $v_k \geq 3$ for all k, $1 \leq k \leq m$, and let every boundary cycle ∂f_k has length 1. Let

$$Q(z)\,dz^2 = -\frac{1}{4\pi^2}\frac{P(z)}{\prod_{k=1}^{n-1}(z-b_k)^2}\,dz^2 \tag{8.1}$$

be the quadratic differential defined by Theorem 2 for the graph Γ. Here $P(z) = z^{2n-1} + \cdots + c_0$ is a polynomial of degree $2n-4$ such that $P(b_k) = \prod'^{\,n-1}_{k=1}(b_k - b_j)^2$.

It is well known, see [14, Chapter 6], that the quadratic differential (8.1) is extremal for the problem of the maximal product $\prod_{k=1}^{n} R(D_k, b_k)$ of conformal radii of non-overlapping simply connected domains $D_1, \ldots, D_n$ such that $b_k \in D_k$ and $b_n = \infty$. Thus, Theorem 2 asserts in this case that every cellular structure, each cell of which has perimeter 1, can be realized as a domain configuration of a quadratic differential $Q(z)\,dz^2$ which is extremal for the problem of the maximal product of conformal radii.

(d) *Linear graphs.* Let Γ have a single vertex v_1 lying at ∞ and $n \ge 2$ loops. Then its dual Γ' is a linear graph with n edges and $n+1$ vertices. We may assume that Γ' has vertices $v'_1, \dots, v'_{n+1}$ such that $0 = v'_1 < v'_2 < \cdots < v'_{n+1} = 1$, edges $l_{k,k+1} = [v'_k, v'_{k+1}]$, and weights $w_{k,k+1} = v'_{k+1} - v'_k$, $1 \le k \le n$. Then, of course, the total weight of the vertex v_1 is 1. Theorem 2 implies that the linear graphs Γ' are in a one-to-one correspondence with quadratic differentials of the form

$$Q(z)\,dz^2 = -\frac{C^2}{4\pi^2}\frac{dz^2}{\prod_{k=1}^{n+1}(z-b_k)^2}, \tag{8.2}$$

where $0 = b_1 < b_2 < \cdots < b_{n+1} = 1$ and $C = w_{1,2}\prod_{j=2}^{n+1} b_j$. One can easily see that the quadratic differential (8.2) has a single zero of order $2n-2$ at ∞. An explicit correspondence between the poles $b_1, \dots, b_{n+1}$ of $Q(z)\,dz^2$ and the weights $w_{1,2}, \dots, w_{n,n+1}$ of the graph Γ is given by the following $n-1$ equations:

$$\begin{aligned} w_{k-1,k} + w_{k,k+1} &= w_{1,2}\prod_{j=2}^{n+1} b_j {\prod_{j=1}^{n+1}}' |b_k - b_j|^{-1}, \qquad 2 \le k \le n-1, \\ w_{n,n+1} &= w_{1,2}\prod_{j=2}^{n+1} b_j \prod_{j=1}^{n} (1-b_j)^{-1}. \end{aligned} \tag{8.3}$$

It follows from Theorem 2 that for every set of positive weights $w_{k,k+1}$, $1 \le k \le n$, such that $\sum_{k=1}^{n} w_{k,k+1} = 1$, the equations (8.3) have a unique solution $b_1, \dots, b_{n+1}$ satisfying the conditions $0 = b_1 < b_2 < \cdots < b_{n+1} = 1$. The latter also can be established by lengthy, routine calculation involving some work with Vandermonde determinants; cf. proof of Theorem 1.3 in [3].

(e) *Cyclic graphs.* If Γ has two vertices $v_1 = 0$ and $v_2 = \infty$ and $n \ge 2$ parallel edges from v_1 to v_2, then its dual Γ' is a cyclic graph with n vertices and n edges. We may assume that Γ' has vertices $b_k = e^{i\theta_k}$, $0 = \theta_1 < \cdots < \theta_n < 2\pi = \theta_{n+1}$ and edges $l_{k,k+1} = \{e^{i\theta} : \theta_k < \theta < \theta_{k+1}\}$, $1 \le k \le n$. In addition, we may assume that $l_{k,k+1}$ carries the weight $w_{k,k+1} = (\theta_{k+1} - \theta_k)/2\pi$. Then, the total weight of each of the vertices v_1 and v_2 is 1. Theorem 2 implies that the cyclic graphs are in a one-to-one correspondence with quadratic differentials of the form:

$$Q(z)\,dz^2 = -\frac{C^2}{4\pi^2}\frac{z^{n-2}}{\prod_{k=1}^{n}(z-e^{i\alpha_k})^2}\,dz^2, \tag{8.4}$$

where $0 = \alpha_1 < \alpha_2 < \cdots < \alpha_n < 2\pi = \alpha_{n+1}$ and $C = (w_{1,2} + w_{n,n+1})\prod_{k=2}^{n}(1 - e^{i\alpha_k})$. An explicit correspondence between weights $w_{k,k+1}$ and poles of $Q(z)\,dz^2$ is given by the system of equations:

$$w_{k-1,k} + w_{k,k+1} = Ce^{i\alpha_k(n-2)/2}{\prod_{j=1}^{n}}'(e^{i\alpha_k} - e^{i\alpha_j})^{-1}, \qquad 1 \le k \le n. \tag{8.5}$$

It follows from Theorem 2 that for every admissible choice of weights the system (8.5) has a unique solution $C \in \mathbb{C}$ and $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ such that $0 = \alpha_1 < \cdots < \alpha_n < 2\pi$.

In the case of equal weights $w_{1,2} = \cdots = w_{n,n+1} = 1/n$, the poles of quadratic differential (8.4) are equally spaced on the unit circle and therefore (8.4) becomes:

$$Q_n(z)\,dz^2 = -\frac{1}{\pi^2}\frac{z^{n-2}}{(z^n-1)^2}\,dz^2. \tag{8.6}$$

By the well-known theorem of Dubinin [4, Theorem 2.17], the quadratic differential $Q_n(z)\,dz^2$ is extremal for the problem of the maximal product $\prod_{k=1}^n R(D_k, b_k)$ of conformal radii of non-overlapping simply connected domains D_k, $1 \le k \le n$, whose centers $b_k \in D_k$ vary on the unit circle. The counterpart of this problem, suggested by G.V. Kuz'mina, for $2n$ non-overlapping domains, n of which have centers varying on the circle $C_r(0)$ and the remaining n have centers varying on $C_{1/r}(0)$, $0 < r < 1$, has remained open for a quite long time. It is conjectured that for every $n \ge 2$ and $0 < r < 1$, the extremal partition of this problem is given by the domain configuration of the quadratic differential

$$Q_{n,r}(z)\,dz^2 = \frac{z^{n-2}(z^n+p^n)(z^n-1/p^n)}{(z^n-r^n)^2(z^n+1/r^n)^2}\,dz^2$$

with some $0 < p < 1$ depending on n and r.

It is worth mentioning that symmetric cyclic graphs with an even number of edges are in a one-to-one correspondence with linear graphs. Therefore, the corresponding quadratic differentials are in a one-to-one correspondence as well. Indeed, let $Q(z)\,dz^2$ be a quadratic differential of the form (8.4), for which the the set of poles is symmetric with respect to the real axis and contains poles at ± 1. Then, scaled Joukowski's mapping $z \mapsto (1/4)(z + 1 + z^{-1})$ transforms $Q(z)\,dz^2$ into a quadratic differential of the form (8.2).

We want to mention also that the Joukowski's mapping $z \mapsto (1/2)(e^{\pi i/2n}z + e^{-\pi i/2n}z^{-1})$ transforms the symmetric quadratic differential $Q_{2n}(z)\,dz^2$ defined by (8.6) into the quadratic differential

$$Q_T(z)\,dz^2 = -\frac{1}{4\pi^2}\frac{dz^2}{(z^2-1)^2T_n^2(z)},$$

where $T_n(z) = \cos(n \arccos(z))$ denotes the classical Chebyshev polynomial of degree n which deviates least from zero on $[-1, 1]$.

(f) *Platonic solids.* Of course, every regular pattern can be represented by the critical graph of some Jenkins-Strebel quadratic differential. The five Platonic polyhedra: tetrahedron, cube, octahedron, dodecahedron, and icosahedron, correspond, respectively, to the following *Platonic quadratic differentials*:

$$Q_t(z)\,dz^2 = -\frac{1}{4\pi^2}\frac{z(z^3+8)}{(z^3-1)^2}\,dz^2,$$

$$Q_c(z)\,dz^2 = -\frac{1}{4\pi^2}\frac{z^8+14z^4+1}{z^2(z^4-1)^2}\,dz^2,$$

$$Q_o(z)\,dz^2 = -\frac{1}{4\pi^2}\frac{(z^6+5\sqrt{2}z^3-1)^2}{z^2(z^6-(7\sqrt{2}/4)z^3-1)^2}\,dz^2,$$

$$Q_d(z)\,dz^2 = -\frac{1}{4\pi^2}\,\frac{z^{20}-228z^{15}+494z^{10}+228z^5+1}{z^2(z^{10}+11z^5-1)^2}\,dz^2,$$

$$Q_i(z)\,dz^2 = -\frac{1}{4\pi^2}\,\frac{(z^{12}+11\sqrt{5}z^9-33z^6-11\sqrt{5}z^3+1)^3}{z^2(z^{18}-\frac{57\sqrt{5}}{8}z^{15}-\frac{57}{2}z^{12}-\frac{247\sqrt{5}}{4}z^9+\frac{57}{2}z^6-\frac{57\sqrt{5}}{8}z^3-1)^2}\,dz^2.$$

For each of the Platonic polyhedra, the above representation is unique up to a Möbius map. The zeros of the Platonic quadratic differential represent the vertices of the corresponding polyhedra while the poles represent its face centers. In each case, the representation above is given via a stereographic projection in such a way that the center of one of the faces is located at ∞.

The first three of these quadratic differentials are easy to compute. To get the explicit expression for the dodecahedral quadratic differential, we used Klein's invariants $V(z) = z(z^{10}+11z^5-1)$ and $F(z) = z^{20}-228z^{15}+494z^{10}+228z^5+1$ for the icosahedral group G_{60}, see [5, Section 5]. Then since the icosahedron and dodecahedron are dual polyhedra, the icosahedron can be represented by the quadratic differential $Q_i(\zeta)\,d\zeta^2 = C(V^3(\zeta)/F^2(\zeta))\,d\zeta^2$ for some constant $C < 0$. Now changing variables via the Möbius map $z = (\zeta - c)/(1 + c\zeta)$ with $c = -\sqrt[5]{5\sqrt{255+114\sqrt{5}}-57-25\sqrt{5}}$ being the smallest, in absolute value, negative zero of $F(z)$, we obtain the desired form $Q_i(z)\,dz^2$ after long but routine simplification.

It is important to emphasize that the critical graph of each of the Platonic quadratic differentials coincides precisely with the stereographic projection of the corresponding Platonic tessellation of the sphere and is not just a homeomorphic representation of it.

(g) *Disconnected graphs.* If Γ is a disconnected graph homeomorphic to the critical graph Γ_Q of some Jenkins-Strebel quadratic differential $Q(z)\,dz^2$, then each face f_k of Γ is necessarily a simply connected domain or doubly-connected domain on $\widetilde{\mathcal{R}}$. Thus, Theorems 1 and 2 can not be extended for arbitrary disconnected graphs, at least not directly. Even more, as simple examples show, an analytic embedding of a disconnected graph, if it exists, is not unique in its homotopy class, not even up to Möbius maps in the case of $\mathcal{R} = \overline{\mathbb{C}}$. So, any study of analytic embedding of disconnected graphs should address these issues.

Acknowledgements

I would like to thank the referee of the first version of this paper for his suggestion to use uniformization to prove the existence assertion of Theorem 1.

I also would like to express my sincere thanks to my colleagues at the Department of Mathematics and Statistics of Texas Tech University Professor Roger W. Barnard and Professor Kent Pearce for their constant help during my work on this and some other projects. In particular, their careful reading of my manuscripts makes them more readable and allows me to avoid many errors. The remaining errors are mine.

References

[1] L.V. Ahlfors, *Lectures on Quasiconformal Mappings.* Van Nostrand, New York, 1966.

[2] J. Akeroyd, Harmonic measures on complementary subregions of the disk, *Complex Variables Theory Appl.* **36** (1998), 183–187.

[3] J. Akeroyd, K. Karber, and A.Yu. Solynin, Minimal kernels, quadrature identities, and proportional harmonic measure on crescents, *Rocky Mountain J. Math.* to appear.

[4] V.N. Dubinin, Symmetrization in geometric theory of functions of a complex variable, *Uspekhi Mat. Nauk* **49** (1994), 3–76; English translation in: *R*ussian Math. Surveys **49** (1994), 1–79.

[5] W. Duke, Continued fractions and modular functions, *Bull. Amer. Math. Soc.* **42** (2005), 137–162.

[6] P. Duren, *Theory of H^p Spaces.* Academic Press, New York, 1970.

[7] E.G. Emel'yanov, Some properties of moduli of families of curves, *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov* (*LOMI*) **144** (1985), 72–82; English transl., *J. Soviet Math.* **38** (1987), no. 4, 2081–2090.

[8] E.G. Emel'yanov, On extremal partitioning problems, *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov* (*LOMI*) **154** (1986), 76–89; English transl., *J. Soviet Math.* **43** (1988), no. 4, 2558–2566.

[9] F.P. Gardiner, *Teichmüller Theory and Quadratic Differentials.* John Wiley & Sons, 1987.

[10] J.A. Jenkins, On the existence of certain general extremal metrics, *Ann. of Math.* **66** (1957), 440–453.

[11] J.A. Jenkins, *Univalent Functions and Conformal Mapping,* second edition, Springer-Verlag, New York, 1965.

[12] J.A. Jenkins, On the existence of certain general extremal metrics II, *Tohoku Math. J.* **45** (1993), 249–257.

[13] J.A. Jenkins, *The method of the extremal metric.* Handbook of Complex Analysis: Geometric Function Theory, Vol. **1**, 393–456, North-Holland, Amsterdam, 2002.

[14] G.V. Kuz'mina, *Moduli of Families of Curves and Quadratic Differentials.* Trudy Mat. Inst. Steklov., **139** (1980), 1–240; English transl., *Proc. V.A. Steklov Inst. Math.* 1982, no. 1.

[15] G.V. Kuz'mina, On extremal properties of quadratic differentials with strip-like domains in their trajectory structure, *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov* (*LOMI*) **154** (1986), 110–129; English transl., *J. Soviet Math,* **43** (1988), no. 4, 2579–2591.

[16] G.V. Kuz'mina, Methods of the geometric theory of functions. I, *Algebra i Analiz* **9** (1997), no. 3, 41–103; English transl., *St. Petersburg Math. J.* **9** (1998), no. 5, 889–930.

[17] G.V. Kuz'mina, Methods of the geometric theory of functions. II, *Algebra i Analiz* **9** (1997), no. 5, 1–50; English transl., *St. Petersburg Math. J.* **9** (1998), no. 3, 455–507.

[18] A.Yu. Solynin, The dependence on parameters of the modulus problem for families of several classes of curves, *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov* (*LOMI*) **144** (1985), 136–145; English transl.; *J. Soviet Math.* **38** (1988), 2131–2139.

[19] A.Yu. Solynin, Moduli and extremal metric problems, *Algebra i Analiz* **11** (1999), 3–86; English translation in: *St. Petersburg Math. J.* **11** (2000), 1–65.

[20] K. Strebel, *Quadratic Differentials.* Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], 5. Springer-Verlag, Berlin, 1984.

[21] P.M. Tamrazov, Tchebotarov's extremal problem, *Centr. Europ. J. Math.* **3(4)** (2005), 591–605.

[22] A. Vasil'ev, *Moduli of Families of Curves for Conformal nd Quasiconformal Mappings.* Lecture Notes in Mathematics **1788**, Springer, 2002.

Alexander Yu. Solynin
Texas Tech University, Box 41042
Lubbock, TX 79409, USA
e-mail: `alex.solynin@ttu.edu`

Analysis and Mathematical Physics
Trends in Mathematics, 507–514
© 2009 Birkhäuser Verlag Basel/Switzerland

Riesz Transforms and Rectifiability

Xavier Tolsa

Abstract. The n-dimensional Riesz transform of a measure μ in $\mathbb{R}^d$ is defined by the singular integral

$$\int \frac{x-y}{|x-y|^{n+1}}\, d\mu(y), \quad x \in \mathbb{R}^d.$$

Let $E \subset \mathbb{R}^d$ with $\mathcal{H}^n(E) < \infty$, where $\mathcal{H}^n$ stands for the n-dimensional Hausdorff measure. In this paper we survey some recent results and open problems about the relationship between the L^2 boundedness and existence of principal values for the Riesz transform of the measure $\mathcal{H}^n_{|E}$, and the rectifiability of E.

Mathematics Subject Classification (2000). 42B20, 28A75.

1. Introduction

Given $x \in \mathbb{R}^d$, $x \neq 0$, we consider the signed Riesz kernel $K^n(x) = x/|x|^{n+1}$, for n such that $0 < n \leq d$. Observe that K^n is a vectorial kernel. The n-dimensional Riesz transform of a finite Borel measure μ on $\mathbb{R}^d$ is defined by

$$R^n\mu(x) = \int K^n(x-y)\, d\mu(y), \qquad x \in \mathbb{R}^d \setminus \operatorname{supp}(\mu).$$

Notice that the integral above may fail to be absolutely convergent for $x \in \operatorname{supp}(\mu)$. For this reason one considers the ε-truncated n-dimensional Riesz transform, for $\varepsilon > 0$:

$$R^n_\varepsilon\mu(x) = \int_{|x-y|>\varepsilon} K^n(x-y)\, d\mu(y), \qquad x \in \mathbb{R}^d.$$

The principal values are denoted by

$$\text{p.v.}R^n\mu(x) = \lim_{\varepsilon\to 0} R^n_\varepsilon\mu(x),$$

whenever the limit exists.

Partially supported by grants MTM2007-62817 (Spain) and 2005-SGR-00774 (Generalitat de Catalunya).

Given $f \in L^1_{\text{loc}}(\mu)$, we also denote $R^n_\mu(f) = R^n(f\,d\mu)$ and $R^n_{\mu,\varepsilon}(f) = R^n_\varepsilon(f\,d\mu)$. Recall the definition of the maximal Riesz transform:

$$R^n_*\mu(x) = \sup_{\varepsilon>0} |R^n_\varepsilon \mu(x)|.$$

We say that the Riesz transform operator R^n_μ is bounded in $L^2(\mu)$ if the truncated operators $R^n_{\mu,\varepsilon}$ are bounded in $L^2(\mu)$ uniformly on $\varepsilon > 0$.

If in the preceding definitions one replaces the kernel $K^n(x)$ by the Cauchy kernel $1/z$, with $z \in \mathbb{C}$, and one considers a Borel measure μ in the complex plane, one gets the Cauchy transform:

$$\mathcal{C}\mu(z) = \int \frac{1}{z-\xi}\,d\mu(\xi), \qquad z \notin \operatorname{supp}(\mu).$$

We have analogous definitions for $\mathcal{C}_\varepsilon\mu$, $\mathcal{C}_\mu(f)$, $\mathcal{C}_{\mu,\varepsilon}(f)$, $\mathcal{C}_*\mu$, etc.

One says that a subset $E \subset \mathbb{R}^d$ is n-rectifiable if there exists a countable family of n-dimensional $\mathcal{C}^1$ submanifolds $\{M_i\}_{i\geq 1}$ such that

$$\mathcal{H}^n\Big(E \setminus \bigcup_i M_i\Big) = 0,$$

where $\mathcal{H}^n$ stands for the n-dimensional Hausdorff measure.

In this paper we are interested in the relationship between rectifiability and Riesz transforms, in particular in the existence of principal values and L^2 boundedness for Riesz transforms. This subject has been object of active research in the last years and there are still many difficult open questions dealing with this topic. In next sections we survey some recent results and open problems in this field. There is no attempt at completeness.

As usual, in the paper the letter 'C' stands for an absolute constant which may change its value at different occurrences. On the other hand, constants with subscripts, such as C_1, retain its value at different occurrences. The notation $A \lesssim B$ means that there is a positive absolute constant C such that $A \leq CB$. Also, $A \approx B$ is equivalent to $A \lesssim B \lesssim A$.

2. Principal values for Riesz transforms and rectifiability

Before talking about principal values of Riesz transforms we recall a fundamental result of geometric measure theory.

Theorem 2.1. *Let $E \subset \mathbb{R}^d$ with $\mathcal{H}^n(E) < \infty$. Then, the density*

$$\Theta^n(x,E) := \lim_{r\to 0} \frac{\mathcal{H}^n(E \cap B(x,r))}{(2r)^n}$$

exists for $\mathcal{H}^n$-almost every $x \in E$ if and only if n is integer and E is n-rectifiable.

The difficult implication in this theorem is

$$\exists\, \Theta^n(x,E) \quad \mathcal{H}^n\text{-a.e. } x \in E \quad \Rightarrow n \text{ is integer and } E\ n\text{-rectifiable.}$$

The fact that n must be integer if the density exists is a result of Marstrand [Mar], and that E must be n-rectifiable is due to Preiss [Pre] (and to Besicovitch in the case $n = 1$, $d = 2$).

Let us remark that if E is n-rectifiable, then $\Theta^n(x, E) = 1$ for $\mathcal{H}^n$-a.e. $x \in E$. Moreover, previously to Preiss' theorem, Mattila [M1] proved that

$$\Theta^n(x, E) = 1 \text{ for } \mathcal{H}^n\text{-a.e. } x \in E \quad \Rightarrow E \text{ is } n\text{-rectifiable.}$$

Concerning principal values for the Riesz transforms we have the following.

Theorem 2.2. *Suppose that n is an integer such that $0 < n \leq d$. Let $E \subset \mathbb{R}^d$ with $\mathcal{H}^n(E) < \infty$. The principal value* p.v.$R^n(\mathcal{H}^n_{|E})(x)$ *exists for $\mathcal{H}^n$-almost every $x \in E$ if and only if E is n-rectifiable.*

Notice the analogies between Theorems 2.1 and 2.2.

The fact that rectifiability implies the existence of principal values was proved first by Mattila and Melnikov [MM] for the Cauchy transforms (with $n = 1$, $d = 2$), and their proof generalizes easily to n-dimensional Riesz transforms. That E must be n-rectifiable if the principal values p.v.$R^n(\mathcal{H}^n_{|E})(x)$ exist for $\mathcal{H}^n$-almost every $x \in E$ was recently proved by the author in [To7].

Under the additional assumption that

$$\liminf_{r\to 0} \frac{\mathcal{H}^n(B(x,r) \cap E)}{r^n} > 0 \qquad \mathcal{H}^n\text{-a.e. } x \in E, \tag{2.1}$$

Mattila and Preiss had previously proved [MPr] that if the principal value p.v.$R^n(\mathcal{H}^n_{|E})(x)$ exists $\mathcal{H}^n$-almost everywhere in E, then E is n-rectifiable. Getting rid of the hypothesis (2.1) was an open problem raised by authors in [MPr].

Let us also remark that in the particular case $n = 1$, Theorem 2.2 was previously proved in [To1] (and in [M3] under the assumption (2.1)) using the relationship between the Cauchy transform and curvature of measures (see Theorem 3.2 below for the details). In higher dimensions the curvature method does not work (see [Fa]) and new techniques were required.

It is not known if Theorem 2.2 holds if one replaces the assumption on the existence of principal values for the Riesz transforms by

$$R^n_*(\mathcal{H}^n_{|E})(x) < \infty \qquad \mathcal{H}^n\text{-a.e. } x \in E.$$

That this is the case for $n = 1$ was shown in [To1] using curvature. However, for $n > 1$ this is an open problem that looks very difficult (probably, as difficult as proving that the L^2 boundedness of Riesz transforms with respect to $\mathcal{H}^n_{|E}$ implies the n-rectifiability of E. See next section for more details).

Given a Borel measure μ on $\mathbb{R}^d$, its upper and lower n-dimensional densities are defined, respectively, by

$$\Theta^{n,*}(x,\mu) = \limsup_{r\to 0} \frac{\mu(B(x,r))}{r^n}, \qquad \Theta^n_*(x,\mu) = \liminf_{r\to 0} \frac{\mu(B(x,r))}{r^n}.$$

The "only if" part of Theorem 2.2 is a particular case of the following somewhat stronger result.

Theorem 2.3. *Let μ be a finite Borel measure on $\mathbb{R}^d$. Suppose that n is an integer such that $0 < n \leq d$, and let $E \subset \mathbb{R}^d$ be such that for all $x \in E$ we have*

$$0 < \Theta^{n,*}(x,\mu) < \infty \quad \text{and} \quad \exists\, \text{p.v.} R^n\mu(x).$$

Then E is n-rectifiable.

The arguments to prove Theorems 2.2 and 2.3 are very different from the ones in [MPr] and [M3], which are based on the use of tangent measures. A fundamental step in the proof of Theorem 2.3 consists in obtaining precise L^2 estimates of Riesz transforms on Lipschitz graphs. In a sense, these L^2 estimates play a role analogous to curvature of measures in [To1]. Loosely speaking, the second step of the proof consists of using these L^2 estimates to construct a Lipschitz graph containing a suitable piece of E, by arguments more or less similar to the ones in [Lé].

To describe in detail the L^2 estimates mentioned above we need to introduce additional terminology. We denote the projection

$$(x_1,\ldots,x_n,\ldots,x_d) \mapsto (x_1,\ldots,x_n,0,\ldots,0)$$

by Π, and we set $\Pi^\perp = I - \Pi$. We also denote

$$R^{n,\perp}\mu(x) = \Pi^\perp(R^n\mu(x)) \quad \text{and} \quad R^{n,\perp}_\varepsilon\mu(x) = \Pi^\perp(R^n_\varepsilon\mu(x)).$$

That is to say, $R^{n,\perp}\mu(x)$ and $R^{n,\perp}_\varepsilon\mu(x)$ are made up of the components of $R^n\mu(x)$ and $R^n_\varepsilon\mu(x)$ orthogonal to $\mathbb{R}^n$, respectively (we are identifying $\mathbb{R}^n$ with $\mathbb{R}^n \times \{(0,\ldots,0)\}$).

Theorem 2.4. *Consider the n-dimensional Lipschitz graph $\Gamma := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^{d-n} : y = A(x)\}$, and let $\mu = \mathcal{H}^n_{|\Gamma}$. Suppose that A has compact support. If $\|\nabla A\|_\infty \leq \varepsilon_0$, with $0 < \varepsilon_0 \leq 1$ small enough, then*

$$\|\text{p.v.} R^{n,\perp}\mu\|_{L^2(\mu)} \approx \|\text{p.v.} R^n\mu\|_{L^2(\mu)} \approx \|\nabla A\|_2.$$

Let us remark that the existence of the principal values p.v.$R^n\mu$ μ-a.e. under the assumptions of the theorem is a well-known fact.

The upper estimate $\|\text{p.v.} R^n\mu\|_{L^2(\mu)} \lesssim \|\nabla A\|_2$ is an easy consequence of some results from [Do] and [To6] and also holds replacing ε_0 by any big constant. The lower estimate $\|\text{p.v.} R^{n,\perp}\mu\|_{L^2(\mu)} \gtrsim \|\nabla A\|_2$ is more difficult. To prove it one uses a Fourier type estimate as well as the quasiorthogonality techniques developed in [To6].

We remark that we do not know if the inequalities

$$\|\text{p.v.} R^n\mu\|_{L^2(\mu)} \geq C_3^{-1}\|\nabla A\|_2 \quad \text{or} \quad \|\text{p.v.} R^{n,\perp}\mu\|_{L^2(\mu)} \geq C_3^{-1}\|\nabla A\|_2$$

in Theorem 2.4 hold assuming $\|\nabla A\|_\infty \leq C_4$ instead of $\|\nabla A\|_\infty \leq \varepsilon_0$, with C_4 arbitrarily large and C_3 possibly depending on C_4.

3. L^2 boundedness of Riesz transforms and rectifiability

In this section we are interested in the following problem.

Question 3.1. *Consider $E \subset \mathbb{R}^d$ with $\mathcal{H}^n(E) < \infty$, with n integer. Suppose that the Riesz transform R^n_μ is bounded in $L^2(\mathcal{H}^n_{|E})$. Is then E n-rectifiable?*

The answer to this question is known (and it is positive in this case) only for $n = 1$. This is due to the relationship between the Cauchy transform and the notion of curvature of measures. Given a measure μ, its curvature is

$$c^2(\mu) = \iiint \frac{1}{R(x, yz)^2}\, d\mu(x) d\mu(y) d\mu(z), \tag{3.1}$$

where $R(x, y, z)$ stands for the radius of the circumference passing through x, y, z. If two among these points coincide, we let $R(x, y, z) = \infty$. Given $\varepsilon > 0$, $c^2_\varepsilon(\mu)$ stands for the ε-truncated version of $c^2(\mu)$, defined as in the right-hand side of (3.1), but with the triple integral over $\{(x, y, z) \in \mathbb{C}^3 : |x - y|, |y - z|, |x - z| > \varepsilon\}$.

The notion of curvature of a measure was introduced by Melnikov [Me] when he was studying a discrete version of analytic capacity, and it is one of the notions which is responsible of the big recent advances in connection with analytic capacity. Curvature is connected to the Cauchy transform by the following result, obtained by Melnikov and Verdera [MeV].

Theorem 3.2. *Let μ be a Borel measure on $\mathbb{C}$ such that $\mu(B(x, r)) \leq C_0 r$ for all $x \in \mathbb{C}$, $r > 0$. We have*

$$\|\mathcal{C}_\varepsilon \mu\|^2_{L^2(\mu)} = \frac{1}{6} c^2_\varepsilon(\mu) + O(\mu(\mathbb{C})), \tag{3.2}$$

where $|O(\mu(\mathbb{C}))| \leq C_1 \mu(\mathbb{C})$, with C_1 depending only on C_0.

Building on some techniques developed by Jones [Jo] and David and Semmes [DS1], Léger proved the following remarkable result (see also [To3] for another different and more recent proof):

Theorem 3.3. *Let $E \subset \mathbb{C}$ be compact with $\mathcal{H}^1(E) < \infty$. If $c^2(\mathcal{H}^1_{|E}) < \infty$, then E is rectifiable.*

Using Theorems 3.2 and 3.3, it follows then easily that if $\mathcal{H}^1(E) < \infty$ and the Cauchy transform is bounded in $L^2(\mathcal{H}^1_{|E})$, then E is 1-rectifiable. However, an identity analogous to (3.2) is missing in dimensions $n > 1$. This is the reason why Question 3.1 is still open for $n > 1$.

Recall that a measure μ such that $\mu(B(x, r)) \approx r^n$ for all $x \in \operatorname{supp}(\mu)$, $0 < r \leq \operatorname{diam}(\operatorname{supp}(\mu))$, is called n-Ahlfors-David (n-AD) regular, or abusing the language, AD regular. A set $E \subset \mathbb{C}$ is called n-AD regular (abusing the language, AD regular) if $\mathcal{H}^n_{|E}$ is AD regular.

A variant of Question 3.1 is the following:

Question 3.4. *Consider $E \subset \mathbb{R}^d$ n-AD regular, with n integer, and set $\mu = \mathcal{H}^n_{|E}$. If R^n_μ is bounded in $L^2(\mu)$, is then E uniformly n-rectifiable?*

We recall now the notion of uniform n-rectifiability (or simply, uniform rectifiability), introduced by David and Semmes in [DS2]. For $n = 1$, an AD regular 1-dimensional measure is uniformly rectifiable if its support is contained in an AD regular curve. For an arbitrary integer $n \geq 1$, the notion is more complicated. One of the many equivalent definitions (see Chapter I.1 of [DS2]) is the following: μ is uniformly rectifiable if there exist $\theta, M > 0$ so that, for each $x \in \text{supp}(\mu)$ and $R > 0$, there is a Lipschitz mapping g from the n-dimensional ball $B_n(0, R) \subset \mathbb{R}^n$ into $\mathbb{R}^d$ such that g has Lipschitz norm $\leq M$ and

$$\mu\big(B(x, R) \cap g(B_n(0, R))\big) \geq \theta R^n.$$

In the language of [DS2], this means that supp(μ) *has big pieces of Lipschitz images of* $\mathbb{R}^n$. A Borel set $E \subset \mathbb{R}^d$ is called uniformly rectifiable if $\mathcal{H}^n_{|E}$ is uniformly rectifiable.

For $n = 1$ the answer to Question 3.4 is true again, because of curvature. The result is from Mattila, Melnikov and Verdera [MMV]. For $n > 1$, in [DS1] and [DS2] some partial answers are given. Let H_n be class of all the operators T defined as follows:

$$Tf(x) = \int k(x - y) f(y)\, d\mu(x),$$

where k is some odd kernel (i.e., $k(-x) = -k(x)$) smooth away from the origin such that $|x|^{n+j} |\nabla^j k(x)| \in L^\infty(\mathbb{R}^d \setminus \{0\})$ for $j \geq 0$. Next result is from [DS1].

Theorem 3.5. *Let $E \subset \mathbb{R}^d$ be n-AD regular, with n integer. E is uniformly n-rectifiable if, and only if, all operators T from the class H_n are bounded in $L^2(\mathcal{H}^n_{|E})$.*

In Theorem 2.1 we mentioned that the existence of the density $\Theta^n(x, E)$ for $\mathcal{H}^n$-almost every $x \in E$ implies that n is integer. If we replace existence of density by L^2 boundedness of Riesz transforms the following holds:

Theorem 3.6. *Let $E \subset \mathbb{R}^d$ with $\mathcal{H}^n(E) < \infty$, and set $\mu = \mathcal{H}^n_{|E}$. Suppose that R^n_μ is bounded in $L^2(\mu)$. We have:*

(a) *$n \notin (0, 1)$.*
(b) *If E is n-AD regular, then n is integer.*

The statement (a) is from Prat [Pra]. It follows by the "curvature method", that is to say, by using a formula analogous to (3.2) which holds for $0 < n < 1$. The statement (b) was proved Vihtila using tangent measures. Her proof can be easily extended to the case where $\Theta^n_*(x, \mathcal{H}^n_{|E}) > 0$ H^n-a.e. in E, instead of the AD regularity assumption. However, it is an open question to prove that this also holds for arbitrary sets E with $\mathcal{H}^n(E) < \infty$. Under the stronger assumption of the existence of p.v.$R^n(\mathcal{H}^n_{|E})(x)$ $\mathcal{H}^n$-almost everywhere, this problem has been recently solved in [RT].

For more information and additional results regarding L^2 boundedness of Riesz transforms and rectifiability we suggest to have a look at [Vo], [MaT], [GPT], [JP], [To4], [To5], [To6], and [ENV], for instance.

References

[DS1] G. David and S. Semmes, *Singular integrals and rectifiable sets in R_n: Beyond Lipschitz graphs,* Astérisque No. 193 (1991).

[DS2] G. David and S. Semmes, *Analysis of and on uniformly rectifiable sets,* Mathematical Surveys and Monographs, 38, American Mathematical Society, Providence, RI, 1993.

[Do] J.R. Dorronsoro, *A characterization of potential spaces,* Proc. Amer. Math. Soc. 95 (1985), 21–31.

[ENV] V. Eiderman, F. Nazarov and A. Volberg, *Hausdorff content of sets with large values of vector Riesz potentials. Estimates from below,* Preprint (2007).

[Fa] H.M. Farag, *The Riesz kernels do not give rise to higher-dimensional analogues of the Menger-Melnikov curvature*, Publ. Mat. 43 (1999), no. 1, 251–260.

[GPT] J. Garnett, L. Prat and X. Tolsa, *Lipschitz harmonic capacity and bilipschitz images of Cantor sets,* Math. Res. Lett. 13 (2006), no. 6, 865–884.

[Jo] P.W. Jones, *Rectifiable sets and the travelling salesman problem,* Invent. Math. 102 (1990), 1–15.

[JP] P.W. Jones and A.G. Poltoratski, *Asymptotic growth of Cauchy transforms*, Ann. Acad. Sci. Fenn. Math. 29 (2004), no. 1, 99–120.

[Lé] J.C. Léger, *Menger curvature and rectifiability,* Ann. of Math. 149 (1999), 831–869.

[Mar] J.M. Marstrand, *The* (φ, s) *regular subsets of* n*-space,* Trans. Amer. Math. Soc. 113 1964 369–392.

[MaT] J. Mateu and X. Tolsa, *Riesz transforms and harmonic* Lip_1*-capacity in Cantor sets*, Proc. London Math. Soc. 89(3) (2004), 676–696.

[M1] P. Mattila, *Hausdorff* m *regular and rectifiable sets in* n*-space,* Trans. Amer. Math. Soc. 205 (1975), 263–274.

[M2] P. Mattila. *Geometry of sets and measures in Euclidean spaces,* Cambridge Stud. Adv. Math. 44, Cambridge Univ. Press, Cambridge, 1995.

[M3] *P. Mattila,* Cauchy Singular Integrals and Rectifiability of Measures in the Plane, Adv. Math. 115 (1995), 1–34.

[MM] P. Mattila, M.S. Melnikov, *Existence and weak type inequalities for Cauchy integrals of general measures on rectifiable curves and sets,* Proc. Amer. Math. **120**(1994), 143–149.

[MMV] P. Mattila, M.S. Melnikov and J. Verdera, *The Cauchy integral, analytic capacity, and uniform rectifiability,* Ann. of Math. (2) 144 (1996), 127–136.

[MPr] P. Mattila and D. Preiss, *Rectifiable measures in* $\mathbb{R}^n$ *and existence of principal values for singular integrals*, J. London Math. Soc. (2) 52 (1995), no. 3, 482–496.

[Me] M.S. Melnikov, *Analytic capacity: discrete approach and curvature of a measure,* Sbornik: Mathematics 186(6) (1995), 827–846.

[MeV] M.S. Melnikov and J. Verdera, *A geometric proof of the* L^2 *boundedness of the Cauchy integral on Lipschitz graphs,* Internat. Math. Res. Notices (1995), 325–331.

[Pra] L. Prat, *Potential theory of signed Riesz kernels: capacity and Hausdorff measure,* Int. Math. Res. Not. 2004, no. 19, 937–981.

[Pre] D. Preiss, *Geometry of measures in $\mathcal{R}^n$: distribution, rectifiability, and densities,* Ann. of Math. **125** (1987), 537–643.

[RT] A. Ruiz de Villa and X. Tolsa, *Non existence of principal values of signed Riesz transforms of non integer dimension*, preprint (2008), to appear in Indiana Univ. Math. J.

[To1] X. Tolsa, *Principal values for the Cauchy integral and rectifiability*, Proc. Amer. Math. Soc. 128(7) (2000), 2111–2119.

[To2] X. Tolsa, *Painlevé's problem and the semiadditivity of analytic capacity*, Acta Math. 190:1 (2003), 105–149.

[To3] X. Tolsa, *Finite curvature of arc length measure implies rectifiability: a new proof*, Indiana Univ. Math. J. 54 (2005), no. 4, 1075–1105.

[To4] X. Tolsa, *Growth estimates for Cauchy integrals of measures and rectifiability,* Geom. funct. anal. 17 (2007), 605–643.

[To5] X. Tolsa, *Analytic capacity, rectifiability, and the Cauchy integral.* International Congress of Mathematicians. Vol. II, 1505–1527, Eur. Math. Soc., Zürich, 2006.

[To6] X. Tolsa, *Uniform rectifiability, Calderón-Zygmund operators with odd kernel, and quasiorthogonality*, Proc. London Math. Soc. 98(2) (2009), 393–426.

[To7] X. Tolsa, *Principal values for Riesz transforms and rectifiability,* J. Funct. Anal., vol. 254(7) 2008, 1811–1863.

[Vi] M. Vihtila, *The boundedness of Riesz s-transforms of measures in $\mathbb{R}^n$,* Proc. Amer. Math. Soc. 124 (1996), no. 12, 3797–3804.

[Vo] A. Volberg, *Calderón-Zygmund capacities and operators on nonhomogeneous spaces.* CBMS Regional Conf. Ser. in Math. 100, Amer. Math. Soc., Providence, 2003.

Xavier Tolsa
Institució Catalana de Recerca i Estudis Avançats (ICREA)
and Departament de Matemàtiques
Universitat Autònoma de Barcelona
08193 Bellaterra (Barcelona), Catalunya
e-mail: xtolsa@mat.uab.cat