Applied and Numerical Harmonic Analysis

John R. Klauder

A Modern Approach to Functional Integration

Birkhäuser

Applied and Numerical Harmonic Analysis

John R. Klauder

A Modern Approach to Functional Integration

Birkhäuser

John R. Klauder
Department of Physics and
Department of Mathematics
University of Florida
Gainesville, FL 32611-8440
USA
john.klauder@gmail.com

ISBN 978-0-8176-4790-2 e-ISBN 978-0-8176-4791-9
DOI 10.1007/978-0-8176-4791-9
Springer New York Dordrecht Heidelberg London

Mathematics Subject Classification (2010): Primary: 28-XX, 28C20, 46Cxx, 46E22, 81Qxx, 81R30, 81S40, 81T25; Secondary: 32A70, 81-08, 81P16, 81T16

Cover design: Cover design by deblik Berlin based on a preliminary sketch by Aaron Spector.

Printed on acid-free paper

www.birkhauser-science.com

To my family

Preface

The formulation and evaluation—sometimes approximate to be sure—of functional integrals has become increasingly important in a variety of disciplines. The material presented in this text offers the interested reader a modern treatment of a number of topics in this important field, and several exercises at the end of most chapters may help clarify certain aspects. Our emphasis is principally on quantum theory, but similar techniques may find application elsewhere as well. A well-known example observes that the position and momentum in quantum mechanics have their natural analogs with time and frequency in signal analysis.

A few general thoughts may be helpful. Integrals involving N integration variables, when $N < \infty$, are familiar and well understood, even if they are often difficult to evaluate. The present textbook deals with integrations when $N = \infty$, i.e., for infinitely many variables. Such integrals have the usual problems of finite-dimensional integrals—such as conditional convergence, finitely additive measures, etc.—and a number of their own special problems. Integrals with an infinite number of variables are often called *functional integrals* or *path integrals*, and they occur in many areas of science, especially quantum mechanics, quantum field theory, and stochastic processes. We will include discussions on all these topics, and of course, the needed background will also be presented. Traditional discussions of these subjects by physicists, regrettably, are all too often rather cavalier and occasionally misleading—we shall try to point out some of the usual weak points in standard presentations, and even more importantly, we shall try to do better in our own discussions.

While there already exist a variety of other sources that offer good treatments of applications of functional integrals, our presentation instead favors conceptual and methodological developments, and as such there are some novel aspects to our discussion. Besides the standard material, it is noteworthy that we also present: (i) a different and more widely applicable approach for the analysis of quantum constraints, and (ii) a potentially new way to overcome conventional—but unacceptable—results for scalar nonrenormalizable relativistic quantum field theories.

A descriptive summary of the contents may help the reader gauge the overall scope and tenor of the material. After some general remarks (Chapter 1), an analysis of the probability of a single random variable and its three pure forms of realization (Chapter 2) form the background for a discussion of the probability aspects of random functions (Chapters 3 and 4), such as the most important and well-known Wiener process and the equally important but less well-known Poisson processes. These processes may be described via functional integrals or by suitable differential equations, each way offering different vantage points. Quantum phenomena are described by similar expressions with the appearance of $i = \sqrt{-1}$ in critical places. This makes for a formal similarity between probabilistic and quantum phenomena, but often masks a substantial mathematical gap that needs to be bridged to relate the two stories (Chapters 5 and 6). Special quantum states, the so-called coherent states, are especially helpful in bridging this mathematical gap (Chapter 7), and in a final version (Chapter 8) of physically motivated formulations, the probabilistic aspects of phase space paths are united with the quantum aspects of phase space paths to offer a natural formulation of the quantum theory that is both rigorous and truly invariant under classical canonical coordinate transformations leading thereby, in the author's view, for the first and only time, to a rigorously defined version of quantum mechanics that is strictly geometric in nature! This happy result has arisen by seeking further developments that are both physical and natural, and constitutes one of the main points of the book. The presence of constraints, which limit the classical phase space and often introduce unphysical (gauge dependent) variables, can be a challenge when any theory is quantized. Several commonplace techniques may lead to incorrect answers; instead we focus on a formulation, the so-called projection operator method (Chapter 9), that although possibly more difficult to apply, leads to acceptable results and does so with just the usual classical variables and nothing else. Infinitely many variables, as are appropriate to discuss field theories, adds several complications of a new kind, such as divergences that are typically handled by renormalized perturbation theory (Chapter 10). Yet some theories, the so-called nonrenromalizable theories, cannot be successfully studied by such conventional techniques. Finally (Chapter 11), we offer a novel way to deal with these insoluble models on the basis of insight gained by studying some soluble nonrenormalizable models within functional integral formulations. Here we see how such integrals help expose the source of unwanted divergences and support procedures that eliminate those divergences.

This book is intended for physicists, mathematical physicists, applied mathematicians, chemists, engineers, and others who seek alternatives to the study of partial differential equations as a means to examine the properties of certain systems. Part of the material in this book has been used for a one-semester graduate course on Functional Integration (including elements of all chapters except Chapter 11). Exercises at the end of most chapters help enforce the principal topics.

Acknowledgements

Over the years the author has profited from various forms of interaction with a number of people, and it is fitting to thank some of them on this occasion. They include: R. Alicki, E. Aslaksen, K. Baumann, B.G. Bodmann, W.R. Bomstad, I. Daubechies, G.G. Emch, H. Ezawa, H.-Y. Fan, H. Gausterer, J.-P. Gazeau, J. Govaerts, H. Grimmer, L. Hartmann, G.C. Hegerfeldt, C.J. Isham, A. Kempf, J.S. Lee, H. Leutwyler, J. Lewandowski, E. Lieb, J.S. Little, P. Maraner, J. McKenna, K. Muttalib, H. Narnhofer, E. Onofri, T. Paul, K.A. Penson, W.P. Petersen, A. Rubin, S.V. Shabanov, L. Shepp, B.-S. Skagerstam, J.J. Stankowicz, R.F. Streater, L. Streit, J.Y. Su, E.C.G. Sudarshan, W. Tomé, T.T. Truong, G. Watson, and C.-J. Zhu.

Thanks are also expressed to Jennifer Klauder for producing several of the figures for this book, and to Glenn Watson for numerous suggestions for text improvement.

Gainesville, Florida *John R. Klauder*
May 2010

Contents

1

Introduction

1.1 Overview and Latest Developments

The subject of functional integration benefits a number of diverse fields, and it has been studied off and on for many years. One of the most important topics is Brownian motion and the associated Wiener measure that describes such processes [Hid70]. Renewed interest in the subject of functional integration arose when Feynman [Fey48] introduced such tools for the description of propagators for time evolution in quantum mechanics. Functional integrals have taken on an immense importance in the functional formulation of quantum fields. It is our intention to discuss all of these topics to some extent.

Specifically, this monograph is broadly divided into three Parts: I. Stochastic Theory; II. Quantum Theory; and III. Quantum Field Theory. It is the author's conviction that a proper understanding of quantum mechanical functional integrals is helped considerably by a proper grounding in the mathematically more sound story associated with probability, stochastic variables, etc., and this conviction accounts for our emphasis of this material. As in all fields of endeavor, there are those who prefer to create new "tools" and there are others who choose to use existing "tools" to study various problems. No value judgment is placed on this division, and both aspects are important in the overall scheme of things. This rough division is no less true in theoretical studies, and it can be seen in existing texts on functional integration. For example, the text of H. Kleinert [Kle03] lays a great deal of emphasis on applications, while the monograph of G. Roepstorff [Roe96] puts greater emphasis on methodology. It is fair to say that the present text stresses the formulation and associated methodology surrounding functional integration, and consequently has less to say about applications.

This monograph began as a one-semester graduate course in 2004 entitled "Functional Integration"; the lecture notes as well as the recorded audio for this course are still available at http://www.phys.ufl.edu/~klauder. In preparing the present typescript for publication several important additions have been

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9_1,

introduced, including several entirely new topics, all of which help to make for a more complete and unified story.

By far and away the most important addition relates to new material developed during the past year that offers *a concrete proposal to obtain a nontrivial quantization of scalar relativistic nonrenormalizable quantum field theories*. At the time of this writing, this new theoretical proposal has not yet received supportive evidence from Monte Carlo calculations that would add significant weight to the validity of the new concepts; it is not that any "evidence" contradicts the proposal, it is rather that any evidence—one way or the other—is not yet in hand. However, we remain optimistic about the prospects for these new ideas. For the junior reader, therefore, we promise an exciting proposal at the end of their journey through this monograph; for the experienced reader, we invite them to jump right into Chapter 11, which to a large extent is fully self-explanatory.

Of course, nonrenormalizable scalar theories are not necessarily of physical interest by themselves. But there is always the hope that progress in the general nonrenormalizable arena may suggest new procedures to apply to quantum gravity, a nonrenormalizable quantum field theory that surely is a subject of importance and significant interest.

1.2 Topics Receiving Particular Emphasis

It is useful to mention that several topics receiving significant emphasis in this monograph are usually addressed either in a minimal fashion or perhaps not at all in other texts. One such topic relates to *coherent state path integrals*. The formulation of conventional phase space path integrals as coherent state path integrals has decided conceptual advantages over the more traditional treatment, and these features are stressed. Moreover, coherent state phase space path integrals—and only coherent state phase space path integrals—admit formulations that involve a *continuous time regularization*, a procedure that has considerable inherent advantages over standard approaches; these advantages will also be emphasized.

A number of classical systems involve constraints that need to be dealt with when those systems are quantized. Although there are several approaches already in the literature regarding the quantization of constraints, such methods have their limitations. Consequently, as was the case in the original lectures, the recently developed *projection operator method* to deal with all kinds of constraints and how they are to be introduced by way of functional integration formulations is treated as well.

Soluble problems have a fundamental role to play in any discussion of quantum theory. The role of the harmonic oscillator and its ability to capture much of the essence of quantum mechanics and even the methodology of quantization is well known. In much the same way, certain soluble, interacting, quantum field theory models can play a role in clarifying the quantization of

other field theories that have some similar features. In Chapter 10 we review the standard treatment of quartic, self-interacting, relativistic scalar quantum field theories, the so-called φ_n^4 models for spacetime dimension $n \geq 2$, including the super-renormalizable models ($n = 2, 3$), the strictly renormalizable model ($n = 4$), and the nonrenormalzable models ($n \geq 5$) for which no present-day technique can lead to a well-defined, nontrivial result for the quantum theory.

These latter, insoluble cases are reexamined in Chapter 11 in the light of two completely soluble, nonrelativistic, nonrenormalizable models, the so-called independent-value model and the somewhat related ultralocal model. Both of these latter models: (i) are nonrenormalizable in the sense that they require an infinite number of distinct, divergent counterterms in a perturbative analysis in a failed attempt to make them finite, or (ii) become trivial when treated nonperturbatively, allowing for only mass and coupling constant modifications; that is, they effectively become a free theory with a vanishing nonlinearity. It is noteworthy that these two extreme behaviors are also found in certain studies of the relativistic φ_n^4, $n \geq 5$, models [Aiz81, Fro82]. Thus, although the relativistic and the soluble nonrelativistic models differ in detail, they nevertheless have very similar "symptoms" of their "quantum illness," and so it is natural to conjecture that the same "simple surgical procedure" that provides a "complete cure" for the soluble, nonrelativistic, nonrenormalizable models can work the same miracle in the case of the relativistic models. This story is the subject of Chapter 11.

1.3 Basic Background

There are a few mathematical facts that are assumed in the monograph and which are collected here for reference. Many of these facts are no doubt familiar to the general reader, but it does no harm to summarize them in any case.

Most integrals that are encountered may be adequately defined in the manner of Riemann, but a wider class of functions can be integrated by the methods of Lebesgue. In the latter case, for example, one deals with measures, sets of measure zero, and equivalence classes of functions that differ only on sets of measure zero. A proper measure has the special property of countable additivity that gives it many of its important attributes. Integrals are properly defined with proper measures, and with such measures, integrals enjoy certain favorable properties.

However, sometimes certain expressions that appear to be integrals are only masquerading as integrals. That is, some expressions that appear in the form of integrals cannot legitimately be interpreted as such. These include expressions with Dirac delta functions, as well as other expressions that involve only finitely additive measures, such as an integral that is only conditionally convergent. These issues apply to series as well as integrals, and oftentimes the difficulty with such expressions is easier to see in a series. For example,

consider the series given by

$$\sum_{n=1}^{\infty}(-1)^{(n-1)}\, n^{-1}\ .$$

This series is composed of a set of positive numbers (all with odd denominators) that add to $+\infty$ and a set of negative numbers (all with even denominators) that separately add to $-\infty$. Thus any finite result of the given sum arises because of a delicate cancellation between plus and minus ∞. Indeed, if the order of the numbers in the series is changed one can get any answer one likes. For example, suppose that $\pi = 3.14159...$ is the sought-for answer. Simply add enough positive terms (in descending order) to go above π; next, add enough negative terms (in descending order) to go below π; and repeatedly alternate this procedure so that the limiting sum will yield π!

Thus the answer for the sum depends on the order of summation and not solely on the values of the terms that make up the series. The default order of summation is generally understood to mean

$$\sum_{n=1}^{\infty} a_n \equiv \lim_{N\to\infty} \sum_{n=1}^{N} a_n\ .$$

If $-\infty < \Sigma_{n=1}^{\infty} a_n < \infty$, then the original series converges. However, when $\Sigma_{n=1}^{\infty}|a_n| < \infty$, then the original series converges independently of the order of the terms in the series, and the series is said to be *absolutely convergent*; otherwise the convergent series is said to be only *conditionally convergent*.

So too for integrals. If $\int f(x)dx$ is finite, but $\int |f(x)|dx$ diverges, then a delicate cancellation of $+\infty$ and $-\infty$ has taken place. Change the order of integration and the answer will generally change. This can readily lead to situations for multiple integrals for which

$$\int dx \int dy f(x,y) \neq \int dy \int dx f(x,y)\ ,$$

since the order of "summation" has been changed. This can only occur whenever

$$\int dx \int dy\, |f(x,y)| = \int dy \int dx\, |f(x,y)| = \infty\ .$$

An example of this behavior is given by

$$\int_1^{\infty} dx \int_1^{\infty} dy \frac{(x-y)}{(x+y)^3} = -\frac{1}{2}\ ,$$

while

$$\int_1^{\infty} dy \int_1^{\infty} dx \frac{(x-y)}{(x+y)^3} = \frac{1}{2}\ .$$

It is natural that in life certain operations occur in a specified order. For example, you put on your socks before you put on your shoes; you open the

garage door before you drive your car out of the garage; etc. However, in mathematics, there are expressions that involve summing an infinite number of terms or integrating certain functions, etc., for which the order of that summation or integration has no physical significance. Under these circumstances, it is plausible that *physical phenomena are represented by mathematical expressions that involve absolutely convergent sums or integrals.* If that view is accepted, then the reader is ready to be bound by the following proposition:

HONOR CODE FOR INTEGRATORS: We pledge to recognize the following principle: *The representation of physical quantities by integrals or sums should involve absolutely convergent expressions.* If *conditionally convergent* integrals or sums are unavoidable, there *must* also be a physically acceptable and mathematically unambiguous definition of the proposed expression.

1.4 Elementary Integral Facts

Some properties of integrals are more basic than others, and a few of these are collected here.

a) Basics

$$\int f(x)\,dx \text{ exists provided } \int |f(x)|\,dx < \infty$$

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx$$

$$\iint f(x,y)\,dA = \int dx \int dy f(x,y) = \int dy \int dx f(x,y)$$

Dominated Convergence Theorem:

Let $\lim_{n\to\infty} f_n(x) = h(x)$ and $|f_n(x)| \le |g(x)|$ for almost all x. Then

$$\lim_{n\to\infty} \int f_n(x)\,dx = \int h(x)\,dx$$

provided $\int |f_n(x)|\,dx \le \int |g(x)|\,dx < \infty$ for some $g(x)$.

b) If $f(x) = g'(x)$, then $\int f(x)\,dx = g(x)\big|$.

Several common integrals appear repeatedly, and their properties should be stored in each reader's subconscious.

In what follows $\mathrm{Re}(a) > 0$.

1a) $\int_0^\infty e^{-ax}\,dx = -(1/a)\int_0^\infty d(e^{-ax}) = -(1/a)e^{-ax}|_0^\infty = 1/a$

1b) $\int_0^\infty x^n e^{-ax}\,dx = \left(-\frac{d}{da}\right)^n \int_0^\infty e^{-ax}\,dx = \left(-\frac{d}{da}\right)^n \frac{1}{a} = n!/a^{n+1}$

2a) Let $I = \int_{-\infty}^{\infty} e^{-ax^2/2}\,dx$, so $I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-a(x^2+y^2)/2}\,dx\,dy$
$= \int_0^{\infty}\int_0^{2\pi} e^{-ar^2/2}\,r\,dr\,d\theta = -2\pi(1/a)\int_0^{\infty} d(e^{-ar^2/2}) = 2\pi/a$, so

$$I = \int_{-\infty}^{\infty} e^{-ax^2/2}\,dx = \sqrt{2\pi/a}$$

2b) $\int_{-\infty}^{\infty} e^{bx - ax^2/2}\,dx = e^{b^2/2a}\int_{-\infty}^{\infty} e^{-(a/2)(x-b/a)^2}\,dx$
$= \sqrt{2\pi/a}\,e^{b^2/2a}$

2c) $\int_{-\infty}^{\infty} x^{2n} e^{-ax^2/2}\,dx = 2^n\big(-\frac{d}{da}\big)^n \int_{-\infty}^{\infty} e^{-ax^2/2}\,dx$
$= \sqrt{2\pi}\,1\cdot 3\cdot 5\cdots(2n-1)\,a^{-(2n+1)/2}$

Integrals are seldom evaluated by adding up the area under a curve; *integrals are generally evaluated by indirect means. Evaluate integrals any way you can!*

Commentary: The issues discussed above are familiar and routine for integrals involving finitely many integration variables. The same issues are no less important for functional integrals, i.e., integrals involving infinitely many integration variables. For functional integrals, however, there are also new issues that are subtle and more complicated, and the methods needed to handle them are less clear than for the usual, finite-dimensional integrals. We will often need to give a precise definition of a formal functional integration before further discussion can even take place!

****

There are no equation numbers in this monograph, and this is the author's preferred style. As a result, the discussion anywhere is intended to be largely self-contained without requiring frequent page flipping to understand some argument. If this approach requires the occasional repetition of some argument, or parts of some argument, so be it. Sometimes there is a reference to a section where supporting material is discussed, but this information is intended only as supplemental support and should not be necessary for a basic understanding of the argument.

Parts of Chapter 5 and Chapter 10 of this monograph are based on material from the author's previous book *Beyond Conventional Quantization* [Kla00].

****

Abbreviations: a.e. = almost everywhere; w.p.o. = with probability one; w.p.z. = with probability zero; a.k.a. = also known as

Part I

Stochastic Theory

2

Probability

2.1 Random Variables

Flip a coin, throw some dice, measure the voltage of a noisy circuit, and the outcomes are realizations of random variables. Let X denote the random variable in question, which we assume takes on strictly real values, and let J denote the set of values X can assume. The set J may be discrete (as for dice), continuous (as for noisy voltages), or the union of the two. Repeated and independent measurements yield a sequence of possible values x_n, where $n = 1, 2, ..., N$, $N < \infty$, that sample the possible outcomes of X in an unbiased way. Approximate averages involving the x_n values can be obtained easily, such as

$$\overline{X} = \frac{1}{N} \sum_{n=1}^{N} x_n \ ,$$

$$\overline{X^2} = \frac{1}{N} \sum_{n=1}^{N} x_n^2 \ ,$$

etc. Experience shows that the larger N becomes, the closer these values generally are to those values of the true averages for these quantities obtained ideally when $N \to \infty$.

We can use this data in yet another way, in particular to build a histogram, namely, an approximate frequency of distribution, which we may call $\sharp(n)$. For a single die, we can record the number of times each number comes up, i.e., $\sharp(n)$, $n \in \{1, 2, 3, ..., 6\}$, in a sample of N throws; see Fig. 2.1. With such information, we can compute the average $\overline{X}$ and $\overline{X^2}$ as

$$\overline{X} = \frac{\sum_{n=1}^{6} n \, \sharp(n)}{\sum_{n=1}^{6} \sharp(n)} \ ,$$

$$\overline{X^2} = \frac{\sum_{n=1}^{6} n^2 \, \sharp(n)}{\sum_{n=1}^{6} \sharp(n)} \ .$$

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9_2,

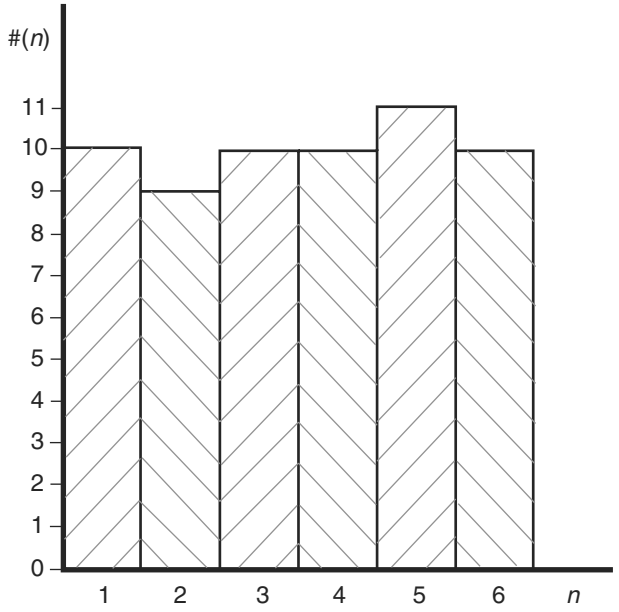

Fig. 2.1. Possible distribution of distinct die values 1-6 obtained in 60 separate and independent throws

Naturally, the larger N is, the closer the result for $\overline{X^m}$ to its idealized value is expected to be. For discrete outcomes, like a coin or a die, it is actually possible that a precise and unbiased histogram arises from a finite number of drawings. If, however, the distribution for X is continuous—at least partly—it is not possible to get an accurate histogram with any finite number of drawings.

This introductory discussion suggests that we choose another but closely related way to characterize the properties of a random variable.

2.1.1 Probability distributions

To cover all possibilities, we suppose that the values that X may assume lie between $-\infty$ and $+\infty$. We introduce [DMS]

$$P(X \leq x) \equiv \mu(x)$$

as the probability that the random variable X takes values between $-\infty$ and x (*including* x itself); the important function $\mu(x)$ is called a *probability measure*. Just from this definition, we have

$$\begin{aligned} &\text{a)} \qquad \mu(-\infty) = 0\,, \qquad \mu(\infty) = 1 \\ &\text{b)} \qquad \mu(x) \leq \mu(x+h)\,, \quad h > 0 \end{aligned}$$

$$\text{c)} \qquad \mu(x_-) \le \mu(x) = \mu(x_+) \,,$$
$$x_\pm = \lim_{\epsilon \to 0^+} x \pm \epsilon \,.$$

The functions $\mu(x)$ that satisfy these conditions are called *probability measures*. It turns out that there are *three* qualitatively different types of functions that obey these conditions. Rather than experimentally determine this distribution, as we discussed above, we now require that μ be given as part of the definition of X in the first place.

The names of the three forms of probability measures μ are (i) absolutely continuous (ac), (ii) discrete (d), and (iii) singular continuous (sc). For applications to the subject of this monograph, the first two forms are by far the most important; however, we briefly discuss the third variety for completeness. For clarity, we discuss examples that are purely of the given variety. More generally, a probability distribution can be a normalized convex linear combination of all three kinds.

Absolutely continuous: In this case

$$\mu_{ac}(x) = \int_{-\infty}^{x} \rho(y)\, dy \,,$$

where $\rho(y) \ge 0$. The function $\rho(y)$ is called the *probability density* and it is necessarily normalized so that

$$\mu_{ac}(\infty) = \int_{-\infty}^{\infty} \rho(y)\, dy = 1 \,.$$

In addition, the function $\mu_{ac}(x)$ is differentiable and $\mu_{ac}(x)' = \rho(y)$, a.e., which stands for 'almost everywhere', namely, up to a set of measure zero (e.g., a set consisting of finitely many points).

Discrete: The prototypical example of this kind of probability distribution is given by $\mu_d(x) = H(x)$, where $H(x)$ is the Heaviside function,

$$H(x) = 1 \,, \qquad x \ge 0 \,,$$
$$H(x) = 0 \,, \qquad x < 0 \,.$$

It follows that

$$\mu_d(x) = \int_{-\infty}^{x} dH(x) = H(x) \,.$$

In modern terms, this measure is often loosely stated as

$$\mu_d(x) = \int_{-\infty}^{x} \delta(x)\, dx \,, \qquad \textit{informal}! \,,$$

where $\delta(x)$ denotes the Dirac δ-function defined by the requirement that $\delta(x) = 0$ for $x \ne 0$, and $\int \delta(x) dx = 1$ so long as the range of integration

includes $x = 0$. However, this latter expression begs the question as to what is the behavior of $\mu_d(x)$ as $x \to 0^+$, $x \to 0^-$, and especially what is the value of $\mu_d(0)$. These delicate issues are all avoided by using the Heaviside function $H(x)$.

The most general discrete probability measure is of the form

$$\mu_d(x) = \sum_{n=1}^{\infty} p_n H(x - x_n) \ ,$$

where the weights $p_n \geq 0$, $\Sigma_{n=1}^{\infty} p_n = 1$, and x_n, for each n, is an arbitrary jump location.

Singular continuous: It is clear that a sequence of absolutely continuous distributions, $\mu_{ac\,n}(x)$, $n = 1, 2, ...$, can converge to a discrete distribution, $\mu_d(x)$. It is less clear, but nevertheless true, that a sequence of discrete distributions, $\mu_{d\,n}(x)$, $n = 1, 2, ...$, can converge to an absolutely continuous distribution, $\mu_{ac}(x)$. A common way to compute the third kind of distribution, the singular continuous distribution, $\mu_{sc}(x)$, is as a suitable limit of a sequence of discrete distributions, $\mu_{d\,n}(x)$. This sequence will be more transparent when we introduce *characteristic functions* in the next section; here, we can only state the canonical example of a singular continuous distribution without making its connection with the other distributions clear.

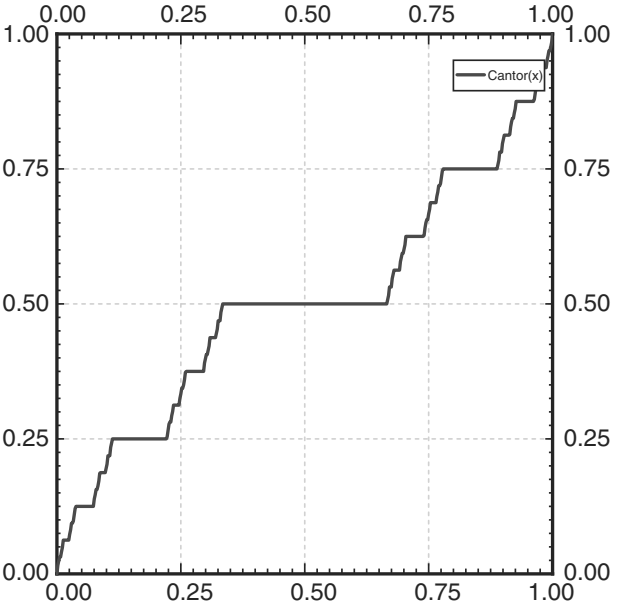

Fig. 2.2. Cantor function

The standard example we have in mind is based on the *Cantor function* $\mu_C(x) = \overline{C}(x)$ illustrated in Fig. 2.2 [Leb07]. The Cantor function is defined as follows: Let x, $0 \le x \le 1$, be given in a *tertiary* representation, i.e., $x = t_1/3 + t_2/3^2 + t_3/3^3 + \cdots$ and map such points to the function $\overline{C}(x)$, $0 \le x \le 1$, given in a *binary* representation, i.e., $\overline{C} = b_1/2 + b_2/2^2 + b_3/2^3 + \cdots$ according to the following rules: if $t_j = 0$, then $b_j = 0$; if $t_j = 2$, then $b_j = 1$; and finally if $t_j = 1$, then $b_j = 1$ *and all further* t_{j+p}, $p > 0$, *are mapped to* $b_{j+p} = 0$. For example, $x \equiv (t_1, t_2, t_3, \ldots)_3 = (2, 1, \ldots)_3$ is mapped to $\overline{C} \equiv (b_1, b_2, b_3, \ldots)_2 = (1, 1, \text{all } 0)_2$. Likewise, $(1, \ldots)_3 \to (1, \text{all } 0)_2$; $(0, 2, 2, 0, 1, \ldots)_3 \to (0, 1, 1, 0, 1, \text{all } 0)_2$, etc. The function $\overline{C}(x)$ has a derivative which is almost everywhere zero. Hence $\overline{C}(x) \ne \int_0^x \overline{C}'(y)\, dy = 0$! Nevertheless, $\overline{C}(x)$ is a continuous function rising from $\overline{C}(0) = 0$ to $\overline{C}(1) = 1$.

There are of course many more examples of singular continuous measures, but this one example serves to give the basic idea.

****

Note that the most general probability measure $\mu(x)$ is the combination

$$\mu(x) = A\mu_{ac}(x) + B\mu_d(x) + C\mu_{sc}(x) \ ,$$

where $A \ge 0$, $B \ge 0$, and $C \ge 0$, as well as $A + B + C = 1$.

2.2 Characteristic Functions

Let us define the function

$$C(t) = \int_{-\infty}^{\infty} e^{itx}\, d\mu(x) \ ,$$

for all real t, $-\infty < t < \infty$, as the *characteristic function* associated with each probability measure $\mu(x)$ [Luk70]. This function has several important properties,

a) $C(0) = 1$,

b) $|C(t)| \le 1$,

c) $\Sigma_{j,k=1}^{K,K} \alpha_j^* \alpha_k\, C(t_k - t_j) \ge 0$, $\quad \alpha_j \in \mathbb{C}$,

d) $C(t)$ is continuous.

The first three properties are evident, so we focus on the fourth property, d). Consider

$$|C(t) - C(s)| \le \int_{-\infty}^{\infty} |e^{itx} - e^{isx}|\, d\mu(x) = \int_{-\infty}^{-A} |e^{itx} - e^{isx}|\, d\mu(x)$$

$$+\int_{-A}^{B}|e^{itx}-e^{isx}|\,d\mu(x)+\int_{B}^{\infty}|e^{itx}-e^{isx}|\,d\mu(x)$$
$$\le 2\left(\int_{-\infty}^{-A}+\int_{B}^{\infty}\right)d\mu(x)+\int_{-A}^{B}|e^{i(t-s)x}-1|\,d\mu(x)\ .$$

The goal is to show that if $|C(t)-C(s)|<\epsilon$, $\epsilon>0$, then $|t-s|<\delta(\epsilon)$, $\delta>0$. Since $\int_{-\infty}^{\infty}d\mu(x)=1$, it follows that we can choose A and B so large that the first term in the last line is less than $\epsilon/2$. With A and B thus fixed and finite, we can clearly choose $|t-s|$ small enough to make the second term in the last line less than $\epsilon/2$. In particular, since

$$|e^{i(t-s)x}-1|=|\textstyle\int_0^{(t-s)x}e^{iy}\,dy|\le|t-s||x|\le|t-s|(|A|+|B|)\ .$$

Thus it suffices to choose $\delta=\epsilon/2(|A|+|B|)$. This concludes the proof that $C(t)$ is continuous.

We note in passing that if $C(-t)=C(t)$, then it follows that $C(t)$ is real.

Bochner's Theorem: It is especially noteworthy that the four properties satisfied by $C(t)$, which are listed above, *automatically* ensure that

$$C(t)=\int_{-\infty}^{\infty}e^{itx}\,d\mu(x)\ ,$$

where $\mu(x)$ is a probability measure [Luk70].

2.2.1 Convergence properties - 1

The Fourier transform of a measure leads to its characteristic function,

$$C(s)\equiv\int e^{isx}\,d\mu(x)\ .$$

Whatever the type of measure involved, the characteristic function is always continuous. However, the three types of measures can be distinguished in the following way. Besides the characteristic function $C(s)$, let us also introduce

$$C_T(s)\equiv T^{-1}\int_0^T C(s+t)\,dt\ ,\qquad\qquad T>0\ ,$$

which amounts to a partial averaging of the function $C(s)$. Then it follows [Luk70] that

(ac) For purely absolutely continuous measures,

$$\lim_{s\to\infty}C(s)=0\ ,\quad\text{and}\quad\lim_{s\to\infty}C_T(s)=0\ .$$

(d) For purely discrete measures,

$$\lim_{s\to\infty}C(s)\neq 0\ ,\quad\text{and}\quad\lim_{s\to\infty}C_T(s)\neq 0\ .$$

(sc) For purely singular continuous measures,

$$\lim_{s\to\infty}C(s)\neq 0\ ,\quad\text{while}\quad\lim_{s\to\infty}C_T(s)=0\ .$$

2.2.2 Convergence properties - 2

It is often extremely convenient to study probability distributions through a study of their corresponding characteristic functions. One aspect deals with the convergence of sequences. Suppose we deal with the sequence

$$C_n(t) = \int_{-\infty}^{\infty} e^{itx}\, d\mu_n(x) \;, \qquad\qquad n = 1, 2, 3, \ldots \;.$$

If

$$\lim_{n\to\infty} C_n(t) = C(t) \;,$$

and the limiting function $C(t)$ obeys all the required four properties—the first three are trivial, so continuity is the only real issue—then $C(t)$ is the characteristic function of *some* probability measure $\mu(x)$. In this case, one says that the sequence $\{\mu_n\}$ of measures converges "weakly" to the measure μ, a property that is denoted by the equation

$$\mathrm{w}-\lim_{n\to\infty} \mu_n(x) = \mu(x) \;.$$

For example, if

$$C_n(t) = e^{-t^2/4n} \;,$$

then

$$\lim_{n\to\infty} C_n(t) = 1 = C(t) \;.$$

This fact shows that the sequence

$$d\mu_n(x) = \sqrt{\pi n}\, e^{-nx^2}\, dx \;, \qquad n = 1, 2, 3, \ldots \;,$$

converges weakly as $n \to \infty$ to the Heaviside measure $\mu(x) = H(x)$, providing an example of a sequence of absolutely continuous measures that converges weakly to a discrete measure.

As a counterexample to weak convergence, consider the sequence

$$C_n(t) = e^{-nt^2}$$

for which

$$\lim_{n\to\infty} C_n(t) = D(t) \;,$$

where $D(0) = 1$, while $D(t) = 0$ if $t \neq 0$, which is *not* a characteristic function since it is not a continuous function.

2.2.3 Characteristic function for a Cantor-like measure

Consider the characteristic function given by

$$C_1(t) = \cos(\frac{t}{3}) \ ,$$

which consists of δ-function distributions located at $x = \pm 1/3$ each with a weight $1/2$. Likewise,

$$C_2(t) = \cos(\frac{t}{3}) \ \cos(\frac{t}{3^2}) = \frac{1}{2}\Big[\cos(\frac{t}{3} + \frac{t}{9}) + \cos(\frac{t}{3} - \frac{t}{9})\Big]$$

has four δ-functions at $x = \pm 1/3 \pm 1/9$ (independent signs), each with a weight $1/4$. A similar kind of discussion applies for

$$C_N(t) = \prod_{n=1}^{N} \cos(t/3^n) \ ,$$

namely, a sequence of δ-functions each with a weight $1/2^N$. Finally, we take the limit

$$C(t) = \lim_{N\to\infty} C_N(t) = \prod_{n=1}^{\infty} \cos(t/3^n) \ ;$$

here the result is infinitely many δ-functions, each with an equal weight: *zero!* This is the characteristic function appropriate to a Cantor-like function, i.e.,

$$\prod_{n=1}^{\infty} \cos(t/3^n) = \int e^{itx} \, d\mu_{Cantor-like}(x) = \int e^{itx} \, d\mu_{sc}(x) \ .$$

A word about infinite products is in order. One says that

$$\prod_{n=1}^{\infty} A_n$$

converges provided that

$$\sum_{n=1}^{\infty} |1 - A_n| < \infty \ .$$

If every $A_n \neq 0$ and yet $\Pi_{n=1}^{\infty} A_n = 0$, one says that the product "diverges to zero." In the case at hand, we want strict convergence, and

$$\sum_{n=1}^{\infty} |1 - \cos(t/3^n)| = \sum_{n=1}^{\infty} | \int_0^{t/3^n} \sin(u) \, du \ |$$
$$\leq \sum_{n=1}^{\infty} | \int_0^{t/3^n} du \, | \leq \sum_{n=1}^{\infty} \frac{|t|}{3^n} \ < \infty \ .$$

An interesting related example refers to

$$C_N(t) = \prod_{n=1}^{N} \cos(t/2^n) \ ,$$

which is also a set of δ-function distributions, all with equal weight 2^{-N}. However, in this case,

$$C(t) = \lim_{N\to\infty} C_N(t) = \prod_{n=1}^{\infty} \cos(t/2^n)$$
$$= \frac{\sin(t)}{t} = \frac{1}{2}\int_{-1}^{1} e^{itx}\, dx = \int e^{itx}\, d\mu_{ac}(x)\ .$$

In other words, the present case leads to an absolutely continuous distribution, while the former case led to a singular continuous distribution.

2.2.4 An application of the characteristic function

As an illustration of such a discussion, let us discuss a *general Gaussian random variable* X. To say that a random variable is Gaussian means that all moments are uniquely determined by the *mean*

$$\langle X\rangle = \int x\, d\mu(x)\ ,$$

and the *variance* $\langle X^2\rangle^c \equiv \langle X^2\rangle - \langle X\rangle^2$, where

$$\langle X^2\rangle = \int x^2\, d\mu(x)\ .$$

Explicitly, the characteristic function has the particular form given by

$$\langle e^{itX}\rangle = \int e^{itx}\, d\mu(x) = e^{it\langle X\rangle - \frac{1}{2}t^2\langle X^2\rangle^c}\ .$$

An expansion in powers of t determines the form of the moments in terms of the mean and the variance. In particular,

$$\langle X\rangle^c = \langle X\rangle\ , \qquad \langle X^2\rangle^c = \langle X^2\rangle - \langle X\rangle^2\ ,$$

etc.

A special case arises if the mean is zero, $\langle X\rangle = 0$, for which the variance is equal to the second moment, $\langle X^2\rangle^c = \langle X^2\rangle$. In that case,

$$\langle e^{itX}\rangle = e^{-\frac{1}{2}t^2\langle X^2\rangle}\ .$$

To distinguish this case, we shall sometimes refer to it as a *normal distribution*, although it is commonly the case in the literature that normal and Gaussian distributions generally refer to the same thing.

Let us next derive a useful property of normal variables X.

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2}\,\langle e^{itX}\rangle\,dt &= \langle e^{-\frac{1}{2}X^2}\rangle \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2}\,e^{-\frac{1}{2}t^2\langle X^2\rangle}\,dt \\ &= \frac{1}{\sqrt{1+\langle X^2\rangle}}\ .\end{aligned}$$

Now, suppose that $\langle X^2\rangle = \infty$. Then it follows that

$$\langle e^{-\frac{1}{2}X^2}\rangle = 0\ .$$

In turn, this equation implies that $X^2 = \infty$ for almost all X values. Instead if $\langle X^2\rangle < \infty$, then it follows—just from this fact—that $X^2 < \infty$ for almost all X values. In summary, if X is a normal variable, then

$$\begin{aligned}\langle X^2\rangle < \infty &\iff X^2 < \infty\ ,\text{a.e.}\ , \\ \langle X^2\rangle = \infty &\iff X^2 = \infty\ ,\text{a.e.}\ .\end{aligned}$$

This property will find good use later on.

2.3 Infinitely Divisible Distributions

Let us examine the product of two characteristic functions

$$C_1(t)C_2(t)\ .$$

This function clearly satisfies all the axioms to be a new characteristic function. In particular, the third axiom follows when we observe that

$$C_1(t)C_2(t) = \langle e^{itX_1}\rangle\langle e^{itX_2}\rangle = \langle e^{it(X_1+X_2)}\rangle\ ,$$

where the final average is over the so-called *product measure*

$$d\mu_1(x_1)\,d\mu_2(x_2) = \rho_1(x_1)\,\rho_2(x_2)\,dx_1\,dx_2\ ,$$

assumed absolutely continuous for clarity. The new random variable $X = X_1 + X_2$ is the sum of two *independent* random variables, and it has a distribution determined by the convolution integral:

$$\begin{aligned}\rho(x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(x-x_1-x_2)\,\rho_1(x_1)\,\rho_2(x_2)\,dx_1\,dx_2 \\ &= \int_{-\infty}^{\infty}\rho_1(x-y)\,\rho_2(y)\,dy = \int_{-\infty}^{\infty}\rho_2(x-y)\,\rho_1(y)\,dy\ .\end{aligned}$$

As a special case, the two characteristic functions, C_1 and C_2, could of course be the same. More generally, one could take any integral power of a

characteristic function to make a new one; that is, if $C(t)$ is a characteristic function, then so too is

$$C_{(p)}(t) \equiv C^p(t) \equiv [C(t)]^p \ ,$$

for every positive integer p. An immediate example of such a discussion involves a Gaussian distribution for which

$$C_{(p)}(t) = [\, e^{it\langle X\rangle - t^2\langle X^2\rangle^c/2} \,]^p = e^{ipt\langle X\rangle - pt^2\langle X^2\rangle^c/2}$$

yields a new Gaussian distribution for which

$$\begin{aligned} \langle X\rangle_{new} &= p\langle X\rangle_{old} \ , \\ \langle X^2\rangle_{new} &= p\langle X^2\rangle_{old} + (p^2-p)\langle X\rangle^2_{old} \ . \end{aligned}$$

2.3.1 Divisibility

Interestingly enough, the pth root—instead of the pth power—applied to a Gaussian characteristic function yields a valid (Gaussian) characteristic function again [Luk70], in particular,

$$C_{(1/p)}(t) = [\, e^{it\langle X\rangle - t^2\langle X^2\rangle/2} \,]^{1/p} = e^{it\langle X\rangle/p - t^2\langle X^2\rangle/2p}$$

also defines a (Gaussian) characteristic function for all $p \in \{1, 2, 3, \ldots\}$. Such a distribution is said to be *infinitely divisible*. As far as random variables go, it means that the original Gaussian random variable may be realized as

$$X = \sum_{k=1}^{p} Y_k \ ,$$

where $Y_1, Y_2, \ldots, Y_p$ denote p *independent, identically distributed* (here Gaussian) random variables for arbitrary positive integer p.

Not all random variables are infinitely divisible—in fact, not all random variables are divisible at all. If a random variable is divisible by two, then

$$X = Y_1 + Y_2 \ ,$$

where Y_1 and Y_2 are two independent and identically distributed random variables each with the same characteristic function, say, $D(t)$. To illustrate a random variable X that is *not* divisible, let X have just two equally weighted values, say, at $x = \pm 1$. Then it follows that the relevant characteristic function $C(t) = \cos(t)$. However, if $X = Y_1 + Y_2$, as suggested, then $C(t) = \cos(t) = D(t)^2$. Since $D(t)$, like $C(t)$, is even they both must be real. However, there is no real $D(t) = \sqrt{\cos(t)}$ for all t.

2.3.2 Infinite divisibility

Let us determine the characteristic function for random variables X that are infinitely divisible. We will see that we can find a fairly explicit formula for these expressions. On the other hand, relatively little is known about the forms of the probability distributions $\mu(x)$ that belong to the class of infinitely divisible distributions [Luk70].

Denote the characteristic function of interest, as usual, by

$$C(t) = \int e^{itx}\, d\mu(x)\ .$$

By definition, $C^{1/q}(t)$ is also a characteristic function for all $q \in \{1, 2, 3, \ldots\}$, which is the definition of infinite divisibility. Integral powers always lead to characteristic functions so $C^{p/q}(t)$ is also a characteristic function. Let the rational ratio p/q tend as a sequence to a nonnegative real number r, and by continuity we learn that $C^r(t)$ is still a characteristic function for all real $r \geq 0$. As such $C^r(t)$ must be nonzero since, for any $P > 1$, it follows that $[C^{r/P}(t)]^P = C^r(t)$ and $C^{r/P}(t) \to 1$ as $P \to \infty$. But if $C^r(t) \neq 0$ for any t, then so too is $C(t)$ itself. Consequently, we learn that $C(t)$ *cannot vanish*. If that is the case, it follows that

$$C(t) = e^{-L(t)} = \int e^{itx}\, d\mu(x) \equiv \langle\, e^{itX}\,\rangle\ ,$$

where L is continuous, $L(0) = 0$, and $\Re\, L(t) \geq 0$.

Next we consider

$$C^r(t) = e^{-r\,L(t)} = \int e^{itx}\, d\mu_r(x)\ ,$$

or stated otherwise,

$$r^{-1}[1 - e^{-r\,L(t)}] = \int [1 - e^{itx}]\, d(r^{-1}\mu_r(x))\ .$$

By assumption, the limit $r \to 0$ exists and leads to

$$\begin{aligned}
L(t) &= \lim_{r\to 0} r^{-1}[1 - e^{-rL(t)}] \\
&= \lim_{r\to 0} \int [1 - e^{itx}]\, d(r^{-1}\mu_r(x)) \\
&= \lim_{r\to 0} \int [1 - e^{itx} + itx/(1+x^2)]\, d(r^{-1}\mu_r(x)) \\
&\qquad -it \lim_{r\to 0} \int [x/(1+x^2)]\, d(r^{-1}\mu_r(x)) \\
&= -ita + \lim_{r\to 0} \int_{|x|\leq r} [1 - e^{itx} + itx/(1+x^2)]\, d(r^{-1}\mu_r(x)) \\
&\qquad + \lim_{r\to 0} \int_{|x|>r} [1 - e^{itx} + itx/(1+x^2)]\, d(r^{-1}\mu_r(x))\ .
\end{aligned}$$

Finally,

$$L(t) = -ita + bt^2 + \int_{|x|>0} [1 - e^{itx} + itx/(1+x^2)]\, d\sigma(x) \ ,$$

where $a \in \mathbb{R}$, $b \geq 0$, and $\sigma(x)$ is a nonnegative measure that satisfies

$$\int_{|x|>0} [x^2/(1+x^2)]\, d\sigma(x) < \infty \ ,$$

although it is quite possible that $\int_{|x|>0} d\sigma(x) = \infty$.

The final expression for $L(t)$ is the most general version that leads to a characteristic function for an infinitely divisible distribution. It follows that the general infinitely divisible random variable X is composed of two parts:

$$X \equiv X_G + X_P \ .$$

The linear and quadratic factors in t correspond to a Gaussian random variable (X_G); the remaining term corresponds to a Poisson random variable (X_P). More specifically, (i) if $\sigma(x)$ has a one-point support, e.g., $d\sigma(x) = k\delta(x-c)\, dx$, then X_P is a *Poisson variable*; (ii) if $\int d\sigma(x) < \infty$ with wider support, then X_P is a *compound Poisson variable*; and (iii) if $\int d\sigma(x) = \infty$, then X_P is a *generalized Poisson variable*.

For many problems, there is a symmetry such that $C(-t) = C(t)$, and thus $L(-t) = L(t)$. Under such circumstances

$$L(t) = bt^2 + \int [1 - \cos(tx)]\, d\sigma(x) \ .$$

If we assume that σ is absolutely continuous, then in this case $d\sigma = U(x)\, dx$, where $U(-x) = U(x) \geq 0$.

2.4 Central Limit Theorem—and Its Avoidance

Consider the random variable

$$X = \sum_{p=1}^{N} Y_p$$

made up of $N < \infty$ independent, identically distributed random variables, Y_p. Let Y denote any one of the identical random variables Y_p, $1 \leq p \leq N$, and for simplicity assume the variable Y is purely absolutely continuous, that all moments $\langle Y^n \rangle$, $n \in \mathbb{N} \equiv \{1, 2, 3, \ldots\}$, exist, and that all *odd* moments vanish:

$$\langle Y^{2l+1} \rangle = 0 \ , \qquad l = 0, 1, 2, \ldots \ .$$

This leads to what is called a symmetric distribution. Moreover, we also assume that the probability distribution for Y, $\mu(y)$, depends on the total number of variables N; thus $\mu(y) = \mu_N(y)$. Likewise the averaging process $\langle(\cdot)\rangle$ over Y depends on N, and making that explicit leads to $\langle(\cdot)\rangle = \langle(\cdot)\rangle_N$.

The characteristic function of X reads

$$\begin{aligned}\langle e^{itX}\rangle_N &= \langle e^{it\Sigma_{p=1}^N Y_p}\rangle_N = [\langle e^{itY}\rangle_N]^N \\ &= [1 - t^2\langle Y^2\rangle_N/2! + t^4\langle Y^4\rangle_N/4! - \cdots]^N \ .\end{aligned}$$

We wish to study the limit of this expression as $N \to \infty$. To obtain a meaningful answer, it is necessary, for large N, that

$$\langle Y^2\rangle_N \propto 1/N \ .$$

This result can be obtained in *two fundamentally different ways.*

Gaussian behavior: In the first way the probability density has the general shape illustrated in Fig. 2.3.

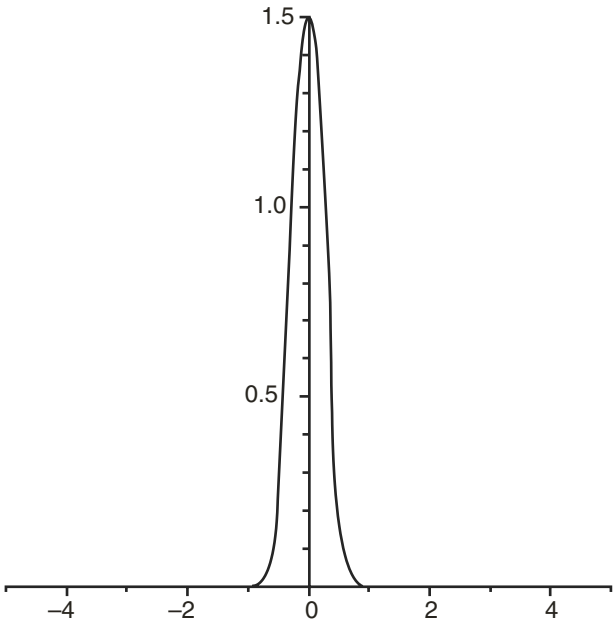

Fig. 2.3. Gaussian-like behavior

Here the principal property is that the width of the distribution should be "narrow," i.e., $\propto 1/\sqrt{N}$. Since there is unit area under the whole curve, it follows that the height is "tall," i.e., $\propto \sqrt{N}$. This shape implies that $\langle Y^2\rangle \propto 1/N$, but it also inevitably implies that $\langle Y^4\rangle_N \propto 1/N^2$, $\langle Y^6\rangle_N \propto 1/N^3$, etc.

As a consequence, the only contribution to $\langle e^{itX} \rangle_N$ that matters is the second moment, $\langle Y^2 \rangle_N$, and therefore

$$\langle e^{itX} \rangle \equiv \lim_{N\to\infty} \langle e^{itX} \rangle_N = e^{-t^2 A/2} ,$$

where

$$A \equiv \lim_{N\to\infty} N \langle Y^2 \rangle_N ,$$

which we assume exists. The resultant distribution, therefore, is a *normal* distribution; *all* features of the original distribution collapse into the single constant A! This behavior illustrates well the central property of the Central Limit Theorem.

Poisson behavior: The second way to achieve an acceptable behavior leads to a *non*-Gaussian result. In this second way the distribution $\rho_N(x) = \mu'_N(x)$ has the general shape illustrated in Fig. 2.4.

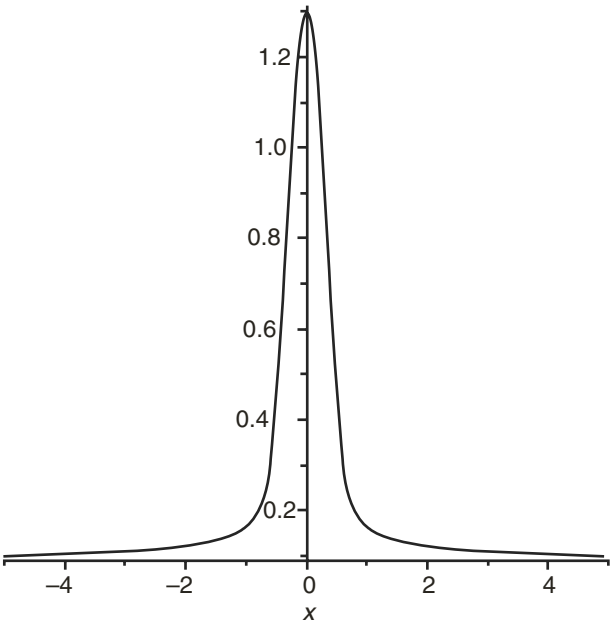

Fig. 2.4. Poisson-like behavior

The essence of this figure is the following: The central "spike" has a width $\propto 1/\sqrt{N}$, as before (or even narrower!), and this spike has an area of $1-O(1/N)$. The "broad shoulder" reaches out to Y values of order 1, i.e., independent of N. However, there is only an area $O(1/N)$ under this broad shoulder. In the

present case, it follows that $\langle Y^2\rangle_N \propto 1/N$, certainly from the shoulder and part possibly from the spike. However, it also follows that $\langle Y^4\rangle_N \propto 1/N$, $\langle Y^6\rangle_N \propto 1/N$, etc., based entirely on the shoulder. Thus *all moments are uniformly small* $\propto 1/N$—save for $\langle 1\rangle_N = 1$. As a consequence,

$$\begin{aligned}\langle e^{itX}\rangle &= \lim_{N\to\infty}[\,1 - t^2\langle Y^2\rangle_N/2! + t^4\langle Y^4\rangle_N/4! - \cdots]^N \\ &= \exp[\,-t^2\langle Y^2\rangle'/2! + t^4\langle Y^4\rangle'/4! - \cdots]\,,\end{aligned}$$

where, for all $m \geq 1$,

$$\langle Y^{2m}\rangle' \equiv \lim_{N\to\infty} N\langle Y^{2m}\rangle_N\,.$$

It is noteworthy that this expression may be written in the form

$$\langle e^{itX}\rangle = e^{\langle e^{itY}-1\rangle'} \equiv e^{-\int[1-\cos(ty)]\,d\sigma(y)}\,,$$

which we recognize as the very same (compound, generalized) Poisson distribution that arose in the study of infinite divisibility. Coupled with the Gaussian answer for the Central Limit Theorem, the result we have obtained is the full range of possibilities found in the study of (even) infinitely divisible distributions. In fact, this should not be a surprise since in both cases we were seeking those random variables that admit the decomposition

$$X = \sum_{p=1}^{\infty} Y_p$$

into infinitely many independent, identically distributed random variables Y_p.

Exercises

2-1 Prove that

$$\textstyle\prod_{n=1}^{\infty}\cos(t/2^n) = \dfrac{\sin(t)}{t}\,.$$

2-2 Expand the expression

$$\langle e^{itX}\rangle = \int e^{itx}\,d\mu(x) = e^{it\langle X\rangle - \frac{1}{2}t^2\langle X^2\rangle^c}$$

in powers of t to determine expressions for $\langle X^3\rangle$ and $\langle X^4\rangle$ in terms of the mean $\langle X\rangle$ and the variance $\langle X^2\rangle^c$.

2-3 When the mean vanishes, the Gaussian distribution becomes

$$\langle e^{itX} \rangle = \int e^{itx}\, d\mu(x) = e^{-\frac{1}{2}t^2 \langle X^2 \rangle} \ ,$$

and we have called X a normal variable. Find the general expression for $\langle X^p \rangle$ in terms of the variance, i.e., the second moment, $\langle X^2 \rangle$.

2-4 Find the characteristic function $C_{\overline{C}}(s)$ for the singularly continuous measure defined by the Cantor function. Show that

$$\lim_{s \to \infty} C_{\overline{C}}(s) \neq 0 \ ,$$

and show that the time-averaged expression

$$C_{\overline{C}\,T}(s) = \frac{1}{T} \int_0^T C_{\overline{C}}(s+t)\, dt \ , \qquad\qquad T > 0 \ ,$$

satisfies

$$\lim_{s \to \infty} C_{\overline{C}\,T}(s) = 0 \ .$$

2-5 Show that the function $F(t) = \exp(-Bt^4)$, with $B > 0$, cannot be the characteristic function of a probability distribution.

2-6 Show that the function $G(t) = \exp(-At^2 - Bt^4)$, with both $A > 0$ and $B > 0$, cannot be the characteristic function of a probability distribution.

2-7 For the characteristic function given, for $1 < \beta < 3$, by

$$C(t) = \exp\{-\textstyle\int [1 - \cos(tu)]\, du/|u|^\beta\} \equiv \displaystyle\int e^{itx}\, \rho_\beta(x)\, dx \ ,$$

determine the asymptotic functional form of the associated probability density function $\rho_\beta(x)$ as $x \to \infty$.

2-8 Let $\{C_n(t)\}$ denote a sequence of characteristic functions and $\{p_n\}$ a sequence of positive numbers. Build the functions

$$D_n(t) \equiv \exp\{p_n[C_n(t) - 1]\} \ .$$

Show that each of the functions $D_n(t)$ corresponds to a characteristic function of an *infinitely divisible* distribution. Show that *any* infinitely divisible distribution for a single random variable can be obtained as the limit of suitable sequences $\{C_n(t)\}$ and $\{p_n\}$. (This latter fact is known as De Finetti's Theorem [DMS].)

Table 2.1. Fertilizer amount (FA) vs. Growth rate (GR)

Sample	FA	GR
1	0.100	1.724
2	0.125	1.812
3	0.150	1.810
4	0.175	2.045
5	0.200	2.187
6	0.225	2.123
7	0.250	2.056
8	0.275	1.823
9	0.300	1.945
10	0.325	1.654

2-9 A farmer measures the growth rate (GR) in 10 different portions of her field as determined by 10 different amounts of fertilizer (FA). The results of the experiment are given in Table 2.1.

Find the average of GR for the experiment. Find the variance of GR, and the standard deviation (= square root of the variance) for the experiment.

If $1 \leq n \leq 10$ denotes the sample number, find the least squares fit of a linear fit to the data for GR given by

$$GR(n) = a + bFA(n) \ ,$$

for suitable constants a and b. Finally, find the least squares fit to a quadratic fit to the data for GR given by

$$GR(n) = c + dFA(n) + f\,[FA(n)]^2 \ ,$$

for suitable constants c, d, and f. On the basis of this information, make an educated guess for the optimal fertilizer amount to achieve the highest growth rate.

3

Infinite-Dimensional Integrals

3.1 Basics

Care must be taken when trying to extend commonplace formulas from many integration variables ($N < \infty$) to infinitely many integration variables ($N = \infty$). The case of independent normal distributions ($\equiv$ mean zero Gaussian) is a useful example from which to learn.

As an initial example, let us consider

$$\begin{aligned} d\mu_N(x) &= \prod_{n=1}^{N} \sqrt{\frac{b}{2\pi}}\, e^{-bx_n^2/2}\, dx_n \\ &= \left(\frac{b}{2\pi}\right)^{N/2} e^{-b\Sigma_{n=1}^{N} x_n^2/2} \prod_{n=1}^{N} dx_n \ , \end{aligned}$$

where $x = \{x_n\}_{n=1}^{N}$, $b > 0$, and $N < \infty$. By construction

$$\int d\mu_N(x) = 1 \ ,$$

where the integration is over the space $\mathbb{R}^N$. Clearly,

$$1 = \lim_{N\to\infty} \int d\mu_N(x) \ ,$$

and one wonders if, in some sense, there is a measure μ such that

$$d\mu(x) = \lim_{N\to\infty} d\mu_N(x) \ .$$

The answer is *yes*, there is a measure on $\mathbb{R}^\infty$ [Sko74], but it is *not* given by

$$d\mu(x) = \left(\frac{b}{2\pi}\right)^{\infty/2} e^{-b\Sigma_{n=1}^{\infty} x_n^2/2} \prod_{n=1}^{\infty} dx_n \qquad (wrong) \ .$$

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9_3,

For example, the first factor is generally ill-defined; if $b < 2\pi$, it is zero; if $b > 2\pi$, it is infinity. That issue can be eliminated if we choose $b = 2\pi$. Thus let us examine the special case

$$d\mu(x) = e^{-\pi \Sigma_{n=1}^{\infty} x_n^2} \prod_{n=1}^{\infty} dx_n \qquad (wrong) \ ,$$

an expression that is also still wrong. [**Remark:** We note that a "right" answer is given by $\Pi_{n=1}^{\infty} e^{-\pi x_n^2}\, dx_n$.]

In order for the "wrong" expressions above to have been "right," it would be necessary for the *prefactor* to be well defined, and thus the measure would have to be concentrated on sequences $\{x_n\}_{n=1}^{\infty}$ such that

$$\sum_{n=1}^{\infty} x_n^2 < \infty \ .$$

However, we can readily disprove this relation.

Since this issue does not depend on the value of b, we return to the original formula with arbitrary b. Let us evaluate (with $c > 0$)

$$\begin{aligned} \int e^{-c\Sigma_{n=1}^{\infty} x_n^2/2}\, d\mu(x) &= \lim_{N\to\infty} \left[\sqrt{\frac{b}{2\pi}} \int e^{-cx^2/2 - bx^2/2}\, dx \right]^N \\ &= \lim_{N\to\infty} \left(\frac{b}{b+c} \right)^{N/2} = \left(\frac{b}{b+c} \right)^{\infty} = 0 \ . \end{aligned}$$

We now seek to understand this result based on the assumption that the measure is concentrated on sequences such that $\Sigma_{n=1}^{\infty} x_n^2 < \infty$ [Sko74]. If this assumption were true, then by the dominated convergence theorem (see Sec. 1.4), it would have to follow that

$$0 = \lim_{c\to 0} \int e^{-c\Sigma_{n=1}^{\infty} x_n^2/2}\, d\mu(x) = \int \lim_{c\to 0} e^{-c\Sigma_{n=1}^{\infty} x_n^2/2}\, d\mu(x) = 1 \ ,$$

which is evidently false!.

We can adopt two points of view toward this result. We can insist that $\mu(x)$ is supported on l^2 sequences, i.e., sequences for which $\Sigma_{n=1}^{\infty} x_n^2 < \infty$. If we do so, then $\mu(x)$ is only a *finitely additive measure*. Or, we can reject the idea that μ is concentrated on l^2 and accept the notion that it is supported on a *larger space* (the details of which will be spelled out in the next section). If we choose this viewpoint, then the conclusion, for $c > 0$, namely, that

$$\int e^{-c\Sigma_{n=1}^{\infty} x_n^2}\, d\mu(x) = 0 \ ,$$

must be understood such that in the new support region $\Sigma x_n^2 = \infty$ holds almost everywhere. Stated otherwise, the space l^2 composed of sequences $\{x_n\}$ such that $\Sigma_{n=1}^{\infty} x_n^2 < \infty$ constitutes a *set of measure zero* in the full support space. At least if we accept this latter view, we can still hold out hope that $\mu(x)$ is a genuine (probability) measure on $\mathbb{R}^{\infty}$ (it is!).

Having accepted that, for any $b > 0$,

$$e^{-b\,\Sigma_{n=1}^{\infty} x_n^2/2} = 0 \ , \quad \text{a.e.} \ ,$$

we can raise the final question about

$$\mathcal{D}x \equiv \prod_{n=1}^{\infty} dx_n \ .$$

Does $\mathcal{D}x$ constitute anything like our familiar Lebesgue measure $\Pi_{n=1}^{N} dx_n$ when $N < \infty$? The answer turns out to be decidedly *no!*

When $N < \infty$, we have the relation

$$\begin{aligned} \int_0^1 \cdots \int_0^1 \prod_{n=1}^{N} dx_n &\equiv \overline{\mu}_N([0,1]^N) \\ &= \sum_{M=0}^{N} \binom{N}{M} \overline{\mu}_N([0,\tfrac{1}{2}]^{N-M}\,[\tfrac{1}{2},1]^M) = 1 \ , \end{aligned}$$

where we have summed over the 2^N sets that make up the partitioning of each interval $[0,1]$ into disjoint segments of $[0,\frac{1}{2}]$ and $[\frac{1}{2},1]$ for each of the N integration variables.

When $N = \infty$, we find that

$$1 = \overline{\mu}_\infty([0,1]^\infty) \neq \sum_{M=0}^{\infty} c_M\, \overline{\mu}_\infty([0,\tfrac{1}{2}]^\infty\,[\tfrac{1}{2},1]^M)$$

for any set of coefficients $\{c_M\}$. Thus, despite appearances, $\mathcal{D}x$ is *not* a measure because it does not satisfy complete additivity. Another argument that $\mathcal{D}x$ is not a measure is, for $0 < c < 1$, based on

$$1 = \overline{\mu}_\infty([0,1]^\infty) \neq \lim_{c\to 1} \overline{\mu}_\infty([0,c]^\infty) = 0 \ .$$

While we have focused primarily on the Gaussian example, the general conclusions are the same. Generally speaking, the usual expressions for a measure for $N < \infty$ do *not* apply when $N = \infty$.

3.2 Support Properties

Let us remain with our example of normal distributions where

$$d\mu_N(x) = \prod_{n=1}^{N} \sqrt{b/2\pi}\, e^{-bx_n^2/2}\, dx_n$$

and $d\mu(x) = \lim_{N\to\infty} d\mu_N(x)$. As a first determination of a relevant space of support, we note that

$$\int x_n^2 \, d\mu(x) = \int_{-\infty}^{\infty} y^2 \sqrt{b/2\pi}\, e^{-by^2/2} \, dy = b^{-1}$$

independently of n. Next introduce a sequence of constants $\{\lambda_n\}$ with the properties that

$$\lambda_n > 0 \,, \quad \text{for all } n \,,$$
$$\textstyle\sum_{n=1}^{\infty} \lambda_n < \infty \,.$$

A suitable set would be given by $\lambda_n = 1/n^2$, or $\lambda_n = e^{-n}$ (but *not* $\lambda_n = 1/n$). Consequently,

$$\int \left(\textstyle\sum_{n=1}^{\infty} \lambda_n x_n^2\right) d\mu(x) = \sum_{n=1}^{\infty} \lambda_n \int x_n^2 \, d\mu(x) = b^{-1} \textstyle\sum_{n=1}^{\infty} \lambda_n < \infty \,.$$

To get a finite answer, it is necessary that the $\{x_n\}_{n=1}^{\infty}$ sequences in the support of μ satisfy

$$\textstyle\sum_{n=1}^{\infty} \lambda_n \, x_n^2 < \infty \,, \qquad \text{a.e.} \,,$$

for *any* set $\{\lambda_n\}$ such that each $\lambda_n > 0$ and $\Sigma_{n=1}^{\infty} \lambda_n < \infty$. Thus, we conclude from the present and former discussions that

$$\mu(\Sigma_{n=1}^{\infty} x_n^2 < \infty) = 0 \,,$$
$$\mu(\Sigma_{n=1}^{\infty} \lambda_n \, x_n^2 < \infty) = 1$$

with $\{\lambda_n\}$ as discussed. Observe that, as presently determined, the space of support does *not* depend on the parameter b. However, this is not the end of the story.

3.3 Characteristic Functional

When $N < \infty$ the characteristic functional of our normal distribution takes the form

$$\int e^{i\Sigma_{n=1}^{N} t_n x_n} \, d\mu(x) = e^{-\Sigma_{n=1}^{N} t_n^2/2b} \,.$$

Note well that, unlike the support of the variables of integration $\{x_n\}$, we are free to choose the space of the sequences $\{t_n\}$ *as we like*. For example, to achieve suitable convergence as $N \to \infty$, we shall initially restrict our "test sequences" $\{t_n\}$ so that

$$\textstyle\sum_{n=1}^{\infty} t_n^2 < \infty \,.$$

When that is the case, we find that

$$e^{-\Sigma_{n=1}^{\infty} t_n^2/2b} = \int e^{i\Sigma_{n=1}^{\infty} t_n x_n}\, d\mu(x) \ ,$$

and in this sense, we may say that

$$\mathrm{w} - \lim_{N\to\infty} d\mu_N(x) = d\mu(x) \ ,$$

where "w" stands for "weak."

However, one last point of concern remains. While the characteristic functional

$$C(t) = e^{-\Sigma_{n=1}^{\infty} t_n^2/2b}$$

is well defined for $\Sigma_{n=1}^{\infty} t_n^2 < \infty$, it is not evident that the factor

$$e^{i\Sigma_{n=1}^{\infty} t_n x_n}$$

is well defined for the same set of $\{t_n\}$. In general, this is not the case, and to fix that we restrict the sequence $\{t_n\}$ even more. Note that

$$\begin{aligned} |\Sigma_{n=1}^{\infty} t_n x_n| &\le \Sigma_{n=1}^{\infty} |t_n||x_n| = \Sigma_{n=1}^{\infty} |t_n|^{1/2} |t_n|^{1/2} |x_n| \\ &\le (\Sigma_{n=1}^{\infty} |t_n|)^{1/2} (\Sigma_{n=1}^{\infty} |t_n||x_n|^2)^{1/2} \ . \end{aligned}$$

Thus, if we insist that $\{t_n\}$ satisfy

$$\sum_{n=1}^{\infty} |t_n| < \infty \ ,$$

then both factors above are finite, almost everywhere, and thus it follows that $\Sigma_{n=1}^{\infty} t_n x_n$ and

$$e^{i\Sigma_{n=1}^{\infty} t_n x_n}$$

are well defined, almost everywhere.

Hence, our conclusion now reads that

$$e^{-\Sigma_{n=1}^{\infty} t_n^2/2b} = \int e^{i\Sigma_{n=1}^{\infty} t_n x_n}\, d\mu(x)$$

represents a completely well-defined expression (on both sides at the same time) when we consider sequences $\{t_n\} \in l^1$, i.e., where

$$\Sigma_{n=1}^{\infty} |t_n| < \infty \ .$$

We emphasize that this space is not all the set of $\{t_n\}$ for which the characteristic function is well defined, but l^1 is dense in l^2 (in the topology of l^2), so suitable limits would readily complete the space for which $C(t)$ is well-defined. (An analog of denseness is that of the rational numbers in the real numbers.)

3.4 Tightest Support Conditions

We can even find significantly tighter conditions on the support of the measure $\mu_b(x) = \mu(x)$, now making explicit its dependence on the parameter b.

To begin with, consider a random variable

$$Y = Y(x_1, x_2, \ldots) = Y(x)$$

with the property, for all suitable $\{t_n\}$, that

$$\int Y(x)\, e^{i\Sigma_{n-1}^{\infty} t_n x_n}\, d\mu_b(x) = C \int e^{i\Sigma_{n-1}^{\infty} t_n x_n}\, d\mu_b(x) \ ,$$

where C denotes a constant, which depends on Y, but appears *outside* the integral and is thus independent of the set $\{t_n\}$. The only way this relation is possible is that although $Y(x)$ superficially appears to be a random variable, therefore apparently assuming different values, $Y(x)$ is in fact an *invariant*, a *constant* with respect to the collection of random variables. In short, despite appearances, $Y(x)$ is *not* random after all, with probability one.

As an interesting example of this behavior consider the expression

$$Y_P(x) = \frac{1}{P} \sum_{p=1}^{P} e^{i\Sigma_{n=1}^{\infty} s_n x_{n+p-1}} \ ,$$

which we observe is uniformly bounded, $|Y_P(x)| \leq 1$. Next, consider

$$\begin{aligned}
&\int Y_P(x)\, e^{i\Sigma_{n=1}^{\infty} t_n x_n}\, d\mu(x) \\
&\quad = \frac{1}{P} \sum_{p=1}^{P} \int e^{i\Sigma_{n=1}^{\infty} (t_n x_n + s_n x_{n+p-1})}\, d\mu(x) \\
&\quad = \frac{1}{P} \sum_{p=1}^{P} \exp\{-(1/2b)[\textstyle\sum_{n=1}^{p-1} t_n^2 + \sum_{n=1}^{\infty} (t_{n+p-1} + s_n)^2]\} \\
&\quad = \frac{1}{P} \sum_{p=1}^{P} \exp[-(1/2b) \textstyle\sum_{n=1}^{\infty} (t_n^2 + s_n^2 + 2 t_{n+p-1}\, s_n)] \ .
\end{aligned}$$

Observe that

$$|\Sigma_{n=1}^{\infty}\, t_{n+p-1}\, s_n| \leq (\Sigma_{l=p}^{\infty}\, t_l^2)^{1/2}\, (\Sigma_{n=1}^{\infty}\, s_n^2)^{1/2} \ ,$$

an expression that tends to vanish as $p \to \infty$ since $\Sigma_{l=1}^{\infty}\, t_l^2 < \infty$. Therefore,

$$\lim_{P \to \infty} \int Y_P(x)\, e^{i\Sigma_{n=1}^{\infty} t_n x_n}\, d\mu_b(x) = \exp[-\textstyle\sum_{n=1}^{\infty} (t_n^2 + s_n^2)/2b] \ .$$

Thanks to the dominated convergence theorem,

$$\lim_{P\to\infty}\int Y_P(x)\,e^{i\Sigma_{n=1}^{\infty}t_n x_n}\,d\mu_b(x) = \int Y(x)\,e^{i\Sigma_{n=1}^{\infty}t_n x_n}\,d\mu_b(x)$$
$$= C(\{s_n\})\,C(\{t_n\})\ .$$

Hence, we draw the important conclusion that

$$Y(x) = \lim_{P\to\infty}\frac{1}{P}\sum_{p=1}^{P} e^{i\Sigma_{n=1}^{\infty}s_n x_{n+p-1}} = e^{-\Sigma_{n=1}^{\infty}s_n^2/2b}$$

with probability one. This equation holds for all sequences $\{s_n\}\in l^1$, i.e., for which $\Sigma_{n=1}^{\infty}|s_n|<\infty$. *Note well that the answer depends on the parameter b!* This means that for the example under consideration, the support of $\mu_b(x)$ *depends* on b—and even more strongly—if b is changed, the support changes *entirely* so in fact, μ_b and $\mu_{b'}$, for $b\neq b'$, have *disjoint support*. One says that μ_b and $\mu_{b'}$ are *mutually singular*, which is denoted by $\mu_b\perp\mu_{b'}$! This is a much stronger support statement than we previously had, namely, that $\Sigma\lambda_n x_n^2<\infty$, for $\lambda_n>0$ and $\Sigma\lambda_n<\infty$.

As a special case of this support condition choose $\{s_n\}$ so that $s_1=s$, and $s_n=0$, $n\geq 2$. Thus

$$\lim_{P\to\infty}\frac{1}{P}\sum_{p=1}^{P} e^{isx_p} = e^{-s^2/2b}\ ,$$

which holds for all s. Consider the expansion of both sides of this equation in powers of s. The term quadratic in s implies that

$$\lim_{P\to\infty}\frac{1}{P}\sum_{p=1}^{P} x_p^2 = b^{-1}\ .$$

Hence, in a sense, the support of μ_b lies on the surface of a hypersphere of infinite (radius)^2 proportional to b^{-1}. Clearly, the nature of the support of infinitely many normal variables is dramatically different from the support of finitely many normal variables!

While we have concentrated on the example of normal distributions in this chapter, it is in fact true that our general conclusions hold much more broadly. In a great many cases, the support of the variables involved is strictly determined by the explicit form of the characteristic functional; change the characteristic functional, however slightly, and the new support properties are mutually singular with respect to those of the original support properties!

3.5 From Sequences to Functions

Working with sequences of random variables $\{x_n\}$ is perfectly acceptable, but there may be a more natural set of variables for a particular problem.

Just as in classical analysis, we can trade sequences for functions with the aid of one or another fixed basis set of real functions, say $\{h_n(x)\}_{n=1}^{\infty}$. It is particularly convenient to choose the $\{h_n(x)\}$ to be a complete orthonormal set of functions, meaning that

$$\begin{aligned} \int h_n(x)\, h_m(x)\, dx &= \delta_{nm} \ , \\ \textstyle\sum_{n=1}^{\infty} h_n(x)\, h_n(y) &= \delta(x-y) \ . \end{aligned}$$

Using such functions, we can introduce

$$\phi(x) \equiv \textstyle\sum_{n=1}^{\infty} x_n\, h_n(x)$$

and the inverse transform given by

$$x_m = \int h_m(x)\, \phi(x)\, dx \ .$$

A perfectly reasonable question to ask is what are the convergence properties involved in defining $\phi(x)$. Assume, first of all, that the $h_n(x)$ are smooth functions, such as the Hermite functions (eigenfunctions of a suitable quantum mechanical harmonic oscillator). These particular orthonormal functions can be summarized in a convenient generating function given by

$$\exp[-s^2 + 2sx - \tfrac{1}{2}x^2] = \pi^{1/4} \textstyle\sum_{n=0}^{\infty} (n!)^{-1/2}\, (s\sqrt{2})^n\, h_n(x) \ .$$

If one were dealing with a finite interval, then one could use sines and cosines. A *finite* sum would lead to a function $\phi(x)$ that had the common properties (e.g., C^{∞}) of the orthonormal set, or if $\Sigma |x_n|^2 < \infty$ then $\int |\phi(x)|^2\, dx < \infty$, but as we have seen, we cannot restrict attention to the space of square summable sequences.

To make progress it is next useful to introduce the idea of *test sequences*. We define a test sequence to be any sequence $\{t_n\}$ with the property that

$$\lim_{n\to\infty} n^r\, |t_n| = 0$$

or, alternatively,

$$\sum_{n=1}^{\infty} n^r\, |t_n|^2 < \infty$$

for all $r \geq 0$. In both cases, we say that $\{t_n\}$ falls to zero faster than any power (or any polynomial). An example of such a sequence is given by $t_n = e^{-\alpha n}$, for any $\alpha > 0$. Such "good" sequences lead to "good" functions, which we call *test functions*, given by

$$\begin{aligned} f(x) &= \sum_{n=1}^{\infty} t_n\, h_n(x) \ , \\ t_m &= \int h_m(x)\, f(x)\, dx \ . \end{aligned}$$

Let us introduce the concept of convergence of a sequence of such functions defined so that a convergent sequence does *not* leave the family of "good" sequences or, equivalently, does not leave the family of "good" functions. Let $\{t_n^j\}$ denote a sequence (indexed by j) of sequences (indexed by n). These lead to a sequence of test functions

$$f^j(x) \equiv \sum_{n=1}^{\infty} t_n^j \, h_n(x) \; .$$

We say that the sequence of coefficients $\{t_n^j\} \to \{t_n\}$, provided that

$$\sup_j \, n^r \, |t_n^j - t_n| = 0 \; ,$$

or

$$\lim_{j \to \infty} \textstyle\sum_{n=1}^{\infty} n^r \, |t_n^j - t_n| = 0 \; ,$$

for all $r \in \{0, 1, 2, 3, \ldots\}$. What role does "$\sup_j$" play? Consider the example where $t_n^j = e^{-n/j}$. As $j \to \infty$, each $t_n^j \to 1$, which is not a test sequence, and not what we want. The $\sup_j$ will prevent this sequence from being allowed.

With test sequences, or equivalently test functions, and a notion of convergence, we have completed the space of test functions to a topological linear vector space. Now, we introduce a dual space, which will be called the associated space of distributions (or, as they are often called, generalized functions). We define "bad" sequences, or generalized sequences, by sequences $\{x_n\}$ that are polynomially bounded, which means that there are constants A, B, and C such that

$$|x_n| \leq A + B n^C \; .$$

This choice is made so that

$$\sum_{n=1}^{\infty} t_n \, x_n$$

is always well defined for any "good" sequence $\{t_n\}$ and any "bad" sequence $\{x_n\}$. The corresponding generalized functions

$$\phi(x) = \sum_{n=1}^{\infty} x_n \, h_n(x)$$

will, in general, *not* converge pointwise, but this sequence converges in the sense of distributions, which means that

$$\sum_{n=1}^{\infty} x_n \int h_n(x) \, f(x) \, dx = \sum_{n=1}^{\infty} x_n \, t_n = \int \phi(x) \, f(x) \, dx \; .$$

The last integral extends the usual notation and meaning of an integral, and it is better to interpret

$$\sum_{n=1}^{\infty} x_n t_n = \phi(f)$$

as a *linear functional* mapping smooth functions into real numbers.

What we have outlined above is the concept of a *nuclear space* $\mathcal{N}$ composed of test sequences (or test functions) and its topological dual $\mathcal{N}'$, the space of associated distributions as sequences (or generalized functions) [GV64]. There are many examples of nuclear spaces, but for our purposes, this simple example illustrates the general idea.

Let's examine our fundamental example in light of this discussion. Consider the function

$$C(\{t\}) = e^{-\Sigma_{n=1}^{\infty} t_n^2/2b} ,$$

which is clearly well defined for all $\{t_n\}$ such that $\Sigma t_n^2 < \infty$. It is a continuous functional in the sense that

$$\lim_{j\to\infty} C(\{t_n^j\}) = C(\lim_{j\to\infty} \{t_n^j\})$$

for suitable sequences. Evidently, if $\Sigma_{n=1}^{\infty}(t_n^j - t_n)^2 \to 0$ as $j \to \infty$, then our function is continuous. This function is also continuous if

$$\lim_{j\to\infty} \sum_{n=1}^{\infty} |t_n^j - t_n| = 0 ,$$

but it would not in general be continuous if

$$\lim_{j\to\infty} \sum_{n=1}^{\infty} |t_n^j - t_n|^3 = 0 ;$$

the reader should become convinced of these facts.

Indeed, the function $C(\{t_n\})$ is also continuous if we insist that

$$\lim_{j\to\infty} \sum_{n=1}^{\infty} n^r (t_n^j - t_n)^2 = 0$$

for all $r = 0, 1, 2, \ldots$, and this is the example we wish to stress. This is clear because we can always limit $C(\{t_n\})$ to test sequences.

3.6 Bochner–Minlos Theorem

Suppose we have a functional $C(\{t_n\})$ depending on sequences $\{t_n\}$ that belong to a nuclear space (such as our $\mathcal{N}$), and which satisfies the following Bochner-Theorem-like (see Sec. 2.2) conditions [GV64]:

$$
\begin{aligned}
&C(\{0\}) = 1 \,, \\
&|C(\{t_n\})| \le 1 \,, \\
&\sum_{p,q=1}^{N,N} \alpha_p^* \, \alpha_q \, C(\{t_{n(p)} - t_{n(q)}\}) \ge 0 \,, \qquad N < \infty \,, \\
&C(\{t_n\}) \text{ is continuous (e.g., on } \mathcal{N}) \,.
\end{aligned}
$$

If these conditions are met, then $C(\{t_n\})$ is a *characteristic functional*,

$$
C(\{t_n\}) = \int e^{i \Sigma_{n=1}^{\infty} t_n x_n} \, d\mu(x) \,,
$$

where μ is a probability distribution with support on sequences $\{x_n\}$ that belong to $\mathcal{N}'$, the dual space of the nuclear space, $\mathcal{N}$.

With reference to our basic example, it is indeed correct that the support of μ is on the space of polynomially bounded sequences since (by choosing $\lambda_n = 1/n^2$) we know that $\Sigma_{n=1}^{\infty} x_n^2/n^2 < \infty$ defines its support. To say that the support is on $\mathcal{N}'$ may be very generous in general, since the real support may be on a much smaller space, but it is not wrong.

There is also a field way (rather than a sequence way) to express the Bochner–Minlos Theorem [GV64]. In particular, if $C\{f\}$—a common notation for a functional—is defined for all $f \in \mathcal{N}$, where $\mathcal{N}$ is some nuclear space, and satisfies the conditions

$$
\begin{aligned}
&C\{0\} = 1 \,, \\
&|C\{f\}| \le 1 \,, \\
&\sum_{p,q=1}^{N,N} \alpha_p^* \, \alpha_q \, C\{f_p - f_q\} \ge 0 \,, \qquad N < \infty \,, \\
&C\{f\} \text{ is continuous for all } f \in \mathcal{N} \,,
\end{aligned}
$$

then $C\{f\}$ is a characteristic functional

$$
\begin{aligned}
C\{f\} &= \int e^{i \int f(x) \, \phi(x) \, dx} \, d\mu(\phi) \\
&= \int e^{i \phi(f)} \, d\mu(\phi) \,.
\end{aligned}
$$

Here the support of μ is limited to generalized functions in the topological dual $\mathcal{N}'$ to the given nuclear space $\mathcal{N}$. This theorem then assures the existence of a wide family of probability measures for function spaces.

The remainder of this chapter is devoted to various ways of dealing with functionals that may prove important later.

3.7 Functional Derivatives

Functional derivatives are familiar in their use in deriving the Euler–Lagrange equations of motion in classical or field mechanics. There is a careful—and a casual—way to define functional derivatives, and it is important that both sides of the issue are presented.

First of all, let us be clear about what a functional is. While a function is normally a map from $\mathbb{R}^N$ to $\mathbb{R}$ (or $\mathbb{C}$), when $N < \infty$, say, a functional has the same purpose when $N = \infty$. In other words, a functional maps functions into real or complex numbers. Invariably, there is a particular space $\mathcal{S}$ of functions, for which the functional is well defined. For example, $\int f(x)^2\,dx$ is well defined for $f \in L^2(\mathbb{R})$; $\int [f'(x)^2 + f(x)^2]\,dx$ requires that both f' and f are L^2; $\int x^4 f(x)^4\,dx$ is well defined for the space of functions for which it is implicitly defined. For such examples, it is clear that $\mathcal{S}$ is a linear vector space; that is, if f_1 and f_2 belong to $\mathcal{S}$, then $\alpha f_1 + \beta f_2 \in \mathcal{S}$. However, this does not have to be the case. For example, suppose we require the finiteness of $\int [f(x)^2 + (1 - e^{-f(x)})^2]\,dx$. Thus $f(x) = |x|^{-1/4} e^{-x^2}$ is acceptable, but $f(x) = -|x|^{-1/4} e^{-x^2}$ fails to be acceptable. At any rate, we assume that along with the given functional $F\{f\}$, we are also given the space of functions $\mathcal{S}$ for which $F\{f\}$ is well-defined.

Let functions of the form $f(x) + \tau h(x)$ for all real τ such that $|\tau| < \epsilon$, for some $\epsilon > 0$, be elements of $\mathcal{S}$. Then,

$$F\{f + \tau h\}$$

is a well-defined function of τ in an open interval about $\tau = 0$. We are interested in the derivative of this function with regard to τ, followed by an evaluation at $\tau = 0$. Namely,

$$\left.\frac{dF\{f + \tau h\}}{d\tau}\right|_{\tau=0} \equiv \int \frac{\delta F\{f\}}{\delta f(x)}\, h(x)\,dx\ .$$

The answer is linear in $h(x)$, and the local coefficient is $\delta F\{f\}/\delta f(x)$, the functional derivative of $F\{f\}$ at that point x, followed by an integration over x. Consider the example

$$F\{f\} = \int f(x)^2\,dx\ .$$

Then

$$F\{f + \tau h\} = \int f(x)^2\,dx + 2\tau \int f(x) h(x)\,dx + \tau^2 \int h(x)^2\,dx\ .$$

Next

$$\frac{dF\{f + \tau h\}}{d\tau} = 2\int f(x) h(x)\,dx + 2\tau \int h(x)^2\,dx\ ,$$

and finally,

$$\left.\frac{dF\{f+\tau h\}}{d\tau}\right|_{\tau=0} = 2\int f(x)\,h(x)\,dx\ .$$

Thus we identify

$$\frac{\delta F\{f\}}{\delta f(x)} = 2\,f(x)\ .$$

A second derivative can also be taken, which leads to

$$\frac{\delta^2 F\{f\}}{\delta f(y)\,\delta f(x)} = 2\delta(x-y)\ ,$$

valid as a distribution. Yes, it is true that

$$\frac{\delta^2 F\{f\}}{\delta f(x)^2} = 2\delta(0) = \infty\ .$$

After all, functions of several variables may also have divergent values. In the field case, however, it is somewhat surprising at first glance that a relatively simple functional can have divergent functional derivatives.

Sometimes one encounters the statement that the functional derivative may be determined by the expression

$$\frac{\delta F\{f\}}{\delta f(x)} = \lim_{\tau\to\infty}\frac{d}{d\tau}\,F\{f+\tau\delta(\cdot-x)\}\Big|_{\tau=0}\ .$$

A casual interpretation of this expression may lead to difficulty if $\delta(x)\notin\mathcal{S}$, the space for which F is well-defined. For example, with this prescription,

$$\frac{\delta\int f(y)^2\,dy}{\delta f(x)} = \lim_{\tau\to\infty}\frac{d}{d\tau}\int[f(y)+\tau\delta(y-x)]^2\,dy\Big|_{\tau=0} = \cdots\ ,$$

where we have cut the calculation short because of the infinity that arises from the square of the δ function in the intermediate step. One is tempted to soften the δ function so its square is finite, but that is exactly what the earlier treatment was all about!

Just as there are ordinary differential equations and partial differential equations, there are also functional differential equations.

3.8 Functional Fourier Transformations

For $N<\infty$ variables, the usual Fourier transform pair can be written as

$$\tilde{u}(p) \equiv \left(\frac{1}{2\pi}\right)^{N/2} \int e^{-i\Sigma_{n=1}^{N} p_n x_n}\, u(x) \prod_{n=1}^{N} dx_n \ ,$$
$$u(x) = \left(\frac{1}{2\pi}\right)^{N/2} \int e^{i\Sigma_{n=1}^{N} p_n x_n}\, \tilde{u}(p) \prod_{n=1}^{N} dp_n \ ,$$

expressions which can be taken as pointwise, well-defined integrals when both $u(x)$ and $\tilde{u}(p)$ are smooth functions of rapid decrease.

We want to find analogous expressions when $N = \infty$. The cleanest and clearest way to do this is simply to take the limit of the given expression as $N \to \infty$. However, we want to offer a more transparent notation for such a limit, albeit one that is formal. Let us at the same time pass from the sequence notation to the function notation. In particular, we have in mind the formal expressions

$$\tilde{u}\{g\} = \mathcal{M}' \int e^{-i\int g(x) f(x)\, dx}\, u\{f\}\, \mathcal{D}f \ ,$$
$$u\{f\} = \mathcal{M}' \int e^{i\int g(x) f(x)\, dx}\, \tilde{u}\{g\}\, \mathcal{D}g \ .$$

These relations involve formal ingredients that should *not* be separately investigated. Observe that—by itself—$\mathcal{M}' = [1/(2\pi)]^{\infty} = 0$. Rather than rendering the whole expression as zero, this factor is compensated by another factor in the expression leading to a meaningful expression overall.

A more proper way to understand the functional Fourier transform is by means of *linear functionals*. All we have to observe is the existence of linear functionals L_g and L_f^{-1} such that

$$\tilde{u}\{g\} = L_g(u\{\cdot\}) \ ,$$
$$u\{f\} = L_f^{-1}(\tilde{u}\{\cdot\}) \ .$$

If we are prepared to accept these expressions, then all we need to do is understand the former, formal "functional integrals" as an alternative notation for linear functionals that map $u\{f\}$ to $\tilde{u}\{g\}$ and vice versa. As such, the "whole" is not considered as made up of its "parts," but is taken only as a "whole."

A crude analog of the former situation refers to the notation

$$\frac{dy}{dx} = \lim_{h \to 0} \frac{y(x+h) - y(x)}{h}$$

in which we don't separately evaluate dy as zero and dx as zero so that the quotient would equal the undefined value $0/0$; rather we keep the ratio intact as a single symbol, the "whole," and do not ask for the values of its top and bottom, the "parts."

If we insert one of the Fourier expressions into the other, we are led to the statement (for $N < \infty$) that

$$u(x) = \left(\frac{1}{2\pi}\right)^N \iint e^{i\Sigma_{n=1}^N p_n(x_n - y_n)}\, u(y) \prod_{n=1}^N dp_n\, dy_n \ ,$$

an expression that involves the N-dimensional δ-function,

$$u(x) = \int \delta^N(x-y)\, u(y) \prod_{n=1}^N dy_n \ .$$

Formally expressing this relation when $N \to \infty$ leads to

$$\begin{aligned} u\{g\} &= \mathcal{M} \iint e^{i\int f(x)[g(x)-h(x)]\,dx}\, u\{h\}\, \mathcal{D}f\, \mathcal{D}h \\ &= \int \delta\{g-h\}\, u\{h\}\, \mathcal{D}h \ , \end{aligned}$$

a relation that introduces and formally defines the *δ-functional*. Once again, these apparent integrals (which they are not!) should be interpreted as linear functionals—nothing more, nothing less.

3.9 Change of Variables

When we change variables inside integrals with finitely many ($N < \infty$) integration variables, we generally introduce a nontrivial Jacobian, i.e., a Jacobian different from unity (or some other constant). Let us first look at three simple, one-dimensional examples, and focus on the measures involved.

The first example involves an absolutely continuous measure, with support on the whole real line $\mathbb{R}$. Let

$$d\mu(x) = (b/2\pi)^{1/2}\, e^{-bx^2/2}\, dx \equiv \rho_1(x)\, dx \ ,$$

where $b > 0$. We consider the change of variable from x to y where $x = (b'/b)^{1/2} y$. It follows that

$$d\mu'(y) = (b'/2\pi)^{1/2}\, e^{-b'\, y^2/2}\, dy \equiv \rho_1'(y)\, dy \ ,$$

which can be regarded as a *passive* change of variables. More interesting is the viewpoint of an *active* change of variables. Specifically, we wish to view the new measure in terms of the old variables, which leads us to

$$d\mu_1'(x) = \xi_1(x)\, d\mu_1(x) \ ,$$

where

$$\xi_1(x) = (b'/b)^{1/2}\, e^{-(b'-b)x^2/2} \ .$$

We observe that this expression is well-defined for all $b' > 0$ and that the support of the new measure is $\mathbb{R}$, just like the original one. Thus, μ and μ',

which have identical supports, correspond to equivalent measures, a fact which is denoted by $\mu' \approx \mu$.

For our second example, which again is absolutely continuous, we suppose that

$$d\mu_2(x) = \rho_2(x)\, dx \ ,$$

where

$$\begin{aligned} \rho_2(x) &= b \ , \qquad |x| \le 1/2b \ , \\ &= 0 \ , \qquad |x| > 1/2b \ , \end{aligned}$$

and $b > 0$. Evidently, the support of μ_2 is $[-1/2b, 1/2b]$. For our change of variables, we let $x = (b'/b)y$, and we immediately adopt the viewpoint of an *active* change of variables. In that case

$$d\mu_2'(x) = \xi_2(x)\, d\mu_2(x)$$

holds *only* when $b' \ge b$, and specifically

$$\begin{aligned} \xi_2(x) &= b' \ , \qquad |x| \le 1/2b' \ , \\ &= 0 \ , \qquad 1/2b' < |x| \le 1/2b \ . \end{aligned}$$

The new measure has support on the interval $[-1/2b', 1/2b'] \subset [-1/2b, 1/2b]$. Thus, in the present case, μ_2' is absolutely continuous with respect to μ_2, which is denoted by $\mu_2'(x) \ll \mu_2(x)$. The converse is not true.

For our third and final example, which involves discrete measures, we suppose that

$$d\mu_3(x) = \delta(x - b^{-1})\, dx \ .$$

The support of this measure is at the single point $x = b^{-1} > 0$, say. For the change of variables, we again choose $x = (b'/b)y$, and, assuming $b' \ne b$, there is no $\xi(x)$ that relates μ' to μ (or vice versa), implying that the two measures μ' and μ are mutually singular, which is denoted by $\mu' \perp \mu$. This is evident when the two measures

$$\begin{aligned} d\mu_3(x) &= \delta(x - b^{-1})\, dx \ , \\ d\mu_3'(x) &= \delta(x - b'^{-1})\, dx \end{aligned}$$

are directly compared with each other. Observe from the point of view of the *probability distributions*, that the three qualitatively different support relationships appear quite distinct. However, that is not the case when we view matters from the *characteristic function* point of view.

Characteristic functions for the three examples are given, respectively, by

$$\begin{aligned} C_1(t) &= e^{-t^2/2b} \ , \\ C_1'(t) &= e^{-t^2/2b'} \ , \end{aligned}$$

for the normal distributions;

$$C_2(t) = \frac{\sin(t/2b)}{(t/2b)} ,$$
$$C_2'(t) = \frac{\sin(t/2b')}{(t/2b')} ,$$

for the rectangular distribution; and

$$C_3(t) = e^{it/b} ,$$
$$C_3'(t) = e^{it/b'} ,$$

for the discrete distributions. Observe that in each instance (i.e., for cases 1, 2, and 3), as $b' \to b$, the corresponding characteristic functions $C_j'(t)$, $j = 1, 2, 3$, pass continuously to the characteristic functions $C_j(t)$. Since each characteristic function behaves continuously as the parameter $b' \to b$, there is no clear evidence of the distinctly different support properties of the separate probability distributions that can be read directly out of the characteristic functions.

As we now pass from one variable—and implicitly any finite number of variables—to infinitely many variables, we will find analogous properties, namely, general continuity within the characteristic functionals accompanying quite diverse support properties of the corresponding distributions.

3.9.1 Change of infinitely many variables

Let us initially consider our standard example of infinitely many, independent identically distributed normal variables. The initial measure

$$d\mu_b(x) = \lim_{N\to\infty} (b/2\pi)^{N/2}\, e^{-b\Sigma_{n=1}^N x_n^2/2}\, \Pi_{n=1}^N\, dx_n$$

has support on the "hypersphere"

$$\lim_{P\to\infty} \frac{1}{P} \sum_{p=1}^{P} x_p^2 = b^{-1} .$$

The active view of the change of variables $x = (b'/b)^{1/2} y$, $b' \neq b$, leads to the measure

$$d\mu_{b'}(x) = \lim_{N\to\infty} (b'/2\pi)^{N/2}\, e^{-b'\Sigma_{n=1}^N x_n^2/2}\, \Pi_{n=1}^N\, dx_n$$

supported, in turn, on the disjoint "hypersphere"

$$\lim_{P\to\infty} \frac{1}{P} \sum_{p=1}^{P} x_p^2 = b'^{-1} .$$

Thus, $\mu_{b'} \perp \mu_b$, whenever $b' \neq b$. On the other hand, the associated characteristic functionals

$$C_b(\{t_n\}) = e^{-\Sigma_{n=1}^{\infty} t_n^2/2b} ,$$
$$C_{b'}(\{t_n\}) = e^{-\Sigma_{n=1}^{\infty} t_n^2/2b'}$$

pass continuously to one another as $b' \to b$.

In the present case, it is also customary to consider another kind of coordinate transformation, namely, $x_n = (b'_n/b)^{1/2} y_n$ in which the parameter b' becomes n *dependent*, and to find the limitations on the family $\{b'_n\}$ in order that μ' remains equivalent to μ, i.e., $\mu' \approx \mu$. We can readily determine the appropriate class of sequences b'_n by reference to the expression for $\xi(x)$ appropriate to the case of just one degree of freedom. Guided by that case, we find that the putative connection reads

$$d\mu'(x) = \xi_\infty(x)\, d\mu(x) ,$$

where

$$\xi_\infty(x) = \textstyle\prod_{n=1}^{\infty} (b'_n/b)^{1/2}\, e^{-(b'_n - b)\, x_n^2/2} .$$

For this expression to be well-defined, and since the x_n are variables, it is necessary that

$$\textstyle\prod_{n=1}^{\infty} (b'_n/b)^{1/2} , \qquad \prod_{n=1}^{\infty} e^{-(b'_n - b)\, x_n^2/2}$$

are *both* well-defined. For the first product, the necessary condition is that

$$\textstyle\sum_{n=1}^{\infty} |1 - (b'_n/b)^{1/2}| < \infty ,$$

while for the second product to have the *same* support we require that

$$\lim_{P\to\infty} \frac{1}{P} \sum_{n=1}^{P} (b'_n - b) x_n^2 = 0 .$$

To satisfy both requirements, the condition

$$\textstyle\sum_{n=1}^{\infty} |b'_n - b| < \infty$$

provides the proper criterion. To see this we observe that

$$\frac{1}{P} \sum_{n=1}^{P} b'_n x_n^2 = \frac{1}{P} \sum_{n=1}^{P} (b'_n - b) x_n^2 + \frac{1}{P} \sum_{n=1}^{P} b x_n^2 .$$

Thus

$$\frac{b}{P} \sum_{n=1}^{P} x_n^2 - \frac{1}{P} \sum_{n=1}^{P} |b'_n - b| x_n^2 \leq \frac{1}{P} \sum_{n=1}^{P} b'_n x_n^2 \leq \frac{b}{P} \sum_{n=1}^{P} x_n^2 + \frac{1}{P} \sum_{n=1}^{P} |b'_n - b| x_n^2 .$$

As $P \to \infty$, and adopting the condition that $\Sigma_{n=1}^{\infty}|b'_n - b| < \infty$, it follows that

$$\lim_{P\to\infty} \frac{b}{P} \sum_{n=1} x_n^2 = 1 = \lim_{P\to\infty} \frac{1}{P} \sum_{n=1}^{P} b'_n x_n^2$$

guaranteeing that μ' and μ have the same support, $\mu' \approx \mu$. It is straightforward to show that $\Sigma_{n=1}^{\infty}|b'_n - b| < \infty$ implies $\Sigma_{n=1}^{\infty}|b_n'^{1/2} - b^{1/2}| < \infty$ since

$$\begin{aligned} b^{1/2}\, \Sigma_{n=1}^{\infty}|b_n'^{1/2} - b^{1/2}| &\le \Sigma_{n=1}^{\infty}|b_n'^{1/2} - b^{1/2}|\,|b_n'^{1/2} + b^{1/2}| \\ &= \Sigma_{n=1}^{\infty}|b'_n - b| < \infty\,. \end{aligned}$$

Our last example is a special case of a more general class of linear coordinate transformations, namely, under what conditions on the transformation $T = \{T_{n\,m}\}$ is it true that

$$x_n = \sum_{m=1}^{\infty} T_{n\,m}\, y_m$$

defines a new measure μ' that is equivalent to the original measure μ. We leave that question to the exercises.

Exercises

3-1 Give a sequence of sequences $\{t_n^j\}$ such that

$$\lim_{j\to\infty} \sum_{n=1}^{\infty} |t_n^j - t_n|^3 = 0$$

as well as

$$\lim_{j\to\infty} \sum_{n=1}^{\infty} |t_n^j - t_n| = 0$$

both hold.

Next, choose another sequence for which the first expression holds but the second expression *fails* to hold.

3-2 Show that the functional

$$C(\{t_n\}) = e^{-\Sigma_{n=1}^{\infty} t_n^2}$$

is continuous for all $\{t_n\}$ such that $\Sigma_{n=1}^{\infty}|t_n| < \infty$ but *fails* to be continuous for all $\{t_n\}$ such that $\Sigma_{n=1}^{\infty}|t_n|^3 < \infty$.

3-3 Show that the functional

$$C\{f\} = e^{-\int |f(x)|^2\, dx} ,$$

where $x \in \mathbb{R}$, is continuous for all $f \in L^2(\mathbb{R})$. What are the continuity properties for: (1) $f \in L^1(\mathbb{R})$, and (2) $f \in L^3(\mathbb{R})$?

Repeat this problem with the change that $x \in [0, 1]$ rather than $x \in \mathbb{R}$.

3-4 For the case of the transformation $x_n = (b_n/b)^{1/2} y_n$, $n = 1, 2, 3, \ldots$, we concluded that the criterion for equivalence of normal measures was that $\Sigma_n |1 - b_n/b| < \infty$. Consider this transformation to be a special case of $x_n = \Sigma_m T_{nm}\, y_m$, and reexpress the criterion on $\{b_n\}$ above as a condition on the matrix $T = \{T_{nm}\}$. Extend that criterion to find a general nondiagonal form of transformation T_{nm} so that the original normal measure and the transformed measure are equivalent to each other.

3-5 Solve the functional differential equation given by

$$\frac{\delta F\{f\}}{\delta f(x)} = b(x)\, f(x)^2\, F\{f\} ,$$

where $b(x)$ is a smooth fixed function.

3-6 Solve the functional differential equation given by

$$\frac{\delta G\{f\}}{\delta f(x)} = [-f''(x) + c\, f(x) + g\, f(x)^3]\, G\{f\} ,$$

where $f''(x) \equiv d^2 f(x)/dx^2$, and c and g are constants.

4

Stochastic Variable Theory

4.1 General Remarks

Let $X(t)$ be a random function of time t, where $t \in [a, b]$, $-\infty < a < b < \infty$, or $t \in [a, \infty)$, or $t \in (-\infty, \infty)$, as the specific case dictates. One way to describe the observable properties of the set of random functions, a.k.a. (also known as) a *stochastic variable*, is by means of a collection of *correlation functions*

$$C_l(t_1, t_2, \ldots, t_l) \equiv \langle X(t_1) X(t_2) \cdots X(t_l) \rangle \ ,$$

for all $l \geq 1$. If the functions in the set $\{C_l\}$ are pointwise defined, then $X(t)$ is called a *stochastic process*. On the other hand, if the functions in the set are distributional in nature, then $X(t)$ is called a *generalized stochastic process*. We will have occasion to discuss both types.

Since the correlation functions are clearly symmetric functions of their arguments, it follows that the set of correlation functions can be conveniently summarized in terms of a generating functional given by the characteristic functional

$$C\{s\} \equiv \langle \, e^{i \int s(t) X(t) \, dt} \rangle \ .$$

Furthermore we shall assume that this function satisfies the requirements of the Bochner–Minlos Theorem (see Sec. 3.6). In that case the characteristic functional may be written as the functional Fourier transform of a genuine probability measure $\mu(x)$ on suitable distributional paths such that

$$C\{s\} = \int e^{i \int s(t) x(t) \, dt} \, d\mu(x) \ .$$

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9_4,

4.1.1 Stationary processes

An important subset of stochastic variables has the property that the correlation functions are *stationary*, which means that all of the correlation functions are invariant under an arbitrary shift of all the time arguments. This is only possible for a situation for which the time runs over the whole real line, i.e., $t \in (-\infty, \infty)$. For example, if we translate every t_n variable by a common fixed amount τ, then the resultant correlation functions are independent of τ. In symbols, a stationary process is one for which

$$\begin{aligned}
&\langle X(t_1+\tau)\rangle = \langle X(t_1)\rangle \,, \\
&\langle X(t_1+\tau)X(t_2+\tau)\rangle = \langle X(t_1)X(t_2)\rangle \,, \\
&\langle X(t_1+\tau)X(t_2+\tau)X(t_3+\tau)\rangle = \langle X(t_1)X(t_2)X(t_3)\rangle \,,
\end{aligned}$$

etc., for all values of τ. In crude terms, the correlation functions measured today are the same as those measured yesterday! This property can be restated as a property of the characteristic functional as

$$C\{s_\tau\} = C\{s\} \,,$$

for all τ, where $s_\tau(t) \equiv s(t-\tau)$.

4.1.2 Ergodic processes

A further specialization of stationary distributions leads to *ergodic distributions*. For such distributions the following limit holds:

$$\lim_{\tau\to\pm\infty} C\{s_\tau + u\} = C\{s\}\, C\{u\} \,.$$

This expression implies that the correlation functions factor in an appropriate manner. For example, ergodicity implies that

$$\lim_{\tau\to\pm\infty} \langle X(t_1+\tau)X(t_2+\tau)X(t_3)X(t_4)\rangle = \langle X(t_1)X(t_2)\rangle \, \langle X(t_3)X(t_4)\rangle \,,$$

etc.

It is also convenient to introduce *truncated* or *connected* correlation functions. The connected correlation functions are algebraic combinations of the ordinary correlation functions discussed above. In particular, the generating functional of connected correlation functions $C^c\{s\}$ is defined by

$$C^c\{s\} \equiv 1 + \ln[C\{s\}] \,;$$

expanding both sides in powers of s provides the desired connections. In particular,

$$
\begin{aligned}
&C_1^c(t_1) = C_1(t_1)\ ,\\
&C_2^c(t_1,t_2) = C_2(t_1,t_2) - C_1(t_1)C_1(t_2)\ ,\\
&C_3^c(t_1,t_2,t_3) = C_3(t_1,t_2,t_3) - C_1(t_1)C_2(t_2,t_3) - C_1(t_2)C_2(t_3,t_1)\\
&\qquad -C_1(t_3)C_2(t_1,t_2) + 2C_1(t_1)C_1(t_2)C_1(t_3)\ ,\\
&C_4^c(t_1,t_2,t_3,t_4) = C_4(t_1,t_2,t_3,t_4) - C_1(t_1)C_3(t_2,t_3,t_4) - C_1(t_2)C_3(t_3,t_4,t_1)\\
&\qquad -C_1(t_3)C_3(t_4,t_1,t_2) - C_1(t_4)C_3(t_1,t_2,t_3) - C_2(t_1,t_2)C_2(t_3,t_4)\\
&\qquad -C_2(t_1,t_3)C_2(t_2,t_4) - C_2(t_1,t_4)C_2(t_2,t_3) + 2C_2(t_1,t_2)C_1(t_3)C_1(t_4)\\
&\qquad +2C_2(t_1,t_3)C_1(t_2)C_1(t_4) + 2C_2(t_1,t_4)C_1(t_2)C_1(t_3)\\
&\qquad +2C_2(t_2,t_3)C_1(t_1)C_1(t_4) + 2C_2(t_2,t_4)C_1(t_1)C_1(t_3)\\
&\qquad +2C_2(t_3,t_4)C_1(t_1)C_1(t_2) - 6C_1(t_1)C_1(t_2)C_1(t_3)C_1(t_4)\ ,
\end{aligned}
$$

etc. This definition has the virtue that

$$\lim_{\tau\to\pm\infty} C_{l+m}^c(t_1+\tau,\ldots,t_l+\tau,t_{l+1},\ldots,t_{l+m}) = 0\ ,$$

for all $l \geq 1$ and $m \geq 1$. Thus we observe that the connected component of the lth correlation function contains the new information in that correlation function that is not already contained in any of the lower order correlation functions.

Another property of connected correlation functions deals with a stochastic variable that is the sum of two completely independent components. For example, suppose that

$$X(t) = Y(t) + Z(t)\ ,$$

where Y and Z denote two independent stochastic processes. In that case, the characteristic functions are connected by

$$C_X\{s\} = C_Y\{s\}\, C_Z\{s\}\ ,$$

which leads to the relations

$$C_{X\,l}^c(t_1,\ldots,t_l) = C_{Y\,l}^c(t_1,\ldots,t_l) + C_{Z\,l}^c(t_1,\ldots,t_l)\ .$$

Finally, we wish to stress a fundamentally important property of stationary, ergodic ensembles. Let us investigate the stochastic variable

$$U_T\{s\} \equiv \frac{1}{2T}\int_{-T}^{T} e^{i\int s(t-\tau)\,X(t)\,dt}\,d\tau\ ,$$

which maps the stochastic variable X to another random variable U_T which depends on the function s and the time T. We first observe that

$$\langle U_T\{s\}\rangle = \frac{1}{2T}\int_{-T}^{T} d\tau \int e^{i\int s(t-\tau)\,x(t)\,dt}\,d\mu(x)\ ,$$

which implies that

$$\langle U_T\{s\}\rangle = \frac{1}{2T}\int_{-T}^{T} C\{s_\tau\}\,d\tau = C\{s\}$$

by virtue of stationarity of the process. Next we consider the expression

$$\langle\,|U_T\{s\} - C\{s\}|^2\,\rangle\,,$$

which when expanded becomes

$$\langle U_T\{s\}^*\,U_T\{s\}\rangle - |C\{s\}|^2\,.$$

The first term in the previous line becomes

$$\begin{aligned}
&\langle U_T\{s\}^*\,U_T\{s\}\rangle \\
&\quad = \frac{1}{(2T)^2}\int_{-T}^{T}\!\!\int_{-T}^{T} d\tau\,d\tau' \int e^{i\int[s(t-\tau)-s(t-\tau')]\,x(t)\,dt}\,d\mu(x) \\
&\quad = \frac{1}{(2T)^2}\int_{-T}^{T}\!\!\int_{-T}^{T} d\tau\,d\tau'\,C\{s_\tau - s_{\tau'}\}\,.
\end{aligned}$$

Next we consider the limit of the final expression as $T\to\infty$. In view of the ergodicity of the ensemble, we first recall that

$$\lim_{\tau\to\pm\infty} C\{s_\tau + u\} = C\{s\}\,C\{u\}\,.$$

As a consequence of this relation it also follows that

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} C\{s_\tau + u\}\,d\tau = C\{s\}\,C\{u\}\,.$$

Therefore,

$$\lim_{T\to\infty}\langle U_T\{s\}^*\,U_T\{s\}\rangle = |\,C\{s\}\,|^2\,,$$

or stated otherwise that

$$\lim_{T\to\infty}\langle\,|U_T\{s\} - C\{s\}|^2\,\rangle = 0\,.$$

The fundamental conclusion of this exercise is that

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{i\int s(t-\tau)\,X(t)\,dt}\,d\tau = C\{s\}\,.$$

More precisely, this limit exists in the *mean square* sense shown in the previous paragraph. To make that distinction, one often asserts that

$$\text{l.i.m.}_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{i\int s(t-\tau)\,X(t)\,dt}\,d\tau = C\{s\}\ ,$$

where "l.i.m." is short for "limit in (the) mean." What this means is that this relation holds with probability one, i.e., there are potentially paths for which this identity fails, but they only constitute a set of measure zero. Ignoring such subtleties for the moment, one may loosely say that "*One path has it all!*" Namely, since

$$U_\infty\{s\} \equiv \lim_{T\to\infty} U_T\{s\} = C\{s\}\ ,$$

the putative stochastic variable $U_\infty\{s\}$ is *not* random and takes the same value for (almost) all paths. This fact implies that the paths are concentrated on that set for which $U_\infty\{s\} = C\{s\}$ for all s. Thus if we consider two such stochastic processes each with their own characteristic function, say, $C_1\{s\}$ and $C_2\{s\}$, then the associated measures are mutually singular, $\mu_1(x) \perp \mu_2(x)$, whenever the two characteristic functionals are not identical for all s.

4.1.3 Gaussian examples

A Gaussian stochastic process $X(t)$ is distinguished by the fact that

$$\begin{aligned} C\{S\} &= \langle e^{i\int s(t)\,X(t)\,dt}\rangle = \int e^{i\int s(t)\,x(t)\,dt}\,d\mu(x) \\ &= e^{i\int s(t)\,\langle X(t)\rangle\,dt - \frac{1}{2}\int s(t)\,s(u)\langle X(t)\,X(u)\rangle^c\,dt\,du} \\ &= e^{i\int s(t)\,m(t)\,dt\,-\,\frac{1}{2}\int s(t)\,s(u)\,G(t,u)\,dt\,du}\ . \end{aligned}$$

This expression holds whenever the time variable $t \in [a,b]$, $-\infty < a < b < \infty$, or $t \in [0,\infty)$, or $t \in (-\infty,\infty)$. Such a Gaussian process is *stationary* provided that $t \in (-\infty,\infty)$ and

$$\langle X(t)\rangle = m\ , \qquad \langle X(t)\,X(u)\rangle = G(t-u)\ ,$$

and therefore

$$C\{s\} = e^{i\,m\int s(t)\,dt - \frac{1}{2}\int s(t)\,s(u)G(t-u)\,dt\,du}\ .$$

If X is a stochastic process, then $\langle X(t)^2\rangle < \infty$, and as a consequence, $G(t,u)$, or $G(t-u)$, as the case may be, is pointwise defined. If X is a generalized stochastic process, then $G(t,u)$, or $G(t-u)$, is distributional which requires that the function s belong to some form of test function space. In either case, since $\int s(t)\,s(u)\,G(t-u)\,dt\,du \geq 0$, it follows that

$$G(t-u) = \int e^{i(t-u)\omega}\,dm(\omega)\ ,$$

by Bochner's Theorem, where $m(\omega)$ is a positive measure. For a stochastic process, $\int dm(\omega) < \infty$, and the function $G(t-u)$ is continuous; for a generalized stochastic process, $\int dm(\omega) = \infty$, and the generalized function $G(t-u)$ need not be continuous.

For a stationary Gaussian process to be *ergodic*, it is only necessary that the measure $m(\omega)$ be absolutely continuous, i.e.,

$$dm(\omega) = \rho(\omega)\, d\omega \ ,$$

where $\rho(\omega) \geq 0$.

4.2 Wiener Process, a.k.a. Brownian Motion

4.2.1 Definition of a standard Wiener process

A standard Wiener process, denoted by $W(t)$ on the interval $0 \leq t < \infty$—which is therefore not a stationary process—is characterized by the following four properties:

a) $W(0) = 0$,
b) $\langle W(t) \rangle = 0$,
c) $\langle W(t) W(u) \rangle = \min(t,u)$,
d) The process is Gaussian. In fact, it is normal.

These four rules may be incorporated into the single expression given by

$$\langle e^{i \int s(t) W(t)\, dt} \rangle = e^{-\frac{1}{2} \int s(t)\, s(u)\, \min(t,u)\, dt\, du} \ .$$

[**Remark:** Sometimes one sees the notation $t \wedge u \equiv \min(t,u)$, as well as $t \vee u \equiv \max(t,u)$. However, we shall continue to use the more transparent min and max.] Expanding both sides of the previous relation in powers of s leads, for example, to

$$\frac{1}{4!} \langle \big(\int s(t) W(t)\, dt \big)^4 \rangle = \frac{1}{8} \big[\int s(t)\, s(u)\, \min(t,u)\, dt\, du \big]^2 \ ,$$

which implies that

$$\begin{aligned} \langle W(t_1) W(t_2) W(t_3) W(t_4) \rangle &= \langle W(t_1) W(t_2) \rangle \langle W(t_3) W(t_4) \rangle \\ &+ \langle W(t_1) W(t_3) \rangle \langle W(t_2) W(t_4) \rangle + \langle W(t_1) W(t_4) \rangle \langle W(t_2) W(t_3) \rangle \ . \end{aligned}$$

When all times are the same, we find

$$\langle W(t)^4 \rangle = 3 \langle W(t)^2 \rangle^2 \ .$$

More generally, we learn that

$$\langle e^{isW(t)} \rangle = e^{-ts^2/2} ,$$

from which it follows that

$$p(x,t) = \langle \delta(x - W(t)) \rangle = \frac{1}{2\pi} \int dk \langle e^{ik(x - W(t))} \rangle = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} .$$

As yet another example we consider

$$\langle e^{is[W(t_1) - W(t_2)]} \rangle = e^{-|t_1 - t_2| s^2/2} ,$$

which leads to

$$\langle [W(t_1) - W(t_2)]^2 \rangle = |t_1 - t_2| ,$$

as well as

$$\langle [W(t_1) - W(t_2)]^4 \rangle = 3|t_1 - t_2|^2 .$$

4.2.2 Continuity of Brownian paths

We start our discussion by assuming that Brownian motion paths may be piecewise continuous, and then show that assumption is false.

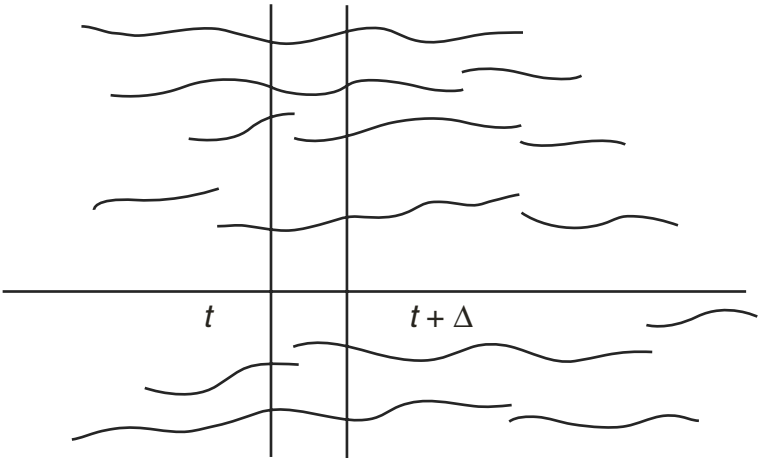

Fig. 4.1. A small sample of hypothetical Brownian paths with their position on the vertical axis vs. time on the horizontal axis. For clarity, the sample paths have been chosen so they do not overlap one another. During the small time interval of width Δ between t and $t + \Delta$, some paths are continuous while other paths have jumps (discontinuities).

Figure 4.1 illustrates several hypothetical paths in a given window of time. Observe that some of the paths are continuous through the short interval from

t to $t+\Delta$, while some of the paths have discontinuities in that same short time interval. Let us evaluate qualitatively the expression

$$\langle [W(t+\Delta) - W(t)]^2 \rangle = A\Delta + B\Delta \, .$$

Here, the factor A comes from the fraction of paths that are *continuous* and its coefficient Δ represents the contribution from the *slope* of such continuous paths within the interval t to $t+\Delta$. Likewise, the factor B comes from the paths that are *discontinuous* and its coefficient Δ reflects the relative *probability* of such paths within the interval t to $t+\Delta$. Next consider the analogous expression

$$\langle [W(t+\Delta) - W(t)]^4 \rangle = A'\Delta^2 + B'\Delta \, .$$

In this case the factor A' comes from the continuous paths and its coefficient Δ^2 comes from the square of the slope. The factor B' comes from the discontinuous paths and its coefficient Δ again reflects the relative probability of such paths in the given time interval.

If $B' = 0$, then there are no discontinuous paths, and as a consequence, all the paths are continuous. For Brownian motion paths, this is exactly what occurs leading to the conclusion that Brownian motion paths are continuous with probability one.

As a notational remark, hereafter, we shall also use the abbreviations: w.p.o. = with probability one, and w.p.z. = with probability zero.

Indeed, there is a more general theorem by Kolmogorov which, in our terminology, asserts that if $X(t)$ is a stochastic process with $t \in [t_0, T]$, and there exist $\alpha > 0$ and $\beta > 0$ and $C > 0$ such that for $t_1, t_2 \in [t_0, T]$,

$$\langle \, |x(t_1) - x(t_2)|^\alpha \rangle \leq C \, |t_1 - t_2|^{1+\beta} \, ,$$

then the process $X(t)$ is continuous w.p.o. [Sko65]

Earlier, in Sec. 2.2.4, we derived the property of a normal random variable X given by

$$\langle e^{-\frac{1}{2}X^2} \rangle = \frac{1}{\sqrt{1 + \langle X^2 \rangle}} \, ,$$

from which we concluded that when $\langle X^2 \rangle = \infty$, it follows that $X^2 = \infty$ w.p.o. On the other hand, if $\langle X^2 \rangle < \infty$, then it follows directly from this fact alone that $X^2 < \infty$ w.p.o. We can apply this reasoning to the normal variable $X = [W(t+h) - W(t)]/h$ for which $\langle ([W(t+h) - W(t)]/h)^2 \rangle = 1/h$. Consequently, it follows that

$$[dW(t)/dt]^2 = \lim_{h \to 0} ([W(t+h) - W(t)]/h)^2 = \infty$$

w.p.o. In other words, Brownian motion paths, while everywhere continuous, are, with probability one, nowhere differentiable!

We can also investigate more carefully issues related to the lack of derivatives. In particular, we now consider the normal variable $X = [W(t) - W(s)]/|t-s|^\gamma$. First observe that

$$\langle ([W(t) - W(s)]/|t-s|^\gamma)^2 \rangle = |t-s|^{1-2\gamma} .$$

Thus, for $|t-s| \leq 1$ and $\gamma = \frac{1}{2} \pm \epsilon$, $\epsilon > 0$, we find that

$$|W(t) - W(s)| \leq C|t-s|^{\frac{1}{2} - \epsilon} , \quad \text{w.p.o.} ,$$
$$|W(t) - W(s)| \leq C'|t-s|^{\frac{1}{2} + \epsilon} , \quad \text{w.p.z.} ,$$

where C and C' denote time-independent stochastic variables.

It is interesting to compare these upper and lower bounds to more traditional classical bounds. In particular,

$$|x(t) - x(s)| = |\int_s^t \dot{x}(u)\, du| \leq |t-s|^{1/2} [\int_s^t \dot{x}(u)^2\, du]^{1/2} ,$$

as follows from the Schwarz inequality. Thus, it would appear that Brownian motion paths fail to have classical behavior by only a "small amount" (ϵ). On the other hand, classical paths may vary quite slowly at certain points, e.g., $x(t) = (t-1)^3$ near $t = 1$; such smoothness is never seen in Brownian paths w.p.o.

4.2.3 Stochastic equivalence

As a preliminary to the present topic, we make a simple remark regarding (in)dependence of two normal random variables X and Y. Let $Z = aX + bY$, and note that

$$\begin{aligned} \langle e^Z \rangle &= e^{\langle Z^2 \rangle} = e^{\langle (aX+bY)^2 \rangle} \\ &= e^{\langle (aX)^2 \rangle}\, e^{\langle (bY)^2 \rangle}\, e^{2ab\langle XY \rangle} \\ &= \langle e^{aX} \rangle\, \langle e^{bY} \rangle\, e^{2ab\langle XY \rangle} . \end{aligned}$$

Thus it follows that X and Y represent independent normal variables if and only if $\langle XY \rangle = 0$.

We say that two stochastic processes $X'(t)$ and $X(t)$ are *stochastically equivalent* if all their correlation functions are equal, i.e.,

$$\langle X'(t_1) X'(t_2) \cdots X'(t_l) \rangle = \langle X(t_1) X(t_2) \cdots X(t_l) \rangle$$

for all $l \geq 1$. When that is the case, we say that $X'(t) \simeq X(t)$. For example, some stochastic processes that are stochastically equivalent to a standard Wiener process are given by

1) $W'(t) = -W(t)$

Brownian motion reflected about the origin.

$\langle W'(t)W'(u)\rangle = \langle W(t)W(u)\rangle = \min(t,u)$;

2) $W'(t) = W(t+t_0) - W(t_0)$, $\quad t_0 > 0$

Brownian motion restarts afresh.

$$\langle W'(t_1)W'(t_2)\rangle = \langle [W(t_1+t_0) - W(t_0)]\,[W(t_2+t_0) - W(t_0)]\rangle$$
$$= \min(t_1,t_2)\ ;$$

3) $W'(t) = W(T) - W(T-t)$, $\quad 0 \le t \le T$

Brownian motion run backwards.

$$\langle W'(t_1)W'(t_2)\rangle = \langle [W(T) - W(T-t_1)]\,[W(T) - W(t-t_2)]\rangle$$
$$= \min(t_1,t_2)\ ;$$

4) $W'(t) = kW(t/k^2)$, $\quad k > 0$

Rescaled Brownian motion, showing fractal character.

$\langle W'(t_1)W'(t_2)\rangle = k^2\langle W(t_1/k^2)W(t_2/k^2)\rangle = \min(t_1,t_2)$;

5) $W'(t) = \lim\limits_{k\to 0} kW(t/k^2)$

Independent Brownian motion derived from initial Brownian motion.

$\langle W'(t_1)W(t_2)\rangle = \lim\limits_{k\to 0} k\langle W(t_1/k^2)W(t_2)\rangle = \lim\limits_{k\to 0} kt_2 = 0$;

6) $W'(t) = \lim\limits_{k\to\infty} kW(t/k^2)$

Second independent Brownian motion derived from initial Brownian motion.

$\langle W'(t_1)W(t_2)\rangle = \lim\limits_{k\to\infty} k\langle W(t_1/k^2)W(t_2)\rangle = \lim\limits_{k\to\infty} t_1/k = 0$.

4.2.4 Independent increments

Imagine dividing the time axis into separate intervals by a sequence of times $t = t_j$ ordered so that $0 < t_1 < t_2 < t_3 \ldots$. From the general expression for the Wiener measure characteristic functional, it follows that

$$\langle e^{i\Sigma_{l=1}^{L} s_l [W(t_l) - W(t_{l-1})]}\rangle$$
$$= \exp\{-\tfrac{1}{2}\textstyle\sum_{l,m=1}^{L,L} s_l\, s_m \langle [W(t_l) - W(t_{l-1})]\,[W(t_m) - W(t_{m-1})]\rangle\}$$
$$= \exp[-\tfrac{1}{2}\textstyle\sum_{l=1}^{L} s_l^2\,|\,t_l - t_{l-1}\,|\,]\ .$$

This result arises from the terms in the double sum where $m = l$; the terms where $m < l$ vanish since

$$\langle [W(t_l) - W(t_{l-1})][W(t_m) - W(t_{m-1})] \rangle = (t_m - t_m) - (t_{m-1} - t_{m-1}) = 0 \,,$$

and likewise when $l < m$. The conclusion we draw from this exercise is that the normal variable $W(t_l) - W(t_{l-1})$ is independent of the normal variable $W(m_l) - W(m_{l-1})$ whenever $m \neq l$. This property of Brownian motion is referred to as *independent increments*.

4.2.5 Some joint and conditional probability densities

In what follows, we offer a few examples of joint and conditional probability densities for Wiener processes. These are presented without proof, but the reader is encouraged to establish them by means of the hints that are offered (see Exercises). Note that $p(a; b)$ denotes the joint probability density of a and b, while $p(a|b)$ denotes the probability density of a given the condition b. Our list includes

- $$p(x_1, t_1) = \langle \delta(x_1 - W(t_1)) \rangle = \langle \delta(x_1 - 0 - [W(t_1) - W(0)]) \rangle \\ \equiv p(x_1, t_1 | 0, 0) \,,$$

- $$p(x_2, t_2; x_1, t_1) = \langle \delta(x_2 - W(t_2))\, \delta(x_1 - W(t_1)) \rangle \\ \equiv p(x_2, t_2 | x_1, t_1)\, p(x_1, t_1) \,,$$

- $$p(x_3, t_3; x_2, t_2; x_1, t_1) = \langle \delta(x_3 - W(t_3))\, \delta(x_2 - W(t_2))\, \delta(x_1 - W(t_1)) \rangle \\ \equiv p(x_3, t_3 | x_2, t_2; x_1, t_1)\, p(x_2, t_2 | x_1, t_1)\, p(x_1, t_1) \,,$$

etc. If

$$p(x_l, t_L | x_{L-1}, t_{L-1}; x_{L-2}, t_{L-2}; \ldots; x_1, t_1) = p(x_L, t_L | x_{L-1}, t_{L-1})$$

for all $L \geq 1$, then the process is called *Markovian*, and one is dealing with a *Markov process*.

For a Wiener process it follows that

$$\begin{aligned} & p(x_l, t_L; x_{L-1}, t_{L-1}; x_{L-2}, t_{L-2}; \cdots; x_1, t_1) \\ & \quad = \langle \prod_{l=1}^{L} \delta(x_l - W(t_l)) \rangle \\ & \quad = \langle \prod_{l=1}^{L} \delta(x_l - x_{l-1} - [W(t_l) - W(t_{l-1})]) \rangle \\ & \quad = \prod_{l=1}^{L} \langle \delta(x_l - x_{l-1} - [W(t_l) - W(t_{l-1})]) \rangle \,, \end{aligned}$$

establishing that a Wiener process is Markovian. Specifically, this leads to the relation that

$$\begin{aligned} p(x_L,t_L;x_{L-1},t_{L-1};x_{L-2},t_{L-2};\ldots;x_1,t_1) \\ &= p(x_L,t_L|x_{L-1},t_{L-1})\cdots p(x_2,t_2|x_1,t_1)\,p(x_1,t_1|0,0) \\ &= \prod_{l=1}^{L}\left\{\frac{1}{\sqrt{2\pi(t_l-t_{l-1})}}\,\exp\left[-\frac{(x_l-x_{l-1})^2}{2(t_l-t_{l-1})}\right]\right\}. \end{aligned}$$

The limit $L\to\infty$ will be discussed later.

4.2.6 Itô calculus

In this section, the goal is to calculate the expression

$$\langle\,[W(t+\Delta)-W(t)]^2\,e^{i\int s(u)W(u)\,du}\,\rangle\,.$$

To reach this goal we start with the expression

$$\begin{aligned} \langle e^{ik[W(t+\Delta)-W(t)]+i\int s(u)W(u)\,du}\rangle \\ &= e^{-\frac{1}{2}k^2\Delta-\frac{1}{2}\int s(u)\,s(v)\,\langle W(u)W(v)\rangle\,du\,dv} \\ &\quad\times e^{-k\int s(u)\,\langle W(u)[W(t+\Delta)-W(t)]\rangle\,du}\,. \end{aligned}$$

The last term in the exponent is of order Δ; specifically,

$$\begin{aligned} k\int s(u)\,\langle W(u)[W(t+\Delta)-W(t)]\rangle\,du \\ &= k\int_t^{t+\Delta} s(u)\,[u-t]\,du + k\Delta\int_{t+\Delta}^{\infty} s(u)\,du \equiv k\mathcal{O}(\Delta;s)\,. \end{aligned}$$

The expressions above can be combined so that

$$\begin{aligned} \langle e^{ik[W(t+\Delta)-W(t)]+i\int s(u)W(u)\,du}\rangle \\ &= e^{-\frac{1}{2}k^2\Delta-k\mathcal{O}(\Delta;s)}\,\langle e^{i\int s(u)W(u)\,du}\rangle\,, \end{aligned}$$

and expanding both sides to second order in k leads to

$$\langle\,[W(t+\Delta)-W(t)]^2\,e^{i\int s(t)W(t)\,dt}\,\rangle = [\Delta+O(\Delta^2;s)]\,\langle e^{i\int s(u)W(u)\,du}\rangle\,.$$

The limit as $\Delta\to dt$ becomes

$$\langle\,dW(t)^2\;e^{i\int s(u)W(u)\,du}\,\rangle = dt\,\langle e^{i\int s(u)W(u)\,du}\rangle\,,$$

and this relation holds for all functions $s(u)$. Superposition over such different functions implies that

$$dW(t)^2 = dt \ , \qquad \text{w.p.o.}$$

This relation is one of the celebrated multiplication rules of Itô. As any differential, it achieves its true meaning inside an integral such as

$$\int F(t)\, dW(t)^2 = \int F(t)\, dt \ .$$

The implication of this equation is that the fractal nature of Brownian paths ensures that $dW(t)^2$ acts in every way as dt.

For any two *independent* Wiener processes, $W_1(t)$ and $W_2(t)$, it follows that $\langle W_1(t)\, W_2(u) \rangle = 0$. Consequently, we have the Itô rule that

$$dW_1(t)\, dW_2(t) = 0 \ .$$

The basic Itô calculus multiplication table [Hid70] is given in Table 4.1.

Table 4.1. Multiplication rules of Itô calculus

$\times$	$dW(t)$	dt
$dW(t)$	dt	0
dt	0	0

4.2.7 Stochastic integrals

We have already encountered the integral $\int s(t) W(t)\, dt$, which for suitable $s(t)$ can be interpreted as

$$\int s(t)\, W(t)\, dt = \int W(t)\, dy(t) \ ,$$

provided $y'(t) = s(t)$. Such an interpretation extends further to any $y(t)$ which is a signed measure. Additionally, there are many more interesting ways in which Brownian paths can enter integrals. For example, consider

$$\int s(t)\, dW(t) = s(t)\, W(t) \Big| - \int W(t)\, ds(t) \ ,$$

which we can interpret more easily after the integration by parts. Not so easily handled is the integral

$$\int W(t)\, dW(t) \ .$$

Here we have our first example of a genuine *stochastic integral*. Since $W(t)$ cannot serve as a measure, this integral cannot be handled by classical methods. Moreover, there are two fundamentally different ways to define this stochastic integral.

Both procedures use a limiting behavior of the integral as defined on a temporal lattice, which, for convenience, we choose to be a lattice of uniform spacing $t_l - t_{l-1} = \epsilon$ for all l. On this lattice we define the values of the Wiener process at the lattice points as

$$W_l \equiv W(t_l) = W(l\epsilon) \ , \qquad 0 \le l \le L \ .$$

The overall time interval in question is given by $t = L\epsilon$. The limit of interest is $L \to \infty$ and $\epsilon \to 0$, with the product held fixed, $L\epsilon = t$.

Following the calculus of Itô [MS74], let us first consider a differential of a function on this lattice of points. In particular,

$$\begin{aligned} \Delta f(W_l) &= f(W_l) - f(W_{l-1}) \\ &= f(W_{l-1} + (W_l - W_{l-1})) - f(W_{l-1}) \\ &= f'(W_{l-1})(W_l - W_{l-1}) + \tfrac{1}{2} f''(W_{l-1})(W_l - W_{l-1})^2 + \cdots \ ; \end{aligned}$$

observe that the differential $(W_l - W_{l-1})$ points toward the future as compared to the coefficients $f'(W_{l-1})$ and $f''(W_{l-1})$. In the limit that the lattice spacing $\epsilon \to 0$, we learn that

$$df(W(t)) = f'(W(t)) \cdot dW(t) + \tfrac{1}{2} f''(W(t)) dt \ ,$$

based on the fact that $dW(t)^2 = dt$ for standard Brownian motion.

However, there is another way to define the differential of a function that leads to a different answer. Thus, following Stratonovich [MS74], we consider the expression

$$\begin{aligned} \Delta f(W_l) &= f(W_l) - f(W_{l-1}) \\ &= \tfrac{1}{2}[f(W_{l-1} + (W_l - W_{l-1}))] - \tfrac{1}{2}[f(W_l - (W_l - W_{l-1}))] \\ &\quad + \tfrac{1}{2}[f(W_l) - f(W_{l-1})] \\ &= \tfrac{1}{2}[f'(W_{l-1})(W_l - W_{l-1}) + \tfrac{1}{2} f''(W_{l-1})(W_l - W_{l-1})^2 + \cdots] \\ &\quad + \tfrac{1}{2}[f'(W_l)(W_l - W_{l-1}) - \tfrac{1}{2} f''(W_l)(W_l - W_{l-1})^2 + \cdots] \\ &= \tfrac{1}{2}[f'(W_l) + f'(W_{l-1})](W_l - W_{l-1}) + \cdots \ . \end{aligned}$$

In the limit that the lattice spacing $\epsilon \to 0$, we learn that

$$df(W(t)) = f'(W(t)) \, dW(t) \ .$$

Thus, according to the prescription of Stratonovich, the ordinary rules of calculus apply. This fact has some decided advantage for us and subsequently we shall exclusively focus on the Stratonovich rule. From the way these two rules

are defined, one refers to the Itô rule as *nonanticipating*, while the Stratonovich rule is referred to as the *midpoint* rule.

We have denoted the Itô rule as $f'(W(t)) \cdot dW(t)$ and the Stratonovich rule as $f'(W(t))\, dW(t)$, i.e., one with a center dot and one without a center dot. However, since the Itô rule was developed initially, it is the one that is traditionally denoted by $f'(W(t))\, dW(t)$; the Stratonovich rule, which was introduced later, is commonly denoted by $f'(W(t)) \circ dW(t)$ when the two rules are discussed simultaneously. We choose the simpler notation for the Stratonovich ("midpoint") rule since it is the only rule we will need.

It is clear that the two rules are fundamentally different. For example, the Itô integral

$$\int_0^T W(t) \cdot dW(t) = \tfrac{1}{2}\left[W(T)^2 - T\right] ,$$

while the Stratonovich integral

$$\int_0^T W(t)\, dW(t) = \tfrac{1}{2} W(T)^2 .$$

Observe, for the Itô integral, that

$$\langle \int_0^T W(t) \cdot dW(t) \rangle = \tfrac{1}{2} \langle [W(T)^2 - T] \rangle = 0 .$$

In fact, a vanishing expectation is generally true for an Itô integral,

$$\langle \int f(W(t)) \cdot dW(t) \rangle = 0 ,$$

thanks to the future pointing nature of the differential $dW(t)$ and the independence of Wiener process increments.

4.3 Wiener Measure

It is now time to illustrate the distribution of paths that are part of a standard Wiener process [Hid70, IMc65]. According to results presented earlier, the probability density for Brownian motion paths to pass through the set of points $\{x_l\}_{l=1}^L$ at times $\{t_l\}_{l=1}^L$, where $0 < t_1 < t_2 < \ldots < t_L$, may be expressed as

$$\begin{aligned} p(x_l, t_L; x_{L-1}, t_{L-1}; \ldots; x_1, t_1) &= \textstyle\prod_{l=1}^L p(x_l, t_l | x_{l-1}, t_{l-1}) \\ &= (1/2\pi\epsilon)^{L/2}\, e^{-\Sigma_{l=1}^L (x_l - x_{l-1})^2/2\epsilon} , \end{aligned}$$

where we have also introduced $x_0 = 0$ and $t_0 = 0$.

What is the result for the probability density when $L \to \infty$? A formal expression arises when we let $L \to \infty$ in the previous equation and write the formal result for continuous and differentiable paths

$$p(x(\cdot)) = \mathcal{M}\, e^{-\frac{1}{2}\int \dot{x}^2(t)\, dt} \ ,$$

where $\mathcal{M}$ is a formal (infinite) normalization factor. Moreover, we have already concluded that Brownian motion paths are continuous but nowhere differentiable, which means that there are no paths that are both continuous and differentiable; this makes the second factor effectively zero w.p.o. We now approach the probability density for Brownian paths by another procedure.

By definition, the characteristic functional for a Wiener process is given by

$$C\{s\} = \langle e^{i\int s(t) W(t)\, dt} \rangle = e^{-\frac{1}{2}\int s(t)\, s(u)\, \min(t,u)\, dt\, du} \ ,$$

which is clearly a continuous functional that satisfies the conditions of the Bochner–Minlos Theorem. As a consequence

$$C\{s\} = \int e^{i\int s(t)\, x(t)\, dt}\, d\mu_W(x) \ ,$$

where $\mu_W(x)$ denotes the measure on Wiener process paths. Formally, we can relate this measure to the formal probability density derived above by

$$C\{s\} = \mathcal{M} \int e^{i\int s(t)\, x(t)\, dt}\, e^{-\frac{1}{2}\int \dot{x}^2(t)\, dt}\, \mathcal{D}x \ .$$

Although the separate ingredients in this equation are not well defined, when the expression is taken as a whole it is meaningful (much like dy/dx).

4.3.1 General Wiener process

The standard Wiener process we have considered so far can be taken as the basis of a general Wiener process, which we denote by $X(t)$, namely,

$$X(t) = \sqrt{\nu}\, W(t - t_0) + x(t_0) \ .$$

Observe that this general process: (i) requires that $t \geq t_0$, (ii) starts at the fixed position $X(t_0) = x(t_0)$, (iii) has a mean $\langle X(t) \rangle = x(t_0)$, and thus (iv) has a variance given by

$$\langle X(t)\, X(u) \rangle^c = \langle X(t)\, X(u) \rangle - \langle X(t) \rangle \langle X(u) \rangle = \nu\, \min(t - t_0, u - t_0) \ ,$$

where the parameter ν is called the *diffusion constant*.

4.3.2 Pinned Brownian motion

Let us start by illustrating pinned Brownian motion using a standard Wiener process [IMc65]. We know that a standard Wiener process starts at $t = 0$ at $W(0) = 0$; in the present context, we seek to distort the paths so that they all pass through a fixed point, here chosen as $x = 0$, at a fixed future time, taken to be $t = 1$. This process leads to what is commonly called a *Brownian bridge*. To this end, we define

$$B(t) = W(t) - t\,W(1)\ .$$

Clearly, $B(0) = 0$ as well as $B(1) = 0$. Moreover,

$$\langle B(t_1)\,B(t_2)\rangle = \langle [W(t_1) - t_1\,W(1)][W(t_2) - t_2\,W(1)]\rangle = t_<(1 - t_>)\ ,$$

where $t_< \equiv \min(t_1, t_2)$ and $t_> \equiv \max(t_1, t_2)$.

We can also force the generalized Wiener process $X(t)$ to take the form of a bridge process as well. For that purpose, and $t_0 \le t \le t_1$, one need only consider

$$B_X(t) = X(t) - (t - t_0)/(t_1 - t_0)\,[X(t_1) - x(t_1)]\ ,$$

a process which obeys $B_X(t_0) = x(t_0)$ and $B_X(t_1) = x(t_1)$.

It is noteworthy that all Brownian bridges are Gaussian processes, being simply a linear combination of Gaussian variables. Indeed, if $d\mu^{\nu}_{x'}(x)$ denotes the differential of a normalized measure for a generalized Wiener process that starts at $X(0) = x'$ at $t = 0$ and has diffusion constant ν, then

$$d\mu^{\nu}_{x'',x'}(x) = \delta(x(T) - x'')\,d\mu^{\nu}_{x'}(x)$$

denotes the (unnormalized) measure for a Brownian bridge connecting $x(t_0) = x'$ and $x(T) = x''$, $T > 0$. This sort of measure will be of fundamental importance in the next section.

4.3.3 Generalized Brownian bridges

Suppose we wanted to force not only $B(1) = 0$ but in addition $B(2) = 0$, along with the initial requirement that $B(0) = 0$. This kind of process could be constructed as

$$E(t) = W(t) - t(2 - t)W(1) - \tfrac{1}{2}t(t - 1)W(2)\ .$$

We leave it as an exercise to work out the variance for this process.

This procedure may be generalized to many pinnings by

$$Z(t) = W(t) - \sum_{j=1}^{J} c_j(t; t_1, t_2, \ldots, t_J)\,W(t_j)\ .$$

Here,

$$c_j(t; t_1, t_2, \ldots, t_J) \equiv \prod_{l=1\,(\neq j)}^{J} (t - t_l)/(t_j - t_l) \; ,$$

with the property that

$$c_j(t_p; t_1, t_2, \ldots, t_J) = \delta_{j,p} \; ;$$

this choice leads to the desired pinning, namely,

$$Z(t) = 0 \; , \qquad t = t_1, t_2, \ldots, t_J \; , \qquad 0 < t_1 < t_2 < t_3 \ldots < t_J \; .$$

It seems natural to refer to such processes as *multispan Brownian bridges!*

4.3.4 Alternative Brownian bridges

Our definition of a Brownian bridge is not uniquely fixed by the requirement that $B(0) = B(1) = 0$ since we could also have chosen

$$B'_r(t) = W(t) - t^r \, W(1) \; , \qquad r > 0, \; r \neq 1 \; .$$

However, these modified processes lack a kind of time-reversal invariance that the process $B(t)$ exhibits. In particular, in the time interval $0 \leq t \leq 1$, it is clear that the process $B(1 - t)$ is stochastically equivalent to the process $B(t)$ within the same time interval. Such an equivalence is not present in the alternative Brownian bridges $B'_r(t)$, $r > 0$, $r \neq 1$. Thus, although they are potential alternatives, they would be useful only in special circumstances where time-reversal invariance is not an issue. Similar constructions exist for the generalized Brownian bridges as well, and the same remark about a kind of time-reversal invariance applies as well.

Once again we observe that all forms of generalized Brownian bridges are Gaussian stochastic processes that are fully determined by their mean and covariance functions.

4.4 The Feynman–Kac Formula

Consider the functional Fourier transform of a pinned Wiener measure with a diffusion constant $\nu = m^{-1}$, given by

$$\int e^{i \int_0^T s(t) x(t) \, dt} \; d\mu^{m^{-1}}_{x'',x'}(x)$$
$$= \mathcal{N} \int \delta(x'' - x(T)) \, e^{i \int_0^T s(t) x(t) \, dt - (m/2) \int_0^T \dot{x}^2 \, dt} \; \mathcal{D}x \; .$$

In general, this expression is *not* a characteristic functional since, for $s(t) = 0$,

$$K(x'', T; x', 0) = \int d\mu^{m^{-1}}_{x'',x'}(x) = (1/\sqrt{2\pi T}\,)\; e^{-m(x''-x')^2/2T}\;.$$

In spite of K not being normalized this expression is valuable in another right since it satisfies the differential equation

$$\frac{\partial}{\partial t} K(x,t;x',0) = \frac{1}{2m}\,\frac{\partial^2}{\partial x^2}\,K(x,t;x',0)\;,$$

which is recognized as a diffusion equation, or as a Schrödinger equation for a free particle of mass m for imaginary time. Moreover, this solution to the diffusion equation satisfies the initial boundary condition that

$$\lim_{T\to 0} K(x'', T; x', 0) = \delta(x'' - x')\;,$$

a boundary condition that leads to the fact that K is often called the *fundamental solution* or a *Green's function* for the given differential equation [Sim79].

Viewed as a solution to the free particle Schrödinger equation (for imaginary time), the question naturally arises: what about a similar equation for a *non*free particle, i.e., a particle that is also in the presence of a potential $V(x)$?

The expression of interest obeys the following chain of equations:

$$\begin{aligned}
K(x'', T; x', 0) &= \mathcal{N} \int \delta(x'' - x(T))\; e^{i\int_0^T s(t)\,x(t)\,dt} \\
&\qquad \times e^{-\int_0^T V(x)\,dt\; -\; (m/2)\int_0^T \dot{x}^2\,dt}\;\mathcal{D}x \\
&= \int e^{i\int_0^T s(t)\,x(t)\,dt}\; e^{-\int_0^T V(x)\,dt}\; d\mu^{m^{-1}}_{x'',x'} \\
&= e^{-\int_0^T V(\delta/i\delta s(t))\,dt}\; \int e^{i\int_0^T s(t)\,x(t)\,dt}\; d\mu^{m^{-1}}_{x'',x'} \\
&= e^{-\int_0^T V(\delta/i\delta s(t))\,dt}\; e^{i\int_0^T s(t)[(1-t/T)x' + (t/T)x'']\,dt} \\
&\qquad \times e^{-(1/2m)\int_0^T\int_0^T s(t_1)\,s(t_2)[t_<(1-t_>/T)]\,dt_1\,dt_2}\;.
\end{aligned}$$

In particular, this function satisfies the differential equation

$$\frac{\partial}{\partial t} K(x,t;x',0) = \left[\frac{1}{2m}\frac{\partial^2}{\partial x^2} - V(x) + is(t)x\right] K(x,t;x',0)\;,$$

subject to the boundary condition

$$\lim_{t\to 0} K(x,t;x',0) = \delta(x - x')\;.$$

4.5 Ornstein–Uhlenbeck Process

The Ornstein–Uhlenbeck (O-U) process $U(t)$ is another Gaussian process that is intimately related to the Wiener process, but it differs in several important ways [Hid70]. For one thing, for the O-U process, the time variable t runs over the whole real line, $-\infty < t < \infty$. As such, it is possible to consider a stationary process, and it is customary to consider a stationary O-U process. There are several ways to characterize the O-U process, and the first procedure is as follows.

As before, let $W(t)$, $0 \le t < \infty$, denote a standard Wiener process which is a Gaussian process with (i) $W(0) = 0$, (ii) $\langle W(t) \rangle = 0$, and (iii) $\langle W(t) W(u) \rangle = \min(t, u)$. In that case, we define the O-U process $U(t)$, $-\infty < t < \infty$, by

$$U(t) \equiv (1/\sqrt{2})\, e^{-t}\, W(e^{2t}) \,.$$

It follows from this definition that $U(t)$ is a Gaussian process with

$$\langle U(t) \rangle = 0 \,,$$

and

$$\begin{aligned} \langle U(t_1) U(t_2) \rangle &= \frac{1}{2}\, e^{-(t_< + t_>)}\, \langle W(e^{2t_1}) W(e^{2t_2}) \rangle \\ &= \frac{1}{2}\, e^{-(t_< + t_>)}\, e^{2t_<} \\ &= \frac{1}{2}\, e^{-|t_1 - t_2|} \\ &= \frac{1}{2\pi} \int e^{i\omega(t_1 - t_2)}\, \frac{d\omega}{1 + \omega^2} \,. \end{aligned}$$

The fact that $\langle U(t_1) U(t_2) \rangle$ depends on the difference $t_1 - t_2$ is a guarantee that $U(t)$ is a stationary process. Since

$$\langle U(t)^2 \rangle = \frac{1}{2\pi} \int \frac{d\omega}{(1 + \omega^2)} = \frac{1}{2} < \infty \,,$$

it follows that the O-U process is pointwise defined. Moreover, since the integrand for the two-point function is an absolutely continuous measure, the O -U process is *ergodic*.

The characteristic functional for an O-U process is readily determined since it is a Gaussian process. Therefore,

$$\begin{aligned} \langle e^{i \int s(t) U(t)\, dt} \rangle &= e^{i \int s(t) \langle U(t) \rangle\, dt \,-\, \frac{1}{2} \int s(t)\, s(u)\, \langle U(t) U(u) \rangle^c\, dt\, du} \\ &= e^{-\frac{1}{4} \int s(t)\, e^{-|t-u|}\, s(u)\, dt\, du} \,, \end{aligned}$$

expressions in which all integrations extend from $-\infty$ to $+\infty$. Expansion of both sides in powers of s leads to various correlation functions of the O-U process.

There are additional ways of writing the characteristic functional for the O-U process that are informative. In particular, it follows that

$$\begin{aligned}\langle e^{i\int s(t)U(t)\,dt}\rangle &= e^{-\frac{1}{2}\int |\tilde{s}(\omega)|^2/(1+\omega^2)\,d\omega} \\ &= \mathcal{N}' \int e^{i\int \tilde{s}(\omega)^*\tilde{x}(\omega)\,d\omega - \frac{1}{2}\int (1+\omega^2)|\tilde{x}(\omega)|^2\,d\omega}\,\mathcal{D}\tilde{x} \\ &= \mathcal{N} \int e^{i\int s(t)x(t)\,dt - \frac{1}{2}\int [\dot{x}(t)^2 + x(t)^2]\,dt}\,\mathcal{D}x\ .\end{aligned}$$

The last two expressions are formal, but instructive nonetheless. Neither integral in the second part of the last exponent is well defined: the first fails because the paths are nowhere differentiable, and both integrals diverge because of the infinite range of the integration. Nevertheless, there is a proper probability measure $\mu_{O-U}(x)$ on continuous but nowhere differentiable paths such that

$$C_{O-U}(s) \equiv \langle e^{i\int s(t)U(t)\,dt}\rangle = \int e^{i\int s(t)x(t)\,dt}\,d\mu_{O-U}(x)\ .$$

4.5.1 Addition of a potential to an O-U process

The path integral expressions just given lend themselves naturally to the inclusion of an additional potential $V(x)$ [Sim79]. Specifically, the addition of such a potential leads to

$$\mathcal{N} \int e^{i\int s(t)x(t)\,dt - \frac{1}{2}\int [\dot{x}(t)^2 + x(t)^2 + V(x(t))]\,dt}\,\mathcal{D}x\ ,$$

a formal expression that can also be written in the form

$$\begin{aligned}&e^{-\int V(\delta/i\delta s(t))\,dt}\,\mathcal{N} \int e^{i\int s(t)x(t)\,dt - \frac{1}{2}\int [\dot{x}(t)^2 + x(t)^2]\,dt}\,\mathcal{D}x \\ &\quad = e^{-\int V(\delta/i\delta s(t))\,dt}\, e^{-\frac{1}{4}\int s(t)e^{-|t-u|}\,s(u)\,dt\,du}\ .\end{aligned}$$

Neither of the last two expressions entails proper normalization. An improved analysis begins by limiting the integration range for the added potential, i.e., by replacing $\int V(x(t))\,dt$ by

$$\int_{-T}^{T} V(x(t))\,dt\ ,$$

and then reserving a limit in which $T \to \infty$ for the final step in the calculation.

4.6 Realization of a General Gaussian Process

Let $\{u_n(t)\}_{n=1}^{\infty}$ denote a set of real, linearly independent functions over the interval $0 < t < \infty$ or $-\infty < t < \infty$, as the case may be. Next, let $\{X_n\}_{n=1}^{\infty}$

denote a set of independent, identical, standard Gaussian random variables with mean and variance given by

$$\langle X_n \rangle = 0 \ , \qquad \langle X_m X_n \rangle = \delta_{mn} \ .$$

Finally, let $m(t)$ be a general real function. Then we define the stochastic variable

$$X(t) = \textstyle\sum_{n=1}^{\infty} X_n \, u_n(t) + m(t) \ ,$$

leading to a Gaussian process $X(t)$ for which

$$\langle X(t) \rangle = m(t) \ ,$$

and

$$\begin{aligned} \langle X(t_1) X(t_2) \rangle^c &= \textstyle\sum_{m,n=1}^{\infty} u_m(t_1) \, u_n(t_2) \, \langle X_m X_n \rangle = \sum_{n=1}^{\infty} u_n(t_1) \, u_n(t_2) \\ &\equiv \ G_2(t_1, t_2) \ . \end{aligned}$$

As an example of this construction, let $\{h_n(t)\}_{n=1}^{\infty}$ denote an orthonormal set of functions for $t \geq 0$. In particular, we have

$$\begin{aligned} \int_0^{\infty} h_m(t) h_n(t) \, dt \ &= \delta_{mn} \ , \\ \textstyle\sum_{n=1}^{\infty} h_n(t_1) h_n(t_2) \ &= \delta(t_1 - t_2) \ . \end{aligned}$$

Next, we choose

$$u_n(t) = \int_0^t h_n(s) \, ds \ .$$

Then it follows that

$$W(t) = \sum_{n=1}^{\infty} X_n u_n(t) = \sum_{n=1}^{\infty} X_n \int_0^t h_n(s) \, ds \ ,$$

leading to $\langle W(t) \rangle = 0$ and

$$\begin{aligned} \langle W(t_1) W(t_2) \rangle \ &= \sum_{n=1}^{\infty} \int_0^{t_1} h_n(t_1') \, dt_1' \int_0^{t_2} h_n(t_2') \, dt_2' \\ &= \int_0^{t_1} \int_0^{t_2} \delta(t_1' - t_2') \, dt_1' \, dt_2' \\ &= \int_0^{\min(t_1, t_2)} dt = \min(t_1, t_2) \ . \end{aligned}$$

Thus, as the notation suggests, $W(t)$, as defined above, is a realization of a standard Wiener process.

An O-U process can also be described in another way. In this case we let $\{h_n(\omega)\}_{n=1}^{\infty}$, where we require that $h_n(-\omega) = h_n^*(\omega)$, denote a set of complex orthonormal functions on the real line, satisfying

$$\sum_{n=1}^{\infty} h_n(\omega)\, h_n^*(\omega') = \delta(\omega - \omega') \ .$$

Let

$$u_n(t) \equiv \frac{1}{\sqrt{2\pi}} \int \frac{h_n(\omega)}{\sqrt{1+\omega^2}}\, e^{i\omega t}\, d\omega \ ,$$

from which we deduce that $u_n^*(t) = u_n(t)$ for all t. Then we define

$$U(t) = \sum_{n=1}^{\infty} X_n\, u_n(t) \ ,$$

which leads to $\langle U(t) \rangle = 0$ as well as

$$\begin{aligned} \langle U(t_1) U(t_2) \rangle &= \sum_{n=1}^{\infty} u_n(t_1)\, u_n(t_2) \\ &= \frac{1}{2\pi} \int \frac{e^{i\omega(t_1 - t_2)}}{1+\omega^2}\, d\omega \\ &= \tfrac{1}{2}\, e^{-|t_1 - t_2|} \ , \end{aligned}$$

as desired from the representative of an O-U process.

4.7 Generalized Stochastic Process

4.7.1 Gaussian white noise

Consider the representation of a Wiener process given by

$$W(t) = \sum_{n=1}^{\infty} X_n \int_0^t h_n(t')\, dt' \ .$$

Now let us take a formal time derivative of this expression, which leads to

$$N(t) \equiv \frac{dW(t)}{dt} = \sum_{n=1}^{\infty} X_n\, h_n(t) \ .$$

The result is a Gaussian process, known as *white noise* [Hid70], and the mean and variance of the process $N(t)$ are $\langle N(t) \rangle = 0$ and

$$\begin{aligned} \langle N(t_1) N(t_2) \rangle &= \sum_{m,n=1}^{\infty} h_m(t_1) h_n(t_2)\, \langle X_m X_n \rangle \\ &= \sum_{n=1}^{\infty} h_n(t_1) h_n(t_2) \\ &= \delta(t_1 - t_2) \ . \end{aligned}$$

Two points about the expression for the variance are in order. For $t_1 \neq t_2$, the vanishing of the variance signifies that the process $N(t)$ is statistically

independent for different values of t. For $t_1 = t_2$, the variance diverges, which implies that white noise is not pointwise defined, but rather that $N(t)$ is a generalized stochastic process, with distributional values, which requires smearing to be well-defined.

Let $s(t)$ denote a smooth function, e.g., an element that is C_0^∞. We define

$$N(s) \equiv (s, N) \equiv \int s(t)\, N(t)\, dt \ .$$

It follows that $\langle N(s) \rangle = 0$ and that

$$\langle N(s_1) N(s_2) \rangle = (s_1, s_2) \equiv \int s_1(t)\, s_2(t)\, dt \ .$$

Moreover, the characteristic functional

$$\langle e^{i \int s(t) N(t)\, dt} \rangle = e^{-\frac{1}{2} \int s(t)^2\, dt} \ ,$$

based on the fact that (s, N) is a normal random variable. The characteristic functional satisfies the conditions of the Bochner–Minlos Theorem, and therefore is the functional Fourier transform of a genuine measure. In particular,

$$\begin{aligned} e^{-\frac{1}{2} \int s(t)^2\, dt} &= \mathcal{N} \int e^{i \int s(t) x(t)\, dt}\, e^{-\frac{1}{2} \int x(t)^2\, dt}\, \mathcal{D}x \\ &= \int e^{i \int s(t) x(t)\, dt}\, d\mu_{WN}(x) \ . \end{aligned}$$

From our earlier discussion we may be sure that the support of the measure μ_{WN} is much larger than $L^2(\mathbb{R})$. In addition, it is true that the characteristic functional is well-defined for all $s \in L^2(\mathbb{R})$. Thus there is an apparent conflict with the allowed arguments of the characteristic functional and the support of the white noise measure. The resolution of this conundrum lies in the observation that the integral representation of the characteristic functional is well-defined for all s in some test function space (e.g., $\mathcal{D}$ composed of C_0^∞ functions), but the *result* of that integral representation admits a larger space of definition (L^2) than the integral representation itself. [**Remark:** An elementary example of this state of affairs is given by the equation

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots \ .$$

This equation is valid for all x such that $|x| < 1$, which is the full range for which the right-hand side is well-defined. However, the left-hand side is well defined everywhere—except at $x = 1$—illustrating that each side of an equation may have a distinct domain of validity.]

Although a Wiener process is defined only for $t \geq 0$, the white noise process can be considered to be defined for all time. In view of the expression for the

variance of white noise, it is common to regard white noise as a stationary process as well. Hence the integrals in the expressions for the characteristic function can be taken to run from $-\infty$ to $+\infty$.

We have used Gaussian white noise to illustrate a generalized stochastic process, and indeed, it is generally regarded as the most important such process. However, any stochastic process which is not pointwise defined and thus requires to be regarded as a distribution-valued stochastic process fits into this category. One closely related set of examples involves so-called *colored noise*. Colored noise may be obtained from white noise by filtering. Specifically, let us define (here C stands for "colored")

$$C(t) = \int Y(t-\tau)\, N(\tau)\, d\tau \ ,$$

where $Y(t)$ denotes the impulse response function for a suitable idealized filter. In particular, we let

$$Y(t) = \frac{1}{\sqrt{2\pi}} \int \tilde{Y}(\omega)\, e^{i\omega t}\, d\omega \ ,$$

and it then follows that

$$\langle C(t_1) C(t_2) \rangle = \frac{1}{2\pi} \int |\tilde{Y}(\omega)|^2\, e^{i\omega(t_1 - t_2)}\, d\omega \ .$$

The general class of such processes is referred to as colored noise, colored because the integrand $|\tilde{Y}(\omega)|^2$ generally weights some frequencies ("colors") differently from others. White noise is so-called because $|\tilde{Y}(\omega)|^2 = 1$, and thus all frequencies are weighted equally. If $\int |\tilde{Y}(\omega)|^2\, d\omega < \infty$, we are dealing with a stochastic process which is pointwise defined; if instead, $\int |\tilde{Y}(\omega)|^2\, d\omega = \infty$, then we are dealing with a generalized stochastic process.

4.8 Stochastic Differential Equations a.k.a. Langevin Equations

Recall our definition of an Ornstein–Uhlenbeck (O-U) process given by

$$U(t) = (1/\sqrt{2}\,)\, e^{-t}\, W(e^{2t}) \ .$$

Let us take the differential of this expression with respect to t. It follows that

$$dU(t) = -U(t)\, dt + (1/\sqrt{2})\, e^{-t}\, dW(e^{2t}) \ ,$$

and we turn our attention to the last term in this relation. We know that $dW(t) = N(t)\, dt$, and therefore

$$dW(e^{2t}) = 2\, N(e^{2t})\, e^{2t}\, dt \ .$$

Clearly $\langle N(e^{2t})\rangle = 0$ and

$$\begin{aligned} 2\langle e^{t_1}\, N(e^{2t_1})\, e^{t_2}\, N(e^{2t_2})\rangle &= 2e^{t_1+t_2}\,\delta(e^{2t_1} - e^{2t_2}) \\ &= e^{t_1+t_2}\, e^{-2t_1}\delta(t_1 - t_2) \\ &= \delta(t_1 - t_2)\ . \end{aligned}$$

Since this variance is the same as that of white noise itself, it follows that

$$\sqrt{2}\, N(e^{2t})\, e^{t} \simeq N'(t)\ ,$$

namely, the left-hand side is stochastically equivalent to white noise itself. Dropping the prime, we are led to the fundamental equation among differentials given by

$$dU(t) = -U(t)\, dt + N(t)\, dt\ ,$$

or, as is more commonly stated,

$$dU(t) = -U(t)\, dt + dW(t)\ .$$

This is an example of a *stochastic differential equation* [IMc65]. As with all expressions involving differentials, we give meaning to this relation by integration, as we shall do in the next paragraph. However, before we leave this equation, consider dividing it by dt, which then leads to the relation

$$\frac{dU(t)}{dt} = -U(t) + N(t)\ .$$

In this form the expression is called a Langevin equation [CKW]. Viewed pointwise it is ill-defined; however, if this relation is regarded as a distribution, then smearing with a test function leads to a meaningful expression.

Regarding the stochastic differential equation as a conventional differential equation, we can attempt to solve it by traditional means. We are then led to the solution

$$U(t) = e^{-t}\, U(0) + e^{-t} \int_0^t e^{t'}\, dW(t')\ .$$

Thus our solution depends on what we choose for $U(0)$. If we seek a stationary solution, then we should set

$$U(0) = \int_{-\infty}^{0} e^{t'}\, dW(e^{t'})\ ,$$

and for our final expression for the solution we adopt

$$U(t) = e^{-t} \int_{-\infty}^{t} e^{t'}\, dW(t')\ .$$

In writing these expressions, we have used the fact that although $W(t)$ cannot be a stationary process, $dW(t)$ can be taken as a stationary process.

It must be admitted that the resultant expression does not appear to be the same as our original definition of an O-U process. Could it be stochastically equivalent? To check this possibility, consider the proposed variance

$$\begin{aligned}\langle U(t_1)U(t_2)\rangle &= e^{-(t_1+t_2)}\int_{-\infty}^{t_1}\int_{-\infty}^{t_2} e^{t'+t''}\,\delta(t'-t'')\,dt'\,dt'' \\ &= e^{-(t_1+t_2)}\int_{-\infty}^{\min(t_1,t_2)} e^{2t'}\,dt' \\ &= \tfrac{1}{2}e^{-|t_1-t_2|}\,.\end{aligned}$$

Thus we have indeed established that the solution to the stochastic differential equation is stochastically equivalent to our original definition of the O-U process!

4.9 Poisson Process

Up to this point we have focused heavily on Gaussian processes of one or another kind. This is perfectly appropriate since Gaussian processes are generally regarded as the most important cases. However, there is another class of processes, the so-called *Poisson processes*, that we wish to discuss briefly here. Just as the Wiener process is fundamental to the Gaussian family of stochastic processes, the *shot noise* stochastic process is basic to the family of Poisson processes [Luk70].

A shot noise stochastic process $S(t)$ may be defined as the limit of a sequence of processes defined by

$$S_N(t) \equiv \sum_{n=1}^{N} X_n\,\delta(t-\Upsilon_n)\,,$$

where $\{X_n\}_{n=1}^N$ are independent identically distributed, random variables, with a measure μ_N, and $\{\Upsilon\}_{n=1}^N$ are likewise independent, identically distributed, random variables conveniently taken to be uniformly distributed in a finite interval of time T. It follows therefore that the characteristic functional for S_N is given by

$$\begin{aligned}\langle e^{i\int s(t)\,S_N(t)\,dt}\rangle &= \langle e^{i\,\Sigma_{n=1}^N X_n\,s(\Upsilon_n)}\rangle \\ &= \langle e^{iX\,s(\Upsilon)}\rangle^N \\ &= \left[\int e^{ix\,s(t)}\,d\mu_N(x)\,dt/T\right]^N \\ &= \left[1-\int[1-e^{ix\,s(t)}]\,d\mu_N(x)\,dt/T\right]^N\,.\end{aligned}$$

We are now in a position to take a limit in which $N \to \infty$, $T \to \infty$ in such a way that $N/T \equiv k$ remains finite and constant. This limit defines the shot noise stochastic process $S(t)$, and the resultant characteristic functional is given by

$$\langle e^{i\int s(t)\,S(t)\,dt} \rangle = e^{im\int s(t)\,dt \,-\, k\int dt \int [1 - e^{ixs(t)} + ixs(t)/(1+x^2)]\,d\sigma(x)} ,$$

where the first integral runs over an infinite time interval (usually taken to be $-\infty$ to ∞), and the nonnegative measure $\sigma(x)$ is required to satisfy

$$\int [x^2/(1+x^2)]\,d\sigma(x) < \infty .$$

There are three distinct forms for σ and they have different names: (i) if σ is a point measure, i.e., $d\sigma(x) = c\delta(x - x_0)\,dx$, $c > 0$, then $S(t)$ is called a Poisson stochastic process; (ii) if σ is not a point measure but $\int d\sigma(x) < \infty$, then $S(t)$ is called a compound Poisson process; and (iii) if $\int d\sigma(x) = \infty$, then $S(t)$ is called a generalized Poisson process. It is clear that $m = \langle S(t) \rangle$. Moreover, it follows from the generating functional, for $p \geq 2$, that

$$\langle S(t_1)\,S(t_2) \cdots S(t_p) \rangle^c = C_p\,\delta(t_1 - t_2)\delta(t_2 - t_3) \cdots \delta(t_{p-1} - t_p) ,$$

a situation which explains why such a process is called a *δ-correlated process*. The coefficients

$$C_p \equiv \int x^p\,d\sigma(x) ,$$

provided such integrals exist.

In many ways shot noise plays the role for Poisson processes that white noise plays for Gaussian processes; indeed, observe that the correlation function

$$\langle S(t_1)\,S(t_2) \rangle \propto \langle N(t_1)\,N(t_2) \rangle .$$

Let us also define a stochastic process

$$P(t) \equiv \int_0^t S(t')\,dt' ,$$

which can be formally restated as

$$P(t) = \sum_{n=1}^{\infty} X_n\,H(t - \Upsilon_n) ,$$

where $H(t) \equiv 1$ for $t \geq 0$ and $H(t) \equiv 0$ for $t < 0$ [the function $H(t)$ defined this way is known as the Heaviside function; see Sec. 2.1.1]. In particular, for $t_2 > t_1$, it follows that

$$\langle e^{is[P(t_2) - P(t_1)]} \rangle = e^{-k\,|t_2 - t_1| \int [1 - e^{isx}]\,d\sigma(x)} ,$$

and therefore

$$\langle [P(t_2) - P(t_1)]^l \rangle = A_1 |t_2 - t_1| + \cdots + A_l |t_2 - t_1|^l \ ,$$

where the values of the positive coefficients A_r, $1 \le r \le l$, are not important save to note that the appearance of the term $|t_2 - t_1|$ implies that the paths $P(t)$ are *not* continuous paths. Moreover, for $t_4 > t_3 \ge t_2 > t_1$, we have

$$\langle e^{is[P(t_4) - P(t_3)] + iu[P(t_2) - P(t_1)]} \rangle$$
$$= \exp\{-|t_4 - t_3| \int [1 - \cos(sx)] d\sigma(x) - |t_2 - t_1| \int [1 - \cos(ux)] d\sigma(x)\} \ ,$$

from which we learn that [like the Wiener process $W(t)$] the process $P(t)$ has independent increments.

As a simple example of a compound Poisson process, we choose

$$d\sigma(x) = \tfrac{1}{2}[\delta(x-1) + \delta(x+1)] \ .$$

In this case a sample path is of the form

$$p(t) = \sum_{n=1}^{\infty} x_n \, H(t - \tau_n) \ ,$$

where $x_n = \pm 1$ with equal probability and the times τ_n are chosen randomly with a uniform density of $k > 0$. Thus the sample paths are strictly piecewise constant. For this process, it follows that

$$\langle e^{is[P(t_2) - P(t_1)]} \rangle = e^{-|t_2 - t_1| \{1 - \cos(s)\}} \ ,$$

and in this case we learn that

$$\langle [P(t_2) - P(t_1)]^{2l} \rangle^c = |t_2 - t_1| \ .$$

Exercises

4-1 For a normal stochastic process, the characteristic functional is given by

$$C\{s\} = \int e^{i \int s(t) x(t) \, dt} \, d\mu(x) = e^{-\frac{1}{2} \int \int s(t) G(t-u) \, s(u) \, dt \, du} \ ,$$

where

$$G(t-u) = \int e^{i(t-u)\omega} \, dm(\omega) \ .$$

Derive the condition on the measure $m(\omega)$ in order that the process is also ergodic.

4-2 The process

$$E(t) = W(t) - t(2-t)W(1) - \tfrac{1}{2}t(t-1)W(2)$$

satisfies the pinning $E(0) = E(1) = E(2) = 0$. Determine the variance of the process $E(t)$.

4-3 Consider the stochastic differential equation given by

$$dG(t) = -G(t)^3\,dt + dW(t)\ .$$

Find the partial differential equation that determines the distribution $p(g,t)$, where

$$p(g,t) \equiv \langle \delta(g - G(t)) \rangle\ .$$

Part II

Quantum Theory

5

Background to an Analysis of Quantum Mechanics

5.1 Hilbert Space and Operators: Basic Properties

5.1.1 Hilbert space

The quantum analog of the classical phase space is a complex Hilbert space, and the quantum analog of a phase-space point is a vector lying in this Hilbert space. A complex Hilbert space has many properties in common with a complex, finite-dimensional, Euclidean space. In particular, such a space is a linear vector space with an inner product. In the elegant notation of Dirac [Dir74], vectors are denoted by "kets," such as $|\psi\rangle$, $|\phi\rangle$, etc. These vectors lie in an abstract Hilbert space $\mathfrak{H}$, as does their sum with arbitrary complex coefficients,

$$a|\psi\rangle + b|\phi\rangle \in \mathfrak{H} , \qquad a, b \in \mathbb{C} .$$

Every linear vector space has a unique zero vector $0 \in \mathfrak{H}$ which has the property that $|\phi\rangle + 0 = |\phi\rangle$ and $0|\phi\rangle = 0$; in this last expression, the first $0 \in \mathbb{C}$ while the second $0 \in \mathfrak{H}$. To each "ket" vector $|\chi\rangle$ there is associated an adjoint form of the vector denoted by $\langle\chi|$, which is referred to as a "bra" vector. The adjoint vector for the given sum is $a^*\langle\psi| + b^*\langle\phi|$ and involves the complex conjugate of the coefficients a and b. The inner product of two vectors involves a bra, say $\langle\chi|$, and a ket, say $|\phi\rangle$, and is a complex number denoted by $\langle\chi|\phi\rangle \in \mathbb{C}$, which is sometimes referred to as a "bra-ket." The inner product satisfies $\langle\phi|\chi\rangle = \langle\chi|\phi\rangle^*$, where * denotes the complex conjugate. Furthermore, the inner product is linear in the right-hand argument, namely,

$$\langle\chi|(a|\psi\rangle + b|\phi\rangle) = a\langle\chi|\psi\rangle + b\langle\chi|\phi\rangle .$$

It follows that the inner product of any vector with the zero vector vanishes (set $b = 0$ in the preceding expression). Hilbert spaces have a positive-definite inner product, which means that the inner product satisfies $0 \leq \langle\phi|\phi\rangle < \infty$ for any $|\phi\rangle \in \mathfrak{H}$; furthermore, the condition $\langle\phi|\phi\rangle = 0$ holds if and only if $|\phi\rangle = 0$, the zero vector.

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9_5,

One advantage of the Dirac notation is that the unadorned symbol $|\ \rangle$ already denotes a ket vector, and thus one is free to place any label inside one likes. Thus we may use $|0\rangle, |n\rangle, |\#\rangle$, etc., all of which denote vectors with one or another label. Normally, the meaning of the label is clear from the context; otherwise it should be explained. The length, or norm, of a vector $|\psi\rangle$ is given by $\| \, |\psi\rangle \, \| \equiv +\sqrt{\langle\psi|\psi\rangle}$, which is always finite. Schwarz's inequality states that $|\langle\phi|\psi\rangle|^2 \leq \langle\phi|\phi\rangle\langle\psi|\psi\rangle$ with equality holding only if $|\phi\rangle = c|\psi\rangle$ for some complex coefficient c. A vector $|\psi\rangle$ is called normalized if $\langle\psi|\psi\rangle = 1$. Two vectors $|\psi\rangle$ and $|\phi\rangle$ are said to be orthogonal if $\langle\phi|\psi\rangle = 0$. A set of vectors $\{|n\rangle\}_{n=1}^{N}$ is called an orthonormal set if it satisfies the property that $\langle n|m\rangle = \delta_{nm}$ for all $1 \leq n, m \leq N,\ N \leq \infty$. Linear sums of the vectors in an N-dimensional orthonormal set span an N-dimensional space. If every vector in $\mathfrak{H}$ may be represented in this fashion, then the orthonormal set is called complete, or more simply a "basis"; it should be noted that there exist many (if $N > 1$, uncountably many!) distinct bases. The number N then becomes the dimension of the Hilbert space, and if $N = \infty$, then the Hilbert space is infinite dimensional. Many Hilbert spaces in quantum theory are infinite dimensional. In summary, for an infinite-dimensional Hilbert space, for example, there exists a basis (a complete orthonormal set) $\{|n\rangle\}_{n=1}^{\infty}$ with the property that every vector in $\mathfrak{H}$ admits a unique representation in the form

$$|\psi\rangle = \sum_{n=1}^{\infty} \psi_n \, |n\rangle, \qquad \psi_n \in \mathbb{C} \, .$$

Due to the orthonormality, it follows that $\psi_n = \langle n|\psi\rangle$. Of course, if the basis is changed, then the coefficients ψ_n are generally different. However, in any orthonormal basis, it is necessary that

$$0 \leq \langle\psi|\psi\rangle = \sum_{m,n=1}^{\infty,\infty} \psi_m^* \psi_n \langle m|n\rangle = \sum_{n=1}^{\infty} |\psi_n|^2 < \infty \, .$$

Thus the coefficients ψ_n, namely, the "coordinates" in the present basis, must be square summable, or as one says, $\{\psi_n\}_{n=1}^{\infty} \in l^2$, in order for the vector $|\psi\rangle \in \mathfrak{H}$. If $|\psi\rangle = 0$, then $\psi_n = 0$ for all $n \geq 1$. This example shows that we can "represent" each abstract Hilbert space vector $|\psi\rangle$ by means of a unique set of square summable complex numbers $\{\psi_n\}_{n=1}^{\infty}$, which are nothing but the coordinates of $|\psi\rangle$ with respect to the basis $\{|n\rangle\}$. If we represent $|\phi\rangle$ in the same basis by $\{\phi_n\}_{n=1}^{\infty}$, then it follows that the inner product is given by

$$\langle\phi|\psi\rangle = \sum_{n=1}^{\infty} \phi_n^* \psi_n \, .$$

An infinite-dimensional Hilbert space that is spanned by a countable number of basis vectors is called "separable."

Although we have indexed our components from $n = 1$ to ∞, there is nothing very special about such a choice. In other cases it may be more convenient

to index components from 0 to ∞, or even from $-\infty$ to ∞. Such changes would make evident modifications in the expressions already given. We begin with just such an example.

5.1.2 Fourier representation

Alternative representations for the vectors in $\mathfrak{H}$ can readily be given. One method for doing so is provided by the usual Fourier series. Let us introduce the transform pair given by[1]

$$f(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} e^{inx} \psi_n \ ,$$
$$\psi_n = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{-inx} f(x)\, dx \ .$$

It readily follows that

$$\langle \psi | \psi \rangle = \sum_{n=-\infty}^{\infty} |\psi_n|^2 = \int_{-\pi}^{\pi} |f(x)|^2\, dx \ ,$$

which asserts that the functions f are square integrable on the interval $-\pi$ to π, i.e., $f \in L^2([-\pi, \pi])$. For two such functions the inner product reads

$$\langle \phi | \psi \rangle = \sum_{n=-\infty}^{\infty} \phi_n^* \psi_n = \int_{-\pi}^{\pi} g(x)^* f(x)\, dx \ ,$$

where g is defined in terms of the coefficients ϕ_n in the same way that f is defined by the coefficients ψ_n.

5.1.3 L^2 representatives

Another set of examples is readily given as well. Let $\{h_n(x)\}_{n=0}^{\infty}$, $x \in \mathbb{R}$, denote a complete set of orthonormal functions on the real line, $-\infty < x < \infty$. When integrated over the real line such functions have the integral property that

$$\int_{-\infty}^{\infty} h_m(x)^* h_n(x)\, dx = \delta_{mn} \ ,$$

and the completeness property that

$$\sum_{n=0}^{\infty} h_n(x)^* h_n(y) = \delta(x - y) \ ,$$

[1] In general, the function $f(x)$ defined by the following series is only convergent in the mean.

where $\delta(x)$ again denotes the Dirac δ-function. The well-known Hermite functions, $h_n(x)$, $n \geq 0$, introduced in Sec. 3.5, and implicitly defined through the expression

$$\exp(-s^2 + 2sx - \tfrac{1}{2}x^2) = \pi^{\frac{1}{4}} \sum_{n=0}^{\infty} (n!)^{-\frac{1}{2}} (s\sqrt{2}\,)^n \, h_n(x) \ ,$$

provide a standard example. Using this or any other complete orthonormal set of functions, we can introduce a functional representation for the vector $|\psi\rangle$ according to the transform pair

$$\begin{aligned} \psi(x) \ &= \sum_{n=0}^{\infty} h_n(x) \, \psi_n \ , \\ \psi_n \ &= \int_{-\infty}^{\infty} h_n(x)^* \, \psi(x) \, dx \ . \end{aligned}$$

We note further that these relations imply that

$$\langle \psi | \psi \rangle = \sum_{n=0}^{\infty} |\psi_n|^2 = \int_{-\infty}^{\infty} |\psi(x)|^2 \, dx \ ,$$

showing that the functions $\psi(x)$ are square integrable on the real line, i.e., $\psi \in L^2(\mathbb{R})$. The inner product of two distinct vectors is given by a similar expression, namely,

$$\langle \phi | \psi \rangle = \sum \phi_n^* \psi_n = \int \phi(x)^* \psi(x) \, dx \ .$$

In the last expression we have left the limits of summation and the limits of integration implicit when they are sufficiently clear from the context. In the present case we say that $\psi \in L^2(\mathbb{R})$. Due to issues involving sets of measure zero, the elements of the Hilbert space are not specific functions but instead are *equivalence classes* of functions which differ from one another on a set of measure zero. For example, the zero vector is represented by a set of functions which includes, for example, (i) $f(x) = 0$, (ii) $f(x) = 0$ for $x \neq 3$ and $f(3) = 17$, (iii) $f(x) = 0$ for $x \neq 2$ and $x \neq 5$, with $f(2) = -4$ and $f(5) = \pi$, etc.

A completely similar representation holds for square-integrable functions $f(x)$ when $x \in \mathbb{R}^N$. To see this we need only generalize the orthonormal set of functions to a set $\{g_n(x)\}_{n=1}^{\infty}$ that satisfies

$$\begin{aligned} \int g_m(x)^* g_n(x) \, d^N x \ &= \delta_{mn} \ , \\ \sum_{n=1}^{\infty} g_n(y)^* g_n(x) \ &= \delta(y - x) \ . \end{aligned}$$

Here $\delta(x)$ denotes an N-dimensional δ-function with the property that it vanishes if $x \neq 0$ and $\int \delta(x) \, d^N x = 1$ when the point $x = 0$ is within the domain of integration. In the present case, the Hilbert space is represented by the space $L^2(\mathbb{R}^N)$, and the Hilbert space vectors are again represented by sets of functions that differ from one another only on sets of measure zero.

5.1.4 Segal–Bargmann representation

Let us illustrate yet another representation of the Hilbert space $\mathfrak{H}$. Let $z = x + iy \in \mathbb{C}$ denote a complex variable, and introduce the transform pair[2]

$$F(z) = \sum_{n=0}^{\infty} \frac{z^n \, \psi_n}{\sqrt{n!}} \, ,$$
$$\psi_n = \int \frac{z^{*n}}{\sqrt{n!}} \, F(z) \, e^{-|z|^2} \, \frac{dx \, dy}{\pi} \, ,$$

where the integration runs over the entire two-dimensional plane. It is straightforward to verify that

$$\langle \psi | \psi \rangle = \sum |\psi_n|^2 = \int |F(z)|^2 \, e^{-|z|^2} \, \frac{dx \, dy}{\pi} \, .$$

If for convenience we introduce the abbreviation

$$d\mu = e^{-|z|^2} \, \frac{dx \, dy}{\pi} \, ,$$

then we see that each of the analytic functions $F \in L^2(\mathbb{R}^2, d\mu)$. It follows that the inner product of two such analytic function representatives is given by

$$\langle \phi | \psi \rangle = \int G(z)^* \, F(z) \, d\mu \, .$$

This kind of example can be readily extended to analytic functions of several complex variables as well, and such representations are often referred to as Segal–Bargmann representations [Seg63, Bar61]. Observe that each vector of the Hilbert space is represented by a unique, analytic—hence continuous and infinitely differentiable—function. There is no need to deal with functions that differ on sets of measure zero in this representation.

5.1.5 Reproducing kernel Hilbert spaces

Another representation that involves Hilbert space vectors represented by continuous functions is that created with the help of a *reproducing kernel*. Let $\mathcal{L}$ represent an L-dimensional label space that is locally isomorphic to the space $\mathbb{R}^L$, and let $l \equiv \{l^1, l^2, \ldots, l^L\} \in \mathcal{L}$, where each $l^j \in \mathbb{R}$. Let $K(l'; l)$ be a function on $\mathcal{L} \times \mathcal{L}$ that is jointly continuous in the variables l' and l. Such a function is called a reproducing kernel if it is a function of positive type, i.e., specifically, if, for arbitrary complex coefficients $\{a_n\}_{n=1}^N$,

[2] Unlike the previous functional representations, the present sum converges absolutely for every vector in $\mathfrak{H}$.

$$\sum_{m,n=1}^{N,N} a_m^* a_n \, K(l_m; l_n) \geq 0 \; , \qquad \text{for all } \; N \, , \; 1 \leq N < \infty \; .$$

The celebrated GNS (Gel'fand, Naimark, Segal) Theorem [Emc74] then asserts that there exists an abstract, separable Hilbert space $\mathfrak{H}$ with elements $|\, l \rangle$, for all $l \in \mathcal{L}$, which are continuously labeled, i.e., if $l_\alpha \in \mathcal{L}$, $\alpha \in \{1, 2, 3, \ldots\}$, is a sequence of points such that $l_\alpha \to l$ as $\alpha \to \infty$, then $\| \, |l_\alpha\rangle - |l\rangle \, \| \to 0$ in the same limit. Moreover, the kernel $K(l'; l) \equiv \langle l' | l \rangle$; namely, it is given by the inner product of two such vectors.

With that background, we next define a functional representation of the abstract Hilbert space $\mathfrak{H}$ [Aro43]. Consider two vectors defined by

$$|\psi\rangle \; = \sum_{n=1}^{N} a_n \, |l_{(n)}\rangle \; , \qquad N < \infty \; ,$$

$$|\phi\rangle \; = \sum_{m=1}^{M} b_m \, |l_{[m]}\rangle \; , \qquad M < \infty \; .$$

We wish to give a *functional representation* of these abstract vectors, and the only set of vectors we know that span the Hilbert space is the set $\{|\, l\rangle\}$ for all $l \in \mathcal{L}$. Thus we define the functional representatives of these two abstract vectors as

$$\psi(l) \; \equiv \sum_{n=1}^{N} a_n \, \langle l | l_{(n)} \rangle \equiv \langle l | \psi \rangle \; ,$$

$$\phi(l) \; \equiv \sum_{m=1}^{M} b_m \, \langle l | l_{[m]} \rangle \equiv \langle l | \phi \rangle \; .$$

The inner product for vectors given by this functional representation is chosen to be

$$(\phi, \psi) \equiv \sum_{m,n=1}^{M,N} b_m^* a_n \, \langle l_{[m]} | l_{(n)} \rangle \equiv \langle \phi | \psi \rangle \; ,$$

which is just the same value of the inner product as in the abstract Hilbert space. With the so-chosen functional representatives and their inner product, we have defined the desired functional representation for a dense set of vectors since we still have the restriction that $N < \infty$. We complete the Hilbert space representation by taking the limit of Cauchy sequences as $N \to \infty$. In particular, a sequence $\{a_n\}_{n=1}^{\infty}$ is admissible provided that the limit $\| \, |\psi\rangle_N - |\psi\rangle_{N'} \, \| < \epsilon$ for any $\epsilon > 0$ for all $N, N' \geq P(\epsilon)$, where the notation $|\psi\rangle_N$ means the sum defining this vector has only N terms. The result of including the limit elements leads to a functional representation of the complete abstract Hilbert space by continuous functions given by

$$\psi(l) = \sum_{n=1}^{\infty} a_n \langle l | l_{(n)} \rangle$$

for suitable sequences $\{a_n\}$. In this representation, each vector in the abstract Hilbert space $\mathfrak{H}$ is represented by a unique continuous function. Observe carefully that everything about such a representation—namely, the functional representatives and the definition of the inner product—is completely defined in terms of the all-important reproducing kernel $K(l'; l) = \langle l' | l \rangle$.

This particular representation [Aro43] gets its name from the following fact. Let $b_1 = 1$ and all other $b_m = 0$. In that case

$$(\phi, \psi) = (K(\cdot, l_{[1]}), \psi) = \sum_{n=1}^{N} a_n \langle l_{[1]} | l_{(n)} \rangle = \psi(l_{[1]}) \; ;$$

in other words, the given inner product has "reproduced" the functional representative $\psi(l)$. [**Remark:** One might argue that the "function" $\delta(x-y)$ acts as a reproducing kernel in that $\int \delta(x-y)\, g(y)\, dy = g(x)$, but the difference lies in the fact that the function $K(l; l')$ is an element of the reproducing kernel Hilbert space while the "function" $\delta(x - y)$ is most certainly *not* an element of the space $L^2(\mathbb{R})$.]

The choice of which representation of a Hilbert space to use very much depends on what particular problem one is faced with. The most commonly used representation for quantum mechanics is undoubtedly the Schrödinger representation, namely, $L^2(\mathbb{R})$ or $L^2(\mathbb{R}^N)$. However, the author has found that reproducing kernel Hilbert spaces can also be very useful.

Some reproducing kernel Hilbert spaces—but not all!—admit a second definition for the inner product of the basic elements. The second definition we have in mind is given by

$$(\phi, \psi) = \int \phi^*(l)\, \psi(l)\, \delta l \; ,$$

namely, by a local integral over the space $\mathcal{L}$ with a positive, absolutely continuous measure $\delta l = \rho(l)\, d^L l$ with $\rho(l) > 0$ almost everywhere. When this case holds, it is common to refer to the set of states $\{|l\rangle\}$ as a set of *coherent states* and the representation they induce, $\psi(l) = \langle l | \psi \rangle$, as a *coherent state representation*; examples of coherent states and coherent state representations will be given in Chapters 7 and 8.

****

The preceding examples illustrate several different functional representations of an infinite-dimensional Hilbert space, and it is clear from their construction that they are all isomorphic to one another, i.e., they are all equivalent to one another under the given equations of transformation. There are, of course, still other functional representations, but the examples we have introduced above will serve our present purposes.

5.1.6 Operators for Hilbert space

Simply put, linear operators are elements of an associative algebra, and thus have the properties of a linear vector space with the extra property of an associative multiplication rule. These properties are just those that are familiar from a set of $N \times N$ matrices. For example, if R, S, and T denote linear operators, then $(aR + bS)T = aRT + bST$ denotes another operator, where $a, b \in \mathbb{C}$ and RT represents the product of two operators. The overworked symbol 0 now also stands for the zero operator with evident properties, and we let $\mathbb{1}$ denote the unit operator for which $\mathbb{1}R = R\mathbb{1} = R$ for every operator R. Following standard convention, however, we sometimes omit the symbol for the unit operator when it is clear from the context exactly what is meant. Associativity means that $(RS)T = R(ST) \equiv RST$, and so the parentheses are unnecessary.

The role of operators in a Hilbert space is to map vectors into other vectors. A linear operator, say R, has the property that

$$R(a|\phi\rangle + b|\psi\rangle) = aR|\phi\rangle + bR|\psi\rangle \ .$$

We shall only be concerned with linear operators and will generally refer to them simply as operators hereafter. If an operator R takes every vector in $\mathfrak{H}$ into a vector in $\mathfrak{H}$, then the operator is defined everywhere. This is true, in particular, if the operator is bounded, that is, if, for a fixed c, where $0 \leq c < \infty$, $\|R|\psi\rangle\| \leq c\| \, |\psi\rangle\|$ holds for all $|\psi\rangle \in \mathfrak{H}$. An operator—let us say T—may also be unbounded, in which case there is no finite c for which the previous inequality holds. In that case the operator can act only on a subset of vectors in $\mathfrak{H}$ and still yield vectors in $\mathfrak{H}$. To deal with this fact it is necessary to ascribe to T a domain $\mathfrak{D}(T) \subset \mathfrak{H}$ composed of vectors that are transformed by T into vectors in $\mathfrak{H}$. An example may help clarify the situation. In the sequence space l^2, the transformation R that takes the vector $\{\psi_n\}$, $n \in \mathbb{N} \equiv \{1, 2, 3, \ldots\}$, into the vector $\{[n/(1+n)]\psi_n\}$ is a bounded operator (for any $c \geq 1$), while the transformation T that takes the vector $\{\psi_n\}$ into the sequence $\{n\psi_n\}$ does not result in a vector for all initial vectors. In the latter case it is necessary that $\sum n^2|\psi_n|^2 < \infty$ in order that $\{n\psi_n\}$ actually represent a vector. For the present example, therefore,

$$\mathfrak{D}(T) = \{|\psi\rangle : \{n\psi_n\}_{n=1}^{\infty} \in l^2\} \ .$$

Unbounded operators play an important role in quantum mechanics, but the need to deal with domains is very often a nuisance, so much so that they are all too often ignored.

Several definitions and properties of operators deserve mention. An operator R has an eigenvalue $\lambda_r \in \mathbb{C}$ and an eigenvector $|r\rangle$, which lies in $\mathfrak{H}$, provided that $R|r\rangle = \lambda_r|r\rangle$.[3] An operator may have a number of distinct

[3] It may also happen that this eigenvalue equation holds not for a vector in $\mathfrak{H}$ but for a generalized eigenvector that strictly speaking is not in $\mathfrak{H}$. We deal with this more general situation below.

eigenvectors and several different eigenvalues. For example, every vector is an eigenvector of the zero operator with eigenvalue zero, and every vector is also an eigenvector of the unit operator with eigenvalue one. If R denotes an operator, then the adjoint operator, denoted by $R^\dagger$, is defined by the equation

$$\langle\psi|R^\dagger|\phi\rangle = \langle\phi|R|\psi\rangle^* \ .$$

A Hermitian operator has the property that $R^\dagger = R$. Thus eigenvectors of Hermitian operators satisfy $\langle r|R|r\rangle = \lambda_r\langle r|r\rangle = \lambda_r^*\langle r|r\rangle$, and so their eigenvalues are real. If $|r\rangle$ and $|s\rangle$ denote two distinct eigenvectors for a Hermitian operator, then it follows that

$$\langle s|R|r\rangle = \lambda_r\langle s|r\rangle = \langle r|R|s\rangle^* = \lambda_s\langle s|r\rangle \ ,$$

and thus if $\lambda_r \neq \lambda_s$, $|r\rangle$ and $|s\rangle$ are orthogonal, $\langle s|r\rangle = 0$. Even when $\lambda_r = \lambda_s$ one may choose a suitable linear combination of the two eigenvectors so that they are orthogonal. Consequently, it is generally asserted that any Hermitian operator, say R, has a complete set of orthonormal eigenvectors $\{|r\rangle\}$, $r \in \mathbb{N}$, namely, a set that constitutes a basis. Indeed, this is one of the most common ways of defining and choosing a basis set with which to work. Based on the representation generated by that basis,

$$\begin{aligned} |\psi\rangle &= \Sigma\,|r\rangle\psi_r = \Sigma\,|r\rangle\langle r|\psi\rangle \ , \\ \langle\phi|\psi\rangle &= \Sigma\,\langle\phi|r\rangle\langle r|\psi\rangle \equiv \langle\phi|\mathbb{1}|\psi\rangle \ , \end{aligned}$$

and it is useful to introduce a representation for the unit operator itself given by

$$\mathbb{1} = \Sigma\,|r\rangle\langle r| \ ,$$

which is called a resolution of unity. Every Hermitian operator leads to a basis with which a resolution of unity may be constructed. Clearly we have

$$\begin{aligned} R = R\,\mathbb{1} &= R\,\Sigma\,|r\rangle\langle r| = \Sigma\,R|r\rangle\langle r| = \Sigma\,\lambda_r|r\rangle\langle r| \ , \\ R^2 &= \Sigma\,\lambda_r^2|r\rangle\langle r| \ , \\ f(R) &= \Sigma\,f(\lambda_r)|r\rangle\langle r| \ , \end{aligned}$$

etc., for any reasonable function f. In such a basis, one has diagonalized the operator R inasmuch as it acts by simple multiplication.

If the eigenvectors of the Hermitian operator, say X in this case, are generalized eigenvectors, then the eigenvalues comprise an open set in $\mathbb{R}$. In that case orthonormality is taken to mean $\langle x|y\rangle = \delta(x-y)$, and the generalized eigenvectors $|x\rangle$ are said to be δ-function normalized. Under these circumstances, one has

$$\mathbb{1} = \int|x\rangle\langle x|\,dx \ ,$$

and therefore

$$\begin{aligned} X = X\,\mathbb{1} &= \int X|x\rangle\langle x|\,dx = \int x|x\rangle\langle x|\,dx\ , \\ X^2 &= \int x^2|x\rangle\langle x|\,dx\ , \\ f(X) &= \int f(x)|x\rangle\langle x|\,dx\ , \end{aligned}$$

where the integral runs over the possible continuous range of eigenvalues of X, i.e., the continuous spectrum of X. These generalized eigenvectors give rise to L^2 representation spaces by means of functions $\psi(x) = \langle x|\psi\rangle$. In this language the function $\psi(x)$ represents the "coordinates" of the abstract vector $|\psi\rangle$, although some care must be exercised in pursuing this analogy with the discretely labeled "coordinates." Once again, a few technical remarks may be in order. In particular, the zero vector may be represented not only by the zero function, but by any other function that is zero almost everywhere so that its square integral vanishes. In point of fact, the representatives in the continuum case are equivalence classes composed of functions that differ from one another on sets of measure zero. This seemingly "minor technicality" is responsible for reducing the uncountable number of dimensions in the basis $\{|x\rangle\}\,, x \in \mathbb{R}$, to the countable number of dimensions in the basis $\{|n\rangle\}\,, n \in \mathbb{N}$, that represents the true dimension of a separable, infinite-dimensional Hilbert space. [**Remark:** There also are Hilbert spaces of uncountably many dimensions, the so-called "nonseparable" Hilbert spaces, but generally, they have a limited physical applicability.]

Another important class of operators is that of the unitary operators. An operator U is unitary provided that $U^\dagger U = UU^\dagger = \mathbb{1}$. A unitary operator has the property that the inner product of $U|\psi\rangle$ and $U|\phi\rangle$ is identical to the inner product of $|\psi\rangle$ and $|\phi\rangle$ for any pair of vectors; the same statement holds with U replaced by $U^\dagger$. It is important to note that every unitary operator can be written in the form $U = e^{-iF}$ for some Hermitian operator F (more precisely, F is a self-adjoint operator; see below). The action of a unitary operator may be thought of as a rigid rotation of Hilbert space; indeed, unitary operators are the natural extension to an infinite-dimensional complex space of the family of orthogonal rotations familiar in three-dimensional space which have the property that the inner product of any two three-vectors is invariant under such rotations. In three dimensions it is evident that one may rotate any orthonormal frame into any other orthonormal frame, including if necessary an improper rotation (e.g., a reflection through any one plane). Likewise, in Hilbert space, a unitary operator can be found that can map any orthonormal basis set into any other orthonormal basis set. Let $|r\rangle$ and $|s'\rangle$ denote two arbitrary vectors of two such bases. Then the operator U defined by the matrix in the $|r\rangle$-basis with components $U_{sr} = \langle s|U|r\rangle \equiv \langle s'|r\rangle$, $r, s \in \mathbb{N}$, generates the required transformation. In particular, as follows from the resolution of unity, the transformed vector components read

$$\phi'_s = \langle s'|\phi\rangle = \Sigma\,\langle s'|r\rangle\langle r|\phi\rangle = \Sigma\,U_{sr}\,\phi_r\ .$$

Normally, in three space, physical quantities of interest cannot depend on the artificial choice of coordinates, and so they must be expressed in a co-

ordinate invariant form. Likewise, in quantum mechanics the physical results will be expressed in terms of inner products and not in terms of the vector representatives by themselves. In other words, the physical answers must be independent of the representation of the Hilbert space. This fact has two consequences. The first consequence is that much of quantum mechanics can be discussed in the language of an abstract Hilbert space, i.e., in terms of abstract operators and bra- and ket-vectors. The second consequence is that in order to compute something—and one generally needs to choose a concrete representation to do so—the physical answers will not depend on the choice of representation. Thus the choice of representation can be made for the convenience of the calculation.

The commutator of any two operators A and B is defined by

$$[A, B] \equiv AB - BA$$

and in general is nonvanishing. Any two operators for which $[A, B] = 0$ are said to commute. By definition, the unit operator, $\mathbb{1}$, commutes with all operators, $[\mathbb{1}, R] = 0$. A set of operators $\{R_n\} \equiv \{R_1, R_2, R_3, \ldots\}$ is called irreducible if the only operator B that commutes with every operator in the set is a multiple of the identity operator, i.e., $\{R_n\}$ is irreducible if and only if $[B, R_n] = 0$ for all n implies that $B = b\mathbb{1}$ for some $b \in \mathbb{C}$. A minimal irreducible set consists of a set of operators from which all other operators can be constructed by algebraic operations or perhaps by suitable limits thereof, or as we shall simply say, as a function of those operators. A subset $\{C_n\}$ of a minimal irreducible set is called a complete set of commuting operators provided that $[C_n, C_m] = 0$ for every n and m in the set, and any other operator that commutes with all other operators in the set is a function of the operators in the complete set of commuting operators. With this definition, a complete set of commuting operators is at the same time a minimal set of commuting operators. All operators in a complete set of commuting operators that can be diagonalized can be simultaneously diagonalized, and for these it follows that

$$C_n C_m |c\rangle = C_n c_m |c\rangle = c_m C_n |c\rangle = c_n c_m |c\rangle \ ,$$

where we have put $|c\rangle \equiv |\{c_n\}\rangle \equiv |c_1, c_2, \ldots\rangle$. For systems with a finite number of degrees of freedom, a minimal irreducible set $\{R_n\}$ contains only a finite number of operators. Likewise, for finitely many degrees of freedom, a complete set of commuting operators contains only a finite number of operators. The simultaneous (possibly generalized) eigenvectors of a complete set of commuting operators may be chosen to be orthonormal and they therefore constitute a suitable basis set of vectors with which to define a representation of Hilbert space.

Additionally, we introduce the concept of the trace. For a finite, square matrix the trace is simply the sum of the diagonal elements. This concept is invariant under orthogonal transformations and so it is an intrinsic property of the matrix and does not depend on the particular representation that has

been chosen. In an infinite-dimensional Hilbert space, each operator may be represented by a semi-infinite, square matrix. In this case the trace is again simply the sum of the diagonal elements, but this sum is independent of the representation and therefore an intrinsic property of the operator only for a subset of operators. Every operator A for which the sum will be intrinsic admits a canonical form given by

$$A = \Sigma_n a_n |\alpha_n\rangle\langle\alpha'_n| \,, \qquad \Sigma_n |a_n| < \infty \,.$$

Here $\{|\alpha_n\rangle\}$ and $\{|\alpha'_n\rangle\}$ denote two, possibly identical, complete orthonormal bases. Such operators are said to be "trace class," and up to trivial phase factors this decomposition itself is intrinsic to the operator A. To test whether an operator A has such a representation it suffices to require that

$$\mathrm{Tr}((A^\dagger A)^{1/2}) \equiv \Sigma_m \langle m|(A^\dagger A)^{1/2}|m\rangle = \Sigma_n |a_n| < \infty \,.$$

Here $\{|m\rangle\}$ denotes any complete orthonormal basis. If the operator of interest satisfies this criterion, then the trace of A is defined by

$$\mathrm{Tr}(A) \equiv \Sigma_r \langle r|A|r\rangle = \Sigma_n a_n \langle\alpha'_n|\alpha_n\rangle \,,$$

which is a number independent of the particular complete orthonormal basis set with which it is evaluated. If the operator in question is not trace class, then the indicated sum may converge but the result will in general depend not only on the operator but also on the particular basis set with which the sum has been evaluated! It is hard to accept that such situations can describe real physics.

For $N \times N$ matrices, where $N < \infty$, it is well known that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$. In the Exercises, we ask under what conditions this cyclic invariance holds when $N = \infty$.

This concludes our initial discussion of Hilbert space and operators. In the following section we offer some deeper insight into some of these issues. These extra facts are not all relevant for the rest of this monograph, but they do represent important issues that are seldom dealt with in the usual textbooks on quantum theory, and in the author's view, should be more widely appreciated.

5.2 Hilbert Space and Operators: Advanced Properties

In the previous section of this chapter, we introduced and discussed various properties of Hilbert spaces and operators suitable for quantum mechanical applications. In all earlier applications we have used the elegant notation of Dirac, and this notation is generally appropriate for most purposes. Unfortunately, the Dirac notation is awkward to use for certain general questions, especially those related to operator adjoints, and as a consequence we shall

also introduce and use a common mathematical notation for the vectors in a Hilbert space $\mathfrak{H}$ as lowercase Greek letters ψ, ϕ, ξ, etc., and especially an inner product denoted by (ψ, ϕ) $[= (\phi, \psi)^*]$ that is linear in the second element (ϕ) and antilinear in the first element (ψ). The Hilbert space vectors form a complex linear vector space with 0 denoting the zero element with the property that $\psi + 0 \equiv \psi$ for all $\psi \in \mathfrak{H}$. If $a \in \mathbb{C}$, then $a\psi \in \mathfrak{H}$ whenever $\psi \in \mathfrak{H}$. It follows from these relations that $(0, \phi) = 0$ for any $\phi \in \mathfrak{H}$. The norm of a vector is given by $\|\psi\| \equiv +\sqrt{(\psi, \psi)}$ with $\|\psi\| = 0$ if and only if $\psi = 0$. Only vectors for which $\|\psi\| < \infty$ belong to the Hilbert space, which as usual is assumed to be separable and complete. The Schwarz inequality reads $|(\phi, \psi)| \leq \|\phi\|\|\psi\|$, while the triangle inequality is given by $\|\phi + \psi\| \leq \|\phi\| + \|\psi\|$. It is convenient to regard $\|\psi - \phi\|$ as the distance between the indicated vectors in $\mathfrak{H}$. There are two common notions of convergence of vectors. Weak convergence of a sequence of vectors $\{\phi_n\}_{n=1}^{\infty}$ to a vector ϕ holds if and only if for all $\lambda \in \mathfrak{H}$, it follows that $\lim_{n\to\infty}(\lambda, \phi_n - \phi) = 0$. Strong convergence of a sequence $\{\phi_n\}_{n=1}^{\infty}$ to a vector ϕ, on the other hand, means that $\lim_{n\to\infty} \|\phi_n - \phi\| = 0$. Equivalently, we may say that the sequence $\{\phi_n\}_{n=1}^{\infty}$ is a Cauchy sequence if for any $\epsilon > 0$ there exists a $P(\epsilon) < \infty$ such that $\|\phi_n - \phi_m\| < \epsilon$ holds provided $m > P$ and $n > P$. Hilbert space is complete in the sense that it already contains the limits of all Cauchy sequences. When we speak of the convergence of vectors without further qualification, we always have strong convergence in mind.

It is convenient at this point to introduce the concept of a *dense set* of vectors. A set $\mathcal{D}$ is a dense set of vectors if for any $\epsilon > 0$ and any $\lambda \in \mathfrak{H}$, there exists a vector $\phi \in \mathcal{D}$ such that $\|\lambda - \phi\| < \epsilon$. Intuitively, this means that for any vector in the Hilbert space (λ) there is a vector in the dense set (ϕ) that is arbitrarily close to it. (The analogy of the rationals being dense in the reals, or the set of continuous functions being dense in the space $L^2([0, 1])$ should help convey the meaning of this concept.) We are primarily interested in operators A for which the domain $\mathfrak{D}(A)$ is dense in the Hilbert space $\mathfrak{H}$, and let us restrict attention initially to that case. We observe that if $(\lambda, \phi) = 0$ for all ϕ in a dense set, then it follows by continuity that $(\lambda, \xi) = 0$ for all $\xi \in \mathfrak{H}$, and hence $\lambda = 0$. For completeness, we define a *total set*. A set $\mathcal{T}$ is a total set if for all $\tau \in \mathcal{T}$, $(\tau, \psi) = 0$ implies that $\psi = 0$. A total set is also commonly called a "complete set", however, we prefer to use the phrase "total set" since in many applications $\mathcal{T}$ contains an *un*countable number of vectors. Note that the finite linear span (the set of vectors given by all finite linear combinations) of a total set leads to a dense set.

Operators will generally (but not universally) be denoted by uppercase Roman letters, e.g., A, B, C, etc., and the domain of an operator, say A, is denoted by $\mathfrak{D}(A)$. We shall restrict attention only to *linear operators*, namely, operators that have the property that $A(a\phi + b\psi) = aA\phi + bA\psi$ for all $a, b \in \mathbb{C}$ and all $\phi, \psi \in \mathfrak{D}(A)$. Linear Hilbert space operators may be divided into two families: bounded and unbounded operators. For a bounded linear operator, say B, it follows that $\|B\psi\| \leq b\|\psi\|$ for some $0 \leq b < \infty$. Since in this case $\|B0\| = 0$ it follows that $B0 = 0$ for all bounded B. The operator norm

of a bounded operator is defined by $\|B\| \equiv \sup \|B\psi\|/\|\psi\| < \infty$ where the supremum is over all $\psi \neq 0$. For any bounded operator B, the domain $\mathfrak{D}(B) = \mathfrak{H}$. For an unbounded operator, say A, $\|A\psi\|$ is not finite for all $\psi \in \mathfrak{H}$. In fact, $\|A\psi\| < \infty$ only for vectors $\psi \in \mathfrak{D}(A)$, although it is sometimes convenient if $\mathfrak{D}(A)$ does not consist of all such vectors. We note that the zero vector is always in the domain of any operator, and $A0 \equiv 0$.

Two notions of the convergence of a sequence of operators are common. On one hand, a sequence of operators B_n converges weakly to an operator B whenever $\lim(\phi, (B_n - B)\psi) = 0$ for all ϕ and ψ in a common dense domain. On the other hand, a sequence of operators C_n converges strongly to an operator C provided that $\lim \|(C_n - C)\psi\| = 0$ for all ψ in a common dense domain. In both cases, if the sequence of operators is uniformly bounded, e.g., if $\|B_n\| \leq b < \infty$ for all n, then it suffices to show convergence for ϕ and ψ in a total set of vectors. This procedure is an important and extremely practical way to show convergence.

Hilbert spaces are either finite dimensional or they are countably infinite dimensional. The former are like finite-dimensional complex Euclidean spaces, and all operators (or matrices, if you prefer) are bounded operators, and one's intuition is almost always correct. In infinite-dimensional Hilbert spaces, on the other hand, very unusual things can transpire that occasionally defy common sense!

As linear operators we might expect that

$$\lim_{n\to\infty} A\,\psi_n = A \lim_{n\to\infty} \psi_n$$

for a sequence of vectors $\{\psi_n\}_{n=1}^{\infty}$ such that each $\psi_n \in \mathfrak{D}(A)$ as well as $\lim_{n\to\infty} \psi_n \in \mathfrak{D}(A)$. Some operators satisfy this equality while others do not. That further divides unbounded operators into two classes, namely, the *closable operators* for which this equality always holds and the *nonclosable operators* for which it sometimes fails. Rather than deal with all possible vector sequences, it is sufficient to examine those situations where $\lim_{n\to\infty} \psi_n \to 0$. For a closable operator, $\lim_{n\to\infty} A\psi_n$ either converges to 0 or it fails to converge in $\mathfrak{H}$. For a nonclosable operator, on the other hand, besides these possibilities, it may happen that $\lim_{n\to\infty} A\psi_n$ converges to a vector in the Hilbert space that is different from 0! A simple example will illustrate this phenomenon. Let us consider the sequence space l^2 composed of square-summable sequences $\{\phi^k\}_{k=1}^{\infty}$; we put the index on the top this time to make room for the sequence label. Choose the operator $A = \{A_{kl}\}$ with entries $A_{kl} = l\delta_{k1}$, namely, the sequence $1, 2, 3, \ldots$ along the first row and zero otherwise. This matrix defines an operator for which the domain consists of all sequences for which $\Sigma_{k=1}^{\infty}\, k\phi^k$ is finite. However, consider the sequence of vectors $\phi_n = \{\phi_n^k\}$ where $\phi_n^k = \delta_{nk}/n$. It follows that $\phi_n \to 0$ but $A\phi_n \to \xi$, where $\xi^k = \delta_{k1}$ which is not the zero vector! In brief, we have illustrated a simple operator and a sequence of vectors such that

$$0 \neq \xi = \lim_{n\to\infty} A\,\phi_n \neq A \lim_{n\to\infty} \phi_n = A\,0 = 0\,.$$

The reader can well imagine the difficulties such operators may cause, but it is reassuring to learn that all operators of physical interest will not have such bizarre properties.

The adjoint of an operator A may be uniquely defined only when $\mathfrak{D}(A)$ is dense. In particular, the adjoint $A^\dagger$ of A is defined as follows. For a suitable $\psi \in \mathfrak{H}$, let ξ be a vector such that $(\xi, \phi) = (\psi, A\phi)$ holds for all vectors $\phi \in \mathfrak{D}(A)$. Then $\psi \in \mathfrak{D}(A^\dagger)$ and $A^\dagger\psi = \xi$. Note that it is necessary for the domain of A to be dense in order for ξ to be uniquely determined. At the very least the vector $0 \in \mathfrak{D}(A^\dagger)$ and $A^\dagger 0 = 0$ as required. For some operators $A^\dagger$, 0 is the *only* vector in $\mathfrak{D}(A^\dagger)$!

A symmetric operator (which is called a Hermitian operator in the physics literature) is one for which $(A\phi, \psi) = (\phi, A\psi)$ holds for all $\phi, \psi \in \mathfrak{D}(A)$. It follows that $A^\dagger = A$ on $\mathfrak{D}(A)$, from which we learn that $\mathfrak{D}(A^\dagger) \supseteq \mathfrak{D}(A)$ [simply because $\mathfrak{D}(A^\dagger)$ is definitely not smaller than $\mathfrak{D}(A)$]. The special case of equality holds for the very important case of self-adjoint operators; i.e., if $A^\dagger = A$ and $\mathfrak{D}(A^\dagger) = \mathfrak{D}(A)$, then we say that the operator A is *self-adjoint*. Notice that a self-adjoint operator is necessarily symmetric, but the converse is not necessarily true. In particular, a symmetric operator may uniquely determine a self-adjoint operator; in this case the operator is called essentially self-adjoint. An example of such an operator is the differential operator $P = -i(\partial/\partial x)$ acting on the space $L^2((-\infty, \infty))$ with $\mathfrak{D}(P) = \mathfrak{D}_o(P) \equiv \{\psi : \psi \in \mathcal{S}\}$, where $\mathcal{S}$ is the space of functions with continuous derivatives of arbitrary order that all fall to zero at infinity faster than any power, or $\mathfrak{D}(P) \equiv \{\psi : \int[|\psi'(x)|^2 + |\psi(x)|^2]\,dx < \infty\}$. The second domain is the maximal domain on which $\mathfrak{D}(P^\dagger) = \mathfrak{D}(P)$, and one speaks of extending the original smaller domain $\mathfrak{D}_o(P)$ to the maximal domain $\mathfrak{D}(P)$. For some symmetric operators it may be possible to extend the operator by extending its domain so that it becomes self-adjoint, but that extension is not unique and therefore various distinct self-adjoint operators may arise. An example of this type is for the operator $P = -i(\partial/\partial x)$ acting this time on the space $L^2([0,1])$. In this case we choose as the initial domain $\mathfrak{D}_o(P) \equiv \{\psi : \psi' \in L^2([0,1]),\ \psi(0) = \psi(1) = 0\}$. Now $\mathfrak{D}_o(P^\dagger) = \{\psi : \psi' \in L^2([0,1])\}$ without any need for boundary conditions. We can extend the original domain of P (thereby shrinking the domain of $P^\dagger$) and arrive at a self-adjoint operator where $\mathfrak{D}(P^\dagger) = \mathfrak{D}(P) = \{\psi : \psi' \in L^2([0,1]), \psi(1) = e^{i\alpha}\psi(0)\}$, where $\alpha \in \mathbb{R}$ is a needed new parameter to define the extension. Each different α (mod 2π) leads to a distinct operator, say P_α, with a spectrum depending on the parameter α. In the third category are symmetric operators that can never be extended to self-adjoint operators. An example of such an operator is $P = -i(\partial/\partial x)$ acting on the Hilbert space $L^2([0,\infty))$ with $\mathfrak{D}(P) = \{\psi : \psi' \in L^2([0,\infty)),\ \psi(0) = 0\}$. In this case $\mathfrak{D}(P^\dagger) = \{\psi : \psi' \in L^2([0,\infty))\}$ without a boundary condition required. A rapid way to see the category into which a symmetric operator fits is the following: Consider the distributional solutions of the two equations $P^\dagger\psi_\pm(x) = \pm i\psi_\pm(x)$. Let (n_+, n_-) denote the number of linearly independent, square-integrable solutions to these two equations; the numbers $n_\pm$ are called

the deficiency indices. If $n_+ = n_- = 0$, the symmetric operator is (essentially) self-adjoint. If $n_+ = n_- > 0$, the operator admits self-adjoint extensions, but n_+^2 real parameters are needed to specify the operator. If $n_+ \neq n_-$, no self-adjoint operator exists. In the examples for P considered in this paragraph, $\psi_\pm(x) = e^{\mp x}$ and the deficiency indices (n_+, n_-) are $(0,0)$ for $L^2((-\infty,\infty))$, $(1,1)$ for $L^2([0,1])$, and $(1,0)$ for $L^2([0,\infty))$.

Recall that a unitary operator U is defined by the fact that $UU^\dagger = U^\dagger U = \mathbb{1}$. Now it is of the utmost importance that every unitary operator U is given by $U = e^{iA}$ and $U^\dagger = e^{-iA}$ for some self-adjoint operator A, and conversely, every self-adjoint operator A generates a unitary operator and its adjoint by way of $e^{iA} \equiv U$ and $e^{-iA} \equiv U^\dagger$. All this seems familiar save for the question of how e^{iA} is to be defined as an operator. We may include in the domain of every self-adjoint operator a dense set of *analytic vectors*. A vector ψ is an analytic vector for an operator A provided that $\Sigma_{n=0}^\infty \|A^n\psi\|/n! < \infty$. For such vectors we define e^{iA} by means of its power series expansion, namely,

$$e^{iA}\,\psi \equiv \sum_{n=0}^{\infty} \frac{i^n A^n}{n!}\,\psi\,,$$

which is well-defined because the series converges strongly to a vector in $\mathfrak{H}$ for all analytic vectors. On the analytic vectors it follows that $\|e^{iA}\,\psi\| = \|e^{-iA}\,\psi\| = \|\psi\|$. Since the set of analytic vectors is dense we can extend the definition of e^{iA} to all vectors in $\mathfrak{H}$ by continuity, namely, for an arbitrary vector ϕ by $e^{iA}\,\phi = \lim e^{iA}\,\psi_n$ where each ψ_n is an analytic vector for A and the sequence is chosen so that $\lim \|\phi - \psi_n\| = 0$.

A *unitary one-parameter group* is a set of operators $\{U(t)\}_{t\in\mathbb{R}}$ that satisfies, for all $t, s \in \mathbb{R}$, the following four properties: (1) $U(0) = \mathbb{1}$; (2) $U(t)U(s) = U(t+s)$; (3) $U^\dagger(t) = U(-t)$; and (4) $U(t)$ is weakly continuous, which means, for all $\psi, \phi \in \mathfrak{H}$, that $(\psi, U(t)\,\phi)$ is a continuous function of t. If a set of operators fulfills these four conditions, then it follows that there exists a self-adjoint operator A such that $U(t) = e^{itA}$. This rule is extremely useful to determine if an operator is self-adjoint or not. The relation of these conditions to representations of an operator A is straightforward. In particular, if we diagonalize A, then it follows for each pair of vectors ψ, ϕ that there exists a complex, countably additive measure (hereafter simply called a measure) $\sigma_{\psi,\phi}$ such that

$$(\psi, U(t)\phi) = \int e^{ita}\, d\sigma_{\psi,\phi}(a)\,.$$

If $\psi = \phi$, then the measure $\sigma_{\phi,\phi}$ is nonnegative, and if ϕ is normalized, then the measure is a probability measure. In the latter case the function

$$C(t) \equiv \int e^{ita}\, d\mu_\phi(a)$$

defines the characteristic function for a probability measure $\mu_\phi \equiv \sigma_{\phi,\phi}$ with $(\phi,\phi) = 1$. As a characteristic function, $C(t)$ is continuous whether the measure μ_ϕ is discrete, absolutely continuous, or singular continuous.

5.3 Basic Lie Group Theory

A group $\mathcal{G}$ consists of elements $g \in \mathcal{G}$ that enjoy several properties. The basic property is that these elements have a combination law—a generalized multiplication—such that if g_1 and g_2 are two arbitrary group elements, then $g_1 g_2$ represents another group element. The identity element $g = e$ is part of the group, where $ge = eg = g$ for all g, and every element g has a unique inverse element g^{-1} such that $gg^{-1} = g^{-1}g = e$. The final property is associativity, namely, for three arbitrary group elements f, g, and h, the relation $(fg)h = f(gh)$ which means that fgh is uniquely defined. Groups can consist of a finite number of elements, a countable infinity of elements, or, as we shall focus on, a continuous infinity of elements. To describe this last case, it is useful to label the group elements by points in an $L < \infty$ dimensional label space $\mathcal{L}$ that is locally isomorphic to $\mathbb{R}^L$. Thus to each point $l = \{l^1, l^2, \ldots, l^L\} \in \mathcal{L}$, we identify a distinct group element $g(l)$. Group multiplication then involves some form of label multiplication that encodes the particular group under consideration. In particular, we define $g(l_1)g(l_2) = g(l_1 \circ l_2)$, where $\circ$ denotes a particular rule of combination to yield another element of the label space. We reserve the label l_0 for the identity e so that $g(l_0)g(l) = g(l)g(l_0) = g(l)$, i.e., $l_0 \circ l = l \circ l_0 = l$. Likewise, we let $g(l^{-1}) \equiv g(l)^{-1}$, the element inverse to $g(l)$.

A subgroup is a new group, composed of a subset of the elements of some group, that obeys the same rules as the original group. Let $\mathcal{H} \subset \mathcal{G}$ denote such a subgroup consisting of elements $h \in \mathcal{H}$ for which h^{-1} and e, the original identity element of $\mathcal{G}$, are also elements of $\mathcal{H}$. Additionally, if h_1 and h_2 are elements of $\mathcal{H}$, then so too are the elements $h_1 h_2$ and $h_2 h_1$ as well as all combinations involving the inverse elements. The elements of $\mathcal{H}$ can also be described by the same labels $h(l)$ that applied to the original group $\mathcal{H}$.

A few simple examples may help make these concepts clear. Suppose we have a one-dimensional label ($L = 1$), and we define the multiplication rule as $g(l_1)g(l_2) = g(l_1 + l_2)$, with $l \in \mathbb{R}$; for this example, $e = g(0)$ and $g(l)^{-1} = g(-l)$. Another example would be to choose $g(l_1)g(l_2) = g(l_1 l_2)$; in this case the parameter $l \in \mathbb{R} - 0$. Here, the identity element $e = g(1)$, and the inverse element is $g(l^{-1}) = g(l)^{-1}$. A subgroup of this example arises when we restrict the labels so that $l > 0$; observe that the subset where $l < 0$ does not form a subgroup since it lacks the identity element. Familiar examples of continuous groups include the following. (i) The real n-dimensional rotation group $\mathrm{O}(n, R)$, which is defined by real $n \times n$ matrices with the property that $O\,O^T = O^T O = \mathbb{1}$, where T denotes transpose. The determinant of such matrices is ± 1. The subgroup of matrices with determinant $+1$ is called $\mathrm{SO}(n, R)$.

(ii) The complex unitary group $\mathrm{U}(n, C)$ defined by $n \times n$ complex matrices such that $U\,U^\dagger = U^\dagger U = \mathbb{1}$. It follows that the determinant of such matrices is given by $e^{i\theta}$, where θ is real. The subgroup of $\mathrm{U}(n, C)$ for which $\theta = 0$ is called $\mathrm{SU}(n, C)$. (iii) The n-dimensional affine group defined as the group of transformations (A, b) such that $x \to x' = Ax + b$, where A is a real, invertible $n \times n$ matrix and x, x', and b are all real n-dimensional vectors. Here the multiplication rule of the group elements becomes $(A', b')\,(A, b) = (A'A, A'b + b')$. All of these examples are examples of Lie groups.

A continuous group is also associated with two group invariant measures which are defined below. For now, suffice it to say that the group is called *compact* if the integral over the group invariant measure is finite, while the group is called *noncompact* if the integral over the group invariant measure diverges.

5.3.1 Lie algebras

Lie groups also give rise to Lie algebras as follows. Let $g(l)$ be chosen so that $e = g(0)$, and consider the set of group elements very close to the identity which we shall denote by

$$g(\delta l) = g(0) + \delta l^j\, \partial g(l)/\partial l^j|_{l=0} \equiv e + \delta l^j\, X_j \ ,$$

where X_j, $j = 1, 2, \ldots, L$, are called the *infinitesimal generators*, and we have adopted the summation convention. Evidently, the inverse element near the identity is given by

$$g(\delta l)^{-1} = e - \delta l^j\, X_j \ ,$$

which is valid to lowest order. The infinitesimal generators form a linear vector space; we shall assume that the elements X_j are all linearly independent and that they represent a complete basis set of elements.

Attention is next focused on the so-called group commutator (GC)

$$GC(\delta l, \delta\tilde{l}) \equiv g(\delta l)^{-1}\, g(\delta\tilde{l})^{-1}\, g(\delta l)\, g(\delta\tilde{l}) \ ,$$

which leads to the expression

$$\begin{aligned} GC(\delta l, \delta\tilde{l}) &= (e - \delta l^j\, X_j)\,(e - \delta\tilde{l}^k\, X_k)\,(e + \delta l^m\, X_m)\,(e + \delta\tilde{l}^r\, X_r) \\ &= e + \delta l^r\, \delta\tilde{l}^s\, [X_r, X_s] \ . \end{aligned}$$

Since the result of this calculation should also be a group element near to the identity, it follows that

$$[X_r, X_s] = C_{rs}{}^t\, X_t \ ,$$

for some constants $C_{rs}{}^t$, which is the celebrated multiplication rule of Lie for the infinitesimal generators. Combining this multiplication rule with the linear

vector space character leads to an algebra, the Lie algebra associated with the Lie group. The constants ${C_{rs}}^t$ are known as the *structure constants*, and they help determine the nature of the Lie algebra itself. These constants satisfy two basic identities: (1) antisymmetry, ${C_{sr}}^t = -{C_{rs}}^t$, and (2) the Jacobi identity

$$ {C_{ab}}^s {C_{cs}}^t + {C_{ca}}^s {C_{bs}}^t + {C_{bc}}^s {C_{as}}^t = 0 \, . $$

Any set of real constants that fulfill these two conditions are the structure constants of some Lie algebra.

Of course, the choice of the coordinates l is not unique, and we can consider coordinate transformations restricted so that when all coordinates vanish, we still deal with the identity element. It follows that ${C_{rs}}^t$ transforms as a tensor of covariant rank 2 and contravariant rank 1. Thus, only tensorial invariants generate coordinate independent classifications. In particular, if (i) ${C_{rs}}^t = 0$, the algebra (and group) are Abelian; if (ii) ${C_{rs}}^s = 0$, we deal with a unimodular group; if (iii) $g_{ra} \equiv {C_{rs}}^t \, {C_{at}}^s$ is nondegenerate, the algebra is semisimple; if (iv) $\{g_{ra}\} < 0$, i.e., if that symmetric matrix is negative definite, then the group is compact—otherwise it is noncompact, etc.

5.3.2 Invariant group measures

Both compact and noncompact Lie groups have invariant group measures. Indeed, there generally are two kinds of group invariant measures to be discussed. Let

$$ dg(l) \equiv g(l + dl) - g(l) $$

denote the difference between two group elements that are infinitesimally close to each other. The expression

$$ g(l)^{-1} \, dg(l) $$

translates that differential back to the neighborhood of the identity, and as a consequence it follows that

$$ g(l)^{-1} \, dg(l) = \lambda^k(l) \, X_k \, , $$

where $\lambda(l)^k$ is a one form given by $\lambda^k(l) = {\lambda^k}_m(l) \, dl^m$.

Let us repeat this argument for the combination $d[g(l_1)\, g(l)] = g(l_1) dg(l)$, where we assert that only the label l is to be varied while the label l_1 is held fixed. Next consider

$$ [g(l_1)\, g(l)]^{-1} \, d[g(l_1)\, g(l)] = g(l)^{-1} g(l_1)^{-1} g(l_1) dg(l) = g(l)^{-1} dg(l) \, , $$

which implies the important relation that

$$ g(l_1 \circ l)^{-1} \, dg(l_1 \circ l) = g(l)^{-1} \, dg(l) \, , $$

or that $\lambda^k(l_1 \circ l)\, X_k = \lambda^k(l)\, X_k$, and since the infinitesimal generators X_k are linearly independent, it follows that

$$\lambda^k(l_1 \circ l) = \lambda^k(l) \ ,$$

which are known as the Maurer–Cartan equations [Che46, Coh61]. Each one form $\lambda^k(l)$ is thus *invariant under left translations* by an arbitrary group element. Exterior products of these one forms retain this property, and the Lth such exterior product yields a volume element on the group given by

$$d\mu_{left}(l) \equiv \wedge_{k=1}^{L} \lambda^k(l) \equiv \det[\lambda^k{}_m(l)]\, d^L l \ .$$

Since, for all l_1,

$$d\mu_{left}(l_1 \circ l) = d\mu_{left}(l)$$

this expression is one of the sought-for group invariant measures, the so-called *left-invariant group measure*.

This whole procedure can be carried out as well for the expression

$$dg(l)\, g(l)^{-1} = d[g(l)\, g(l_1)]\, [g(l)\, g(l_1)]^{-1} = \rho^k(l)\, X_k = \rho^k(l \circ l_1)\, X_k \ ,$$

where $\rho^k(l)$ are one forms given by $\rho^k(l) = \rho^k{}_m(l)\, dl^m$. These one forms ultimately lead to the *right-invariant group measure* given by

$$d\mu_{right}(l) \equiv \wedge_{k=1}^{L} \rho^k(l) \equiv \det[\rho^k{}_m(l)]\, d^L l \ ,$$

which enjoys the property that

$$d\mu_{right}(l \circ l_1) = d\mu_{right}(l)$$

for all l_1.

For compact groups $d\mu_{left}(l) = d\mu_{right}(l) \equiv d\mu(l)$, but that equality need not be true for noncompact groups. For compact groups the *group volume*

$$\int d\mu(l) < \infty \ ,$$

while for noncompact groups the group volume diverges as calculated by either the left- or the right-invariant measure.

5.3.3 Group representations

A representation of a Lie group consists of an associative set of linear transformations acting on a representation space, typically a linear vector space [Gil06]. In forming this representation we map the elements of the group onto transformations, such as matrices, in the manner $g \to W(g)$ or, using the labels, $g(l) \to W(l)$, with the property that this map is a homomorphism

(many-to-one) or an isomorphism (one-to-one). The transformations satisfy the rules of the group, such as $W(l_1)W(l_2) = W(l_1 \circ l_2)$.

In addition, representations of the Lie algebra are of importance. Here, one considers linear maps $X_k \to M_k = \rho(X_k)$ (this is not the same ρ used above!) such that

$$\rho([X_k, X_l]) = [\rho(X_k), \rho(X_l)] \,,$$

as required by the Lie algebra multiplication rule. Although the infinitesimal generators are linearly independent, under the homomorphism induced by $\rho(\,\cdot\,)$, this no longer need be the case. As an extreme example, the map $\rho(X_k) = M_k \equiv 0$ constitutes a possible (but not very interesting) representation. When the representation transformations M_k themselves are linearly independent, one speaks of a faithful representation; when they are not all linearly independent, the representation is said to be nonfaithful.

Examples can help clarify these matters. Consider the Lie group SO(3) defined by 3×3 orthogonal matrices of unit determinant. There are three basic rotations that together generate the group, namely, the three matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix}, \begin{pmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{pmatrix}, \begin{pmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

which we call R_1, R_2, and R_3, and which represent rotations about the 1st, 2nd, and 3rd axes, respectively. For very small rotation angles, it is clear, to lowest order, that

$$R_k(\delta\theta) = \mathbb{1} + \delta\theta\, X_k + \cdots \,,$$

and in these three cases we have

$$X_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \,.$$

In the same vein, it should be true that a normal-sized rotation should be given by a huge number of tiny, identical, repeated rotations, namely, that

$$R_k(\theta) = \lim_{N\to\infty} [\,\mathbb{1} + (\theta/N)\, X_k\,]^N = e^{\theta\, X_k} \,,$$

which is true for the matrices given above, namely, that

$$R_1(\alpha) \;=\; \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix} \;=\; \exp\left\{ \alpha \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \right\} ,$$

$$R_2(\beta) = \begin{pmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{pmatrix} = \exp\left\{ \beta \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \right\},$$

$$R_3(\gamma) = \begin{pmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{pmatrix} = \exp\left\{ \gamma \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\}.$$

These relations illustrate a basic truth about the relation between Lie groups and Lie algebras: the Lie algebra infinitesimal generators arise as quantities near the identity group element, and the exponential of the generators leads to group elements, at least that subgroup that is continuously connected to the identity. For the example chosen, all elements of the group SO(3) are given in this exponential form; however, for the group O(3), which contains elements with determinant minus one as well, those latter elements cannot be reached in this exponential fashion. For many applications it suffices to consider the subgroup that is continuously connected to the identity, and all those group elements may be reached by exponentiating the infinitesimal generators.

Moreover we can evaluate the commutator of the infinitesimal generators in this case to learn that

$$[X_k, X_l] = C_{kl}{}^m X_m = \epsilon_{klm} X_m \ ,$$

where ϵ_{klm} is the totally antisymmetric symbol with $\epsilon_{123} = 1$. This relation determines the structure constants of the group SO(3), which are also the structure constants for the group O(3) since near the identity, these groups are identical. Further properties of the rotation group are reserved for the Exercises.

We include a discussion about two other examples. The first is the one-dimensional, two-parameter affine group composed of transformations of the real line x given by $x \to x' = a\,x + b$, where $a > 0$ and $b \in \mathbb{R}$. This restriction on a leads to the subgroup that is continuously connected to the identity element. The transformations of this group may be represented as 2×2 matrices given by

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$$

with the combination rule given by simple matrix multiplication

$$\begin{pmatrix} a' & b' \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a'a & a'b + b' \\ 0 & 1 \end{pmatrix}.$$

The identity element is clearly given by $a = 1$ and $b = 0$, and the inverse element is given by $a' = 1/a$ and $b' = -b/a$. The two infinitesimal generators are given by

$$X_a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} , \qquad X_b = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} ,$$

and the commutator of these two generators is

$$[X_a, X_b] = X_b .$$

Indeed, every non-Abelian, two-parameter Lie algebra can be brought to this form by a suitable coordinate transformation. Moreover it follows that

$$\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} = \exp\left\{ \ln(a) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right\} , \qquad \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \exp\left\{ b \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \right\} .$$

Even more enlightening is the relation

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} = e^{b X_b} e^{\ln(a) X_a} = e^{\ln(a) X_a} e^{(b/a) X_b} = e^{\ln(a) X_a + b \frac{(a-1)}{\ln(a)} X_b} ,$$

which the reader is asked to verify in the Exercises.

Let us find the group invariant measures for the affine group. The starting point for the left-invariant group measure is given by

$$\begin{pmatrix} 1/a & -b/a \\ 0 & 1 \end{pmatrix} \left[\begin{pmatrix} da & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & db \\ 0 & 0 \end{pmatrix} \right] = \frac{da}{a} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{db}{a} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} ,$$

which leads to

$$d\mu_{left}(a, b) = \frac{da\, db}{a^2} .$$

A similar argument for the right-invariant group measure involves

$$\left[\begin{pmatrix} da & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & db \\ 0 & 0 \end{pmatrix} \right] \begin{pmatrix} 1/a & -b/a \\ 0 & 1 \end{pmatrix} = \frac{da}{a} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + db \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} ,$$

which leads to

$$d\mu_{right}(a, b) = \frac{da\, db}{a} .$$

Observe that these measures are not equal, but they both integrate to ∞, asserting that the affine group is noncompact.

The final example we study is the Heisenberg group formed by 3×3 matrices of the form

$$\begin{pmatrix} 1 & 0 & 0 \\ q & 1 & 0 \\ a & p & 1 \end{pmatrix} ,$$

which enjoy the group multiplication rule that

$$\begin{pmatrix} 1 & 0 & 0 \\ q' & 1 & 0 \\ a' & p' & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ q & 1 & 0 \\ a & p & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ q'+q & 1 & 0 \\ a'+a+p'q & p'+p & 1 \end{pmatrix} .$$

The identity element is given by $q = p = a = 0$ and the inverse element is given by $q' = -q$, $p' = -p$, $a' = -a + qp$. The infinitesimal generators for this group are given by

$$X_q = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \; X_p = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} , \; X_a = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} ,$$

with the commutation relations given by

$$[X_q, X_p] = -X_a , \qquad [X_q, X_a] = 0 , \qquad [X_p, X_a] = 0 .$$

It is straightforward to show (see Exercises) that the left and right group invariant measures are equal and are given by

$$d\mu(p, q, a) = dp\, dq\, da ;$$

this group is also noncompact.

The two principal examples we have discussed—the affine group and the Heisenberg group—have been based on 2×2 and 3×3 matrices, respectively. However, it is important to appreciate that there are other realizations of these groups by matrices (or operators) with different dimensions. Even when this is the case, several aspects of our discussion also apply to all representations regardless of their dimension or mode of realization. In particular, assuming the group elements are parametrized in the same manner, the commutation relations of the infinitesimal generators will always be the same, having, in particular, the same structure constants. The second feature that will always be the same whatever representation is chosen (using the same coordinates) is the left and right group invariant measures. Thus it is often useful to choose a simple representation to evaluate these invariant aspects since their determination may be easier in certain representations.

In many applications, and especially for quantum mechanics, one is often interested in *unitary representations* of the group elements. We will examine this feature when we study quantum theory.

5.4 Outline of Abstract Quantum Mechanics

Armed with all the ammunition presented in this chapter, it is quite simple to state the basic abstract properties of quantum mechanics [Dir74, BEH]. The state of a quantum system is described as either a pure state or a mixed state. Either case can be represented by a *density matrix* ρ which is a Hermitian, nonnegative, trace class operator that is given by

$$\rho = \Sigma_{n=1}^{\infty} \rho_n |\psi_n\rangle\langle\psi_n| , \qquad \langle\psi_m|\psi_n\rangle = \delta_{mn} ,$$
$$\mathrm{Tr}(\rho) = \Sigma_{n=1}^{\infty} \rho_n \equiv 1 , \qquad 0 \le \rho_n \le 1 .$$

For any observable A, which needs to be a self-adjoint operator, the expectation of A in the state ρ is given by

$$\langle A\rangle \equiv \mathrm{Tr}(A\rho) = \Sigma_{n=1}^{\infty} \rho_n \langle\psi_n| A |\psi_n\rangle ,$$

and this number is interpreted as the expected value of the quantity represented by A in the state ρ averaged over many independent, identical measurements. A pure state exists if (say) $\rho_1 = 1$ and all other $\rho_n = 0$. In that case, $\rho = |\psi\rangle\langle\psi|$, a projection operator, and

$$\langle A\rangle \equiv \langle\psi| A |\psi\rangle .$$

Stated otherwise, in general $\rho^2 \le \rho$: if $\rho^2 = \rho$, then ρ describes a pure state; if $\rho^2 < \rho$, then the state is mixed.

As a self-adjoint operator, the observable A admits a spectral representation given by

$$A = \int a \, d\mathbf{E}(a) ,$$

where the integral runs over the spectrum of the operator A. The one-parameter family of unitary operators $\exp[itA]$ for all real t has an expectation given by

$$C(t) \equiv \mathrm{Tr}[e^{itA}\rho] = \int e^{ita} d\mu(a) ,$$

where $\mu(a)$ is a probability measure defined by $\mu(a) = \mathrm{Tr}[\mathbf{E}(a)\,\rho]$, which makes $C(t)$ into a characteristic function.

5.4.1 Schrödinger picture

Dynamics in quantum mechanics is generated by a self-adjoint operator $\mathcal{H}$, the system Hamiltonian. In turn, this operator is the generator of the unitary *evolution operator*

$$U(t) = e^{-i\mathcal{H}t/\hbar} ,$$

which defines a one-parameter unitary group for all t, $-\infty < t < \infty$. Adopting the Schrödinger picture, this evolution operator advances the state of the system from (say) time 0 to time t as

$$\rho(t) = U(t)\,\rho\,U(t)^{\dagger} .$$

As a consequence, the density operator satisfies the equation of motion given by

$$i\hbar\,\partial\rho(t)/\partial t = [\mathcal{H}, \rho(t)] \ .$$

For a pure state, this equation leads to the abstract Schrödinger equation

$$i\hbar\,\partial|\psi(t)\rangle/\partial t = \mathcal{H}|\psi(t)\rangle \ ,$$

as well as its adjoint.

It is important to stress that in this abstract form the description of quantum mechanics is completely *representation independent*. All representational versions of quantum mechanics must, ultimately, conform with such a representation-independent form in order to be correct. This fact shows why the bra-ket notation of Dirac is so very useful since it offers a simple prescription to ensure that any representation of the abstract equations will be fully compatible with them. As just one example of that prescription, let us introduce the basis vectors $\{|x\rangle\}$ that are the formal eigenvectors of the coordinate operator Q, namely, $Q|x\rangle = x\,|x\rangle$ for all $x \in \mathbb{R}$. In that case, Schrödinger's equation becomes

$$i\hbar\,\partial\langle x|\psi(t)\rangle/\partial t = \langle x|\mathcal{H}|\psi(t)\rangle = \int \langle x|\mathcal{H}|y\rangle\langle y|\psi(t)\rangle\,dy \ ,$$

using the resolution of unity that these states provide. If, in particular,

$$\mathcal{H} = (1/2m)\,P^2 + V(Q) \ ,$$

where P is the conjugate operator to Q and $[Q, P] = i\hbar\mathbb{1}$, as so often is the case, then the integral kernel $\langle x|\mathcal{H}|y\rangle$ has a differential operator representation that leads to the usual form of Schrödinger's equation given by

$$i\hbar\,\partial\psi(x,t)/\partial t = -(\hbar^2/2m)\,\partial^2\psi(x,t)/\partial x^2 + V(x)\,\psi(x,t) \ ,$$

where $\psi(x,t) \equiv \langle x|\psi(t)\rangle$.

When we come to discuss quantum mechanics in the next chapter, we shall make every effort to maintain close contact with the abstract operator formulation so that proper representations are ensured.

5.4.2 Heisenberg picture

Since we shall also make use of the Heisenberg picture of quantum mechanics in our further studies, it is appropriate to complete our brief survey of quantum mechanics by discussing this important concept. The time dependence of the expectation of an observable is given in the Schrödinger picture by the relation

$$\langle A\rangle(t) = \mathrm{Tr}[A\rho(t)] = \mathrm{Tr}[A\{e^{-i\mathcal{H}t/\hbar}\,\rho\,e^{i\mathcal{H}t/\hbar}\}] \ ,$$

which, thanks to the cyclic invariance of the trace operation, can be rewritten as

$$\langle A\rangle(t) = \mathrm{Tr}[\{e^{i\mathcal{H}t/\hbar}\, A\, e^{-i\mathcal{H}t/\hbar}\}\,\rho] \equiv \mathrm{Tr}[A(t)\,\rho]\ ,$$

where we now have removed time dependence from the density matrix and introduced time dependence to the operator A, as given by

$$A(t) \equiv e^{i\mathcal{H}t/\hbar}\, A\, e^{-i\mathcal{H}t/\hbar}\ .$$

Note well the different order of the evolution operators for $\rho(t)$ and for $A(t)$. It follows that a time-dependent operator satisfies the differential equation

$$i\hbar\,\partial A(t)/\partial t = [A(t), \mathcal{H}]\ .$$

In this fashion temporal evolution has been passed completely from the states of the system to the operators of the system. In a similar way, it is also possible to have intermediate pictures as well, in which part of the temporal behavior (say the "interaction" part) resides with the states, and the rest of the temporal behavior (say the "free" part) resides with the operators.

Exercises

5-1 Consider the semi-infinite square matrix T acting on the sequence space l^2 defined as

$$T = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & \cdot & \cdot \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}\ ,$$

which is composed of $n \times n$ unit matrix blocks, $n = 1, 2, 3, \ldots$, arranged in order and with their uppermost unit being in the first row. Find the domain of the matrix T, $\mathfrak{D}(T)$. Find the domain of the adjoint matrix $T^\dagger$, $\mathfrak{D}(T^\dagger)$.

5-2 The Euler angles parametrize a general proper rotation according to the expression

$$R(\psi, \theta, \phi) \equiv e^{\psi X_3}\, e^{\theta X_2}\, e^{\phi X_3}\ ,$$

despite the fact that this parametrization becomes degenerate when $\theta = 0$. Find the expression for the left- and right-invariant group measures in these

coordinates. If $0 \leq \psi < 2\pi$, $0 \leq \theta \leq \pi$, and $0 \leq \phi < 2\pi$, determine the volume of the proper rotation group SO(3) and the volume of the full rotation group O(3).

5-3 For the nondegenerate parametrization of the proper rotation group given by

$$\tilde{R}(\alpha, \theta, \phi) \equiv e^{\alpha X_1}\, e^{\theta X_2}\, e^{\phi X_3} \ ,$$

determine the range of the angles involved, and then find the left- and right-invariant group measures as well as the volume of this form of the group SO(3).

5-4 Determine the left- and right-invariant group measures for the three-parameter Heisenberg group.

5-5 For the two-parameter affine group, verify the relations

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} = e^{bX_b}\, e^{\ln(a)\, X_a} = e^{\ln(a)\, X_a}\, e^{(b/a)\, X_b} = e^{\ln(a)\, X_a \,+\, b\frac{(a-1)}{\ln(a)}\, X_b} \ ,$$

where

$$X_a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \ , \qquad X_b = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \ .$$

5-6 For $N \times N$ matrices, $N < \infty$, show that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ and how this relation implies that $\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA) = \mathrm{Tr}(CAB)$. When $N = \infty$, find conditions on such matrices so that the same permutation invariance holds. Discuss how these conditions apply, or fail to apply, to the trace of the Heisenberg commutation relation $[Q, P] = i\hbar\mathbb{1}$.

6

Quantum Mechanical Path Integrals

6.1 Configuration Space Path Integrals

6.1.1 Schrödinger equation (special case)

The fundamental equation of quantum mechanics for a single particle is generally taken to be a Schrödinger equation of the form

$$i\hbar \frac{\partial}{\partial t}\psi(x,t) = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x,t) \right] \psi(x,t) \ ,$$

subject to some initial condition such as

$$\lim_{t \to t'} \psi(x,t) = \psi(x,t') \ .$$

One particular solution, generally called the propagator, refers to the solution of the Schrödinger equation subject to the initial condition

$$\lim_{t \to t'} \psi(x,t) = \delta(x - x') \ ,$$

and the solution subject to this initial condition is normally called

$$K(x,t;x',t') \ .$$

Since Schrödinger's equation is linear, the solution for a general initial condition can be derived from the propagator by means of the relation

$$\psi(x,t) = \int K(x,t;x',t')\, \psi(x',t')\, dx' \ .$$

Some basic properties of this important quantity are given by:

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9_6,

$$1) \qquad \lim_{t''\to t'} K(x'',t'';x',t') = \delta(x''-x') \ ,$$

$$2) \qquad K(x'',t'';x',t') = \int K(x'',t'';x,t)\, K(x,t;x',t')\, dx \ ,$$

$$3) \qquad K(x'',t'';x',t')^* = K(x',t';x'',t'') \ ,$$

$$4) \qquad \delta(x''-x') = \int K(x'',t';x,t)\, K(x,t;x',t')\, dx' \ .$$

In 1948 [Fey48] Richard Feynman, guided in part by ideas of Dirac [Dir33], proposed that the propagator could be written as

$$K(x'',t'';x',t') = \mathcal{N} \int_{x(t')=x'}^{x(t'')=x''} e^{(i/\hbar)\int_{t'}^{t''} [\,\frac{1}{2} m\dot{x}(t)^2 - V(x(t))]\, dt}\ \mathcal{D}x \ .$$

This expression is very similar to analogous expressions from Part 1 dealing with functional integrals such as the Feynman–Kac formula; on the other hand, the appearance of $i = \sqrt{-1}$ and the fact that the integrand is entirely oscillatory, as opposed to everywhere positive, makes for a vast difference in the mathematical properties in the two cases. As we shall see, in strong contrast to the Feynman–Kac formula, there is *no* underlying measure on paths that directly relates to the foregoing path integral expression.

While the path integral expression is strictly formal and cries out for some form of definition, it is nevertheless not without interest. From a heuristic viewpoint, it is true that the path integral expression involves the classical action functional for a particle of mass m moving under the influence of the potential $V(x,t)$. As such, the path integral formulation of quantum mechanics was the first and is the most well-known procedure to use the Lagrangian rather than the Hamiltonian in approaching quantization. Nevertheless, the path integral expression is so formal that it is hard to argue that it is anything more than a heuristic, semiclassical formula; consequently, to proceed further we need to provide an honest definition of what this formal expression represents.

The usual procedure to give meaning to the formal path integral is to first impose a lattice regularization in which the time axis is broken into $L+1$ segments of equal length—chosen equal for convenience—given by $\epsilon = (t''-t')/(L+1)$, and the path integral is then formulated as an integral over the L values of the coordinate, i.e., $x_l = x(l\epsilon)$, where $x_0 = x'$ and $x_{L+1} = x''$. In so doing we replace the formal paths by piecewise linear paths and interpret the integration over all paths as an integration over the path values at the lattice points. This formulation leads to reinterpreting Feynman's formal path integral as

$$K(x'',t'';x',t') = \lim_{L\to\infty} N_L \int \exp\{(i/\hbar)\Sigma_{l=0}^{L}[(m/2\epsilon)(x_{l+1}-x_l)^2 - \epsilon V(x_l, l\epsilon)]\}\ \Pi_{l=1}^{L}\, dx_l \ ,$$

where as $L \to \infty$, $\epsilon \to 0$ so that $\epsilon(L+1) = t'' - t'$. In addition, the normalization constant

$$N_L = \left(\frac{m}{2\pi i\epsilon\hbar}\right)^{(L+1)/2} .$$

6.1.2 The free particle

In this case we set $V = 0$, and as a result

$$K(x'', t''; x', t') = \lim_{L\to\infty} N_L \int e^{(i/\hbar)\Sigma_{l=0}^{L}(m/2\epsilon)(x_{l+1}-x_l)^2}\, \Pi_{l=1}^{L} dx_l \, .$$

This expression involves a number of repeated convolutions, and suggests that we invoke the Fourier transformation. In particular,

$$\int e^{ip_{l+1/2}(x_{l+1}-x_l)/\hbar - i\epsilon p_{l+1/2}^2/2m\hbar}\, dp_{l+1/2}/(2\pi\hbar) = (m/2\pi i\epsilon\hbar)^{1/2} e^{im(x_{l+1}-x_l)^2/2\epsilon\hbar} \, .$$

If we use this relation in the lattice form of the path integral for the free particle, we are led to consider

$$K(x'', t''; x', t') = \lim_{L\to\infty} \int e^{(i/\hbar)\Sigma_{l=0}^{L} p_{l+1/2}(x_{l+1}-x_l)/\hbar - i\epsilon p_{l+1/2}^2/2m\hbar} \times \Pi_{l=0}^{L} dp_{l+1/2}/(2\pi\hbar)\ \Pi_{l=1}^{L} dx_l \, .$$

(**Remark:** This equation offers an accidental preview of a (fairly simple) path integral as an integral over *phase space paths*, but we do not pursue that direction here. The subject of phase space path integrals will be addressed in detail in later sections.) Integrate over the x variables first, and we are led to the expression

$$\begin{aligned} K(x'', t''; x', t') &= (1/2\pi\hbar) \int e^{i(x'' p_{L+1/2} - x' p_{1/2})/\hbar} \Pi_{l=1}^{L} \delta(p_{l+1/2} - p_{l-1/2}) \\ &\qquad \times e^{-i\epsilon\Sigma_{l=0}^{L} p_{l+1/2}^2/2m\hbar}\ \Pi_{l=0}^{L} dp_{l+1/2} \\ &= (1/2\pi\hbar) \int e^{ip(x''-x')/\hbar - i(t''-t')p^2/2m\hbar}\, dp \\ &= [m/2\pi i(t''-t')\hbar]^{1/2}\ e^{im(x''-x')^2/2\hbar} \, . \end{aligned}$$

These calculations have much in common with those carried out for the Wiener process. However, there is a genuine distinction between the two cases as we shall demonstrate subsequently.

6.1.3 Quadratic path integrals

There are several other path integrals that involve Gaussian integrals such as that for the free particle. It is useful, therefore, to offer a rather general discussion of such integrals that is widely applicable.

Consider multivariable integrals of the form

$$I_L = \int \cdots \int e^{\Sigma_{l=1}^{L} c_l x_l - \frac{1}{2}\Sigma_{l,m=1}^{L,L} x_l A_{lm} x_m} \; \Pi_{l=1}^{L} dx_l \; ,$$

where $L < \infty$. Here $A_{lm} = A_{ml}$ is a symmetric, positive-definite matrix. As such it is possible to diagonalize the matrix $A = \{A_{lm}\}$ by an orthogonal transformation. Let $x_l = \Sigma_r O_{lr} y_r$, where $O = \{O_{lr}\}$ is an orthogonal matrix, and therefore $O^T O = \mathbb{1}$, where the superscript T denotes transpose and $\mathbb{1}$ denotes the unit matrix. Let O be chosen so that $O^T A O = D$, where D denotes a diagonal matrix with elements $\{d_l\}_{l=1}^{L}$ on the diagonal. Since A is positive definite, it follows that $d_l > 0$ for all l. The Jacobian of the transformation from the x variables to the y variables is unity, and thus

$$I_L = \int \cdots \int e^{\Sigma_{l=1}^{L} k_l y_l - \frac{1}{2}\Sigma_{l=1}^{L} d_l y_l^2} \; \Pi_{l=1}^{L} dy_l \; ,$$

where we have set $k_r \equiv \Sigma_l c_l O_{lr}$. In this form the integral is a product of L independent Gaussians which is evaluated as

$$\begin{aligned} I_L &= \Pi_{l=1}^{L} (2\pi/d_l)^{1/2} \; e^{\frac{1}{2}\Sigma_{l=1}^{L} d_l^{-1} k_l^2} \\ &= \sqrt{(2\pi)^L / \det(A)} \; e^{\frac{1}{2}\Sigma_{l,m=1}^{L} c_l A_{lm}^{-1} c_m} \; , \end{aligned}$$

the last line following from the fact that

$$\det(O^T A O) = \det(O^T) \det(A) \det(O) = \det(A) \; .$$

Observe in this calculation that the exponent term $\frac{1}{2}\Sigma_{l,m=1}^{L} c_l A_{lm}^{-1} c_m$ in the final result can be obtained as the extremal element in the integral. As an example, consider the exponent of the original integral

$$\Sigma_{l=1}^{L} c_l x_l - \tfrac{1}{2}\Sigma_{l,m=1}^{L,L} x_l A_{lm} x_m \; ,$$

and seek the extremal value of this expression. To find that extremal, we first set the derivative of the initial exponent with respect to x_l to zero, which leads to

$$c_l - \Sigma_{m=1}^{L} A_{lm} x_m = 0 \; ,$$

with the solution given by

$$x_m = \Sigma_{l=1}^{L} A_{lm}^{-1} c_l \; ,$$

and evaluate the initial exponent for this extremal value, leading to

$$\tfrac{1}{2}\Sigma_{l,m=1}^{L,L} c_l\, A_{lm}^{-1}\, c_m \ ,$$

which is exactly the exponent in the final answer.

As an application of this formula, let us consider the free particle. The original exponent in the integrand (in continuum form) is

$$\tfrac{1}{2}\int_{t'}^{t''} (m/\hbar)\dot{x}(t)^2\, dt \ .$$

The extremal of this expression is just the familiar

$$\ddot{x}(t) = 0 \ ,$$

with the solution

$$x(t) = x'(t'' - t)/(t'' - t') + x''(t - t')/(t'' - t') \ ,$$

consistent with the boundary conditions, which leads to

$$\dot{x}(t) = \frac{(x'' - x')}{(t'' - t')} \ .$$

The evaluation of the exponent for this extremal path gives

$$\frac{m}{2\hbar}\int_{t'}^{t''} \dot{x}(t)^2\, dt = \frac{m}{2\hbar}\,\frac{(x'' - x')^2}{(t'' - t')} \ .$$

This result applies to quantum mechanical path integrals for which the matrix is effectively pure imaginary. One only needs to introduce a small damping factor to ensure convergence of the integrals. Thus, on the basis of the previous discussion, the free particle path integral is

$$\mathcal{N}\int e^{\frac{1}{2}i(m/\hbar)\int \dot{x}(t)^2\, dt}\, \mathcal{D}x = A\, e^{i(m/2\hbar)(x'' - x')^2/(t" - t')} \ ,$$

where A denotes an amplitude factor. The correct expression for the amplitude factor can be determined as follows. If we denote the exponent on the right-hand side by

$$iS(x'', t''; x', t')/\hbar \ ,$$

then the correct amplitude factor is given by

$$A = \sqrt{\frac{i}{2\pi\hbar}\,\frac{\partial^2 S(x'', t''; x', t')}{\partial x''\, \partial x'}} \ .$$

Thus, for the free particle, we are led to

$$A = \sqrt{\frac{m}{2\pi i\hbar(t'' - t')}} \ .$$

6.1.4 Harmonic oscillator

Obtaining the propagator for the harmonic oscillator is an important exercise. Formally, the path integral for the harmonic oscillator is given by

$$\mathcal{N} \int e^{(i/\hbar)\int \frac{1}{2}[m\dot{x}(t)^2 - k\,x(t)^2]\,dt}\, \mathcal{D}x \ .$$

As a quadratic path integral, we can appeal to the general results described above. In particular, the extremal equation for the oscillator is

$$\ddot{x}(t) + \omega^2 x(t) = 0 \ ,$$

where $\omega = \sqrt{k/m}$, and the solution that fits the boundary conditions is given by

$$x(t) = [x'' \sin((t-t')\omega) + x' \sin((t''-t)\omega)]/[\sin((t''-t')\omega)] \ .$$

The time derivative yields the velocity

$$\dot{x}(t) = \omega[x'' \cos((t-t')\omega) - x' \cos((t''-t)\omega)]/[\sin((t''-t')\omega)] \ .$$

Evaluation of the exponent for the extremal path is made easier by the observation that

$$\int [\dot{x}^2 - \omega^2 x^2]\,dt = x\dot{x}| - \int [x(\ddot{x} + \omega^2 x)]\,dt = x\dot{x}| \ .$$

Use of this equation leads to

$$\int [\dot{x}^2 - \omega^2 x^2]\,dt = \omega[(x''^2 + x'^2)\cos((t''-t')\omega) - 2x''x']/\sin((t''-t')\omega) \ ,$$

and this leads to an exponent of the answer given by

$$\frac{im\omega}{\hbar \sin((t''-t')\omega)} \left[\frac{(x''^2 + x'^2)}{2} \cos((t''-t')\omega) - x''x' \right] .$$

The amplitude factor for the harmonic oscillator follows from this expression as

$$A = \sqrt{\frac{m\omega}{2\pi i\hbar \sin((t''-t')\omega)}} \ ,$$

and this leads to the propagator for the harmonic oscillator given by

$$\begin{aligned} K(x'',t'';x',t') &= \sqrt{\frac{m\omega}{2\pi i\hbar \sin((t''-t')\omega)}} \\ &\quad \times e^{(im\omega/\hbar)[\frac{1}{2}(x''^2 + x'^2)\cos((t''-t')\omega) - x''x']/\sin((t''-t')\omega)} \ . \end{aligned}$$

6.1.5 Eigenfunctions and eigenvalues

The interpretation of the propagator for a dynamical system with a discrete spectrum (for convenience) is given by the chain (with $T = t'' - t'$)

$$\begin{aligned} K(x'', t''; x', t') &= \langle x'', t''|x', t'\rangle \\ &= \langle x''|e^{-i\mathcal{H}T/\hbar}|x'\rangle \\ &= \sum_n \langle x''|n\rangle e^{-iE_nT/\hbar}\langle n|x'\rangle \\ &= \sum_n \phi_n(x'')\phi_n^*(x')e^{-iE_nT/\hbar} . \end{aligned}$$

A similar relation holds for imaginary time quantum mechanics, which arises when the variable T is rotated in the complex plane to $-iT$. In that case one deals with the expression

$$\begin{aligned} \langle x''|e^{-\mathcal{H}T/\hbar}|x'\rangle &= \sum_n \langle x''|n\rangle e^{-E_nT/\hbar}\langle n|x'\rangle \\ &= \sum_n \phi_n(x'')\phi_n^*(x')e^{-E_nT/\hbar} . \end{aligned}$$

Assuming that the eigenvectors are nondegenerate and that the energy eigenvalues are ordered so that

$$E_0 < E_1 < E_2 < \ldots < \ldots ,$$

it follows when $T \gg 1$, that

$$\sum_n \phi_n(x'')\phi_n^*(x')e^{-E_nT/\hbar} \simeq \phi_0(x'')\phi_0^*(x')e^{-E_0T/\hbar} + \text{l.o.t.} ,$$

where l.o.t. denotes lower order terms. This relation shows that one can identify the low-order eigenfunctions and eigenvalues by taking appropriate limits of the imaginary time propagator.

To see such a formula in action, let us consider the case of the harmonic oscillator for which

$$\begin{aligned} &\sqrt{\frac{m\omega}{2\pi\hbar\sinh(T\omega)}}\, e^{-(m\omega/\hbar)[\frac{1}{2}(x''^2 + x'^2)\coth(T\omega) - x''x'\operatorname{csch}(T\omega)]} \\ &\qquad = \sum_{n=0}^{\infty} \phi_n(x'')\phi_n^*(x')\, e^{-TE_n/\hbar} . \end{aligned}$$

For large T, we can set $\sinh(T\omega) \simeq \frac{1}{2}e^{T\omega}$, $\coth(T\omega) \simeq 1$, and $\operatorname{csch}(T\omega) \simeq 2e^{-T\omega}$, leading to

$$\sum_{n=0}^{\infty} \phi_n(x'')\,\phi_n^*(x')\,e^{-TE_n/\hbar}$$
$$\simeq \sqrt{\frac{m\omega}{\pi\hbar}}\,e^{-T\omega/2}\,e^{-(m\omega/\hbar)[(x''^2+x'^2)/2\,-\,2x''x'e^{-T\omega}]}\,.$$

Expanding to leading powers we find that

$$\sum_{n=0}^{\infty} \phi_n(x'')\,\phi_n^*(x')\,e^{-TE_n/\hbar}$$
$$\simeq \sqrt{\frac{m\omega}{\pi\hbar}}\,e^{-T\omega/2}\,e^{-(m\omega/\hbar)[(x''^2+x'^2)/2]}$$
$$\times\{1+2(m\omega/\hbar)\,x''x'\,e^{-T\omega}+\cdots\}\,,$$

from which we conclude that

$$\phi_0(x)=\left(\frac{m\omega}{\pi\hbar}\right)^{1/4}e^{-(m\omega/2\hbar)x^2}\,,\qquad E_0=\hbar\omega/2\,,$$
$$\phi_1(x)=2^{1/2}\left(\frac{m\omega}{\pi\hbar}\right)^{3/4}x\,e^{-(m\omega/2\hbar)x^2}\,,\qquad E_1=3\hbar\omega/2\,.$$

Further eigenfunctions and eigenvalues follow from a more complete analysis of the large T behavior.

6.1.6 Connection with operators

The abstract Schrödinger equation for select Hamiltonian operators reads

$$i\hbar\,\partial|\psi\rangle/\partial t\;=[(1/2m)\,P^2+V(Q)]\,|\psi\rangle$$
$$\equiv\mathcal{H}|\psi\rangle\,.$$

The solution to this equation may formally be expressed by

$$|\psi(t)\rangle=e^{-i\mathcal{H}t/\hbar}|\psi\rangle\,,$$

or in the Schrödinger representation as

$$\psi(x,t)=\langle x|e^{-i\mathcal{H}t/\hbar}|\psi\rangle\equiv\int\langle x|e^{-i\mathcal{H}t/\hbar}|x'\rangle\langle x'|\psi\rangle\,dx'\,.$$

The integral kernel

$$K(x'',t'';x',t')\equiv\langle x''|e^{-i\mathcal{H}(t''-t')/\hbar}|x'\rangle$$

is the propagator which takes the wave function at an initial time t' and yields the wave function at a final time t''. A multi-integral formula for the

propagator may be given by breaking the time interval $T \equiv t'' - t'$ into $N+1$ equal steps each of length $\epsilon \equiv T/(N+1)$. This leads to the expression (setting $\hbar = 1$ for this calculation)

$$\begin{aligned} K &= \langle x''|e^{-i\mathcal{H}\epsilon}\, e^{-i\mathcal{H}\epsilon} \cdots e^{-i\mathcal{H}\epsilon}|x'\rangle \\ &= \int \cdot \int \langle x''|e^{-i\mathcal{H}\epsilon}|x_N\rangle\langle x_N|e^{-i\mathcal{H}\epsilon}|x_{N-1}\rangle \cdots \langle x_1|e^{-i\mathcal{H}\epsilon}|x'\rangle \; \Pi_{n=1}^{N}\, dx_n \;, \end{aligned}$$

where $x_{N+1} \equiv x''$ and $x_0 \equiv x'$. This equation is valid for all N. For large N—and therefore for small ϵ—we may approximate the integrand as follows:

$$\begin{aligned} \langle x_{n+1}|e^{-i\epsilon\mathcal{H}}|x_n\rangle &\simeq \langle x_{n+1}|e^{-i\epsilon P^2/2m}\, e^{-i\epsilon V(Q)}|x_n\rangle \\ &= \int \langle x_{n+1}|e^{-i\epsilon P^2/2m}|p_{n+1/2}\rangle\langle p_{n+1/2}|e^{-i\epsilon V(Q)}|x_n\rangle \, dp_{n+1/2} \\ &= (1/2\pi)\int e^{ip_{n+1/2}(x_{n+1}-x_n) - i\epsilon p_{n+1/2}^2/2m}\, e^{-i\epsilon V(x_n)}\, dp_{n+1/2} \\ &= \sqrt{(m/2\pi i\epsilon)}\; e^{im(x_{n+1}-x_n)^2/2\epsilon - i\epsilon V(x_n)} \;. \end{aligned}$$

This approximation is valid for small ϵ and for a certain class of potentials V; just what this special class of potentials is will be discussed in the following section.

Combining our approximate formula for the short time propagator and restoring $\hbar$ to its proper place leads to the expression

$$\begin{aligned} &K(x'', t''; x', t') \\ &= \lim_{N\to\infty} A^{N+1} \int \cdot \int e^{(i/\hbar)[\Sigma_{n=0}^{N} m(x_{n+1}-x_n)^2/2\epsilon - \epsilon V(x_n)]}\; \Pi_{n=1}^{N}\, dx_n \;, \end{aligned}$$

where $A \equiv \sqrt{m/2\pi i\hbar\epsilon}$, and it has been necessary to introduce the limit in which $N \to \infty$ and $\epsilon \to 0$ such that $(N+1)\epsilon = T = t'' - t'$ because of the approximation made to the short time propagator. Thus we have arrived at the same result for the propagator conjectured earlier, but this time we have been guided by following where the abstract operator connection has led us.

It is certainly tempting—and physicists cannot resist such a temptation—to interchange the limit and the many integrals in the former expression. Such an interchange is indeed valid for the Wiener process discussed in Sec. 4.2, but it is strictly speaking completely invalid in the present case. Nevertheless, we too find it irresistible to interchange those operations and we do so and write for the integrand the form it would take for continuous and differentiable paths. This then leads to the formal expression given by

$$K(x'', t''; x', t') = \mathcal{N} \int e^{(i/\hbar)\int [m\dot{x}(t)^2/2 - V(x(t))]\, dt}\; \mathcal{D}x \;.$$

Such an expression has only heuristic meaning—which by itself is not insignificant—but it begs for a clear-cut definition in order to give a precise meaning. The formulation above, which amounts to making a Riemann

sum approximation to the integral in the exponent and then integrating over the points in that approximation, is called a *lattice space regularization* or a *temporal lattice regularization*. Later in this monograph we shall have occasion to introduce an additional form of regularization that is fundamentally different from the lattice space regularization.

6.1.7 Validity of the lattice space regularization

In the previous subsection we made use of the approximation that

$$\lim_{N\to\infty}\left[e^{-i\epsilon P^2/2m}\,e^{-i\epsilon V(Q)}\right]^{N+1} = e^{-i[P^2/2m+V(Q)]T}\ ,$$

where $(N+1)\epsilon \equiv T$. This result is known as the Trotter product formula [Tro58], and it is true whenever $\mathfrak{D}(P^2)\cap\mathfrak{D}(V(Q))$ is a *core* for the operator $P^2/2m+V(Q)$. Here, $\mathfrak{D}(P^2)$ refers to the *domain* of Hilbert space vectors for the operator P^2, and $\mathfrak{D}(V(Q))$ likewise refers to the domain of Hilbert space vectors for the operator $V(Q)$. The intersection of those domains, i.e., the set of vectors in both domains, should be a core for the final operator which means that the action of the operator on vectors in that intersected domain should be sufficient to uniquely define the operator. A few examples may help make the definition more clear, and for these we let $2m=1$ for simplicity. As a first example, we observe that for P^2+Q^4 the result is correct, namely, that the operator P^2+Q^4 restricted to the domain $\mathfrak{D}(P^2)\cap\mathfrak{D}(Q^4)$ is uniquely determined. However, the operator P^2-Q^4 is *not* uniquely determined by its action on the limited domain $\mathfrak{D}(P^2)\cap\mathfrak{D}(Q^4)$. In a related terminology one says that P^2+Q^4 is essentially self-adjoint on the domain $\mathfrak{D}\equiv\mathfrak{D}(P^2)\cap\mathfrak{D}(Q^4)$, while the operator P^2-Q^4 is *not* essentially self-adjoint on that domain. The reason for this different behavior arises because for the sum operator the domain $\mathfrak{D}$ is the maximal domain, while for the difference operator that is not the case and the information given by its action on $\mathfrak{D}$ does not tell one how the operator should act on the rest of its true domain. Indeed, that ambiguity leads to various possible extensions of the difference operator with different actions on the additional domain of its definition. These amount to different self-adjoint extensions, which have, generally speaking, different eigenfunctions and eigenvalues.

If one ignores these niceties and nevertheless uses the lattice space prescription as described above when it is technically not allowed, the result may well be one in which probability is lost as time proceeds; it is as if the particle is being continually absorbed as time increases, and this implies a *non*unitary time evolution [Nel64].

6.1.8 Classical symptoms of quantum illnesses

It is noteworthy that this different behavior for the sum and difference operators, $P^2\pm Q^4$, has an image in the associated *classical* dynamical behavior

of these two quantum systems. For the sum system, the classical Hamiltonian reads $p^2 + q^4$, and the equations of motion lead to

$$E = \tfrac{1}{4}\dot{q}^2 + q^4 \ ,$$

where $E > 0$ denotes the energy which is a constant of the motion. The integral

$$t - t_0 = \int dt = \int dq/\sqrt{4(E - q^4)}$$

yields an implicit formula for $q(t)$. It is clear that this expression leads to oscillatory motion for $q(t)$ between two turning points where $q^2 = \sqrt{E}$. On the other hand, for the difference system, the energy equation reads

$$E = \tfrac{1}{4}\dot{q}^2 - q^4 \ ,$$

and E can have any value. The integral

$$t - t_0 = \int dt = \int dq/\sqrt{4(E + q^4)}$$

leads to an implicit formula for $q(t)$, but there is a profound difference between the present case and the former one. Suppose that $q^4 > -E$, and for the sake of illustration that $q > 0$. It follows that q will continue to grow and at some point

$$t^* - t_0 = \int dt = \int^{\infty} dq/\sqrt{4(E + q^4)} \ ,$$

namely, q will reach *infinity*, $q = \infty$, at a *finite time* t^*. The classical theory has reached an impasse and some choice must be made, e.g., relaunching the particle back from infinity toward the origin at $t = t^*$, and repeating this or some other such procedure when the particle again reaches plus or minus infinity in order to ensure a full time solution for $q(t)$. The ambiguity in this classical procedure to obtain a full time classical solution is reflected in the quantum theory where boundary conditions at $x = \pm\infty$ are needed to fully specify the chosen self-adjoint extension of the operator $P^2 - Q^4$.

Of course, one is also free in the classical domain, when a hypothetical stream of particles evolves and at a various times some of them reach either plus or minus spatial infinity, to declare that (say) 10% of the particles remain there and are thus lost to the system; if this loss procedure is maintained as time proceeds, then one has the analog of a nonunitary quantum evolution exhibited in the classical realm.

This sort of comparison of the problem that both a classical and quantum theory may share has been more extensively studied in reference [ZK93].

6.1.9 The question of a measure for the Feynman path integral

Ever since the introduction of the path integral for quantum systems by Feynman in 1948, there has been the question of whether there exists an underlying measure on path space that defines such an integral. For example, analogous to the Feynman–Kac formula for imaginary time path integrals, is there a genuine measure $\tau(x)$ such that the propagator can be written as

$$K(x'', t''; x', t') = \int e^{-(i/\hbar)\int V(x(t))\,dt}\; d\tau(x)\ ?$$

It is straightforward to demonstrate that no such genuine measure $\tau(x)$ can exist. The putative measure in question would then require a well-defined existence, for suitable sets S, of the expression

$$\tau(S) = \lim_{N\to\infty} A^{N+1} \int_S e^{(i/\hbar)\Sigma_{n=0}^{N} m(x_{n+1}-x_n)^2/2\epsilon}\; \Pi_{n=1}^{N}\, dx_n\ .$$

For such a measure to exist it must involve *absolute convergence*, i.e., eliminating the possibility that a finite result may have arisen as a delicate cancellation between positive and negative infinities. However, absolute convergence requires that

$$\begin{aligned}\lim_{N\to\infty} |A|^{N+1} \int_S |\, e^{(i/\hbar)\Sigma_{n=0}^{N} m(x_{n+1}-x_n)^2/2\epsilon}\,|\; \Pi_{n=1}^{N}\, dx_n \\ = \lim_{N\to\infty} |A|^{N+1} \int_S \Pi_{n=1}^{N}\, dx_n\end{aligned}$$

should be *finite*; but for any set S, other than a set of measure zero, this integral diverges due to the prefactor.

Note well, that in analyzing formal path integrals, such as

$$\mathcal{N} \int e^{(i/\hbar)\int [m\dot{x}(t)^2/2 - V(x(t))]\,dt}\; \mathcal{D}x\ ,$$

many authors refer to the "paths" and their "properties" despite the fact that there are no true "paths"! This situation is in complete contrast to the Wiener measure and the integrals in the Feynman–Kac formula for which there do exist paths and their properties are a legitimate topic of discussion, one about which much can be said.

6.1.10 Proposal of Gel'fand and Yaglom to introduce a measure

Since the raw definition of the path integral itself lacks the properties to directly define a genuine measure, efforts have been made over the years to introduce a genuine measure through a suitable form of regularization, and this discussion introduces us to the subject of a *continuous-time regularization*

of the Feynman path integral. The first such discussion of this type is that of Gel'fand and Yaglom [GY60]. Their proposal was to consider the expression

$$\lim_{\nu\to\infty} \mathcal{N} \int e^{(i/\hbar)\int [m\dot{x}(t)^2/2 - V(x(t))]\,dt}\; e^{-(1/2\nu)\int \dot{x}(t)^2\,dt}\; \mathcal{D}(x)\;.$$

Observe that these authors have introduced the distribution appropriate for a Wiener measure into the integrand in which the real parameter ν plays the role of the *diffusion constant*, and sought to study the limit in which $\nu \to \infty$, a limit which may be called the *ultra-diffusive limit*. Heuristically, the inserted factor should reduce to unity as $\nu \to \infty$ and the result ought to yield the desired propagator for the quantum system. It was the thought of these authors to combine the kinetic energy factor and the inserted term, both of which have a similar integral, into a common factor and use that to define a new Wiener measure with a modified diffusion constant. This procedure would clearly work if the original factor in the integrand had a minus sign in place of the factor $i = \sqrt{-1}$, for then one would be combining two positive factors to yield a new Wiener measure distribution with a positive diffusion constant. Their extension of this argument was to see if a Wiener measure could be defined with a *complex* diffusion constant.

The issue boils down to whether the expression

$$\sigma(S) = \lim_{N\to\infty} \sigma_N(S) = \lim_{N+1} \tilde{A}^{N+1} \int_S e^{-(1/2\tilde{\nu})\,\Sigma_{n=0}^{N}(x_{n+1}-x_n)^2/\epsilon}\; \Pi_{n=1}^{N}\, dx_n$$

leads to a genuine measure. In this expression we have introduced $\tilde{\nu}$ defined by

$$1/\tilde{\nu} \equiv 1/\nu - im/\hbar$$

and

$$\tilde{A} \equiv 1/\sqrt{2\pi\tilde{\nu}\epsilon}\;;$$

recall that $A = \sqrt{m/2\pi i\hbar\epsilon}$ was the quantity being modified. For any $N < \infty$, it follows that $\sigma_N(\mathbb{R}^N) = 1$ and thus $\lim_{N\to\infty}\ \sigma_N(\mathbb{R}^N) = 1$ as well. Just as we did for the direct lattice space formulation, it is natural to ask whether this result is obtained through a cancellation of infinities or whether it is an absolutely convergent result. To that end, consider

$$\begin{aligned} |\tilde{A}^{N+1}| \int_{\mathbb{R}^N} & e^{-\frac{1}{2}\,[\Re\,\tilde{\nu}^{-1}]\,\Sigma_{n=0}^{N}(x_{n+1}-x_n)^2/\epsilon}\; \Pi_{n=1}^{N}\, dx_n \\ &= |\tilde{A}|\,\{|\tilde{\nu}|/[\Re\,\tilde{\nu}]\}^{N/2} \\ &= |\tilde{A}|\,\{1 + [\Im\,\tilde{\nu}]^2/[\Re\,\tilde{\nu}]^2\}^{N/4}\;, \end{aligned}$$

an expression which diverges as $N \to \infty$ whenever $[\Im\,\tilde{\nu}] \neq 0$; in this expression we have introduced the real and imaginary parts of $\tilde{\nu}$, namely, $[\Re\,\tilde{\nu}]$ and

$[\Im\tilde{\nu}]$, respectively. This argument, first given by Cameron [Cam62], demonstrates that the Gel'fand–Yaglom proposal does *not* lead to a genuine, countably additive measure but rather only to a finitely additive—and therefore nongenuine—measure, which sadly does not solve the problem.

6.1.11 Proposal of Itô to introduce a measure

To overcome the difficulties of the Gel'fand–Yaglom procedure, Itô [Itô61] suggested a modification of their proposal. One way to describe what Itô introduced can be seen in the expression (with $\hbar = 1$)

$$\lim_{\nu\to\infty} \mathcal{N} \int e^{i\int[(m/2)\dot{x}^2 - V(x)]\,dt}\, e^{-(1/2\nu)\int[\ddot{x}^2 + \dot{x}^2]\,dt}\, \mathcal{D}x \ .$$

This expression involves an Ornstein–Uhlenbeck (see Sec. 4.5) velocity process, a process that has continuous velocities $\dot{x}(t)$, but for which $\ddot{x}(t)$ is either $\pm\infty$ for almost all time. Such a process has smoothed out the paths so that the kinetic energy is finite for almost all paths for every $\nu < \infty$, thus overcoming the problem that proved insoluble for Gel'fand and Yaglom, and ensuring that in the continuum limit there is a genuine measure for all $0 < \nu < \infty$. To pin the measure at the initial and final times, it is necessary to specify $x(T), \dot{x}(T)$ and $x(0), \dot{x}(0)$, namely, *four* boundary conditions rather than the normal two, i.e., just $x(T), x(0)$. This conditioning does not conform with the Schrödinger representation which involves just the position and not the position and velocity together. However, at this intermediate stage, it is not necessary to solve the Schrödinger equation nor even be consistent with the Schrödinger representation. In the limit that $\nu \to \infty$, and for suitable potentials, it turns out that the dependence on the end point velocities, $\dot{x}(T)$ and $\dot{x}(0)$, can be chosen as functions of the end point positions, $x(T)$ and $x(0)$, to yield the desired result. Itô proved this relation yields a propagator $K(x'', T; x', 0)$ that satisfies the Schrödinger equation for potentials of the form $V(x) = bx + \frac{1}{2}kx^2 + U(x)$, where $b \in \mathbb{R}$, $k \geq 0$, and where—analyzed by means of a perturbation expansion—the function $U(x)$ should be the Fourier transform of a bounded, possibly complex, measure, i.e.,

$$U(x) = \int e^{ix\kappa}\, dm(\kappa) \ , \qquad \int d|m(\kappa)| < \infty \ .$$

The reason this approach succeeds is fairly clear. Consider the case of a free particle plus a quadratic potential including a linear term. As a quadratic integral over all, the exponent of the result of the functional integration can be obtained just by finding the solution of the extremal of the action. When $\nu < \infty$, that extremal depends on all four pinned data points. The values of the end point velocities, $\dot{x}(T)$ and $\dot{x}(0)$, can be chosen so that in the limit as $\nu \to \infty$, only the physically important, ν-independent, terms contribute. The proper amplitude factor then follows from this exponent as shown above (Sec. 6.1.3). The perturbation expansion of the potential $U(x)$ involves only

additional linear terms in the potential, which suggests that suitable potentials can be included in this fashion.

In summary, Itô's prescription does offer a way to introduce a measure into the formal configuration expression. In Chapter 8 we will offer another way to introduce a measure into path integral expressions that is much more general and also, we believe, physically more appealing.

6.2 Phase Space Path Integrals

In 1951 Feynman [Fey51] proposed a second formal path integral expression for the propagator that is an integral over paths in phase space given by

$$K(q'', t''; q', t') = \mathcal{M} \int e^{(i/\hbar)\int [p\dot{q} - H(p,q)]\, dt}\, \mathcal{D}p\, \mathcal{D}q \; .$$

In this expression one is instructed to integrate over all paths $q(t)$ subject to the conditions that $q(t') = q'$ and $q(t'') = q''$. Furthermore, the paths $p(t)$ are to be integrated over without any values being held fixed. The phase space path integral implicitly includes the original, configuration space path integral since if the Hamiltonian has the form

$$H(p,q) = p^2/2m + V(q) \; ,$$

then a Gaussian integral over the momentum variables formally leads to the original, configuration space path integral

$$K(q'', t''; q', t') = \mathcal{N} \int e^{(i/\hbar)\int [m\dot{q}(t)^2/2 - V(q(t))]\, dt}\, \mathcal{D}q \; .$$

However, although the phase space path integral as given is formal and needs to be defined, it has the implicit virtue that it may apply to a wider class of dynamical systems since there is no restriction that the Hamiltonian have the form used above to derive the configuration space form. As we shall see, the phase space path integral is indeed much more general than the configuration space version.

Just like the configuration space path integral, the phase space version has a suggestive passage to the classical theory on the basis of a stationary phase argument. In particular, as $\hbar$ approaches zero, the major contributions to either form of the path integral are formally given by paths for which the action is stationary. Such a rule leads directly to the classical equations of motion, and thus to classical mechanics in either the Lagrangian or the Hamiltonian form.

6.2.1 Momentum space propagator

It is instructive to repeat the discussion about the phase space path integral, but this time do it for the momentum space propagator. Let us start with the Schrödinger equation in the momentum representation, which reads

$$i\hbar\partial\psi(p,t)/\partial t = \mathcal{H}(p, i\hbar\partial/\partial p)\,\psi(p,t)\ .$$

As a first-order equation with respect to time, the solution to this equation can be given with the help of the momentum space propagator. In particular, it follows that

$$\psi(p'',t'') = \int K(p'',t'';p',t')\,\psi(p',t')\,dp'\ ,$$

where $K(p'',t'';p',t')$ denotes the desired propagator, which has a formal phase space path integral representation given (for $T = t'' - t'$) by

$$K(p'',T;p',0) = \mathcal{M}\int e^{-(i/\hbar)\int[q\dot{p}+H(p,q)]\,dt}\,\mathcal{D}p\,\mathcal{D}q\ .$$

As a boundary condition, we impose

$$\lim_{T\to 0} K(p'',T;p',0) = \delta(p''-p')\ .$$

The most direct way to derive an operator expression for the proposed propagator is to simply Fourier transform the two end points of the configuration space version of the phase space path integral. In particular, we propose to use

$$K(p'',T;p',0) = \frac{1}{2\pi\hbar}\int e^{-ip''q''/\hbar}\,K(q'',T;q',0)\,e^{ip'q'/\hbar}\,dq''\,dq'\ .$$

Combining this expression with the earlier formulation of the configuration space form of the phase space path integral leads to the expression (setting $\hbar = 1$ for now)

$$\begin{aligned}K(p'',T;p',0) &= \langle p''|e^{-i\mathcal{H}\epsilon}\,e^{-i\mathcal{H}\epsilon}\cdots e^{-i\mathcal{H}\epsilon}|p'\rangle\\ &= \frac{1}{2\pi}\int e^{-i(p''q_{N+1}-p'q_0)}\,dq_{N+1}\,dq_0\int\cdots\int\\ &\quad\times\langle q_{N+1}|e^{-i\mathcal{H}\epsilon}|q_N\rangle\langle q_N|e^{-i\mathcal{H}\epsilon}|q_{N-1}\rangle\cdots\langle q_1|e^{-i\mathcal{H}\epsilon}|q_0\rangle\ \mathit{\Pi}_{n=1}^{N}\,dq_n\ ,\end{aligned}$$

where $T = \epsilon(N+1)$. Inserting intermediate resolutions of unity, $\mathbb{1} = \int|p\rangle\langle p|\,dp$, we are led to

$$\begin{aligned}K(p'',T;p',0) &= \int\cdots\int\langle p_{N+3/2}|q_{N+1}\rangle\langle q_{N+1}|e^{-i\mathcal{H}\epsilon}|p_{N+1/2}\rangle\cdots\\ &\quad\times\langle q_1|e^{-i\mathcal{H}\epsilon}|p_{1/2}\rangle\langle p_{1/2}|q_0\rangle\langle q_0|p_{-1/2}\rangle\ \mathit{\Pi}_{n=0}^{N}\,dp_{n+1/2}\ \mathit{\Pi}_{n=0}^{N+1}\,dq_n\ ,\end{aligned}$$

where $T = (N+1)\epsilon$, $p'' = p_{N+3/2}$, and $p' = p_{-1/2}$. For small ϵ we introduce the approximation

$$\langle q|e^{-i\mathcal{H}\epsilon}|p\rangle \simeq \langle q|\,[1-i\epsilon\mathcal{H}]\,|p\rangle = \langle q|p\rangle\,[1-i\epsilon H(p;q)] \simeq \langle q|p\rangle\,e^{-i\epsilon H(p;q)}\ ,$$

where $H(p;q) \equiv \langle q|\mathcal{H}|p\rangle/\langle q|p\rangle$.

Insertion of this last expression into the multiple integral for the propagator leads, effectively, to

$$K(p'',T;p',0) = \lim_{N\to\infty} \frac{1}{(2\pi)^{N+2}} \int\cdots\int \times e^{-i\Sigma_{n=-1}^{N}[q_{n+1}(p_{n+3/2}-p_{n+1/2})+\epsilon H(p_{n+1/2};q_{n+1})]} \times \Pi_{n=0}^{N}\, dp_{n+1/2}\, \Pi_{n=0}^{N+1}\, dq_n \ .$$

Note in this case that there is one more q integration than p integration, which is just the opposite situation that we found for the configuration space version of the phase space path integral.

Interchanging the limit and integration, and writing for the integrand the form it takes for continuous and differentiable paths leads to the formal expression (restoring $\hbar$)

$$K(p'',T;p',0) = \mathcal{M}\int e^{-(i/\hbar)\int[q\dot{p}+H(p,q)]\,dt}\ \mathcal{D}p\,\mathcal{D}q \ ,$$

where we have set $H(p,q) = H(p;q)$.

6.2.2 Physical interpretation of path integrals

We have focused much attention on the mathematical interpretation of the path integral expressions for the quantum mechanical propagator in both its coordinate and momentum space formulations. However, we have paid scant attention to the physical interpretation of these expressions.

We note first of all that the physical interpretation of the symbol q is that of a sharp eigenvalue as stems from the equation $Q|q\rangle = q|q\rangle$; likewise for p, namely, $P|p\rangle = p|p\rangle$. The Heisenberg uncertainty principle asserts that it is not possible to diagonalize both P and Q at the same time, and it is for this very reason that we have carefully indexed both p and q in our lattice version of the path integrals in distinct ways indicating that they indeed do not pertain to the same time. In the lattice versions, therefore, the approximate paths run from sharp q, to sharp p, to sharp q, to sharp p, etc. Whenever there is a sharp q there is a *completely unsharp* p, and vice versa. Thus the q paths, for example, go from sharp, to unsharp, to sharp, to unsharp, etc. Likewise for the p paths. This alternating behavior persists for all values of $\epsilon > 0$ and all values of $\hbar > 0$. In the continuum limit the paths $q(t)$ and $p(t)$ alternate every instant, and when the value of q is known, the value of p is completely unknown, and vice versa. This holds for any value of $\hbar$ and thus continues to hold in the limit as $\hbar \to 0$.

As a consequence, we are forced to conclude that, despite appearances, the physical nature of the "paths" $q(t)$ and $p(t)$ which enter the phase space version of either the formal configuration space or the formal momentum

space path integrals, is *not* connected with the physical nature of the usual classical paths as we would like to believe. (**Remark:** In Chapter 7, using an *alternative* phase space path integral construction, we will establish that the associated paths in that case do have a natural connection with the usual classical paths.)

The reader may question whether it is really important that we have a close physical connection; isn't a mathematically consistent formulation that leads to the correct results all that is needed?

To address this natural question, let us recall the *theory of epicycles*. Before Newton's theory of the motion of the planets, there was a former theory that lasted for over 1000 years, known as the theory of epicycles. In this theory the other planets circled the Earth (not the Sun) and their unusual motion relative to the fixed stars was explained by the planet moving on a circle that moved on another circle (and if necessary) which also moved on yet another circle, etc. This theory explained the data very well. And there is a very good reason that this theory worked so well, and that is because it was effectively dealing with the Fourier series representation of the relative motion. For example, it was observed that any given planet did have a periodic pattern in its back-and-forth motion relative to the fixed stars. If the deviation from the normal star motion was analyzed in a Fourier series manner, there would be certain amplitudes for the leading Fourier components that could be represented by circles moving on other circles the radius of which is related to the frequency of one of the Fourier components. The data was good, and its physical interpretation in terms of multicircular motion was not "wrong" because it reproduced the correct results, and allowed for higher precision by adding further circles when necessary. However, the epicycle explanation was not "natural," especially when compared to the far more natural idea of the mutual attraction of all bodies introduced by Newton.

By analogy, it is our view that the realization of the phase space path integral by "paths" that are alternately, and repeatedly, sharp and unsharp, as described above, is not "natural." In Chapters 7 and 8 we will offer a far more natural formulation of phase space path integrals than the one developed in the present chapter.

6.2.3 Selected applications

Free particle motion in a random potential. Let us examine the average behavior of a free particle in an array of potentials $V(x)$ that are distributed along a one-dimensional line with a Poisson distribution and a fixed density. Specifically, we introduce the propagator for a given sample of the process by the relation

$$\langle x''|e^{-i\mathcal{H}T/\hbar}|x'\rangle = \mathcal{N}\int e^{(i/\hbar)\int[(m/2)\dot{x}^2 - V(x(t)-Y)]\,dt}\,\mathcal{D}x\,,$$

where Y denotes the stochastic variable representing the location of the potential. As a Poisson process, it follows that an average over the distribution

of Y, with N elements within a finite length L, is given by

$$\mathcal{N}\int e^{(i/\hbar)\int (m/2)\dot{x}^2\,dt}\left[\int e^{-i\int V(x(t)-y)\,dt}\,dy/L\right]^N \mathcal{D}x$$
$$= \mathcal{N}\int e^{(i/\hbar)\int (m/2)\dot{x}^2\,dt}\left[1-\int\{1-e^{-i\int V(x(t)-y)\,dt}\}\,dy/L\right]^N \mathcal{D}x\ .$$

In the limit that $L\to\infty$ and $N\to\infty$ such that $N/L=\rho$, the density of potentials V along the line, the averaged propagator becomes

$$\mathcal{N}\int e^{(i/\hbar)\int (m/2)\dot{x}(t)^2\,dt}\ e^{-\rho\int[1-e^{-(i/\hbar)\int V(x(t)-y)\,dt}]\,dy}\,\mathcal{D}x\ .$$

This example shows how a system composed of two parts can lead to a reduced system involving just one part by integrating out the influence of the other system. We can perhaps see this fact more clearly if we reformulate the same problem as follows. We assume that the free particle is moving in a field of N very heavy particles with mass $M\gg m$, and the full propagator is defined (with $\hbar=1$) by

$$\langle x'', p_1'', p_2'', \ldots, p_N''|e^{-iT\mathcal{H}}|x', p_1', p_2', \ldots, p_N'\rangle\ ,$$

where

$$\mathcal{H} = P^2/2m + \Sigma_{n=1}^N[P_n^2/2M + V(Q-Q_n)]\ ,$$

where P and Q refer to the light particle and P_n and Q_n, $1\le n\le N$, refer to the N heavy particles. The quantum mechanics of this coupled system is given by

$$\mathcal{M}\int e^{i\int\{(m/2)\dot{x}^2-\Sigma_{n=1}^N[q_n\dot{p}_n+\frac{1}{2}p_n^2/M+V(x-q_n)]\}\,dt}$$
$$\times\,\mathcal{D}x\ \Pi_{n=1}^N\,\mathcal{D}p_n\,\mathcal{D}q_n\ .$$

To make contact with our first calculation we let $M\to\infty$ and choose the momenta values $p_n''=p_n'=0$ for all n. The resultant integration over all the p_n variables leads to N δ-functionals as given by

$$\mathcal{M}\int e^{i\int\{(m/2)\dot{x}^2-\Sigma_{n=1}^N V(x-q_n)\}\,dt}\ \Pi_{n=1}^N\delta\{\dot{q}_n\}\,\mathcal{D}x\ \Pi_{n=1}^N\,\mathcal{D}q_n\ ,$$

which in turn becomes

$$\mathcal{M}\int e^{i\int\{(m/2)\dot{x}^2-\Sigma_{n=1}^N V(x-q_n)\}\,dt}\,\mathcal{D}x\ \Pi_{n=1}^N\,dq_n\ .$$

When $N\to\infty$, this expression can be rewritten (with $\hbar$ restored) as

$$\mathcal{N}\int e^{(i/\hbar)\int (m/2)\dot{x}(t)^2\,dt - \rho\int [1 - e^{-(i/\hbar)\int V(x(t)-q)\,dt}]\,dq}\ \mathcal{D}x\ ,$$

just as we had earlier obtained.

The previous exercise can be turned into an experimental prediction for the density of states for the light particle moving in the random potential we have described. To obtain this relation we turn to the expression

$$\begin{aligned}\lim_{\epsilon\to 0}\Re\int_0^\infty e^{(iE-\epsilon)T}\int\langle x|e^{-i\mathcal{H}T}|x\rangle\,dx\,dT &= \lim_{\epsilon\to 0}\Re\,\mathrm{Tr}\{[\epsilon + i(E-\mathcal{H})]^{-1}\}\\ &= \pi\,\mathrm{Tr}[\delta(E-\mathcal{H})]\\ &\equiv \pi\,\sigma(E)\ ,\end{aligned}$$

where $\sigma(E)$ denotes the density of states for the light particle moving in the random potential. In the expression above we have used the identity that

$$\lim_{\epsilon\to 0}[\epsilon + iu]^{-1} = \lim_{\epsilon\to 0}[(\epsilon - iu)/(\epsilon^2+u^2)] = \pi\,\delta(u) - i\mathcal{P}u^{-1}\ ;$$

here, the first expression arises because

$$\lim_{\epsilon\to 0}\epsilon/(\epsilon^2+u^2) = 0 \quad (u\neq 0), \qquad \int \epsilon/(\epsilon^2+u^2)\,du = \pi\ ,$$

and for the second expression, $\mathcal{P}$ means the Cauchy principal-value integral, namely,

$$\mathcal{P}\int u^{-1}\,f(u)\,du \equiv \lim_{\epsilon\to 0}\left(\int_{-\infty}^{-\epsilon} + \int_\epsilon^\infty\right)u^{-1}\,f(u)\,du\ ,$$

when the limit exists.

With this example one can see how the path integral for this system can be put to use. Indeed, this kind of expression has been used in experimental studies for the motion of electrons in crystals with random charged potentials.

Relativistic free particle. In units in which the speed of light $c = 1$, the Lagrangian for such a system is given by

$$L = -m\sqrt{1-\dot{x}^2}\ ,$$

and it might be thought that the propagator for such a system could be written as

$$K(x'',T;x',0) = \mathcal{N}\int e^{-(i/\hbar)\int m\sqrt{1-\dot{x}^2}\,dt}\ \mathcal{D}x\ .$$

Besides the fact that this is a formal integral, there is the added question of whether one should impose a limiting velocity $\dot{x}^2 \leq 1$ as required by relativity,

and if so, how is that to be accomplished. This leads to a major issue in formulating this expression via the Lagrangian path integral—and after all there may be no operator connection with such a proposal in any case. However, all problems disappear when this system is studied by a phase space path integral.

The classical momentum for the free relativistic particle is $p = \partial L/\partial \dot{x} = m\dot{x}/\sqrt{1-\dot{x}^2}$, and thus the classical Hamiltonian is given by

$$H = \sqrt{p^2 + m^2}\,,$$

and the phase space path integral reads

$$K(q'', T; q', 0) = \mathcal{M} \int e^{(i/\hbar)\int [p\dot{q} - \sqrt{p^2+m^2}]\, dt}\, \mathcal{D}p\, \mathcal{D}q\,.$$

To evaluate this integral we first integrate the kinematic term by parts to yield

$$K(q'', T; q', 0) = \mathcal{M} \int e^{(i/\hbar)\{\int [-q\dot{p} - \sqrt{p^2+m^2}]\, dt + pq|\}}\, \mathcal{D}p\, \mathcal{D}q\,.$$

Integration over the q variables leads to

$$K(q'', T; q', 0) = \mathcal{N} \int e^{(i/\hbar)\int [-\sqrt{p^2+m^2}]\, dt + (i/\hbar)pq|}\, \delta\{\dot{p}\}\mathcal{D}p\,,$$

which, just as in the calculation for the nonrelativistic free particle, leads to

$$\begin{aligned} K(q'', T; q', 0) &= \frac{1}{2\pi\hbar} \int e^{(i/\hbar)[p(q''-q') - \sqrt{p^2+m^2}\,T]}\, dp \\ &= \frac{1}{\pi\hbar}\, \frac{imT}{\sqrt{(q''-q')^2 - T^2}}\, K_1((m/\hbar)\sqrt{(q''-q')^2 - T^2}\,)\,, \end{aligned}$$

where K_1 is one of the standard Bessel functions.

This problem causes no difficulties when formulated as a phase space path integral, and that very ease of calculation suggests we look again at the configuration space path integral once more. Is there no way to interpret the formal configuration space path integral so it too yields the same answer as the phase space form? To make sense of the formal path integral we need to give it a well-defined lattice space formulation. For that purpose we suggest (for $\hbar = 1$) that

$$\begin{aligned} &\mathcal{N} \int e^{-(i/\hbar)\int m\sqrt{1-\dot{x}^2}\, dt}\, \mathcal{D}x \\ &\quad = \lim_{N\to\infty} \int \prod_{l=0}^{N} \Big[\frac{im\epsilon}{\pi\sqrt{(x_{l+1}-x_l)^2 - \epsilon^2}}\, K_1(m\sqrt{(x_{l+1}-x_l)^2 - \epsilon^2}) \Big] \\ &\qquad \times \prod_{l=1}^{N} dx_l\,. \end{aligned}$$

This prescription works, but it has been carefully "designed'' to do the job. It is unlikely that such a short time propagator would have arisen naturally, i.e., unaided by the fact that we know what answer we want to obtain. In that sense, this proposal cannot be accepted as a natural choice, and we conclude that the best way to approach this problem is still by means of the phase space path integral, rather than with the configuration space path integral, since no prior knowledge of the proper answer is used to choose its short time propagator.

****

It is noteworthy that there is a *Handbook of Feynman Path Integrals* [GS98]. These path integrals are generally evaluated *not* by actual integration but by *indirect means* such as summing suitable expressions involving both eigenfunctions and eigenvalues of the associated Hamiltonian.

6.2.4 Choice of canonical coordinates

Recall that the "old" quantum theory of Bohr and Sommerfeld relied heavily on the quantization condition that

$$\oint p\,dq = 2\pi(n + 1/2)\hbar\ .$$

This condition is all well and good, but there is no instruction as to which set of canonical coordinates are to be used. Recall that two sets of canonical coordinates are connected by the equation

$$\overline{p}d\overline{q} = pdq + dF(\overline{q}, q)\ ,$$

where the function F is known as the generator of the coordinate transformation. The closed line interval of the transformation rule leads to

$$\oint \overline{p}d\overline{q} = \oint pdq + \oint dF(\overline{q}, q) = \oint pdq\ ,$$

since the closed line integral of a total derivative vanishes when the enclosed area is topologically trivial. As a consequence, the closed line integral is *invariant* under canonical coordinate transformations; or stated otherwise, it does not matter what choice of coordinates is used: the answer is the same whatever set of canonical coordinates are involved. When questions can be formulated in a coordinate-independent fashion, it follows that there is an underlying geometric interpretation. In the present case, such a result is of fundamental importance since it demonstrates that the Bohr–Sommerfeld condition of the "old" quantum mechanics is a geometric prescription.

It is clear that the "new" quantum mechanics of Heisenberg and Schrödinger does not share the same invariance under canonical coordinate transformations. For example, in one set of coordinates the classical harmonic oscillator

is given by $\frac{1}{2}(p^2+q^2)$, which would become $\frac{1}{2}(-\hbar^2\,\partial^2/\partial x^2+x^2)$; in another set of coordinates the classical harmonic oscillator reads $\overline{p}$, which would become $-i\hbar\partial/\partial\overline{x}$. Clearly, the spectrum of these two operators is quite different—the first has a discrete spectrum, the second has a continuous one—and thus the physical aspects of the quantum theory, embodied in the spectrum of the Hamiltonian operator, are not invariant under canonical coordinate transformations.

The "new" quantum mechanics of Heisenberg and Schrödinger is *not* invariant under canonical coordinate transformations despite the fact that it agrees better with experiment than the "old" quantum mechanics. By the "new" quantum mechanics we mean the rule by which a classical Hamiltonian is converted to a quantum Hamiltonian, specifically, the general rule in which

$$p \to P = -i\hbar\partial/\partial x\ , \qquad q \to Q = x\ ,$$

acting on square integrable wave functions $\psi(x)$, and which is extended to the classical Hamiltonian by means of

$$H(p,q) \to \mathcal{H} = H(P,Q) = H(-i\hbar\partial/\partial x, x)\ ,$$

modulo factor ordering ambiguities. Of course, a great deal of ambiguity can be hidden under the umbrella of factor ordering, as, for example, the addition to any classical Hamiltonian of the additional term

$$0 = (qp-pq)^2 Y(p,q) \to (QP-PQ)^2 Y(P,Q) = -\hbar^2\, Y(-i\hbar\partial/\partial x, x)\ ,$$

which could change everything. Let us therefore restrict our attention to "natural" factor ordering ambiguities that arise within the classical Hamiltonian itself. Given that restriction, the general rule of quantization in the "new" quantum mechanics applies only to *Cartesian* canonical coordinates and not to more general curvilinear coordinates; see page 114 of [Dir74]. The implication of this remark is that a proper quantization depends on the choice of coordinates, and specific applications confirm the truth of this statement. Two reactions to this remark are possible: The first response is to accept that limitation and just use it in applications; the second response would be to argue that something so physical as quantization should not depend on something so arbitrary as the choice of coordinates, and furthermore, that Cartesian coordinates depend on the existence of a *metric* and a metric on phase space is not part of the classical theory of mechanics which includes a symplectic structure but not that of a metric. With respect to the first viewpoint, it should be appreciated that any lattice form of path integrals that we have considered has implicitly been assumed to be in Cartesian coordinates. And regarding the second viewpoint, the program of Geometric Quantization [Sni80] was based on the assumption that quantum mechanics must be independent of the choice of canonical coordinates and so proponents of that

theory created a mathematical formulation that satisfied this fundamental requirement; unfortunately, the resultant theory did not agree with experiment nor with predictions of the Schrödinger equation which already did agree with experiment. In its original formulation, Geometric Quantization was a mathematically consistent theory that enjoyed full covariance under canonical coordinate transformations; however, it did not describe the quantum mechanics of the real world.

On the other hand, there is a perfectly natural reason why the "new" quantum mechanics requires Cartesian coordinates for its formulation. Unlike the classical theory, the algebraic structure of the Hamiltonian operator carries real physics; for example, we identify the expression P^2 as that of a free particle, $P^2 + Q^2$ as that of a harmonic oscillator, $P^2 + Q^2 + Q^4$ as that of an oscillator plus a quartic interaction, etc. Under a unitary transformation in which $P \to P' = U^\dagger P U$ and $Q \to Q' = U^\dagger Q U$, it follows that the free particle becomes P'^2, the oscillator becomes $P'^2 + Q'^2$, the addition of a quartic interaction becomes $P'^2 + Q'^2 + Q'^4$, etc., and in particular *the algebraic structure has remained unchanged for each physically distinct Hamiltonian.* To pass from the classical Hamiltonian to the special quantum form requires that the classical form already carry some physical meaning of its mathematical expression, and as we shall demonstrate in the following two chapters, that property requires that the classical Hamiltonian is expressed in Cartesian canonical coordinates, a physical property which involves a phase space metric that the quantum theory freely bequeathes to the classical theory!

6.3 Action Principle—and Equations of Motion

It is heuristically useful to review how the action principle is used to derive equations of motion in classical mechanics. This should be familiar territory to the reader, but there is some aspect that needs clarification and emphasis.

For a single degree of freedom the classical action is normally given by

$$I = \int_{t'}^{t''} [p\dot{q} - H(p,q)]\, dt \ .$$

The equations of motion arise from a stationary variation of the action for independent variations of p and q. Such a variation leads to

$$\delta I = \int_{t'}^{t''} [\dot{q} - \partial H/\partial p]\, \delta p\, dt + \int_{t'}^{t''} [-\dot{p} - \partial H/\partial q]\, \delta q\, dt + p\, \delta q \Big|_{t'}^{t''} \ .$$

Stationarity requires that $\delta I = 0$, and this can be secured first by insisting that the variation of q vanishes at the end points, i.e., $\delta q(t'') = 0$ and $\delta q(t') = 0$, which then ensures that the final term above vanishes. To ensure that $\delta I = 0$, the coefficients of δq and δp should also vanish thereby leading to the usual equations of motion

$$\dot{q} - \partial H/\partial p = 0 \,, \qquad -\dot{p} - \partial H/\partial q = 0 \,.$$

This is the usual argument to *derive* the equations of motion, but that leaves open the question about the suitability of this procedure regarding the finding of *solutions* to these very equations of motion.

For example, with the variation of q fixed at the initial time t' and the final time t'', it is the general assumption that a unique solution is thereby determined. After all, it is plausible that two fixed data values should determine the solution of a second-order equation for q. We will comment on this assumption below, but for the moment we will assume it to be true.

Nothing much changes if a different form for the action is chosen, namely,

$$I = \int_{t'}^{t''} [-q\dot{p} - H(p,q)]\, dt \,.$$

In this case

$$\delta I = \int_{t'}^{t''} [-\dot{p} - \partial H/\partial q]\delta q\, dt + \int_{t'}^{t''} [\dot{q} - \partial H/\partial p]\delta q\, dt - q\,\delta p\Big|_{t'}^{t''} \,,$$

which leads to the same equations of motion provided that we impose the conditions that $\delta p(t'') = 0$ and $\delta p(t') = 0$. Again, one makes the assumption that with p fixed at the initial time t' and the final time t'', there is a unique solution to the equations of motion. Again, this is a reasonable assumption just as it was when q was held fixed initially and finally.

Matters change significantly when we consider the action

$$I = \int_{t'}^{t''} [\tfrac{1}{2}(p\dot{q} - q\dot{p}) - H(p,q)]\, dt \,,$$

which leads to

$$\delta I = \int_{t'}^{t''} [\dot{q} - \partial H/\partial p]\, dt + \int_{t'}^{t''} [-\dot{p} - \partial H/\partial q]\, dt + \tfrac{1}{2}(p\,\delta q - q\,\delta p)\Big|_{t'}^{t''} \,.$$

Now to ensure that $\delta I = 0$ requires that $\delta p(t'') = \delta p(t') = 0$ as well as $\delta q(t'') = \delta q(t') = 0$. This requires *four* fixed data values which would seem to be in conflict with finding any solution to the equations of motion. The same situation seems to arise when we consider

$$I = \int_{t'}^{t''} [p\dot{q} + \dot{G}(p,q) - H(p,q)]\, dt \,,$$

which includes a general total derivative. This latter form also requires that four values are held fixed.

First, we wish to stress the great difference between *deriving* the equations of motion and the process of *solving* those equations of motion. It is not true

that holding only q fixed at two different times leads to a unique solution as witnessed by the example where

$$\ddot{q}(t) = -q(t) \ ,$$

subject to the boundary conditions that $q(0) = 0$ and $q(\pi) = 0$ as well. Every solution of the form $q(t) = A \sin(t)$ satisfies those conditions for all A; in this case there is an infinite number of solutions to the given boundary conditions. For the same equation of motion, the conditions $q(0) = 0$ and $q(\pi) = c$, $c \neq 0$, admit *no* solution whatsoever. A similar story holds when two values of p are to be held at different times. Thus we see that the number of solutions is not even continuous in the boundary data in the two-value case. Of course, when we insist that *four* values, namely, $p(t'')$, $q(t'')$, $p(t')$, and $q(t')$ are held fixed, there is generally *no* solution compatible with such conditions. However, and this is the important point, in the four data points case there is always a fixed *subset* of the requested data to hold fixed that leads to a unique solution. For example, in our case the two data points $p(t')$, $q(t')$ by themselves yield a unique solution. It may happen that the unique solution determined by these values passes through the points $p(t'')$, $q(t'')$ in the future, but that is far more the exception than the rule. Most likely, the solution determined by $p(t')$, $q(t')$ will not pass through the prefixed final points $p(t'')$, $q(t'')$. So be it. The point to celebrate in all this is that there exists a fixed subset of the required data points that does ensure a unique solution in all cases. Of course, there is no guarantee that such a solution will be singularity free for all time; such properties are determined by the Hamiltonian.

The moral of the story is as follows. The procedure used to derive the equations of motion from the action principle is one thing; the boundary data needed to determine a unique solution of those equations of motion is quite another thing.

Exercises

6-1 Offer a proof that the amplitude factor A for the free particle propagator,

$$\mathcal{N} \int e^{i(m/2\hbar)\int \dot{x}(t)^2 \, dt} \, \mathcal{D}x = A \, e^{i(m/2\hbar)(x'' - x')^2/(t'' - t')} \ ,$$

where the exponent on the right-hand side is denoted by $i S(x'', t''; x', t')/\hbar$, can be determined by the expression

$$A = \sqrt{i/(2\pi\hbar)} \, \sqrt{\partial^2 S(x'', t''; x', t')/\partial x'' \, \partial x'} \ .$$

6-2 Show that the phase space path integral for the relativistic particle has a close connection with the abstract operator construction.

7

Coherent State Path Integrals

7.1 Canonical Coherent States and Their Properties

7.1.1 Coherent states—what are they?

The usual, irreducible Heisenberg operators Q and P are the basic building blocks of the operators in quantum mechanics. Each of the basic operators can be chosen self-adjoint and each admits formal eigenvectors and eigenvalues, e.g., $Q|q\rangle = q|q\rangle$ as well as $P|p\rangle = p|p\rangle$, for which $\langle q'|q\rangle = \delta(q'-q)$ and $\langle p'|p\rangle = \delta(p'-p)$, and which lead to standard resolutions of unity in the canonical form

$$\mathbb{1} = \int |q\rangle\langle q|\, dq = \int |p\rangle\langle p|\, dp\ .$$

However, such resolutions of unity are by no means the only kind that have a role to play in quantum formulations.

Let us first focus on the Weyl operators

$$W(p,q) \equiv e^{-iqP/\hbar}\, e^{ipQ/\hbar}\ .$$

These operators have a multiplication rule given by

$$W(p',q')\, W(p,q) = e^{ip'q/\hbar}\, W(p'+p, q'+q)\ ,$$

which is a reflection of the usual Heisenberg commutation relation $[Q,P] = i\hbar\mathbb{1}$. Observe that if the Weyl operators are augmented by a phase factor, i.e.,

$$W(p,q,a) \equiv e^{-iqP/\hbar}\, e^{ipQ/\hbar}\, e^{ia/\hbar}\ ,$$

then the multiplication rule becomes

$$W(p',q',a')\, W(p,q,a) = W(p'+p, q'+q, a'+a+p'q)\ ,$$

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9_7,

which is recognized as the same multiplication rule as that of the Heisenberg group discussed in Sec. 5.3. (**Remark:** Our definition of the Weyl operators uses the phase $a = 0$; oftentimes, one defines the Weyl operators with a different choice for a.)

Next, we introduce a normalized, specific Hilbert space vector $|0\rangle$, which satisfies the relation

$$(Q + iP)\,|0\rangle = 0\ .$$

This vector is just the ground state eigenvector of the standard harmonic oscillator Hamiltonian $(P^2 + Q^2)/2$. Finally, we combine the two ingredients above and introduce a family of unit vectors

$$|p, q\rangle \equiv W(p, q)\,|0\rangle = e^{-iqP/\hbar}\, e^{ipQ/\hbar}\,|0\rangle\ ,$$

defined for all $(p, q) \in \mathbb{R}^2$. These states are known as *coherent states* [KlS85], and more particularly as the canonical coherent states. The distinguished vector $|0\rangle$ is called the fiducial vector for the coherent states.

Let us find the representation of the coherent states in the Schrödinger representation. In particular, we consider

$$\begin{aligned}\langle x|p, q\rangle &= \langle x|e^{-iqP/\hbar}\, e^{ipQ/\hbar}\,|0\rangle \\ &= \langle x - q|e^{ipQ/\hbar}\,|0\rangle \\ &= e^{ip(x-q)/\hbar}\,\langle x - q|0\rangle\ ,\end{aligned}$$

which, when combined with $\langle x|0\rangle = (1/\pi\hbar)^{1/4}\,\exp(-x^2/2\hbar)$, leads to

$$\langle x|p, q\rangle = (1/\pi\hbar)^{1/4}\, e^{ip(x-q)/\hbar}\, e^{-(x-q)^2/2\hbar}\ .$$

[**Remark:** As an incidental observation at this point we note that

$$\int \langle x|p, q\rangle\, dp = (1/\pi\hbar)^{1/4}\, 2\pi\hbar\,\delta(x - q)\ ;$$

we shall come back to this relation at a later point.]

Introducing $\Delta q \equiv (q' - q)/2$, the expression for $\langle x|p, q\rangle$ allows us to compute the coherent state overlap function (with $\hbar = 1$ until the end)

$$\begin{aligned}&\langle p', q'|p, q\rangle \\ &= (1/\pi)^{1/2} \int e^{-ip'(x-q') + ip(x-q) - (x-q')^2/2 - (x-q)^2/2}\, dx \\ &= (1/\pi)^{1/2} \int e^{-ip'(x-\Delta q) + ip(x+\Delta q) - (x-\Delta q)^2/2 - (x+\Delta q)^2/2}\, dx \\ &= e^{i\frac{1}{2}(p'+p)(q'-q) - \frac{1}{4}[(p'-p)^2 + (q'-q)^2]}\ .\end{aligned}$$

Furthermore, it follows that

$$\begin{aligned}\int \langle p'', q''|p, q\rangle\langle p, q|p', q'\rangle \, dp\, dq/2\pi\hbar \\ = \int e^{i\frac{1}{2}(p''+p)(q''-q)/\hbar - \frac{1}{4}[(p''-p)^2+(q''-q)^2]/\hbar} \\ \times e^{i\frac{1}{2}(p+p')(q-q')/\hbar - \frac{1}{4}[(p-p')^2+(q-q')^2]/\hbar} \, dp\, dq/2\pi\hbar \\ = e^{i\frac{1}{2}(p''+p')(q''-q')/\hbar - \frac{1}{4}[(p''-p')^2+(q''-q')^2]/\hbar} ,\end{aligned}$$

or stated otherwise,

$$\int \langle p'', q''|p, q\rangle\langle p, q|p', q'\rangle \, dp\, dq/2\pi\hbar = \langle p'', q''|p', q'\rangle \, .$$

Finally, in the Hilbert space spanned by the coherent states, the previous equation implies the all-important *resolution of unity for coherent states* given by

$$\mathbb{1} = \int |p, q\rangle\langle p, q| \, dp\, dq/2\pi\hbar \, .$$

This resolution of unity is fundamentally different from the usual ones with continuous labels such as $\mathbb{1} = \int |q\rangle\langle q| \, dq$ because: (i) the states $|p, q\rangle$ are true vectors in the Hilbert space while the "vectors" $|q\rangle$ are not true vectors, (ii) the "vectors" $|q\rangle$ are orthogonal to one another for different q values while the vectors $|p, q\rangle$ are not mutually orthogonal, and (iii) the vectors $|p, q\rangle$ are *overcomplete* while the "vectors" $|q\rangle$ are strictly complete. Moreover, the coherent state vectors are continuously labeled in the sense that as $(p, q) \to (p', q')$ it follows that $\| \, |p, q\rangle - |p', q'\rangle \, \| \to 0$, where $\| \, |\psi\rangle \, \| \equiv \sqrt{\langle\psi|\psi\rangle}$.

Although the coherent states $|p, q\rangle$ are very different from the states $|q\rangle$, the fact that both sets of states admit resolutions of unity in a similar fashion permits the introduction of a *coherent state representation of Hilbert space*. Recall that the Schrödinger representation can be viewed as arising from the abstract vectors $|\psi\rangle$ by means of the relations

$$\psi(q) \equiv \langle q|\psi\rangle \, , \qquad (\psi, \psi) \equiv \int |\psi(q)|^2 \, dq = \langle\psi|\psi\rangle \, .$$

Less well known, or at least, less appreciated—but nonetheless important—is the fact that the Schrödinger representation is burdened with the issue of sets of measure zero and the concomitant fact that the vectors of the Schrödinger representation are strictly speaking not functions but rather equivalence classes of functions all equal to one another up to sets of measure zero. In contrast, the representation induced by the coherent states has no such cumbersome baggage. In analogy with the expressions above, we introduce the coherent state representation given by

$$\psi(p, q) \equiv \langle p, q|\psi\rangle \, , \qquad (\psi, \psi) \equiv \int |\psi(p, q)|^2 \, dp dq/2\pi\hbar = \langle\psi|\psi\rangle \, .$$

We note that the functional representatives $\psi(p,q)$ are all continuous functions in the phase space variables (p,q), and in addition they are all bounded functions in the sense that $|\psi(p,q)| \leq \|\psi\| \equiv \||\psi\rangle\|$ thanks to the Schwarz inequality. In contrast to the Schrödinger case, the representatives are not only continuous functions, but they are all C^∞ thanks to the fact that

$$(-i\hbar\partial/\partial q)^n \psi(p,q) = \langle 0| e^{-ipQ/\hbar} e^{iqP/\hbar} P^n |\psi\rangle \ ,$$

and

$$\begin{aligned}(i\hbar\partial/\partial p)^n \psi(p,q) &= \langle 0| e^{-ipQ/\hbar} Q^n e^{iqP/\hbar} |\psi\rangle \\ &= \langle 0| e^{-ipQ/\hbar} e^{iqP/\hbar} (Q-q)^n |\psi\rangle \ .\end{aligned}$$

These equations teach us two things. First, since $|0\rangle$ is in the domain of Q^n and P^n for all n, the resultant derivatives are continuous functions. Second, they teach us the important relation that the coherent state representatives of the abstract operators P and Q are given by

$$\begin{aligned}(-i\hbar\partial/\partial q)\psi(p,q) &= (P\psi)(p,q) \ , \\ (q + i\hbar\partial/\partial p)\psi(p,q) &= (Q\psi)(p,q) \ .\end{aligned}$$

This information can be directly used to find the Schrödinger equation in the coherent state representation in the form

$$i\hbar\partial\psi(p,q,t)/\partial t = \mathcal{H}(-i\hbar\partial/\partial q, q + i\hbar\partial/\partial p)\, \psi(p,q,t) \ .$$

On one hand, this equation seems "far" from the Schrödinger equation, and yet on the other hand, it is rather "close." To emphasize the latter fact, observe that the standard Schrödinger representation is obtained from the coherent state representation (in the present choice of phase convention) essentially by integrating out the unwanted p variable; specifically (see Remark above), we find that

$$\psi(q) = (\pi^{1/4}/2\pi\hbar^{3/4}) \int \psi(p,q)\, dp \ .$$

For example, assume that the Hamiltonian operator is a polynomial in both P and Q with some suitable ordering having been chosen. In that case, consider the integral over p of the Schrödinger equation in the coherent state representation, namely,

$$(i\hbar\partial/\partial t) \int \psi(p,q,t)\, dp = \int \mathcal{H}(-i\hbar\partial/\partial q, q + i\hbar\partial/\partial p)\, \psi(p,q,t)\, dp \ .$$

Integration by parts on the right-hand side, as many times as necessary, eliminates all contributions of $\partial/\partial p$, and leads directly to

$$(i\hbar\partial/\partial t) \int \psi(p,q,t)\, dp = \mathcal{H}(-i\hbar\partial/\partial q, q) \int \psi(p,q,t)\, dp \ ,$$

which (modulo a constant factor) is just the usual version of Schrödinger's equation.

The preceding paragraph explains how the action of P and Q can be introduced as differential operators in the coherent state representation, but there is another important way to introduce the action of operators and that is by means of integral kernels. In particular, let $\mathcal{G} = \mathcal{G}(P, Q)$ denote a generic operator, and consider the kernel

$$\langle p', q' | \mathcal{G} | p, q \rangle \ ,$$

which can be used to represent

$$\langle p', q' | \mathcal{G} | \psi \rangle = \int \langle p', q' | \mathcal{G} | p, q \rangle \langle p, q | \psi \rangle \, dp dq / 2\pi\hbar \ ,$$

yielding a representation of the action of $\mathcal{G}$ on the vector $|\psi\rangle$. Examples of such kernels are given by

$$\begin{aligned}
&\langle p', q' | P | p, q \rangle = (-i\hbar\partial/\partial q') \langle p', q' | p, q \rangle \\
&= (-i\hbar\partial/\partial q') \, e^{i\frac{1}{2}(p'+p)(q'-q)/\hbar - \frac{1}{4}[(p'-p)^2 + (q'-q)^2]/\hbar} \\
&= \tfrac{1}{2}\left[(p'+p) + i(q'-q)\right] e^{i\frac{1}{2}(p'+p)(q'-q)/\hbar - \frac{1}{4}[(p'-p)^2 + (q'-q)^2]/\hbar} \ ,
\end{aligned}$$

and

$$\begin{aligned}
&\langle p', q' | Q | p, q \rangle \\
&= \tfrac{1}{2}\left[(q'+q) - i(p'-p)\right] e^{i\frac{1}{2}(p'+p)(q'-q)/\hbar - \frac{1}{4}[(p'-p)^2 + (q'-q)^2]/\hbar} \ .
\end{aligned}$$

It may be noticed that the coherent state representation we have been discussing is also a reproducing kernel Hilbert space for which the reproducing kernel is given by the coherent state overlap

$$K(p', q'; p, q) = \langle p', q' | p, q \rangle \ .$$

While we have focussed on the inner product being given by a local integral, it is also the case that the inner product for the coherent state representation may be given by the usual formula for inner products for reproducing kernel Hilbert spaces as described in Sec. 5.1.5.

7.1.2 Diagonal coherent state matrix elements

It is important to examine the diagonal coherent state matrix elements of various operators. For example, it follows from the previous paragraphs that

$$\langle p, q | P | p, q \rangle = p \ , \qquad \langle p, q | Q | p, q \rangle = q \ ,$$

asserting that the *mean value* of P in the coherent state $|p,q\rangle$ is p, and likewise for Q and q. One can interpret this result as saying that the physical interpretation of the label p and q are as mean values of the operators P and Q, respectively. Note well, therefore, that the meaning of p and q in the coherent states $|p,q\rangle$ is that of mean values and definitely not sharp eigenvalues as is the case for the label p in the state $|p\rangle$ and for the label q in the state $|q\rangle$.

For more general operators, such as the Hamiltonian $\mathcal{H} = \mathcal{H}(P,Q)$, the diagonal coherent state matrix elements become

$$\begin{aligned} H(p,q) &\equiv \langle p,q|\mathcal{H}(P,Q)|p,q\rangle \\ &= \langle 0|\mathcal{H}(P+p,Q+q)|0\rangle \\ &= \mathcal{H}(p,q) + \mathcal{O}(\hbar;p,q)\ . \end{aligned}$$

The last term, $\mathcal{O}(\hbar;p,q)$, vanishes as $\hbar \to 0$, and leads to the assertion that

$$H_c(p,q) \equiv \lim_{\hbar\to 0}\langle p,q|\mathcal{H}(P,Q)|p,q\rangle = \mathcal{H}(p,q)\ ,$$

which confirms that the true classical limit (in which $\hbar \to 0$) for the quantum operator $\mathcal{H}(P,Q)$ becomes $\mathcal{H}(p,q)$, i.e., *the very same functional form*, provided one uses the same choice of canonical coordinates (p,q) used to define the coherent states.

To elaborate on this last remark, let us briefly consider at this point a change of canonical coordinates from $(p,q) \to (\overline{p},\overline{q})$, for which we define the transformation of the coherent states as $|p,q\rangle \to |\overline{p},\overline{q}\rangle \equiv |p(\overline{p},\overline{q}),q(\overline{p},\overline{q})\rangle$ under a canonical coordinate transformation. As a consequence, the diagonal coherent state matrix elements become

$$\overline{H}(\overline{p},\overline{q}) \equiv H(p,q) = \mathcal{H}(p(\overline{p},\overline{q}),q(\overline{p},\overline{q})) + \mathcal{O}(\hbar;p,q) \equiv \overline{\mathcal{H}}(\overline{p},\overline{q}) + \mathcal{O}(\hbar;p,q)\ ,$$

which would lead to the true classical limit being given by $\overline{\mathcal{H}}(\overline{p},\overline{q})$ which is generally not the same functional form as $\mathcal{H}(\overline{p},\overline{q})$.

Here is the first clue that there is something special about the original choice of coordinates used in defining the coherent states; we will find further confirmation of this fact as we proceed.

7.2 Coherent State Propagator

Armed with the information that the coherent states can be used to form a representation of Hilbert space in a manner similar to the more familiar orthonormal bases, we next study the coherent state propagator

$$K(p'',q'',T;p',q',0) \equiv \langle p'',q''|\,e^{-i\mathcal{H}T/\hbar}|p',q'\rangle\ ,$$

which has the virtue of propagating the coherent state wave function forward in time as

$$\psi(p'', q'', T) = \int K(p'', q'', T; p', q', 0)\, \psi(p', q', 0)\, d\mu(p', q') \ ,$$

where

$$d\mu(p, q) \equiv dp\, dq/2\pi\hbar$$

and the integration extends over the entire plane.

Earlier, in discussing the phase space propagator in the configuration space representation, we were able to provide a construction that conformed with the formal expression

$$K(q'', T; q', 0) = \mathcal{M} \int e^{(i/\hbar)\int [p\dot{q} - H(p,q)]\, dt}\, \mathcal{D}p\, \mathcal{D}q \ ,$$

integrated over all paths $p(t)$ and all paths $q(t)$, $0 \leq t \leq T$, subject to the boundary conditions that $q(T) = q''$ and $q(0) = q'$. In the coherent state case, we deal with a very different set of basis states, and even a very different physical interpretation of the variables themselves. Nevertheless, we will find that we can again show, from a formal perspective, that

$$K(p'', q'', T; p', q', 0) = \mathcal{M} \int e^{(i/\hbar)\int [p\dot{q} - H(p,q)]\, dt}\, \mathcal{D}p\, \mathcal{D}q \ ,$$

where, in the coherent state case, the integration is over all paths $p(t), q(t)$, $0 \leq t \leq T$, subject to the boundary conditions $p(T), q(T) = p'', q''$ and $p(0), q(0) = p', q'$, respectively. Thus, from a formal perspective, there seems to be little to choose between the two different ways that lead to the propagator.

Next, we provide an operator-based derivation of the coherent state propagator. We begin in a familiar fashion, namely (with $\hbar = 1$ temporarily),

$$\langle p'', q''| e^{-iT\mathcal{H}} |p', q'\rangle = \langle p'', q''| e^{-i\epsilon\mathcal{H}} e^{-i\epsilon\mathcal{H}} \cdots e^{-i\epsilon\mathcal{H}} |p', q'\rangle \ ,$$

where there are $N+1$ factors involving ϵ, and $\epsilon = T/(N+1)$. Between each pair of such factors, we introduce unity expressed as a coherent state resolution of unity to yield

$$\begin{aligned} &\langle p_{N+1}, q_{N+1}| e^{-i\epsilon\mathcal{H}} e^{-i\epsilon\mathcal{H}} \cdots e^{-i\epsilon\mathcal{H}} |p_0, q_0\rangle \\ &\quad = \int \cdots \int \textstyle\prod_{n=0}^{N} \langle p_{n+1}, q_{n+1}| e^{-i\epsilon\mathcal{H}} |p_n, q_n\rangle \prod_{n=1}^{N} d\mu(p_n, q_n) \ , \end{aligned}$$

where $p_{N+1}, q_{N+1} = p'', q''$ and $p_0, q_0 = p', q'$. The previous expression holds for all N. However, for large N, and hence small ϵ, we can make an approximation such that

$$\begin{aligned} &\langle \tilde{p}, \tilde{q}| e^{-i\epsilon\mathcal{H}} |p, q\rangle \simeq \langle \tilde{p}, \tilde{q}| [1 - i\epsilon\mathcal{H}] |p, q\rangle \\ &\quad \equiv \langle \tilde{p}, \tilde{q}|p, q\rangle [1 - i\epsilon H(\tilde{p}, \tilde{q}; p, q)] \simeq \langle \tilde{p}, \tilde{q}|p, q\rangle\, e^{-i\epsilon H(\tilde{p}, \tilde{q}; p, q)} \ , \end{aligned}$$

where

$$H(\tilde{p},\tilde{q};p,q) \equiv \langle \tilde{p},\tilde{q}|\,\mathcal{H}\,|p,q\rangle / \langle \tilde{p},\tilde{q}|p,q\rangle \ .$$

Thus we are led to the expression

$$\begin{aligned} &K(p'',q'',T;p',q',0) = \\ &\lim_{N\to\infty} \int \textstyle\prod_{n=0}^{N} \langle p_{n+1},q_{n+1}|p_n,q_n\rangle \, e^{-i\epsilon H(p_{n+1},q_{n+1};p_n,q_n)} \prod_{n=1}^{N} d\mu(p_n,q_n) \\ &= \lim_{N\to\infty} \int e^{i\frac{1}{2}\Sigma_{n=0}^{N}(p_{n+1}+p_n)(q_{n+1}-q_n)} \\ &\qquad \times e^{-\frac{1}{4}\Sigma_{n=0}^{N}[(p_{n+1}-p_n)^2+(q_{n+1}-q_n)^2]} \\ &\qquad \times e^{-i\epsilon\Sigma_{n=0}^{N} H(p_{n+1},q_{n+1};p_n,q_n)} \textstyle\prod_{n=1}^{N} dp_n\, dq_n/2\pi \ . \end{aligned}$$

This latter expression is a meaningful lattice regularization of the formal phase space path integral that leads to a coherent state representation for the propagator provided the approximations that have raised $H(p_{n+1},q_{n+1};p_n,q_n)$ to the exponent lead to convergent integrals. If that is not the case, one should replace the final expression above with one in which the term

$$e^{-i\epsilon\Sigma_{n=0}^{N} H(p_{n+1},q_{n+1};p_n,q_n)}$$

is replaced by

$$\Pi_{n=0}^{N}[1 - i\epsilon H(p_{n+1},q_{n+1};p_n,q_n)] \ .$$

In either case, it is noteworthy that the coherent state overlap function provides a damping factor due to the coherent state overlap function that was not present in the traditional phase space path integral discussed earlier. Interchanging the continuum limit and the integrations leads to a formal expression for the propagator, which for continuous and differentiable paths has the standard formal form for phase space path integrals, as was to be shown.

The introduction of coherent states, their resolution of unity, and the construction of the coherent state phase space path integral was done by the author [Kla60].

7.2.1 Change of coordinates

This section discusses a number of issues regarding a change of canonical coordinates from one set to another set in a preparatory analysis to a discussion of such changes for the path integral itself. A change of canonical coordinates is basically a one-to-one map from (p,q) to $(\overline{p},\overline{q})$, which may be written as $\overline{p} = \overline{p}(p,q)$ and $\overline{q} = \overline{q}(p,q)$, or its inverse $p = p(\overline{p},\overline{q})$ and $q = q(\overline{p},\overline{q})$. These are not arbitrary coordinate transformation, but are subject to certain conditions which are embodied in the equation

$$\overline{p}\, d\overline{q} = p\, dq + dF(\overline{q}, q) \ ,$$

where $F(\overline{q}, q)$ is known as the generator of the canonical transformation. It is also traditional to discuss other forms of such transformations, such as

$$-\overline{q} d\overline{p} = p\, dq + dF_2(\overline{p}, q) \ .$$

Such additional forms are useful when a single form does not sufficiently distinguish the coordinates. For example, consider the case where $\overline{p} = cp$ and $\overline{q} = q/c$, where c is a constant. For this case $F = 0$, but $F_2 = -\overline{p}q/c$, which distinguishes between different c values. For the most part we shall content ourselves with the simple form given initially.

We shall also employ yet another notation for the canonical transformation. In particular, we shall use the relation

$$\overline{p}\, d\overline{q} = p dq + dF(\overline{q}, q) = p\, dq + dF(\overline{q}(p, q), q) \equiv p\, dq + dG(p, q) \ ,$$

which introduces an alternative form for the generator.

It is convenient to characterize the phase space $\mathcal{M}$ (not to be confused with our use of the same symbol as a normalization factor for phase space path integrals!) by means of a set of coordinates, which serve to label the phase space points. An alternative set of coordinates may be laid down as well, and the map given by means of the generating function provides the map from one set of coordinates to another that is used to identify the very same points in phase space.

A coherent state $|p, q\rangle$ is identified with a point in the classical phase space. We define the transformation of a coherent state under canonical coordinate transformations to be that of a scalar so that

$$|p, q\rangle \rightarrow |\overline{p}, \overline{q}\rangle \equiv |p(\overline{p}, \overline{q}), q(\overline{p}, \overline{q})\rangle = |p, q\rangle \ .$$

One should not be confused by the notation here, a confusion that is common when considering Dirac bras and kets. The argument of a ket is generally quite insufficient to completely define the vector. For example, the state $|17\rangle$ in no way clarifies the meaning that should be attached to the number 17 that is its label, and without that knowledge, the meaning of the vector $|17\rangle$ itself is quite unknown. In the same way, the meaning of the ket $|\overline{p}, \overline{q}\rangle$ requires an independent knowledge of the canonical coordinate system involved.

Our choice of transformation property makes coherent states into geometrical entities, namely, quantities labeled by phase space points themselves. There are many favorable consequences that follow from this choice, and we now develop several of those consequences. The first quantity we construct will be a one form

$$\theta \equiv i\hbar \langle p, q| d|p, q\rangle \ .$$

Note that

$$
\begin{aligned}
i\hbar\, d|p,q\rangle &= i\hbar\, d\, e^{-iqP/\hbar}\, e^{ipQ/\hbar}\,|0\rangle \\
&= (dq\,P)\,|p,q\rangle - e^{-iqP/\hbar}\, e^{ipQ/\hbar}\, dp\,Q\,|0\rangle \ ,
\end{aligned}
$$

which leads directly to

$$
\begin{aligned}
i\hbar\langle p,q|\, d|p,q\rangle &= \langle p,q|\, P\,|p,q\rangle\, dq - \langle 0|\, Q\,|0\rangle\, dp \\
&= p\, dq \ .
\end{aligned}
$$

In a different canonical coordinate system we find that

$$
i\hbar\langle \overline{p},\overline{q}|\, d|\overline{p},\overline{q}\rangle = p(\overline{p},\overline{q})\, dq(\overline{p},\overline{q}) = \overline{p}\, d\overline{q} + d\overline{G}(\overline{p},\overline{q}) \ .
$$

The second quantity we consider is the diagonal coherent state matrix elements of an operator $\mathcal{G}$ given by

$$
G(p,q) \equiv \langle p,q|\, \mathcal{G}(P,Q)\,|p,q\rangle = \langle 0|\, \mathcal{G}(P+p,Q+q)\,|0\rangle \ .
$$

As observed earlier, for operators $\mathcal{G}$ that are polynomials, it follows that

$$
G(p,q) = \mathcal{G}(p,q) + \langle 0|\,[\mathcal{G}(P+p,Q+q) - \mathcal{G}(p,q)]\,|0\rangle \equiv \mathcal{G}(p,q) + \mathcal{O}(p,q;\hbar) \ ,
$$

showing that as $\hbar \to 0$,

$$
G_c(p,q) \equiv \lim_{\hbar\to 0} G(p,q) = \mathcal{G}(p,q) \ .
$$

In different canonical coordinates it follows that

$$
\begin{aligned}
\overline{G}(\overline{p},\overline{q}) &\equiv \langle \overline{p},\overline{q}|\, \mathcal{G}(P,Q)\,|\overline{p},\overline{q}\rangle = \mathcal{G}(p(\overline{p},\overline{q}),q(\overline{p},\overline{q})) + \mathcal{O}(p(\overline{p},\overline{q}),q(\overline{p},\overline{q});\hbar) \\
&= \overline{\mathcal{G}}(\overline{p},\overline{q}) + \overline{\mathcal{O}}(\overline{p},\overline{q};\hbar) \ .
\end{aligned}
$$

It is important to observe that since the coherent states transform as scalars, i.e., $|\overline{p},\overline{q}\rangle = |p,q\rangle$, it follows that

$$
\overline{G}(\overline{p},\overline{q}) \equiv G(p,q) \ ;
$$

moreover, *no* change of the quantum operator $\mathcal{G} = \mathcal{G}(P,Q)$ occurs under a canonical coordinate transformation of the phase space labels of the coherent states.

We next demonstrate that the diagonal coherent state matrix elements of a general operator uniquely determine the operator. It will be sufficient to show that if $G(p,q) = 0$ for all (p,q), then it follows that the operator $\mathcal{G} = 0$ follows as a consequence. Consider the construction

$$
\mathcal{G} = \int u(k,x)\, e^{-ixP/\hbar}\, e^{ikQ/\hbar}\, d\mu(k,x) \ ,
$$

which is a slight variant of the Weyl construction [KS68] of an operator. It can be shown that

$$\mathrm{Tr}[\mathcal{G}^\dagger\,\mathcal{G}] = \int |u(k,x)|^2\, d\mu(k,x)\ .$$

Therefore, if follows that if $u(k,x) = 0$ almost everywhere, then $\mathcal{G} = 0$, and conversely. We can put this property to use for us. Consider the diagonal matrix elements

$$\begin{aligned}
\langle p,q|\,\mathcal{G}\,|p,q\rangle &= \int u(k,x)\,\langle p,q|e^{-ixP/\hbar}\,e^{ikQ/\hbar}\,|p,q\rangle\,d\mu(k,x)\\
&= \int u(k,x)\,\langle 0|e^{-ix(P+p)/\hbar}\,e^{ik(Q+q)/\hbar}\,|0\rangle\,d\mu(k,x)\\
&= \int u(k,x)\,e^{i(kq-xp)/\hbar}\,\langle 0|e^{-ixP/\hbar}\,e^{ikQ/\hbar}|0\rangle\,d\mu(k,x)\\
&= \int u(k,x)\,e^{i(kq-xp)/\hbar}\,e^{i\frac{1}{2}kx/\hbar}\,e^{-\frac{1}{4}(k^2+x^2)/\hbar}\,d\mu(k,x)\ .
\end{aligned}$$

Now we make use of the fact that $\langle p,q|\,\mathcal{G}\,|p,q\rangle = 0$ for all (p,q). This implies that the last expression in the previous equation chain vanishes, but it is expressed as the Fourier transform over both k and x of a certain expression which therefore must vanish almost everywhere, namely,

$$u(k,x)\,e^{i\frac{1}{2}kx/\hbar}\,e^{-\frac{1}{4}(k^2+x^2)/\hbar} = 0\ .$$

Since the other factors are nonvanishing for all (k,x), it follows that

$$u(k,x) = 0$$

almost everywhere, which according to our earlier discussion implies that $\mathcal{G} = 0$, completing the argument that the diagonal coherent state matrix elements of a given operator uniquely determine that operator, as was to be shown.

The fact that the diagonal coherent state matrix elements determine the operator uniquely, may not seem so surprising until one examines the analogous issue for a traditional orthonormal basis, say $|q\rangle$. For example, the operator $\exp(ixP/\hbar)$ has general matrix elements in this basis given by

$$\langle q'|e^{ixP/\hbar}|q\rangle = \langle q'|q+x\rangle\ ,$$

which, if $x \neq 0$, has diagonal matrix elements given by

$$\langle q|e^{ixP/\hbar}|q\rangle = \langle q|q+x\rangle = 0$$

for all q, thereby demonstrating that the diagonal elements in such a basis generally do *not* uniquely determine the operator in question.

In complex notation [KS68] , the diagonal coherent state matrix elements are associated with *normal* ordering (all creation operators, $a^\dagger$, to the left of all annihilation operators, a),

$$\langle z| \Sigma_{m,n}\, c_{m,n} a^{\dagger\, m}\, a^n\, |z\rangle = \Sigma_{m,n}\, c_{m,n} z^{*\, m}\, z^n \ ,$$

as follows directly from the basic properties that $a|z\rangle = z|z\rangle$, its adjoint $\langle z|a^\dagger = z^*\langle z|$, and the fact that $\langle z|z\rangle = 1$.

The diagonal coherent state matrix elements of an operator lead to a phase space function associated with that operator. This functional association goes by various names: (1) the Husimi representation [Hus40]; (2) the Q representation [Gla63]; (3) the lower symbol [Ber72]; and (in the author's early papers) [Kla64], the upper symbol hereby changed hereafter to the uppercase symbol to avoid confusion and because of its use of uppercase letters.

The third quantity of interest refers to the use of coherent state projection operators to construct various operators. We already know of one such example, namely,

$$\mathbb{1} = \int |p,q\rangle\langle p,q|\, d\mu(p,q) \ .$$

We generalize this kind of construction so that

$$\mathcal{G} = \int g(p,q)\, |p,q\rangle\langle p,q|\, d\mu(p,q) \ ,$$

and our goal is to see how the phase space function $g(p,q)$ is related to the operator $\mathcal{G}$. To generate this connection, we take the diagonal coherent state matrix elements of the previous relation to find that

$$\begin{aligned} G(k,x) &= \langle k,x|\mathcal{G}|k,x\rangle = \int g(p,q)\, |\langle k,x|p,q\rangle|^2\, d\mu(p,q) \\ &= \int g(p,q)\, e^{-\frac{1}{2}[(k-p)^2 + (x-q)^2]/\hbar}\, d\mu(p,q) \ , \end{aligned}$$

which is immediately seen to be a convolution integral. The Fourier transform of this latter equation over both k and x leads to

$$\begin{aligned} \tilde{G}(q,p) &= \int e^{i(qk - px)/\hbar}\, G(k,x)\, d\mu(k,x) \\ &= \tilde{g}(q,p)\, e^{-\frac{1}{2}[p^2 + q^2]/\hbar} \ , \end{aligned}$$

which thus provides the relation

$$\tilde{g}(q,p) = e^{\frac{1}{2}[p^2 + q^2]/\hbar}\, \tilde{G}(q,p) \ .$$

Although this relation is correct it requires additional comment. It is logical that for certain operators, the behavior of the function $\tilde{G}(q,p)$, the Fourier transform of the diagonal coherent state matrix elements of the operator $\mathcal{G}$, may not fall off so fast as $p^2 + q^2 \to \infty$, in which case the expression $\tilde{g}(q,p)$ may grow rather than decrease as the arguments increase. This feature has

the consequence that the function $\tilde{g}(q,p)$ should be viewed as a distribution, i.e., a generalized function, of the type $\mathcal{D}'$. More specifically, this generalized function is to be regarded as belonging to the *dual* space of the space of test functions $\mathcal{D}$ composed of C_0^∞ functions of two variables, i.e., functions that vanish outside a compact set and that, along with all their derivatives, are continuous. Let $\tilde{u}(q,p) \in \mathcal{D}$ be such a test function; then it is clear that the integral

$$\int \tilde{u}(q,p)^*\, \tilde{g}(q,p)\, d\mu(p,q)$$

is well-defined because the integral is in reality confined to a compact set outside of which $\tilde{u}(q,p) \equiv 0$. In like fashion, the double Fourier transform implies that

$$\int u(k,x)^*\, g(k,x)\, d\mu(k,x)$$

is likewise well-defined. Here $u(k,x)$, the Fourier transform of $\tilde{u}(q,p)$, is an element of a test function space called $\mathcal{Z}(\equiv \widetilde{\mathcal{D}})$, and as a consequence the most general element $g(k,x)$ is an element of the dual space to $\mathcal{Z}$, i.e., to the space $\mathcal{Z}'(\equiv \widetilde{\mathcal{D}}')$ of such distributions.

Historically, the third relation was introduced as the *diagonal representation* by Sudarshan [Sud63] in association with studies in quantum optics. It was later called the P-representation by Glauber, and today it is commonly called the Glauber–Sudarshan representation. The weight function has also been called the upper symbol [Ber72], and the present author [KMC] had called it the lower symbol but has now changed that to the lowercase symbol for clarity and to avoid confusion.

In complex notation [KS68], it follows that the diagonal representation is associated with *anti-normal* ordering as follows from the simple argument that

$$\begin{aligned}\Sigma_{m,n} c_{m,n}\, a^m\, a^{\dagger n} &= \int \Sigma_{m,n} c_{m,n}\, a^m\, |z\rangle\langle z|\, a^{\dagger n}\, d^2z/\pi \\ &= \int \Sigma_{m,n} c_{m,n}\, z^m\, z^{*n}\, |z\rangle\langle z|\, d^2z/\pi\ ,\end{aligned}$$

where we have inserted a resolution of unity in the form

$$\mathbb{1} = \int |z\rangle\langle z|\, d^2z/\pi$$

between the creation and the annihilation operators and set $d^2z \equiv d(\Re z)\, d(\Im z)$; the integral is over the entire complex plane.

7.2.2 Metrics from coherent states

The last topic in this section refers to how coherent states lead to a *phase space metric*. Let us consider the expression

$$d\sigma(p,q)^2 \equiv 2\hbar\,[\,\|\,d|p,q\rangle\,\|^2 - |\langle p,q|\,d|p,q\rangle|^2\,]\ .$$

This combination of vector components is known as the Fubini–Study metric for the problem at hand. From our earlier discussion regarding the one form θ, we find that

$$\begin{aligned} d|p,q\rangle &= de^{-iqP/\hbar}\,e^{ipQ/\hbar}\,|0\rangle \\ &= -(i/\hbar)\,(dqP)\,|p,q\rangle + (i/\hbar)\,e^{-iqP/\hbar}\,e^{ipQ/\hbar}\,dpQ\,|0\rangle\ , \end{aligned}$$

which leads to the expression

$$\begin{aligned} \|\,d|p,q\rangle\,\|^2 &= (1/\hbar^2)[dp^2\,\langle 0|Q^2|0\rangle + dq^2\,\langle p,q|P^2|p,q\rangle \\ &\qquad -dp\,dq\,\langle 0,q|\,(PQ+QP)\,|0,q\rangle\,] \\ &= (1/\hbar^2)[dp^2\,\langle 0|Q^2|0\rangle + dq^2\,\langle 0|P^2|0\rangle + p^2\,dq^2\,]\ . \end{aligned}$$

When combined with the fact that $|\langle p,q|\,d|p,q\rangle| = |p\,dq/\hbar|$, as shown previously, and the fact that $\langle 0|P^2|0\rangle = \langle 0|Q^2|0\rangle = \hbar/2$, this expression leads to

$$d\sigma(p,q)^2 = dp^2 + dq^2\ .$$

Thus, we observe that a metric on phase space has arisen from the coherent states.

However, this metric is a rather special one. Note that it is expressed in Cartesian coordinates. Moreover, the fact that it is expressed in Cartesian coordinates implies that the metric leads to a *flat* (i.e., Euclidean) phase space. In brief, the canonical coherent states have bequeathed to the classical phase space the geometry of a flat space expressed in Cartesian coordinates. The fact that they are Cartesian coordinates is a consequence of the coordinate choice made to define the coherent states; if we change those coordinates, we will in general no longer be led to Cartesian coordinates for the metric. For example, let us consider

$$\begin{aligned} d\overline{\sigma}(\overline{p},\overline{q})^2 &\equiv 2\hbar[\,\|\,d|\overline{p},\overline{q}\rangle\,\|^2 - |\langle\overline{p},\overline{q}|\,d|\overline{p},\overline{q}\rangle|^2\,] \\ &= dp(\overline{p},\overline{q})^2 + dq(\overline{p},\overline{q})^2 \\ &\equiv A(\overline{p},\overline{q})\,d\overline{p}^2 + B(\overline{p},\overline{q})\,d\overline{p}\,d\overline{q} + C(\overline{p},\overline{q})\,d\overline{q}^2\ , \end{aligned}$$

for some suitable coefficients A, B, and C. Although the metric is no longer expressed in terms of Cartesian coordinates, it is noteworthy that it still describes a *flat* space, albeit now in terms of curvilinear coordinates. Since we are confining attention to canonical coordinate transformations, it does follow

that the Jacobian of the transformation is unity since, for such transformations, Liouville's Theorem asserts that

$$d\overline{p}\, d\overline{q} = dp\, dq \ ,$$

a fact that follows directly from the exterior derivative of our initial equation in the previous section, i.e.,

$$d\overline{p} \wedge d\overline{q} = d[\overline{p} d\overline{q}] = d[p\, dq + dF(\overline{q}, q)] = dp \wedge dq \ .$$

7.2.3 Coherent state path integrals—a one form and a metric

In calculating the coherent state propagator above, we arrived at the important intermediate formula

$$\begin{aligned} &K(p'', q'', T; p', q', 0) \\ &\quad = \lim_{N\to\infty} \int \cdot \int \textstyle\prod_{n=0}^{N} \langle p_{n+1}, q_{n+1} | p_n, q_n \rangle \, e^{-i\epsilon H(p_{n+1}, q_{n+1}; p_n, q_n)} \\ &\qquad\qquad \times \textstyle\prod_{n=1}^{N} dp_n\, dq_n / 2\pi \ , \end{aligned}$$

which we now carry forward in a different fashion than before. In particular, we use the fact that if the phase space point $(p_{n+1}, q_{n+1}) \equiv$ "2" is very close to the phase space point $(p_n, q_n) \equiv$ "1", then we may rewrite the expression for $\langle 2|1\rangle$, with normalized states $|2\rangle$ and $|1\rangle$, as

$$\begin{aligned} \langle 2|1\rangle &= 1 - \tfrac{1}{2}[(\langle 2| - \langle 1|)(|2\rangle - |1\rangle)] + \tfrac{1}{2}[\langle 2|1\rangle - \langle 1|2\rangle] \\ &\simeq e^{\frac{1}{2}[\langle 2|1\rangle - \langle 1|2\rangle]} - \tfrac{1}{2}[(\langle 2| - \langle 1|)(|2\rangle - |1\rangle)] - \tfrac{1}{8}[\langle 2|1\rangle - \langle 1|2\rangle]^2 \ , \end{aligned}$$

correct to second order, and which, when we assert that $|2\rangle = |1\rangle + d|1\rangle$, reads

$$\langle 2|1\rangle \simeq e^{i\, i\langle 1|d|1\rangle - \frac{1}{2}[\|d|1\rangle\|^2 - |\langle 1|d|1\rangle|^2]} \ .$$

Applied to a C^1 path in phase space, $p(t), q(t)$, and the associated trajectory of unit vectors, $|p(t), q(t)\rangle$, this latter relation leads to

$$\lim_{N\to\infty} \textstyle\prod_{n=0}^{N} \langle p_{n+1}, q_{n+1} | p_n, q_n \rangle = e^{(i/\hbar)\int i\hbar \langle p, q| \, d|p, q\rangle} \ ,$$

which shows how the one form appears in the path integral. When combined with the Hamiltonian expression, it follows that the coherent state path integral assumes the formal form

$$\begin{aligned} &K(p'', q'', T; p', q', 0) \\ &\quad = \mathcal{M} \int e^{(i/\hbar)\int [\, i\hbar \langle p, q| (d/dt) |p, q\rangle - \langle p, q| \mathcal{H} |p, q\rangle \,] \, dt} \, \mathcal{D}p\, \mathcal{D}q \ . \end{aligned}$$

If one observes this expression just before the continuum limit is taken, one can even see the metric induced by the coherent states appear in an ephemeral fashion, i.e., an appearance which vanishes in the continuum limit as we have taken it. Of course, to make sense of this formal phase space path integral it is necessary to regularize the expression, for example, by means of a proper lattice space regularization. Once again, we emphasize that the physical meanings of the variables p and q as used here are those of *mean values* in the coherent states, i.e., $p = \langle p, q| P |p, q\rangle$ and $q = \langle p, q| Q |p, q\rangle$, thus permitting both to be specified simultaneously.

The point of this section has been to show how the ingredients of the classical theory emerge smoothly out of the coherent state family of vectors and the diagonal matrix elements of the Hamiltonian operator, along with an acceptable physical interpretation of the basic variables p and q.

7.2.4 Alternative coherent state path integral construction

It is pedagogically useful to show that another lattice form can be given to a formal phase space path integral that also leads to the propagator. This construction is based on the diagonal representation discussed above. Let us assume that the self-adjoint Hamiltonian operator $\mathcal{H}(P, Q)$ is given in the form

$$\mathcal{H} = \int h(p, q)\, |p, q\rangle\langle p, q|\, d\mu(p, q)\ .$$

Along with this expression, we also recall the coherent state resolution of unity

$$\mathbb{1} = \int |p, q\rangle\langle p, q|\, d\mu(p, q)\ .$$

We next let $\epsilon = T/(N + 1)$ be a small time interval, and observe that

$$\mathbb{1} - i\epsilon\mathcal{H}/\hbar = \int [1 - i\epsilon h(p, q)/\hbar]\, |p, q\rangle\langle p, q|\, d\mu(p, q)$$

follows by a simple combination of the two former expressions. This relation holds for all ϵ, but now we make an approximation to both sides that is valid only to first order in ϵ:

$$e^{-i\epsilon\mathcal{H}/\hbar} = \int e^{-i\epsilon h(p,q)/\hbar}\, |p, q\rangle\langle p, q|\, d\mu(p, q) + O(\epsilon^2)\ .$$

We next multiply this equation by itself $N + 1$ times and then take coherent state matrix elements of that result to yield

$$\begin{aligned}&\langle p'', q''| e^{-iT\mathcal{H}/\hbar} |p', q'\rangle \\ &\quad = \int \textstyle\prod_{n=0}^{N} \langle p_{n+1}, q_{n+1}|p_n, q_n\rangle \prod_{n=1}^{N} e^{-i\epsilon h(p_n,q_n)/\hbar}\, d\mu(p_n, q_n) + O(\epsilon)\,,\end{aligned}$$

where, as has been the case previously, $p_{N+1}, q_{N+1} = p'', q''$ and $p_0, q_0 = p', q'$. Finally, we can eliminate the error term by taking the limit as $N \to \infty$ leading to the expression

$$\begin{aligned} &K(p'', q'', T; p', q', 0) \\ &\quad = \lim_{N\to\infty} \int \prod_{n=0}^{N} \langle p_{n+1}, q_{n+1} | p_n, q_n \rangle \prod_{n=1}^{N} e^{-i\epsilon h(p_n, q_n)/\hbar} \, d\mu(p_n, q_n) \\ &\quad = \lim_{N\to\infty} \int \prod_{n=0}^{N} e^{(i/\hbar)\frac{1}{2}(p_{n+1}+p_n)(q_{n+1}-q_n)} \\ &\qquad \times e^{-\frac{1}{4}[(p_{n+1}-p_n)^2 + (q_{n+1}-q_n)^2]/\hbar} \prod_{n=1}^{N} e^{-i\epsilon h(p_n, q_n)/\hbar} \, d\mu(p_n, q_n) \, . \end{aligned}$$

If one formally interchanges the limit and the integrals and writes for the integrand the form it takes for continuous and differentiable paths, it then formally follows that

$$K(p'', q'', T; p', q', 0) = \mathcal{M} \int e^{(i/\hbar)\int [p\dot{q} - h(p,q)]\, dt} \, \mathcal{D}p \, \mathcal{D}q \, .$$

It is noteworthy that this version of the coherent state path integral involves the lowercase symbol $h(p, q)$ rather than the uppercase symbol $H(p, q)$ as was the case previously. The fact that two such formulas yield the same result shows that the formal phase space path integrals are *not* well defined; they are both valid in the sense that they each enjoy distinct lattice formulations accounting for their different formal structure.

The present formulation of coherent state path integrals using the lowercase symbol for the Hamiltonian operator is due to Berezin, Lieb, and Onofri [Ber75, Lie73, On80].

7.3 Many Degrees of Freedom

Up to this point we have focused on examples that involve a single degree of freedom, namely, one p and one q. This was reasonable because the big issues of formulation and proper definition of path integrals arise already in full force in the case of a single degree of freedom. As we shall see, the extension to many degrees of freedom is quite straightforward and we shall not need to spend too much time and effort in discussing this aspect. In this chapter we purposely do not treat examples with an infinite number of degrees of freedom; some examples involving infinitely many degrees of freedom are discussed in Chapters 10 and 11.

Let us begin with configuration space path integrals.

7.3.1 Configuration space path integrals

In Chapter 6 we dealt with the configuration space path integrals for a single degree of freedom of the form

$$K(x'', T; x', 0) = \mathcal{N} \int e^{(i/\hbar)\int_0^T [(m/2)\dot{x}(t)^2 - V(x(t))]\, dt}\, \mathcal{D}x \ .$$

We defined this expression as the continuum limit of a lattice space regularization, and we even discussed what may go wrong if one considers a particular potential outside a special class.

In extending this discussion from a single variable x to many variables, say $x \equiv \{x^l\}$, $1 \leq l \leq L < \infty$, there are a number of generalizations to consider. It is convenient to first discuss the several generalizations in the form of the associated classical action I. In particular, we may consider

$$I = \int [(m/2)\dot{x}^2 - V(x)]\, dt \ ,$$

in which

$$\dot{x}^2 \equiv \Sigma_{l=1}^{L} \dot{x}^{l\,2} \equiv \dot{x}^m\, \delta_{l\,m}\, \dot{x}^m \ ,$$

an expression that has an evident rotational symmetry invariance of the kinetic energy, and in its final form entails the summation convention on indices. As another example, there is

$$I = \int [\tfrac{1}{2}\, \dot{x}^l\, M_{l\,m}(x)\, \dot{x}^m - V(x)]\, dt \ ,$$

in which the positive-definite mass matrix $M_{l\,m}(x)$ may be position dependent. These examples may be extended to incorporate a magnetic-field-like interaction, such as

$$I = \int [\tfrac{1}{2}\, \dot{x}^l\, M_{l\,m}(x)\, \dot{x}^m + A_l(x) \dot{x}^l - V(x)]\, dt \ .$$

Apart from the variety of potentials $V(x)$ that may be considered, these few examples constitute the modest family of classical actions that can be treated by natural lattice space limits of configuration space path integrals.

7.3.2 Phase space path integrals

All the examples cited above are contained within the family of possible multivariable, phase space path integrals which involve $p \equiv \{p_l\}$ and $q \equiv \{q^l\}$, $1 \leq l \leq L < \infty$, and are based on classical actions of the form

$$I = \int [p_l\, \dot{q}^l - H(p, q)]\, dt \ .$$

This single form encompasses many possibilities, and because it also covers the configuration space examples discussed above [for suitable choices of the Hamiltonian $H(p, q)$], it is this expression that we carefully formulate in the following discussion.

In quantizing the classical system characterized by the last action functional, it is necessary to assume that the coordinates in which it is presently expressed are Cartesian. In that case, we may define the propagator associated with this system by means of the expressions

$$K(q'', T; q', 0) = \mathcal{M} \int e^{(i/\hbar)\int [p_l \dot{q}^l - H(p,q)]\, dt} \; \Pi_{l=1}^{L} \mathcal{D}p_l \, \mathcal{D}q^l \; ,$$

or even more compactly by

$$K(q'', T; q', 0) = \mathcal{M} \int e^{(i/\hbar)\int [p \cdot \dot{q} - H(p,q)]\, dt} \; \mathcal{D}p \, \mathcal{D}q \; .$$

In any case, the lattice space formulation of this problem is given (with $\hbar = 1$) by

$$\begin{aligned} K(q'', T; q', 0) \; = \; & \lim_{N \to \infty} \; (2\pi)^{-(N+1)\, L} \\ & \times \int e^{i \Sigma_{n=0}^{N} [\frac{1}{2} p_{n+1/2} \cdot (q_{n+1} - q_n) - \epsilon H(p_{n+1/2}, q_n)]} \\ & \times \prod_{l=1}^{L} \prod_{n=0}^{N} dp_{l,\, n+1/2} \prod_{n=1}^{N} dq_n^l \; , \end{aligned}$$

where, as before, $\epsilon = T/(N+1)$, and $q_{N+1} = q''$ and $q_0 = q'$.

7.3.3 Coherent state path integrals

For L degrees of freedom, the coherent states of interest are given by

$$|p, q\rangle \equiv e^{-i\, q \, \cdot \, P/\hbar} \, e^{i p \, \cdot \, Q/\hbar} \, |0\rangle \; ,$$

where

$$(Q_l + i\, P^l) \, |0\rangle = 0 \; ,$$

for all l, $1 \leq l \leq L$. In fact, these states are just product states for the L separate degrees of freedom,

$$|p, q\rangle = \bigotimes_{l=1}^{L} |p_l, q^l\rangle \; .$$

Coherent state path integrals for L degrees of freedom are natural generalizations of those for a single degree of freedom. We start with the same classical action, but, just as for the single degree of freedom case, we take a very different lattice formulation. In particular, the multivariable propagator in the coherent state representation is

$$K(p'', q'', T; p', q', 0) = \mathcal{M} \int e^{(i/\hbar)\int [p \cdot \dot{q} - H(p,q)]\, dt} \; \mathcal{D}p \, \mathcal{D}q \; ,$$

which is defined (for $\hbar = 1$) by

$$K(p'', q'', T; p', q', 0) = \lim_{N\to\infty} (2\pi)^{-NL}$$
$$\times \int e^{\Sigma_{n=0}^{N}\{\frac{1}{2}i(p_{n+1}+p_n)\cdot(q_{n+1}-q_n) - \frac{1}{4}[(p_{n+1}-p_n)^2 + (q_{n+1}-q_n)^2]\}}$$
$$\times e^{-i\epsilon\,\Sigma_{n=0}^{N} H(p_{n+1}, q_{n+1}; p_n, q_n)} \prod_{l=1}^{L}\prod_{n=1}^{N} dp_{l,n}\, dq_n^l \ ,$$

where, as before, $p_{N+1}, q_{N+1} = p'', q''$ and $p_0, q_0 = p', q'$, and

$$H(\tilde{p}, \tilde{q}; p, q) \equiv \langle \tilde{p}, \tilde{q}|\, \mathcal{H}(P, Q)\, |p, q\rangle / \langle \tilde{p}, \tilde{q}|p, q\rangle \ .$$

The caveats noted in the case of the single degree of freedom case apply equally well in the multivariable case. Note that the connection between the classical and quantum Hamiltonian expressions is based on the uppercase symbol

$$H(p, q) = \langle p, q|\, \mathcal{H}(P, Q)\, |p, q\rangle \ .$$

Alternatively, we can choose a different form of lattice approximation that ultimately involves the lowercase symbol $h(p, q)$. In particular (for $\hbar = 1$),

$$K(p'', q'', T; p', q', 0) = \lim_{N\to\infty} (2\pi)^{-NL}$$
$$\times \int e^{\Sigma_{n=0}^{N}\{\frac{1}{2}i(p_{n+1}+p_n)\cdot(q_{n+1}-q_n) - \frac{1}{4}[(p_{n+1}-p_n)^2 + (q_{n+1}-q_n)^2]\}}$$
$$\times e^{-i\epsilon\,\Sigma_{n=1}^{N} h(p_n, q_n)} \prod_{l=1}^{L}\prod_{n=1}^{N} dp_{l,n}\, dq_n^l \ ,$$

with again $p_{N+1}, q_{N+1} = p'', q''$ and $p_0, q_0 = p', q'$. Lastly, we observe in the present case that

$$\mathcal{H}(P, Q) = \int h(p, q)\, |p, q\rangle\langle p, q|\, \prod_{l=1}^{L} d\mu(p_l, q^l) \ .$$

7.4 Spin Coherent State Path Integrals

The coherent state path integrals for canonical degrees of freedom discussed above have entirely analogous constructions for coherent states based on other groups besides the Heisenberg group. In this section we discuss path integrals based on the kinematical variables that arise from the groups SU(2) and SO(3).

7.4.1 Spin coherent states

The Lie algebra for the group SU(2) involves three Hermitian generators (S_1, S_2, S_3) $\equiv \mathbf{S}$, which are related to the infinitesimal generators X_k by $S_k = i\hbar X_k$. Hence, the generators S_k fulfill the commutation relations

$$[S_j, S_k] = i\hbar\epsilon_{jkl} S_l \ ,$$

where ϵ_{jkl} is the totally antisymmetric symbol with $\epsilon_{123} = 1$. As the generator of rotations, $\mathbf{S}$ commutes with any rotationally invariant combination, e.g., $[S_k, \mathbf{S}^2] = 0$, and within inequivalent, irreducible representations [Gil06],

$$\mathbf{S}^2 \equiv S_1^2 + S_2^2 + S_3^2 = s(s+1)\hbar^2 \, \mathbb{1}_s \ ,$$

where the spin s takes on values $s \in \{0, 1/2, 1, 3/2, 2, \ldots\}$, and $\mathbb{1}_s$ is the unit matrix in a $(2s+1)$-dimensional Hilbert space. If $s = 0$, the generators are given by $S_k = 0$; for spin values $s > 0$, the generators may be represented by faithful $(2s+1) \times (2s+1)$ matrices. For each s, we introduce eigenvectors and eigenvalues of S_3 given by

$$S_3 |m; s\rangle = m\hbar |m; s\rangle \ , \qquad -s \leq m \leq s \ ,$$

where these vectors satisfy $\langle m'; s | m; s\rangle = \delta_{m',m}$. The state $|s; s\rangle$ is a maximal weight unit vector and we choose it as the fiducial vector for the associated spin coherent states. In particular, for each spin $s > 0$, the spin coherent states

$$|\theta, \phi\rangle \equiv e^{-i\phi S_3/\hbar} \, e^{-i\theta S_2/\hbar} \, |s; s\rangle$$

are defined for all $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$, and it is customary that the appropriate spin value is left implicit.

It is straightforward to show (see Exercises) that these coherent states admit a resolution of unity given by

$$\mathbb{1}_s \equiv \int |\theta, \phi\rangle\langle\theta, \phi| \, d\mu_s(\theta, \phi) \ ,$$

where

$$d\mu_s(\theta, \phi) = [(2s+1)/4\pi] \, \sin(\theta) \, d\theta \, d\phi \ .$$

The validity of this weighting for the integration measure follows trivially if one simply takes the trace of the resolution of unity formula.

The overlap function for two different spin coherent states is given by

$$\langle\theta', \phi' | \theta, \phi\rangle = [\cos(\tfrac{1}{2}\theta')\cos(\tfrac{1}{2}\theta) e^{i\frac{1}{2}(\phi' - \phi)} + \sin(\tfrac{1}{2}\theta')\sin(\tfrac{1}{2}\theta) e^{-i\frac{1}{2}(\phi' - \phi)}]^{2s}.$$

Observe that this overlap function can serve as a reproducing kernel for a reproducing kernel Hilbert space. Different phase conventions for these states have been used by different authors.

Although the spin coherent states are completely different from the canonical coherent states of the first sections of this chapter, they have several general properties in common with those states. These properties include the fact that the diagonal coherent state matrix elements

$$G(\theta,\phi) \equiv \langle \theta,\phi|\mathcal{G}(\mathbf{S})\,|\theta,\phi\rangle$$

uniquely determine the operator $\mathcal{G}(\mathbf{S})$, and that every operator [i.e., every $(2s+1)\times(2s+1)$ matrix] can be represented in the diagonal form

$$\mathcal{G}(\mathbf{S}) = \int g(\theta,\phi)\,|\theta,\phi\rangle\langle\theta,\phi|\,d\mu_s(\theta,\phi)\ .$$

However, for the spin coherent states, the weight function $g(\theta,\phi)$ is not unique, unlike the case for the canonical coherent states.

7.4.2 Spin dynamics and the spin coherent state path integral

The abstract Schrödinger equation for a spin system is given by

$$i\hbar\,\partial|\psi(t)\rangle/\partial t = \mathcal{H}(\mathbf{S})\,|\psi(t)\rangle\ ,$$

which in the spin coherent state representation reads

$$i\hbar\,\partial\psi(\theta,\phi,t) = \int \langle\theta,\phi|\mathcal{H}(\mathbf{S})|\theta',\phi'\rangle\,\psi(\theta',\phi',t)\,d\mu_s(\theta',\phi')\ ,$$

where $\psi(\theta,\phi)\equiv\langle\theta,\phi|\psi\rangle$. The solution of this equation is given by

$$\psi(\theta',\phi',T) = \int K(\theta',\phi',T;\theta,\phi,0)\,\psi(\theta,\phi,0)\,d\mu_s(\theta,\phi)\ ,$$

where

$$K(\theta',\phi',T;\theta,\phi,0) \equiv \langle\theta',\phi'|\,e^{-i\mathcal{H}T/\hbar}|\theta,\phi\rangle$$

denotes the propagator in the spin coherent state representation.

Following exactly the same steps which led to the canonical coherent state path integral, it follows that the propagator can be represented by the expression

$$\begin{aligned} K(\theta'',\phi'',T;\theta',\phi',0) \\ = \lim_{N\to\infty}\int\cdot\int \prod_{n=0}^{N}\langle\theta_{n+1},\phi_{n+1}|\theta_n,\phi_n\rangle\,e^{-i\epsilon H(\theta_{n+1},\phi_{n+1};\theta_n,\phi_n)} \\ \times\prod_{n=1}^{N} d\mu_s(\theta_n,\phi_n)\ , \end{aligned}$$

where $(\theta'',\phi'')=(\theta_{N+1},\phi_{N+1})$, $(\theta',\phi')=(\theta_0,\phi_0)$, $\epsilon = T/(N+1)$, and

$$H(\theta_{n+1},\phi_{n+1};\theta_n,\phi_n) \equiv \frac{\langle\theta_{n+1},\phi_{n+1}|\,\mathcal{H}(\mathbf{S})\,|\theta_n,\phi_n\rangle}{\langle\theta_{n+1},\phi_{n+1}|\theta_n,\phi_n\rangle}\ .$$

Formally interchanging the limit and integrations and writing for the integrand the form it assumes for continuous and differentiable paths leads to the formal expression

$$K(\theta'', \phi'', T; \theta', \phi', 0) = \int e^{(i/\hbar)\int [i\hbar \langle \theta, \phi | (d/dt) | \theta, \phi \rangle - H(\theta, \phi)]\, dt}\, \mathcal{D}\mu_s(\theta, \phi)$$

where

$$H(\theta, \phi) \equiv \langle \theta, \phi | \mathcal{H}(\mathbf{S}) | \theta, \phi \rangle \ .$$

As one forms, it follows that

$$i\hbar \langle \theta, \phi | d | \theta, \phi \rangle = s\hbar \cos(\theta)\, d\phi \ ,$$

and thus our final, formal, expression for the spin coherent state path integral is

$$K(\theta'', \phi'', T; \theta', \phi', 0) = \mathcal{N} \int e^{(i/\hbar)\int [s\hbar \cos(\theta)\, \dot{\phi} - H(\theta, \phi)]\, dt} \prod_t \sin(\theta)\, d\theta\, d\phi \, .$$

Just as with the canonical coherent states, there is an alternative formulation of the spin coherent state path integral that uses the lowercase symbol for the Hamiltonian. If we introduce

$$\mathcal{H}(\mathbf{S}) = \int h(\theta, \phi)\, |\theta, \phi\rangle \langle \theta, \phi|\, d\mu_s(\theta, \phi) \ ,$$

and follow the pattern developed earlier, we are led to the expression

$$\begin{aligned} K(\theta'', \phi'', T; \theta', \phi', 0) \ = \ & \lim_{N \to \infty} \int \cdot \int \prod_{n=0}^{N} \langle \theta_{n+1}, \phi_{n+1} | \theta_n, \phi_n \rangle \\ & \times \prod_{n=1}^{N} e^{-i\epsilon h(\theta_n, \phi_n)/\hbar}\, d\mu_s(\theta_n, \phi_n) \ . \end{aligned}$$

Interchanging the limit and the integrations, and writing the integrand for continuous and differentiable paths leads to the formal expression

$$K(\theta'', \phi'', T; \theta', \phi', 0) = \mathcal{N} \int e^{(i/\hbar)\int [s\hbar \cos(\theta)\, \dot{\phi} - h(\theta, \phi)]\, dt} \prod_t \sin(\theta)\, d\theta\, d\phi \ .$$

It is of interest to observe there is an argument from the program of geometric quantization that restricts s to its proper quantized values. In the integrand of expressions above there appears the quantity

$$e^{(i/\hbar)\int s\hbar \cos(\theta)\, d\phi} \ ,$$

which, if we specialize to a closed orbit where $(\theta'', \phi'') = (\theta', \phi')$, can be written as either one of two separate, two-dimensional integrals, giving rise to the equality

$$e^{i\, s \int_{top} \sin(\theta)\, d\theta\, d\phi} = e^{-i\, s \int_{bottom} \sin(\theta)\, d\theta\, d\phi} \ ,$$

where "*top*" and "*bottom*" refer to the two portions of the spherical shell as bounded by the closed path of interest. This equation can be recast as

$$e^{i\,s\int_{all}\,\sin(\theta)\,d\theta\,d\phi}=1\ ,$$

where "*all*" refers to an integral over the entire spherical surface. Since that surface integral is 4π it follows that

$$e^{i\,4\pi s}=1\ ,$$

or stated otherwise that $s\in\{0,1/2,1,3/2,2,\ldots\}$ as desired.

We will revisit the spin coherent state path integral in the following chapter.

7.5 Affine Coherent State Path Integrals

The affine Lie algebra has been introduced in Sec. 5.3 as a two-parameter, non-Abelian Lie algebra with the only nonvanishing commutator being $[X_a,X_b]=X_b$. Let us derive the same Lie algebra in a different way. Consider the Heisenberg pair Q and P which satisfy the usual commutation relation, $[Q,P]=i\hbar\mathbb{1}$. The Schrödinger representation is such that the spectrum of both Q and P runs from $-\infty$ to $+\infty$. Suppose, however, one dealt with an application in which the classical variable q corresponding to Q satisfied the bound $q>0$. There is no conflict in the classical theory with such a restriction on the basic phase space variables, but there indeed is a conflict within the quantum theory.

7.5.1 Affine coherent states

One way around this conundrum is to multiply the Heisenberg commutator by Q leading to the relation

$$[Q,P]Q=i\hbar Q=[Q,(QP+PQ)/2]\equiv[Q,D]\ ,$$

which suggests that we examine the commutation relation given by

$$[Q,D]=i\hbar Q\ ,$$

which generates a non-Abelian, two-parameter Lie algebra, which we call the affine Lie algebra. Unlike the Heisenberg Lie algebra, the affine Lie algebra admits three, inequivalent, irreducible self-adjoint realizations: one for which the spectrum of $Q>0$, one for which $Q<0$, and a third for which $Q=0$. Our interest lies in the first representation for which $Q>0$. Observe, for $q>0$, that

$$e^{i\,\ln(q)\,D/\hbar}Q\,e^{-i\,\ln(q)\,D/\hbar}=qQ\ ,$$

which shows that such transformations preserve the positivity of the spectrum of Q. Let us simplify matters and set $\hbar=1$ for the further discussion of the affine story. Consequently, the operator

$$A(p,q) \equiv e^{ipQ}\, e^{-i\,\ln(q)\,D}$$

is a two-parameter unitary transformation for all $p \in \mathbb{R}$ and $q > 0$. As affine coherent states, we choose

$$|p,q\rangle \equiv e^{ipQ}\, e^{-i\,\ln(q)\,D}\, |\eta\rangle \ ,$$

and pick the fiducial vector $|\eta\rangle$ so that the coherent state overlap function is

$$\langle p', q' | p, q\rangle = \left[\frac{(q'\,q)^{-1/2}}{\frac{1}{2}(q'^{-1} + q^{-1}) + i\frac{1}{2}\beta^{-1}(p'-p)} \right]^{2\beta} .$$

In particular, this means that the fiducial vector has been chosen, in the basis $|x\rangle$ where $Q|x\rangle = x|x\rangle$, $x > 0$, so that

$$\langle x|\eta\rangle = N\, x^{\beta - 1/2}\, e^{-\beta x} \ .$$

For any $\beta > 0$, the coherent state overlap function defines a function of positive type and can serve as a reproducing kernel for a reproducing kernel Hilbert space. On the other hand, if $\frac{1}{2} \geq \beta > 0$, the coherent states lack a local integral resolution of unity, but when $\beta > \frac{1}{2}$ we are assured that a local integral resolution of unity for the coherent states exists.

When $\beta > \frac{1}{2}$, the affine coherent state resolution of unity takes the form

$$\mathbb{1} = \int |p,q\rangle\langle p,q|\, d\mu_A(p,q)$$

in which the integration extends over the half plane $\mathbb{R} \times \mathbb{R}^+$, and

$$d\mu_A(p,q) \equiv \{[1 - 1/(2\beta)]/2\pi\}\, dp\, dq \ .$$

Observe, in the chosen coordinates, that the left-invariant group measure is proportional to $dp\,dq$, which is the proper version to ensure a resolution of unity provided an appropriate admissibility condition is satisfied [AK69].

We note further that the diagonal affine coherent state matrix elements determine the operator, i.e., that the information within

$$G(p,q) \equiv \langle p,q|\mathcal{G}(D,Q)|p,q\rangle = \langle \eta|\mathcal{G}(D + pqQ, qQ)|\eta\rangle$$

is sufficient to fully determine the operator $\mathcal{G}(D,Q)$. In a related manner, we note that the diagonal representation holds for affine coherent states in the fashion

$$\mathcal{G}(D,Q) = \int g(p,q)\, |p,q\rangle\langle p,q|\, d\mu_A(p,q) \ ,$$

and furthermore, that the diagonal representation is one-to-one.

7.5.2 Affine dynamics and the affine coherent state path integral

At this point, the pattern of construction is familiar, and we essentially quote only the results. In the affine coherent state representation the evolution equation (with $\hbar = 1$ still) reads

$$\psi(p'', q'', T) = \int K(p'', q'', T; p', q', 0)\, \psi(p', q', 0)\, d\mu_A(p', q') \ ,$$

and the path integral affine coherent state representation for the propagator is given [for $\epsilon = T/(N+1)$, $(p_{N+1}, q_{N+1}) = (p'', q'')$, and $(p_0, q_0) = (p', q')$] by

$$\begin{aligned} K(p'', q'', T; p', q', 0) &= \langle p'', q'' | e^{-i\mathcal{H}T} | p', q' \rangle \\ &= \lim_{N\to\infty} \int \cdot \int \textstyle\prod_{n=0}^{N} \langle p_{n+1}, q_{n+1} | p_n, q_n \rangle \, e^{-i\epsilon H(p_{n+1}, q_{n+1}; p_n, q_n)} \\ &\qquad\qquad \times \textstyle\prod_{n=1}^{N} d\mu_A(p_n, q_n) \\ &= \int e^{i\int [i\langle p,q|(d/dt)|p,q\rangle - \langle p,q|\mathcal{H}|p,q\rangle]\, dt} \textstyle\prod_t d\mu_A(p,q) \\ &= \mathcal{M} \int e^{i\int [-q\dot{p} - H(p,q)]\, dt}\, \mathcal{D}p\, \mathcal{D}q \ , \end{aligned}$$

where the last two relations are formal, and

$$\begin{aligned} H(p_{n+1}, q_{n+1}; p_n, q_n) &= \frac{\langle p_{n+1}, q_{n+1} | \mathcal{H} | p_n, q_n \rangle}{\langle p_{n+1}, q_{n+1} | p_n, q_n \rangle} \ , \\ H(p,q) &= \langle p, q | \mathcal{H} | p, q \rangle \ , \\ i\langle p, q | d | p, q \rangle &= -q\, dp \ . \end{aligned}$$

Although the formulas and even the notation are similar to those of the canonical coherent states, it is important to keep in mind that the range of the variables in these expressions is always $(p, q) \in \mathbb{R} \times \mathbb{R}^+$.

An alternative construction of the path integral uses the lowercase symbol relating the system Hamiltonian in the fashion

$$\mathcal{H}(D, Q) = \int h(p,q)\, |p,q\rangle\langle p,q|\, d\mu_A(p,q) \ .$$

As was the case previously, we can incorporate the lowercase symbol with the construction

$$\begin{aligned} K(p'', q'', T; p'q', 0) &= \langle p'', q'' | e^{-i\mathcal{H}T} | p', q' \rangle \\ &= \lim_{N\to\infty} \int \cdot \int \textstyle\prod_{n=0}^{N} \langle p_{n+1}, q_{n+1} | p_n, q_n \rangle \prod_{n=1}^{N} e^{-i\epsilon h(p_n, q_n)}\, d\mu_A(p_n, q_n) \\ &= \int e^{i\int [i\langle p,q|(d/dt)|p,q\rangle - h(p,q)]\, dt} \textstyle\prod_t d\mu_A(p,q) \\ &= \mathcal{M} \int e^{i\int [-q\dot{p} - h(p,q)]\, dt}\, \mathcal{D}p\, \mathcal{D}q \ . \end{aligned}$$

Despite appearances, these expressions refer to affine variables and not canonical variables. For the quantum theory this means we use $\mathcal{H}(D, Q)$ instead of $\mathcal{H}(P, Q)$. For the integration variables, it means we use $(p, q) \in \mathbb{R} \times \mathbb{R}^+$ instead of $(p, q) \in \mathbb{R}^2$.

This completes our preliminary discussion of coherent state path integrals. We shall revisit the subject again in the next chapter from a rather different point of view.

7.6 Coherent State Path Integrals without a Resolution of Unity

We have noted that when the parameter β that appears in the fiducial vector for affine coherent states is too small, the resolution of unity expressed as a local integral over affine coherent state projection operators is not possible. This lack would seem to cause a fatal blow to the construction of an affine coherent state path integral; nevertheless, it is still possible to do so! We shall not review the construction here, but instead point the interested reader to [Kl01a, HK04].

Exercises

7-1 We have focused on canonical coherent states of the form

$$|p, q\rangle = e^{-iqP/\hbar}\, e^{ipQ/\hbar}\, |\eta\rangle \ ,$$

where $|\eta\rangle = |0\rangle$ with $(Q + iP)|0\rangle = 0$. As a first exercise, let $|\eta\rangle$ be *any* unit vector in Hilbert space, and let $\eta(x) \equiv \langle x|\eta\rangle$ be the Schrödinger representative of the fiducial vector. With the choice of $|\eta\rangle$ left implicit in the definition of the coherent states, show that whatever the coherent state overlap function $\langle p'', q''|p', q'\rangle$ is, it satisfies the integral equation

$$\langle p'', q''|p', q'\rangle = \int \langle p'', q''|p, q\rangle\langle p, q|p', q'\rangle d\mu(p, q) \ ,$$

where $d\mu(p, q) = dpdq/2\pi\hbar$.

7-2 Consider spin coherent states of the form (with $\hbar = 1$)

$$|\theta, \phi\rangle = e^{-i\phi S_3}\, e^{-i\theta S_2}\, |s; s\rangle \ ,$$

where $S_3\,|s; s\rangle = s|s; s\rangle$ for arbitrary spin $s > 0$. Show that the coherent state overlap function $\langle\theta'', \phi''|\theta', \phi'\rangle$ satisfies the integral equation

$$\langle\theta'',\phi''|\theta',\phi'\rangle = \int \langle\theta'',\phi''|\theta,\phi\rangle\,\langle\theta,\phi|\theta',\phi'\rangle\,d\mu_s(\theta,\phi)\;,$$

where $d\mu_s(\theta,\phi) = [(2s+1)/4\pi]\,\sin(\theta)\,d\theta\,d\phi$.

7-3 Consider the spin coherent states for spin one (with $\hbar = 1$) given by

$$|\theta,\phi\rangle \equiv e^{-i\phi S_3}\,e^{-i\theta S_2}\,|\eta\rangle\;,$$

where the operators S_2 and S_3 are 3×3 matrices. Find what conditions, if any, on $|\eta\rangle$ are required so that the coherent state overlap function $\langle\theta'',\phi''|\theta',\phi'\rangle$ satisfies the integral equation

$$\langle\theta'',\phi''|\theta',\phi'\rangle = \int \langle\theta'',\phi''|\theta,\phi\rangle\,\langle\theta,\phi|\theta',\phi'\rangle\,d\mu_1(\theta,\phi)\;,$$

where, for the spin one case, $d\mu_1(\theta,\phi) = (3/4\pi)\,\sin(\theta)\,d\theta\,d\phi$.

7-4 For a harmonic oscillator with quantum Hamiltonian $\mathcal{H} = \frac{1}{2}(P^2+Q^2)$, determine the propagator in a coherent state representation based on the coherent states

$$|p,q\rangle = e^{-iqP/\hbar}\,e^{ipQ/\hbar}\,|0\rangle\;,$$

where $|0\rangle$, as usual, satisfies $(Q+iP)\,|0\rangle = 0$, and which, coincidentally, means that $|0\rangle$ is the ground state of the Hamiltonian $\mathcal{H}$ that generates the dynamics.

7-5 Repeat problem **7-4** with the modification that the quantum Hamiltonian is changed to $\mathcal{H} = \frac{1}{2}(P^2+\omega^2 Q^2)$, $\omega \neq 1$, while the coherent states remain unchanged. In particular, note that the fiducial vector for the coherent states is no longer the ground state of the Hamiltonian that generates the temporal evolution.

7-6 The affine coherent state overlap function is given (with $\hbar = 1$) by

$$\langle p',q'|p,q\rangle = \left[\frac{(q'q)^{-1/2}}{\frac{1}{2}(q'^{-1}+q^{-1}) + i\frac{1}{2}\beta^{-1}(p'-p)}\right]^{2\beta}\;,$$

defined for (p,q) & $(p',q') \in (\mathbb{R},\mathbb{R}^+)$. Change coordinates $q \to q+\sqrt{\beta}$ and $q' \to q'+\sqrt{\beta}$ so that in the new coordinates $q > -\sqrt{\beta}$ and $q' > -\sqrt{\beta}$. Then, take the original affine coherent state overlap function given above, now reexpressed in the translated coordinates, and determine the limit as the parameter $\beta \to \infty$. Interpret your answer in terms of a corresponding limit of the original affine commutation relation $[Q,D] = iQ$, $Q > 0$.

8

Continuous-Time Regularized Path Integrals

8.1 Wiener Measure Regularization of Phase Space Path Integrals

Up to this point we have focused on temporal lattice regularizations in order to give meaning to the traditional configuration and phase space path integrals including coherent state phase space path integrals. In this chapter, we take up the question of introducing a *continuous-time regularization procedure* into a formal phase space path integral.

Already in Sec. 6.1.10, we discussed the proposal of Gel'fand and Yaglom for inserting a Wiener measure factor in a formal configuration space path integral and taking a limit in which that factor became unity after the evaluation of the functional integral. That procedure failed, but a modification of their proposal by Itô, sketched in Sec. 6.1.11, provided a successful realization of their ideas, up to a point. We now carry that kind of regularization further, inserting Wiener measure factors into a formal phase space path integral, and show that good results emerge.

In particular, we consider the expression

$$\lim_{\nu\to\infty} \mathcal{M} \int e^{(i/\hbar)\int [p\,\dot{q} - h(p,q)]\,dt}\; e^{-(1/2\nu)\int [\dot{p}^2 + \dot{q}^2]\,dt}\, \mathcal{D}p\,\mathcal{D}q\ .$$

Here we have inserted Wiener measure regularizations for both p and q. This expression raises several immediate questions: (i) does the limit exist for some Hamiltonian expressions $h(p,q)$; (ii) does that limit solve Schrödinger's equation, and if so, what representation is involved, and (iii) for what class of Hamiltonian operators $\mathcal{H} = \mathcal{H}(P,Q)$ does this expression lead to the correct propagator? Before entering into a detailed discussion of these questions, let us briefly summarize the answers: the limit exists for a wide class of lowercase symbols, and it solves Schrödinger's equation in a *coherent-state representation* for a wide class of Hamiltonians including all semibounded, Hermitian Hamiltonians that are polynomials in the variables P and Q [DK85]. In other words, for the wide class of Hamiltonians for which the formula applies,

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9_8,

$$K(p'',q'',T;p',q',0) = \langle p'',q''|e^{-iT\mathcal{H}/\hbar}|p',q'\rangle$$
$$= \lim_{\nu\to\infty} \mathcal{M} \int e^{(i/\hbar)\int[p\dot{q}-h(p,q)]\,dt}\, e^{-(1/2\nu)\int[\dot{p}^2+\dot{q}^2]\,dt}\, \mathcal{D}p\,\mathcal{D}q$$
$$= \lim_{\nu\to\infty} 2\pi\hbar e^{\nu T/2\hbar} \int e^{(i/\hbar)\int[p\,dq-h(p,q)\,dt]}\, d\mu_W^\nu(p,q)\ ,$$

where $\mu_W^\nu(p,q)$ is a two-variable Wiener process on a flat plane expressed in Cartesian coordinates which is pinned so that $p(T),q(T) = p'',q''$ and $p(0),q(0) = p',q'$. In the last, nonformal, line of this expression, the variables are Brownian motion paths for all $\nu < \infty$, where ν represents the diffusion constant for the Brownian motion. As discussed in Sec. 4.2.2, such paths are continuous but nowhere differentiable. The integral $\int p\,dq$ is interpreted as

$$\int p\,dq \equiv \lim_{N\to\infty} \Sigma_{n=0}^N \tfrac{1}{2}(p_{n+1}+p_n)(q_{n+1}-q_n)\ ,$$

which is the Stratonovich form of the stochastic integral $\int p\,dq$ defined as a lattice limit. As noted in Sec. 4.2.7, the Stratonovich form for a stochastic integral obeys the ordinary rules of calculus, a property we shall soon put to good use. It is also significant that the lowercase symbol $h(p,q)$ is the expression that represents the Hamiltonian; as a consequence, the connection between $\mathcal{H}(P,Q)$ and $h(p,q)$ is given by

$$\mathcal{H}(P,Q) = \int h(p,q)\,|p,q\rangle\langle p,q|\,d\mu(p,q)\ .$$

Finally, the expression for the propagator

$$\lim_{\nu\to\infty} 2\pi\hbar e^{\nu T/2\hbar} \int e^{(i/\hbar)\int[p\,dq-h(p,q)\,dt]}\, d\mu_W^\nu(p,q)$$

provides a *continuous-time regularization scheme* for which, for all $\nu < \infty$, the path integral is *absolutely well-defined*; in addition, this formulation has the virtue that it involves *genuine paths* $p(t),q(t)$ *which are everywhere continuous*. A description with genuine paths is possible because the insertion of the Wiener measure regularizing factor has *automatically* given rise to a coherent state representation in which the physical meaning of the variables involved is that of mean momentum and mean position. It is hard to imagine any more satisfying path integral formulation than the one given in this subsection!

Some additional background information may be instructive. One may view the construction given so far as merely inserting a regularizing factor into the integrand of a formal phase space path integral, and then removing that regularization after all the integrations have been carried out. A simple but analogous situation occurs if we consider the conditionally convergent one-dimensional integral

$$I = \int_{-\infty}^{\infty} e^{iy^2/2}\,dy$$

to which we impart some meaning by defining the answer to be that given by

$$I = \lim_{\nu\to\infty} \int e^{iy^2/2 - y^2/2\nu}\, dy = \lim_{\nu\to\infty} [(2\pi\nu)/(1-i\nu)]^{1/2} = \sqrt{2\pi i}\ .$$

Regarding the regularization of the formal phase space path integral in this manner (and setting $\hbar = 1$), we are favored by the facts that: (1) the limit as $\nu \to \infty$ actually converges for a wide class of Hamiltonian functions $h(p,q)$; (2) the limit defines a solution to Schrödinger's equation in a coherent state representation based on a set of vectors $\{\,|p,q\rangle\,\}$ for all $(p,q) \in \mathbb{R}^2$; (3) further, those coherent states are defined by $|p,q\rangle = e^{-iqP}\, e^{ipQ}\, |0\rangle$; (4) the self-adjoint operators Q and P satisfy the commutation relation $[Q,P] = i\mathbb{1}$, and, lastly, the fiducial vector $|0\rangle$ is a unit vector that satisfies the equation $(Q + iP)\,|0\rangle = 0$.

Let us make sure that the content of this last paragraph has been fully appreciated. Starting with the formal, improperly defined phase space path integral, and adding the factors appropriate for a Wiener measure regularization on the two-dimensional plane, *the rest of the story automatically follows.* That is: (i) the existence of the limit for suitable $h(p,q)$, (ii) the representation of the propagator in a coherent state representation, the coherent states involved being the canonical coherent states based on the kinematical variables P and Q with $[Q,P] = i\mathbb{1}$, (iii) the fiducial vector $|0\rangle$ being a normalized solution of $(Q + iP)\,|0\rangle = 0$, and finally, (iv) the Hamiltonian operator $\mathcal{H}(P,Q)$ having a diagonal representation with the weight being the symbol $h(p,q)$. In brief, the Wiener measure has decided everything!

[**Remark:** The reader would be wise at this point to raise the question: what happens if instead of the Wiener measure factor being chosen as $\dot{p}^2 + \dot{q}^2$, it was chosen as $\omega^{-1}\dot{p}^2 + \omega\dot{q}^2$, with $0 < \omega \neq 1$? In that case, the story is the same except that the fiducial vector now obeys the equation $(\omega Q + iP)\,|0\rangle = 0$, and thus the canonical coherent states are changed as well. How could such a modification—which formally disappears in the ultra-diffusive limit in which $\nu \to \infty$—change things? The answer recalls Chapter 1, where we stressed that a conditionally convergent series (or integral) may converge to different limiting values if the order of summation (or integration) is changed. That is essentially what happens when one passes from $\omega = 1$ to $\omega \neq 1$! For the present, let us revert to our standard choice of regularization for which $\omega = 1$.]

8.1.1 Covariance under canonical coordinate transformations

The transformation of the propagator presented above under a canonical coordinate transformation basically follows the story as presented earlier. In particular, under a coordinate transformation from the canonical coordinates (p,q) to $(\overline{p},\overline{q})$, it follows that

$$\overline{K}(\overline{p}'',\overline{q}'',T';\overline{p}',\overline{q}',0) = \langle \overline{p}'',\overline{q}''|e^{-iT\mathcal{H}/\hbar}|\overline{p}',\overline{q}'\rangle$$

$$= \lim_{\nu\to\infty} \mathcal{M} \int e^{(i/\hbar)\int [\overline{p}\,\dot{\overline{q}} + \dot{\overline{G}}(\overline{p},\overline{q}) - \overline{h}(\overline{p},\overline{q})]\,dt} \times e^{-(1/2\nu)\int [A\dot{\overline{p}}^2 + B\dot{\overline{p}}\,\dot{\overline{q}} + C\dot{\overline{q}}^2]\,dt}\, \mathcal{D}\overline{p}\,\mathcal{D}\overline{q}$$
$$= \lim_{\nu\to\infty} 2\pi\hbar\, e^{\nu T/2\hbar} \int e^{(i/\hbar)\int [\overline{p}\,d\overline{q} + d\overline{G}(\overline{p},\overline{q}) - \overline{h}(\overline{p},\overline{q})\,dt]}\, d\overline{\mu}^{\nu}_{W}(\overline{p},\overline{q})\,,$$

where the Wiener measure $\overline{\mu}^{\nu}_{W}$ describes Brownian motion paths on a flat plane but now expressed generally in terms of curvilinear coordinates. Under a coordinate transformation, the Hamiltonian symbol $h(p,q)$ transforms as a scalar, while the one form $p\,dq$ transforms as in the classical analysis.

In Sec. 8.4 we shall argue that the feature of invariance under canonical coordinate transformations displayed here leads to a truly geometric version of the quantization procedure!

8.1.2 Proof of Wiener measure path integral regularization

We begin with the observation that all elements of the coherent state representation satisfy a complex polarization condition. In particular, it follows (for $\hbar = 1$) that

$$\begin{aligned}(-\partial/\partial q + i\,\partial/\partial p + ip)\langle p,q|\psi\rangle &= (-\partial/\partial q + i\,\partial/\partial p + ip)\,\langle 0|\, e^{-ipQ}\, e^{iqP}\,|\psi\rangle \\ &= \langle 0|(Q - iP)\, e^{-ipQ}\, e^{iqP}\,|\psi\rangle \\ &= 0\end{aligned}$$

due to the fact that $(Q + iP)|0\rangle = 0$. Let $B = (-\partial/\partial q + i\,\partial/\partial p + ip)$, for which $B\,\langle p,q|\psi\rangle = 0$ for all $|\psi\rangle$. Then, it also follows that

$$A\,\langle p,q|\psi\rangle = 0$$

for all $|\psi\rangle$, where

$$\begin{aligned}A \equiv B^{\dagger}B &= (\partial/\partial q + i\,\partial/\partial p - ip)\,(-\partial/\partial q + i\,\partial/\partial p + ip) \\ &= (-i\partial/\partial p)^2 + (-i\partial/\partial q - p)^2\,,\end{aligned}$$

which is a *nonnegative*, second-order differential operator. The operator A can be viewed as the Hamiltonian for a *two-dimensional* free particle—now with "coordinates" p and q—in the presence of a magnetic field since it has the form $(1/2m)[(P_p - A_p)^2 + (P_q - A_q)^2]$, where $2m = 1$, and the "vector potential" is $A_p = 0$ and $A_q = p$. Consequently, the "magnetic field" is $B = \nabla \times A = 1$, a constant value in the direction perpendicular to the two-dimensional (p,q) plane.

The spectrum of A is readily found since it involves the Landau levels, and that spectrum is $\mathrm{spec}(A) = \{1,3,5,\ldots\}$. In particular, the eigenvectors and eigenvalues of A, which satisfy $A\,\psi_{n,\kappa}(p,q) = E_{n.\kappa}\,\psi_{n,\kappa}(p,q)$—where $n \in \{0,1,2,\ldots\}$ and $\kappa \in \mathbb{R}$—are given by

$$\psi_{n,\kappa}(p,q) = \frac{1}{\sqrt{2\pi}}\, e^{i\kappa q}\, h_n(p-\kappa)\ , \qquad E_{n,\kappa} = (2n+1)\ ,$$

where the functions $h_n(p)$ are the usual Hermite functions (see Sec. 3.5), for which, in particular,

$$h_0(p-\kappa) = \pi^{-1/4}\, e^{-\frac{1}{2}(p-\kappa)^2}\ .$$

It is convenient to introduce kets $|p;q)$ for the two-dimensional problem for which $Q_p\,|p;q) = p\,|p;q)$, $Q_q\,|p;q) = q\,|p;q)$, and

$$(p;q|p';q') = \delta(p-p')\,\delta(q-q')\ .$$

For the eigenvectors of A, we introduce the kets $|n;\kappa)$ for which $A|n;\kappa) = (2n+1)\,|n;\kappa)$. Thus the eigenfunctions $\psi_{n,\kappa}(p,q) = (p;q|n;\kappa)$, and the projection operator $\mathbb{P}_n$ onto the nth Landau level is given by

$$\mathbb{P}_n = \int |n;\kappa)(n;\kappa|\, d\kappa\ .$$

In this language, we learn that

$$\begin{aligned}
2\pi \lim_{\nu\to\infty} e^{-\frac{1}{2}\nu T(A-1)}(p;q|p';q') &= 2\pi \lim_{\nu\to\infty} \textstyle\sum_{n=0}^{\infty}(p;q|e^{-\nu T n}\,\mathbb{P}_n\,|p';q') \\
&= 2\pi(p;q|\,\mathbb{P}_0\,|p';q') \\
&= 2\pi \int (p;q|0;\kappa)(0;\kappa|p';q')\, d\kappa \\
&= \pi^{-1/2} \int e^{i\kappa(q-q') - \frac{1}{2}(\kappa-p)^2 - \frac{1}{2}(\kappa-p')^2}\, d\kappa \\
&= e^{i\frac{1}{2}(p+p')(q-q') - \frac{1}{4}[(p-p')^2 + (q-q')^2]} \\
&= \langle p,q|p',q'\rangle\ .
\end{aligned}$$

Note carefully that the result of this projection onto the lowest Landau level has led to the *coherent state overlap function for the one-dimensional problem!*

We use this result now to derive the one-dimensional phase space path integral by a "two-dimensional configuration space" construction following the pattern developed in Sec. 6.1.6. In particular, setting $C \equiv \frac{1}{2}\nu(A-1)$ for convenience, we consider

$$\begin{aligned}
&(p'';q''|\, e^{-TC - iTh}\,|p';q') \\
&\qquad = (p'';q''|\, e^{-\epsilon C - i\epsilon h}\, e^{-\epsilon C - i\epsilon h} \cdots e^{-\epsilon C - i\epsilon h}|p';q') \\
&\qquad = \int \cdot \int \textstyle\prod_{n=0}^{N}(p_{n+1};q_{n+1}|e^{-\epsilon C - i\epsilon h}\,|p_n;q_n)\, \prod_{n=1}^{N} dp_n\, dq_n\ .
\end{aligned}$$

As usual, for small ϵ, we shall approximate the exponent

$$[e^{-\epsilon C - i\epsilon h}] \simeq e^{-\epsilon C}\, e^{-i\epsilon h}\ ,$$

so that [when $(N+1)\epsilon = T$]

$$e^{-iTC - iTh} = \lim_{N\to\infty} [e^{-\epsilon C}\, e^{-i\epsilon h}]^{(N+1)} ,$$

which is valid, just like the case for the usual Trotter product formula (see Sec. 6.1.7), when $\mathfrak{D}(C) \cap \mathfrak{D}(h)$ is a core for the operator $C + ih$; recall the action of an operator on vectors in a core suffices to define the operator uniquely. In [DK85] this approximation was shown to hold for all expressions $h(p,q)$ that are bounded below (or above) and polynomially bounded above (or below). Accepting this limitation, we are led to the expression [with $(p'';q'') = (p_{N+1};q_{N+1})$ and $(p';q') = (p_0;q_0)$]

$$\begin{aligned}
&(p'';q''|\,[e^{-TC - iTh}]\,|p';q') \\
&\quad = \lim_{N\to\infty} \int\cdot\int \textstyle\prod_{n=0}^{N} (p_{n+1};q_{n+1}|\, e^{-\epsilon C} e^{-i\epsilon h}\,|p_n;q_n) \prod_{n=1}^{N} dp_n\, dq_n \\
&\quad = \lim_{N\to\infty} \int\cdot\int \textstyle\prod_{n=0}^{N} (p_{n+1};q_{n+1}|\, e^{-\epsilon C}\,|p_n;q_n)\, e^{-i\epsilon h(p_n,q_n)} \\
&\qquad\qquad \times \textstyle\prod_{n=1}^{N} dp_n\, dq_n \\
&\quad = \lim_{N\to\infty} \int\cdot\int \textstyle\prod_{n=0}^{N} \{e^{\frac{1}{2}\epsilon\nu}\, e^{i\frac{1}{2}(p_{n+1}+p_n)(q_{n+1}-q_n)} \\
&\qquad\qquad \times (2\pi\nu\epsilon)^{-1}\, e^{-i\epsilon h(p_n,q_n)} \\
&\qquad\qquad \times e^{-(1/2\nu\epsilon)[(p_{n+1}-p_n)^2 + (q_{n+1}-q_n)^2]}\} \textstyle\prod_{n=1}^{N} dp_n\, dq_n \\
&\quad = e^{\frac{1}{2}\nu T} \int e^{i\int [p\,dq - h(p,q)\,dt]}\, d\mu_W(p,q) ,
\end{aligned}$$

with the stochastic integral $\int p\,dq$ understood as a Stratonovich integral (see Sec. 4.2.7).

To complete the story, we recall a theorem of Chernoff [Che68] which asserts that

$$\lim_{\nu\to\infty} e^{-\frac{1}{2}\nu T(A-1) - iTh} = \mathbb{P}_0\, e^{-iT(\mathbb{P}_0\, h\, \mathbb{P}_0)}\, \mathbb{P}_0 ,$$

leading to

$$\begin{aligned}
&\lim_{\nu\to\infty} (p'';q''|\, e^{-\frac{1}{2}\nu T(A-1) - iTh}\,|p';q') \\
&\qquad = (p'',q''|\, \mathbb{P}_0\, e^{-iT(\mathbb{P}_0\, h\, \mathbb{P}_0)}\, \mathbb{P}_0\,|p',;q') \\
&\qquad = (1/2\pi)\langle p'',q''|\, e^{-iT\mathcal{H}}\,|p',q'\rangle ,
\end{aligned}$$

where

$$\mathcal{H} \equiv \mathbb{P}_0\, h\, \mathbb{P}_0 .$$

It then follows that

$$\langle p'', q''|\mathcal{H}|p', q'\rangle = 2\pi \int (p''; q''|\mathbb{P}_0|p; q)\, h(p, q)\, (p; q|\mathbb{P}_0|p'; q')\, dp dq \ ,$$

or stated otherwise, that

$$\langle p'', q''|\mathcal{H}|p', q'\rangle = \int \langle p'', q''|p, q\rangle\, h(p, q)\, \langle p, q|p'q'\rangle\, dp dq/2\pi \ .$$

Finally, as announced, this relation implies that

$$\mathcal{H}(P, Q) = \int h(p, q)\, |p, q\rangle\langle p, q|\, d\mu(p, q) \ .$$

As examples of Hamiltonian operators that are covered by this construction we mention arbitrary, Hermitian, lower bounded polynomials such as

$$\mathcal{H} = \Sigma_{m,n=0}^{M,N}\, c_{m,n}\, [P^m Q^n + Q^n P^m] \ ,$$

for real coefficients $\{c_{m,n}\}$ (such that $\mathcal{H} \geq -C\mathbb{1}$), and finite values for M and N.

This concludes our discussion of continuous time, Wiener measure regularization of phase space path integrals for single particle canonical systems.

8.1.3 Multivariable Wiener measure regularization of path integrals

As an alternative to the definition of multivariable path integrals by a temporal lattice regularization followed by a continuum limit after the integrals have been performed, we revisit the Wiener measure regularization scheme discussed above for a single variable path integral. In a formal sense, one wants to take the expression

$$\mathcal{M} \int e^{(i/\hbar)\int [p\,\cdot\,\dot{q} - h(p,q)]\, dt}\ \mathcal{D}p\, \mathcal{D}q \ ,$$

interpreted to apply for L degrees of freedom, and insert the formal Wiener measure regularization factors to yield

$$K(p'', q'', T; p', q', 0)$$
$$= \lim_{\nu\to\infty} \mathcal{M} \int e^{(i/\hbar)\int [p\,\cdot\,\dot{q} - h(p,q)]\, dt}\ e^{-(1/2\nu)\int [\dot{p}^2 + \dot{q}^2]\, dt}\ \mathcal{D}p\, \mathcal{D}q \ ,$$

appropriate again to L degrees of freedom. Greater clarification is offered if we note that the preceding equation may be written precisely as

$$K(p'', q'', T; p', q', 0)$$
$$= \lim_{\nu\to\infty} (2\pi\hbar)^L\, e^{L\nu T/2\hbar} \int e^{(i/\hbar)\int [p\,\cdot\, dq - h(p,q)\, dt]}\ d\mu_W^\nu(p, q) \ ,$$

where $\mu_W^\nu(p,q)$ is a Wiener measure on a $2L$-dimensional Euclidean (flat) space expressed in Cartesian coordinates and pinned so that $p(T), q(T) = p'', q''$ and $p(0), q(0) = p', q'$. Additionally, the integral $\int p \cdot dq$—where the " $\cdot$ " here denotes a scalar product of two vectors (and not the Itô product)—is defined as a lattice limit given by

$$\int p \cdot dq \equiv \lim_{N\to\infty} \sum_{n=0}^{N} \tfrac{1}{2}\,(p_{n+1} + p_n) \cdot (q_{n+1} - q_n)\ ,$$

which is the Stratonovich interpretation of this stochastic integral. Just as was the case for a single degree of freedom, this procedure automatically leads to a canonical coherent state representation for the propagator in terms of states $|p,q\rangle$ for L degrees of freedom where the fiducial vector satisfies $(Q^l + i\,P_l)\,|0\rangle = 0$ for each degree of freedom, $1 \le l \le L$. Finally, the quantum Hamiltonian associated with this construction is given by

$$\mathcal{H} = \int h(p,q)\,|p,q\rangle\langle p,q|\,\textstyle\prod_{l=1}^{L} dp_l\,dq^l/2\pi\ .$$

8.2 Continuous-Time Regularization of Spin Variable Path Integrals

In Sec. 7.4.2 we introduced the coherent state phase space path integral for spin degrees of freedom. We now wish to demonstrate that this form of path integral can also be regularized by a Wiener measure much as was the case for the canonical degrees of freedom. To start that procedure, it is pedagogically useful to learn how spin coherent states give rise to a metric on the spin variable version of phase space.

The spin coherent states, introduced in Sec. 7.4.1, were defined (for $\hbar = 1$) as

$$|\theta,\phi\rangle \equiv e^{-i\,\phi\,S_3}\,e^{-i\,\theta\,S_2}\,|s;s\rangle\ ,$$

where θ, ϕ may be interpreted as standard angular variables on the surface of a sphere. Following the prescription by which the set of canonical coherent states gave rise to a metric on phase space, we introduce the quantity

$$d\sigma(\theta,\phi)^2 \equiv 2[\|\,d|\theta,\phi\rangle\,\|^2 - |\langle\theta,\phi|\,d|\theta,\phi\rangle|^2]\ ,$$

which leads directly to the expression

$$d\sigma(\theta,\phi)^2 = s[\,d\theta^2 + \sin(\theta)^2\,d\phi^2]\ ,$$

which we recognize is exactly the metric for the surface of a sphere of radius $r = \sqrt{s}$. In other words, the geometry associated with the spin coherent states is the natural one associated with the rotation group, namely, the surface of a sphere, a surface of constant positive curvature! In different coordinates for

the coherent states, the form of the metric would change, but it would still refer to a surface of constant positive curvature, and more particularly, the surface of a sphere of the same radius.

Consequently, when it comes to taking the formal phase space path integral for spin variables given by

$$\mathcal{M} \int e^{i\int [s\,\cos(\theta)\,\dot{\phi} - h(\theta,\phi)]\,dt} \prod_t \sin(\theta)\,d\theta\,d\phi\ ,$$

and introducing a Wiener measure regularization, it is natural that we propose

$$\begin{aligned} K(\theta'',\phi'',T;\theta',\phi',0) \\ = \lim_{\nu\to\infty} \mathcal{M} \int e^{i\int [s\,\cos(\theta)\,\dot{\phi} - h(\theta,\phi)]\,dt}\ e^{-(s/2\nu)\int [\dot{\theta}^2 + \sin(\theta)^2\,\dot{\phi}^2]\,dt} \\ \times \prod_t \sin(\theta)\,d\theta\,d\phi\ , \end{aligned}$$

an expression that introduces Brownian motion paths on the surface of a sphere, namely, paths that are continuous but nowhere differentiable, just as those on flat space. This fact implies that the former expression is still formal, but using the fact that there is a bona fide Wiener measure on the surface of a sphere with diffusion constant ν, permits us to give a precise formulation of the proposed Wiener measure regularization as

$$\begin{aligned} K(\theta'',\phi'',T;\theta',\phi',0) \\ = \lim_{\nu\to\infty} \frac{4\pi}{(2s+1)}\, e^{\nu T/2} \int e^{i\int [s\,\cos(\theta)\,d\phi - h(\theta,\phi)\,dt]}\ d\mu_s^{\nu}(\theta,\phi)\ . \end{aligned}$$

In this expression, the integral

$$\int \cos(\theta)\,d\phi = \lim_{\epsilon\to 0} \tfrac{1}{2}\,[\cos(\theta_{l+1}) + \cos(\theta_l)]\,(\phi_{l+1} - \phi_l)$$

is interpreted as a Stratonovich stochastic integral. The normalization for the Wiener measure is sufficiently complicated that we send the reader to the literature [DK85].

Just as in the canonical case, it is noteworthy that this very form of Wiener measure regularization *automatically* leads to a spin coherent state representation,

$$K(\theta'',\phi'',T;\theta',\phi',0) \equiv \langle\theta'',\phi''|\,e^{-i\mathcal{H}T}\,|\theta',\phi'\rangle\ ,$$

where

$$|\theta,\phi\rangle \equiv e^{-i\phi S_3}\,e^{-i\theta S_2}\,|s;s\rangle\ ,$$

defined just as before, and

$$\mathcal{H}(\mathbf{S}) \equiv \frac{(2s+1)}{4\pi} \int h(\theta,\phi)\,|\theta,\phi\rangle\langle\theta,\phi|\,\sin(\theta)\,d\theta\,d\phi\ .$$

In turn, this rigorous construction involving continuous paths for all finite ν values works for all Hermitian Hamiltonians for spin s since, on this $(2s+1)$-dimensional Hilbert space, all Hermitian operators are automatically self-adjoint.

Although we have carried out the entire discussion for the spin phase space path integral in terms of spherical angles there is some value in also presenting the main ingredients in another coordinate system. In particular, let us introduce

$$p \equiv \sqrt{s}\cos(\theta)\ , \qquad\qquad q \equiv \sqrt{s}\,\phi\ ;$$

we observe that the range of p is such that $-\sqrt{s} \leq p \leq \sqrt{s}$ and for ϕ, recognizing that it is a periodic angle variable, we choose the range for q so that $-\sqrt{s}\,\pi < q \leq \sqrt{s}\,\pi$. In terms of these variables, the basic measure $s\sin(\theta)\,d\theta\,d\phi$ becomes $dp\,dq$. In the new variables, the metric becomes

$$s[d\theta^2 + \sin(\theta)^2\,d\phi^2] = [1-(p^2/s)]^{-1}\,dp^2 + [1-(p^2/s)]\,dq^2\ ,$$

which leads to the formal, Wiener measure regularized, phase space path integral given by

$$\mathcal{M}\int e^{i\int[p\dot{q}-h(p,q)]\,dt}\,e^{-(1/2\nu)\int\{[1-(p^2/s)]^{-1}\dot{p}^2+[1-(p^2/s)]\dot{q}^2\}\,dt} \times \mathcal{D}p\mathcal{D}q\ ,$$

where, once again, the range of the integration variables is $p \in [-\sqrt{s}, \sqrt{s}]$ and $q \in (-\sqrt{s}\,\pi, \sqrt{s}\,\pi]$. Properly interpreted, this expression has exactly the same status as the previous form expressed in terms of θ, ϕ.

Moreover, the formulation in terms of p and q makes evident quite another feature of the spin coherent state phase space path integral, namely, that *in the limit $s \to \infty$, the spin coherent state path integral passes smoothly to the canonical coherent state phase space path integral!* Indeed, if we let the Hamiltonian symbol vanish altogether, then the result of such a path integral is nothing but the coherent state overlap, and we observe that

$$\lim_{s\to\infty}[\cos(\tfrac{1}{2}\theta'')\cos(\tfrac{1}{2}\theta')\,e^{i\frac{1}{2}(\phi''-\phi')} + \sin(\tfrac{1}{2}\theta'')\sin(\tfrac{1}{2}\theta')\,e^{-i\frac{1}{2}(\phi''-\phi')}]^{2s}$$
$$= \exp\{i\tfrac{1}{2}(p''+p')(q''-q') - \tfrac{1}{4}[(p''-p')^2+(q''-q')^2]\}\ ,$$

as required, which is a good exercise for the reader to verify.

8.3 Continuous-Time Regularization of Affine Variable Path Integrals

In this section we are going to extend the Wiener measure regularization scheme to affine coherent state path integrals. However, it is first necessary to

determine what metric is given to phase space by the affine coherent states. Recall that the affine coherent states are based on the two generators of the affine Lie algebra, namely, D and $Q > 0$, which satisfy (with $\hbar = 1$) the relation $[Q, D] = iQ$. The coherent states are defined by

$$|p, q\rangle \equiv e^{ipQ}\, e^{-i \ln(q)\, D}\, |\eta\rangle \ ,$$

where $-\infty < p < \infty$, $0 < q < \infty$, and $|\eta\rangle$ is a normalized fiducial vector that satisfies

$$[Q - 1 + i\beta^{-1} D]\, |\eta\rangle = 0 \ ,$$

where $\beta > \frac{1}{2}$. As we have already seen in Sec. 7.5, the one form induced by such states is $i\langle p, q|\, d|p, q\rangle = -q\, dp$, and it is an easy exercise to show that the metric on phase space induced by the affine coherent states is given by

$$\begin{aligned} d\sigma(p, q)^2 &\equiv 2[\, \| \, d|p, q\rangle \, \|^2 - |\langle p, q|\, d|p, q\rangle|^2] \\ &= \beta^{-1} q^2\, dp^2 + \beta q^{-2}\, dq^2 \ , \end{aligned}$$

which is a small modification of the standard Poincaré form for the metric of a Lobachevsky plane, i.e., a two-dimensional space of constant negative curvature.

Incorporating a change in the time derivative term, we start with the formal phase space path integral

$$\mathcal{M} \int e^{i \int [-q\dot{p} - h(p,q)]\, dt}\, \mathcal{D}p\, \mathcal{D}q \ ,$$

which looks much like the starting point for the formal canonical phase space path integral, save for the fact that we now insist that $q > 0$ at all times. To handle such a restriction, we insert a Wiener measure regularization which is well adapted to the $q > 0$ requirement by having complete geodesics, and which takes the form

$$\begin{aligned} K(p'', q'', T; p', q', 0) = \lim_{\nu \to \infty} \mathcal{M} \int \\ \times e^{i \int [-q\dot{p} - h(p,q)]\, dt}\, e^{-(1/2\nu) \int [\beta^{-1} q^2 \dot{p}^2 + \beta q^{-2} \dot{q}^2]\, dt}\, \mathcal{D}p\, \mathcal{D}q \ . \end{aligned}$$

In turn, this integral can rigorously be described in terms of a Wiener measure for Brownian motion paths on a surface of constant negative curvature, and thus leads to a precise formulation given by

$$K(p'', q'', T; p', q', 0) = \lim_{\nu \to \infty}\, e^{\nu T/2} \int e^{i \int [-q\, dp - h(p,q)\, dt]}\, d\mu_A^\nu(p, q) \ ,$$

where μ_A^ν is a pinned measure on paths such that $(p'', q'') = (p(T), q(T))$, and $(p', q') = (p(0), q(0))$, the integral of which is rather complicated and again we send the reader to the literature [DKP].

In the present case, the very act of inserting the (constant negative curvature) Wiener measure regularizing factor also dictates: (i) that the propagator is automatically given in an affine coherent state representation; (ii) that the affine coherent states themselves are (up to unitary equivalence) given by

$$|p,q\rangle = e^{ip\,Q}\, e^{-i\ln(q)D}\,|\eta\rangle\ ;$$

(iii) that D and $Q > 0$ satisfy $[Q,D] = iQ$; and (iv) that $|\eta\rangle$ corresponds to the fiducial vector chosen earlier. Finally, we observe that the connection between the symbol $h(p,q)$ and the Hamiltonian operator $\mathcal{H}$ is automatically given by

$$\mathcal{H}(D,Q) = \int h(p,q)\,|p,q\rangle\langle p,q|\,d\mu_A(p,q)\ .$$

In brief, everything that is needed to fully characterize the quantum theory in the case of affine variables is a direct consequence of the chosen form of the Wiener measure regularization!

8.4 Quantization as Geometry

The attentive reader will have observed that in the preceding three sections of this chapter we have started with a formal phase space path integral and by introducing three fundamentally different forms of Wiener measure regularization we have been led to rigorous path integral constructions for three fundamentally different kinematical sets of quantum variables. Specifically, regularization involving: (i) Brownian motion paths on a *flat two-dimensional space* leads to a quantum theory in which the basic variables are the *Heisenberg variables* P and Q for which $[Q,P] = i\hbar\mathbb{1}$; (ii) Brownian motion paths on a sphere (of an appropriate radius) corresponding to a *constant positive curvature two-dimensional space* leads to a quantum theory in which the basic variables are the *spin variables* S_1, S_2, and S_3 for which $[S_j,S_k] = i\hbar\epsilon_{jkl}S_l$; and (iii) Brownian motion paths on a Lobachevsky plane, a *constant negative curvature two-dimensional space* leads to a quantum theory in which the basic variables are the *affine variables* D and $Q > 0$ for which $[Q,D] = i\hbar Q$. In brief, *the geometry of the space that carries the Wiener measure regularization has explicitly determined the choice of the kinematical variables of quantization!* Not only does the introduction of a *mathematically necessary* form of regularization render the mathematical formulation precise, but the very same regularization factor—depending on the geometry of the associated phase space—has uniquely determined the choice, among several inequivalent choices, of the *basic kinematical variables* relevant to the quantization itself!

This delightful result elevates the very process of quantization to the level of pure geometry! A discussion of how this picture can be presented in a truly

geometric form without coordinates is given in [Kla88]. An elegant coordinate-free discussion, strongly featuring the probabilistic aspects of the problem, has been given by Bodmann [Bod03], which also contains many related references.

****

The favorable and geometric outcome discussed above immediately raises the question about what might be the quantum consequences of yet another choice for the geometry of the Wiener measure regularization. It is of course perfectly acceptable mathematically to consider other geometries to support the Brownian motion paths, and one example could be the surface of a non-spherical ellipse, or a crumply, corrugated sheet, etc. In general, if one takes a two-dimensional surface of reduced symmetry—unlike our three favored surfaces of constant curvature—and sets the Hamiltonian symbol to zero, the resultant kernel in the limit of diverging diffusion constant is a reproducing kernel for a *one-dimensional Hilbert space!* Mathematically, that may be interesting, but it is highly unlikely to be interesting from a physical point of view. Further investigation [Mar92, AKL] shows that there is a way around the outcome of a one-dimensional Hilbert space, but there is a price to be paid. In each of the three cases of constant curvature, there was a prefactor of the form (with $\hbar = 1$)

$$e^{\nu T[\frac{1}{2}]} ,$$

which, in the physical analog of the quantum Hall effect, arises from the zero-point energy in the Landau Hamiltonian. When the geometry is no longer of the constant curvature variety, there indeed is a zero-point energy to be subtracted, *but the value of that zero-point energy depends on the location in phase space.* In other words, the zero-point energy becomes *position dependent*, and as a consequence the corresponding factor, crudely speaking, becomes

$$e^{\nu \int [\frac{1}{2}](p(t),q(t))\,dt} ,$$

and therefore *this factor must stand inside the path integral* rather than outside. This change restores the reproducing kernel Hilbert space to a proper dimension, but the price to be paid is that the heuristic weighting of each of the phase space paths is no longer uniform, and this property conflicts with the traditional and physically appealing view of path integrals as having uniform weighting for all paths.

For several degrees of freedom, appropriate to higher-dimensional phase space manifolds, the favored geometries need only possess a constant scalar curvature in order for them to have a uniform weighting of each phase space path in the Wiener measure regularized phase space path integral [WK02].

Exercises

8-1 Given the expression

$$\begin{aligned}\langle x'|U[p',q']|x\rangle &= \langle x'|e^{-iq'P}\,e^{ip'Q}|x\rangle \\ &= \langle x'-q'|x\rangle\, e^{ip'x} \\ &= \delta(x'-q'-x)\,e^{ip'x}\ ,\end{aligned}$$

show that

$$\mathrm{Tr}(U[p,q]^\dagger U[p',q']) = 2\pi\,\delta(p-p')\,\delta(q-q')\ .$$

8-2 From the definition of the spin coherent states (with $\hbar = 1$),

$$|\theta,\phi\rangle = e^{-i\phi S_3}\,e^{-i\theta S_2}\,|s;s\rangle\ ,$$

where $S_3\,|s;s\rangle = s\,|s;s\rangle$, show that

$$d\sigma(\theta,\phi)^2 \equiv 2[\,\|\,d|\theta,\phi\rangle\,\|^2 - |\langle\theta,\phi|\,d|\theta,\phi\rangle|^2] = s[\,d\theta^2 + \sin(\theta)^2\,d\phi^2]\ .$$

8-3 For $p = \sqrt{s}\,\cos(\theta)$ and $q = \sqrt{s}\,\phi$, with $0 \le \theta \le \pi$ and $-\pi < \phi \le \pi$, show (for $\hbar = 1$) that

$$\begin{aligned}&\lim_{s\to\infty}[\cos(\tfrac{1}{2}\theta'')\cos(\tfrac{1}{2}\theta')\,e^{i\frac{1}{2}(\phi''-\phi')} + \sin(\tfrac{1}{2}\theta'')\sin(\tfrac{1}{2}\theta')\,e^{-i\frac{1}{2}(\phi''-\phi')}]^{2s} \\ &\qquad = \exp\{\,i\tfrac{1}{2}(p''+p')(q''-q') - \tfrac{1}{4}[(p''-p')^2 + (q''-q')^2]\,\}\ .\end{aligned}$$

8-4 In terms of the affine coherent states $|p,q\rangle = e^{ipQ}\,e^{-i\ln(q)\,D}\,|\eta\rangle$ (with $\hbar = 1$) show that

$$\begin{aligned}d\sigma(p,q)^2 &\equiv 2[\,\|\,d|p,q\rangle\,\|^2 - |\langle p,q|\,d|p,q\rangle|^2] \\ &= \beta^{-1}q^2\,dp^2 + \beta q^{-2}\,dq^2\end{aligned}$$

based on the fact that

$$[Q - 1 + i\beta^{-1}D]\,|\eta\rangle = 0\ .$$

8-5 Repeat Exercises **8-2** and **8-4** where the fiducial vectors in each case are now allowed to be general vectors $|\eta\rangle$. Determine the corresponding metrics $d\sigma^2$ for each case. Determine the curvature of the associated phase spaces in each case.

9

Classical and Quantum Constraints

9.1 Classical Systems with Constraints

Many systems involve constraints that generally lead to dynamical motion that is constrained in one way or another. For example, three-dimensional dynamical motion that is constrained to lie on a fixed two-dimensional surface, e.g., motion effectively lying on the surface of the Earth, would constitute constrained motion. Oftentimes, constraints are accompanied by superfluous degrees of freedom, variables whose values are not determined by the equations of motion but whose values must be chosen to fully determine the temporal evolution of the remaining variables. Such variables whose values are left undetermined by the equations of motion are called "gauge" variables; and the somewhat arbitrary choice of such variables to fix the behavior of all the variables is called "choosing a gauge."

A few, very simple but nontrivial examples can help illuminate the concepts described above, as well as illustrate their formulation. As the first example, suppose we choose the classical action

$$I = \int_0^T [\, p\dot{q} - \tfrac{1}{2}p^2 \,]\, dt \ ,$$

subject to the constraint that $q = 0$. Varying the action leads to

$$\delta I = \int_0^T [-\dot{p}]\, \delta q\, dt + \int_0^T [\dot{q} - p]\, \delta p\, dt = 0 \ ,$$

assuming the end points are held fixed as needed. One is tempted to conclude that

$$\dot{q} - p = 0 \ , \qquad \dot{p} = 0 \ ;$$

however, the derivation of the last equation of motion violates our requirement that $q = 0$ and thus q *cannot* vary. If q cannot vary, the action collapses to

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9_9,

$$I = -\tfrac{1}{2}\int_0^T p^2\,dt\ ,$$

with variation

$$\delta I = -\int_0^T p\,\delta p\,dt = 0\ .$$

This requires that $p = 0$, and along with the constraint leads to

$$p(t) = 0\ , \qquad q(t) = 0$$

as the solution.

A somewhat more systematic way to deal with constraints is to use Lagrange multipliers where necessary. In particular, for the case at hand, we augment our action to read

$$I = \int [p\dot{q} - \tfrac{1}{2}p^2 - \lambda q]\,dt\ ,$$

which includes an auxiliary variable $\lambda = \lambda(t)$, called the *Lagrange multiplier* variable, which multiplies the constraint function $\phi = \phi(p, q)$, which in our case is just $\phi = q$. Now we vary p, q, and λ freely to deduce that

$$\delta I = \int [\dot{q} - p]\delta p\,dt - \int [\dot{p} + \lambda]\delta q\,dt - \int q\,\delta\lambda\,dt = 0\ ,$$

again with suitable end points held fixed. Thus our equations of motion become

$$\dot{q} = p\ , \qquad \dot{p} = -\lambda\ , \qquad q = 0\ .$$

These equations have as their solution

$$q = 0\ , \qquad 0 = \dot{q} = p\ , \qquad 0 = \dot{p} = -\lambda\ .$$

In this case $q = p = \lambda = 0$, and thus the equations of motion have determined the value of the Lagrange multiplier. We will revisit this example later.

As a second example, we choose the same original action but now say that $p = 1$ (instead of $q = 0$). The relevant classical action is now

$$I = \int_0^T [p\dot{q} - \tfrac{1}{2}p^2]\,dt\ ,$$

subject to the constraint that $p = 1$. Imposing the constraint leads to the action

$$I = \int_0^T [\dot{q} - \tfrac{1}{2}]\,dt\ ,$$

which has the variation

$$\delta I = \delta q\big|_0^T = 0 \ ,$$

from which no equation of motion results. The solution is thus given by

$$p(t) = 1 \ , \qquad q(t) = \text{arbitrary save for the boundary conditions} \ .$$

With the Lagrange multiplier, the relevant classical action is

$$I = \int [p\dot{q} - \tfrac{1}{2}p^2 - \lambda(p-1)]\, dt \ .$$

The equations of motion that arise for this example are

$$\dot{q} - p - \lambda = 0 \ , \qquad \dot{p} = 0 \ , \qquad p = 1 \ .$$

The solution to these equations is

$$p(t) = 1 \ , \qquad q(t) = t + \int_0^t \lambda(s)\, ds + q(0) \ , \qquad \lambda(t) = \text{arbitrary} \ ;$$

in short, p is fixed while q can be anything that agrees with the boundary data. To derive these equations of motion, we have assumed variations that left $q(0)$ and $q(T)$ fixed. But that leaves much room for ambiguity. In particular, $\lambda(t)$ can be chosen so that $q(t) = q(0)(1 - t/T) + q(T)(t/T)$, or $q(t) = q(0)(1 - t/T)^3 + q(T)(t/T)^4 + A(t)\sin(t)\sin(t-T)$, etc. This arbitrariness is a prime example of an unphysical degree of freedom, i.e., a gauge degree of freedom. To determine any answer (as in our two examples) some choice for $\lambda(t)$—some choice of "gauge"—must be made. The fact that the answer depends on λ, and thus can be changed, renders such a variable a gauge variable, i.e., as being unphysical.

9.1.1 General classical construction

With the few examples offered above as appetizers, we now proceed to the full story regarding classical systems with constraints. Before the constraints are imposed the full phase space is called $\mathcal{M}$; after the constraints are enforced the reduced phase space is called the constraint hypersurface $\mathcal{C}$. For a system with L canonical degrees of freedom, $\{p_l, q^l\}_{l=1}^L$, A constraints, $\{\phi_\alpha(p,q)\}_{\alpha=1}^A$, and a basic Hamiltonian, $H(p,q)$, we adopt the classical action

$$I = \int [p_l\, \dot{q}^l - H(p,q) - \lambda^\alpha \phi_\alpha(p,q)]\, dt \ .$$

We assume that the set of constraints is "complete" in a manner that will be clarified below. Assuming the proper variables are held fixed at the initial and final times leads to the equations of motion

$$\begin{aligned} \dot{q}^l &= \partial H/\partial p_l + \lambda^\alpha \partial \phi_\alpha/\partial p_l \ , \\ \dot{p}_l &= -\partial H/\partial q^l - \lambda^\alpha \partial \phi_\alpha/\partial q^l \ , \\ \phi_a(p,q) &= 0 \ . \end{aligned}$$

In terms of Poisson brackets, these equations read

$$\begin{aligned} \dot{q}^l &= \{q^l, H\} + \lambda^\alpha \ \{q^l, \phi_\alpha\} \ , \\ \dot{p}_l &= \{p_l, H\} + \lambda^\alpha \{p_l, \phi_\alpha\} \ , \\ \phi_a(p,q) &= 0 \ , \end{aligned}$$

where as usual

$$\{A, B\} \equiv \frac{\partial A}{\partial q^l}\frac{\partial B}{\partial p_l} - \frac{\partial A}{\partial p_l}\frac{\partial B}{\partial q^l}$$

with summation over the repeated index l understood. Using the equations of motion above, we can find the time derivative of any phase space function $A(p,q)$ as

$$\dot{A}(p,q) = \{A(p,q), H(p,q)\} + \lambda^\alpha \{A(p,q), \phi_\alpha(p,q)\} \equiv \{A(p,q), H_T(p,q)\} \ ,$$

based on the so-called total Hamiltonian

$$H_T(p,q) \equiv H(p,q) + \lambda^\alpha \phi_\alpha(p,q) \ .$$

The constraint functions vanish and must do so for all time. Thus

$$\dot{\phi}_\alpha(p,q) = \{\phi_\alpha(p,q), H(p,q)\} + \lambda^\beta \{\phi_\alpha(p,q), \phi_\beta(p,q)\} = 0 \ .$$

This equation is of fundamental importance, and it can be satisfied in two significantly different fashions.

First, we assume that the Poisson bracket between any two constraints vanishes on the constraint hypersurface $\mathcal{C}$. As a consequence, such a Poisson bracket may be written in the form

$$\{\phi_\alpha(p,q), \phi_\beta(p,q)\} = C_{\alpha\beta}{}^\gamma(p,q)\, \phi_\gamma(p,q) \ ,$$

which vanishes on $\mathcal{C}$ because the constraints do. The coefficients $C_{\alpha\beta}{}^\gamma(p,q)$—which are forbidden from being such as to cancel the zero of the constraint on $\mathcal{C}$—have the superficial appearance of structure constants for a Lie algebra (see Sec. 5.3.1). If the coefficients $C_{\alpha\beta}{}^\gamma$ are *constants*, i.e., independent of p and q, then they are called *structure constants*; if those coefficients are *not* constants, then they are called *structure functions*. If the second bracket $\{\phi_\alpha, \phi_\beta\}$ vanishes by itself, then it must follow that the first bracket also vanishes by itself, or in other words,

$$\{\phi_a(p,q), H(p,q)\} = h_\alpha{}^\gamma(p,q)\phi_\gamma(p,q) \ ,$$

which is a form that vanishes on $\mathcal{C}$. [**Remark:** If the first bracket does not vanish on $\mathcal{C}$ when the second one does, then we have an additional constraint, implying that the original set of constraints was not complete; by assuming the original set is complete, we tacitly assume that the first bracket vanishes on $\mathcal{C}$ whenever the second bracket does.] Constraints of the kind discussed in this paragraph are called *first-class constraints*; indeed, they are called *closed* first-class constraints if one deals with structure constants, and they are called *open* first-class constraints if one deals with structure functions. For first-class constraints, the Lagrange multiplier functions λ_α are *not* determined by the equations of motion; they remain freely specifiable and thereby lead to gauge-dependent quantities which are unphysical variables. Stated otherwise, once a first-class constraint system is put onto the constraint hypersurface, the dynamics are such that it remains on that surface whatever choice is made for the Lagrange multipliers. At the beginning of the dynamical evolution, the relation among the variables that restricts the system to the constraint hypersurface is known as the *initial value equation*.

Second, we assume that neither of the two brackets that arise in the equation $\dot{\phi}_\alpha(p,q) = 0$ vanishes separately. Rather, to take the extreme case, we assume that the elements

$$\Delta_{\alpha,\beta} \equiv \{\phi_\alpha(p,q), \phi_\beta(p,q)\}$$

form a matrix Δ for which $\det[\Delta] \neq 0$ on $\mathcal{C}$, and therefore Δ has an *inverse* matrix $\widetilde{\Delta}$ on $\mathcal{C}$ with matrix elements $\widetilde{\Delta}^{\alpha,\beta}$ defined so that

$$\widetilde{\Delta}^{\alpha,\beta}\,\Delta_{\beta,\gamma} = \delta^\alpha_\gamma \ .$$

In that case, to satisfy the requirement that $\dot{\phi}_\alpha(p,q) = 0$, it is necessary that

$$\lambda^\alpha \equiv -\widetilde{\Delta}^{\alpha,\beta}\,\{\phi_\beta(p,q), H(p,q)\} \ .$$

In this way the Lagrange multipliers are *fully determined* by the equations of motion—they provide necessary forces to ensure that the system stays on the constraint hypersurface once it has been put there. Those forces may be zero, but they are not arbitrary as they were for the case of first-class constraints. This new situation where the Lagrange multipliers are fixed occurs when the constraints are *second-class constraints*. This concludes the general discussion of the two principal ways in which the constraint equations can be maintained for all time.

Of course, there are intermediate situations when some of the constraints are first class and the others are second class. That situation arises when the matrix

$$\Delta \equiv \{\Delta_{\alpha,\beta}\}$$

is not of full rank; in that case only some of the Lagrange multipliers are determined while the rest of them remain free. It may be noted that the first

example of a constraint, $q = 0$ for the free particle, was in effect a second-class constraint since the Lagrange multiplier was fixed; the second example, $p = 1$ for the free particle, was a first-class constraint. Moreover, in light of our general discussion, the original treatment of the first example was in a sense incomplete. If we assert that our constraint hypersurface is $q = 0$ as originally assumed, then

$$\dot{q} = \{q, H\} + \lambda\{q, q\} = \{q, \tfrac{1}{2}p^2\} = p \ ,$$

which implies that $p = 0$ is actually a secondary constraint that should have been included along with $q = 0$. With that modification, the new form of the classical action is

$$I = \int [p\dot{q} - \tfrac{1}{2}p^2 - \lambda_1 q - \lambda_2 p]\, dt \ ,$$

including two Lagrange multipliers and both constraints. The equations of motion become

$$\dot{q} = p + \lambda_2 \ , \qquad \dot{p} = -\lambda_1 \ , \qquad q = 0 \ , \qquad p = 0 \ ,$$

and the solutions to these equations are

$$q(t) = 0 \ , \qquad p(t) = 0 \ , \qquad \lambda_1(t) = 0 \ , \qquad \lambda_2(t) = 0 \ .$$

The result of this analysis coincides with the former answer. However, this latter form of the problem respects the full set of constraints and stresses its second-class nature since the Poisson bracket among the constraints, specifically $\{q, p\} = 1$, does *not* vanish on the constraint hypersurface—in fact, it doesn't vanish anywhere.

9.1.2 Anomalous constraint situations

The general analysis given above covers many cases, but there are certain situations among them that could potentially cause difficulties when it comes to quantizing such systems.

Reducible constraints: The first of the potential difficulties is when the set of constraints involves *linear dependencies*, namely, when

$$c^\alpha \phi_\alpha(p, q) = 0$$

does *not* imply that the coefficients c^α all vanish. If that is the case, one says that the set of constraints is *reducible*. A simple example would be two identical constraints, $\phi_1(p, q) = \phi_2(p, q)$. This situation may seem unlikely, but linear dependence over subsets of phase space may possibly lead to trouble as well. For example, consider the two constraints

$$\phi_1 = q_1 \ , \qquad \phi_2 = q_1 + q_3 (p_2^2 + q_2^2 - c) \ .$$

These two constraints are linearly independent save on the subset where $p_2^2 + q_2^2 = c$. This region amounts to a set of measure zero in the classical phase space. However, the quantum operator $P_2^2 + Q_2^2 - c$ has a discrete spectrum, and therefore, for the right choice of c, this factor may vanish on a nonzero subspace of the Hilbert space, which would then lead to a reducible constraint situation in that nonzero subspace of the Hilbert space. We will discuss further examples of reducible constraints when we take up the issue of the quantization of constrained systems.

Irregular constraints. The second of the potential difficulties arises when the constraints vanish to higher order. Implicitly, we have focused on regular constraints for which although $\phi_\alpha(p,q) = 0$ for all α on the constraint hypersurface $\mathcal{C}$, we have assumed that $\partial\phi_\alpha(p,q)/\partial p_l \neq 0$ and $\partial\phi_\alpha(p,q)/\partial q^l \neq 0$ for all α and l, and throughout $\mathcal{C}$. For irregular constraints this condition is violated. To see what happens, let us consider the two examples of the early part of this chapter taken as irregular constraints. Thus we consider

$$I = \int [p\dot{q} - \tfrac{1}{2}p^2]\, dt \ ,$$

where now our constraint is $\phi_1 = q^3 = 0$ rather than $q = 0$. The equation of motion that follows is $\dot{q} = p$, which thus leads to the solutions $p(t) = 0$ and $q(t) = 0$, just as in the regular case. Now let us repeat the problem using the Lagrange multiplier and the irregular form of the constraint. Thus

$$I = \int [p\dot{q} - \tfrac{1}{2}p^2 - \lambda q^3]\, dt \ ,$$

which leads to the equations of motion

$$\dot{q} = p \ , \qquad \dot{p} = -3\lambda q^2 \ , \qquad q^3 = 0 \ .$$

In this case, the solution becomes

$$p(t) = 0 \ , \qquad q(t) = 0 \ , \qquad \lambda(t) = \text{arbitrary} \ ,$$

which effectively has reproduced the same solutions for the phase space variables p and q, but has given a very different answer for the Lagrange multiplier. Again, $p = 0$ should be interpreted as a secondary constraint, and so we are also led to consider the classical action

$$I = \int [p\dot{q} - \tfrac{1}{2}p^2 - \lambda_1 q^3 - \lambda_2 p]\, dt \ ,$$

which leads to the equations of motion given by

$$\dot{q} = p + \lambda_2 \ , \qquad \dot{p} = -3\lambda_1 q^2 \ , \qquad q^3 = 0 \ , \qquad p = 0 \ .$$

The solution to these equations becomes

$$p(t) = 0 \,, \qquad q(t) = 0 \,, \qquad \lambda_1(t) = \text{arbitrary} \,, \qquad \lambda_2(t) = 0 \,.$$

Again, the solutions for the phase space variables p and q are the same as before, but the Lagrange multipliers are not the same. On the face of it, the difference seems not to be so great; perhaps that difference can be bridged.

Consider the two formulations of the problem, namely,

$$I = \int [p\dot{q} - \tfrac{1}{2}p^2 - \lambda_1 q - \lambda_2 p]\, dt \,,$$
$$I = \int [p\dot{q} - \tfrac{1}{2}p^2 - \lambda_1' q^3 - \lambda_2 p]\, dt \,.$$

To pass from one form to another it seems only necessary to let $\lambda_1 = \lambda_1' q^2$, or, equivalently, $\lambda_1' = \lambda_1 q^{-2}$. But in view of the very constraint being considered, it is fair to ask if $\lambda_1' = \lambda_1 q^{-2}$ is a proper transformation, i.e., an allowed transformation. If this is considered to be an allowed transformation, why not allow the transformation $\lambda' = \lambda'' q^{-3}(q-1)$; this would then take us from

$$I = \int [p\dot{q} - \tfrac{1}{2}p^2 - \lambda' q^3]\, dt$$

to

$$I = \int [p\dot{q} - \tfrac{1}{2}p^2 - \lambda''(q-1)]\, dt \,,$$

which completely changes the intent of the original constraint. Thus, we are forced to conclude that a transformation of the Lagrange multiplier that changes the nature of the original constraint is *not* allowed.

Let us reexamine the second elementary example from the start of this chapter but now with an irregular constraint. In particular, we consider

$$I = \int [p\dot{q} - \tfrac{1}{2}p^2 - \lambda(p-1)^3]\, dt \,,$$

which is the previous example save for the fact that the constraint has been changed from $p - 1 = 0$ to $(p-1)^3 = 0$. In this case the equations of motion read

$$\dot{q} = p - 3\lambda(p-1)^2 \,, \qquad \dot{p} = 0 \,, \qquad (p-1)^3 = 0 \,.$$

The solutions to these equations are then

$$p(t) = 1 \,, \qquad q(t) = t + q(0) \,, \qquad \lambda(t) = \text{arbitrary} \,.$$

In this case the variable q is *not* gauge dependent and thus is an observable, which is unlike the case when the constraint was $p - 1 = 0$; additionally, the

solution requires that $q(T) = T + q(0)$ to satisfy the prescribed boundary conditions. Thus the "price" to be paid to have q raised to the level of an observable, is that only one boundary condition [say $q(0)$] can be freely specified. These aspects are part of the nature of the given irregular constraint. Of course, one can envisage a transformation that takes one from

$$I = \int [p\dot{q} - \tfrac{1}{2}p^2 - \lambda'(p-1)^3]\, dt$$

to

$$I = \int [p\dot{q} - \tfrac{1}{2}p^2 - \lambda(p-1)]\, dt\ ,$$

simply by setting $\lambda' = \lambda(p-1)^{-2}$. But in light of the previous discussion this transformation must be regarded as unallowed; otherwise we could change the constraint to anything else simply by setting $\lambda' = \lambda''\phi(p,q)/(p-1)^3$.

Before turning to a general analysis of classical constraints, we consider the example

$$I = \int [p\dot{q} - \lambda q(2-q)^3]\, dt\ ,$$

which leads to the equations of motion

$$\dot{q} = 0\ , \qquad \dot{p} = -\lambda(2-q)^3 + 3\lambda q(2-q)^2\ , \qquad q(2-q)^3 = 0\ .$$

The solution to this set of equations is

$$\begin{aligned} &q(t) = 0\ , \qquad p(t) = p(0) - 8\int_0^t \lambda(s)\, ds\ , \qquad \lambda(t) = \text{arbitrary}\ ; \\ &q(t) = 2\ , \qquad p(t) = p(0)\ , \qquad\qquad\qquad\quad \lambda(t) = \text{arbitrary}\ . \end{aligned}$$

Thus there are two possible acceptable solutions: For $q = 0$, it follows that p is a gauge variable; for $q = 2$, it follows that p is not a gauge variable.

With regard to a general discussion about constraints, in which the total Hamiltonian is taken as

$$H_T(p,q) = H(p,q) + \lambda^\alpha \phi_\alpha(p,q)\ ,$$

we are also forced to exclude transformations of the form

$$\lambda^\alpha = \lambda'^\alpha \psi_\alpha(p,q)/\phi_a(p,q) \qquad \text{(NO summation)}\ ,$$

which have the effect of exchanging one set of constraints $\{\phi_\alpha(p,q)\}$ for another set of constraints $\{\psi_\alpha(p,q)\}$; to permit such transformations would strip any particular set of constraints of its significance. We also interpret this limitation to exclude any and all transformations that would change the (ir)regular character of any constraint.

Of course, there are many transformations of Lagrange multipliers that are allowed, namely, those that do not change the basic character of the original set of constraints in any way. By such transformations we have in mind those of the form

$$\lambda^\alpha = \lambda'^\alpha W_\alpha(p,q) \qquad \text{(NO summation)}\ ,$$

where the smooth functions $W_\alpha(p,q)$ do *not* vanish or diverge on the constraint hypersurface $\mathcal{C}$ defined as the subset for which $\phi_\alpha(p,q) = 0$ for all α. Such transformations do not have the effect of changing the basic set of constraints as originally given, nor of changing the (ir)regular character of any of the constraints.

Although our examples of irregular constraints have mostly been very elementary, that need not always be the case. For example, our discussion also covers an example such as

$$I = \int \{p_1\dot{q}_1 + p_2\dot{q}_2 - \lambda[p_1^2(p_1-1)^4(p_1-3)^6 - q_2(q_2-1)^3(q_2-2)^4]\}\,dt$$

in which various regular and irregular constraint situations arise.

To briefly summarize our discussion of classical constraints, it is important to understand not only the basic irreducible and regular cases, but it is potentially important to be aware of reducible and/or irregular sets of constraints as well. These situations may well arise when one considers quantizing a system with constraints, and it will be important to work with a quantization procedure that can deal with all possible forms of constraint situations.

9.2 Quantum Theory of Constrained Systems

There are a number of different schemes used to quantize classical systems with constraints. It is not our goal to discuss all such schemes, but rather to focus on one scheme—the so-called *projection operator method*—that can be first analyzed from an abstract operator viewpoint and second cast into a path integral formulation, which is the principal subject of this monograph. Alternative quantization schemes may or may not agree with each other or with the scheme we shall emphasize. In addition, some of the alternative schemes lack the generality that the projection operator method enjoys.

As emphasized before, the canonical quantization of any system should be carried out in special coordinates, namely, Cartesian coordinates, and for that to be possible one must assume that the original phase space is Euclidean and thus admits global Cartesian coordinates. At the very least we must make that assumption about our system before any quantization or reduction is attempted. However, it stands to reason that if the original phase space $\mathcal{M}$ was Euclidean before any constraints have been imposed, there is a good chance that the constraint hypersurface $\mathcal{C}$—the subset of phase space on

which the constraints are satisfied—is no longer Euclidean and therefore cannot admit global Cartesian coordinates. This situation suggests, as Dirac has emphasized, that one should always *quantize first* and *reduce second*. Other schemes that *reduce first* and *quantize second* run the serious risk of not having a Euclidean reduced phase space so that quantization thereafter becomes highly ambiguous. Both schemes provide answers, but the answers may well differ from one another. Despite these facts, there are some workers who claim that quantization and reduction are commuting operations; if that is true it is because the systems under discussion have been purposely self-limited to achieve that outcome. Elsewhere we have offered a very simple example of a system that leads to different answers depending on whether one quantizes first or reduces first, demonstrating clearly that these two procedures generally do not commute [Kla01].

9.2.1 Dirac's procedure for quantization of systems with constraints

Let us start with a general classical system for L phase space degrees of freedom, $(p,q) = \{p_l, q^l\}_{l=1}^{L}$, a basic Hamiltonian $H(p,q)$, and A classical constraints, $\{\phi_\alpha(p,q)\}_{\alpha=1}^{A}$. As the classical action in this case we again choose

$$I = \int [p_l \dot{q}^l - H(p,q) - \lambda^\alpha \phi_\alpha(p,q)]\, dt \ ,$$

including A different Lagrange multipliers λ^α.

Dirac [Dir64] has proposed that the quantization of this system proceed by first promoting the Cartesian phase space variables to canonical operators, $p_l \to P_l$, $q^l \to Q^l$, subject to the canonical commutation relations $[Q^l, P_m] = i\hbar\delta^l_m$. These operators are realized in the original or kinematic Hilbert space $\mathfrak{H}$. Next the Hamiltonian $H(p,q)$ is promoted to a self-adjoint operator $\mathcal{H}(P,Q)$, and likewise for the several constraint functions, $\phi_\alpha(p,q) \to \Phi_\alpha(P,Q)$. All of these exist as self-adjoint operators on $\mathfrak{H}$. That completes the quantization part of the story; now we take up the reduction part.

Dirac proposes that we define vectors $|\psi\rangle_{phys}$ in the physical Hilbert space $\mathfrak{H}_{phys}$, $|\psi\rangle_{phys} \in \mathfrak{H}_{phys} \subset \mathfrak{H}$ by the condition that, for all α,

$$\Phi_\alpha(P,Q)\,|\psi\rangle_{phys} = 0 \ .$$

It is clear that if two such vectors satisfy this equation, so too does any linear combination of them. Thus they clearly form a linear subspace of the original Hilbert space. There are two potential complications with this prescription: (1) one or more of the constraint operators may have the spectral value zero in their continuous spectrum; thus, the "eigenvectors" above would have infinite norm and strictly speaking they would not be vectors in the Hilbert space at all; (2) consistency of the given condition requires that

$$[\Phi_\alpha(P,Q),\Phi_\beta(P,Q)]\,|\psi\rangle_{phys}=0$$

for all α,β pairs. This condition suggests we consider cases where the constraint operators Φ_α form the generators of a Lie algebra, in which case

$$[\Phi_\alpha(P,Q),\Phi_\beta(P,Q)]=iC_{\alpha\beta}{}^{\gamma}\,\Phi_\gamma(P,Q)\ ,$$

where the coefficients $C_{\alpha\beta}{}^{\gamma}$ are the *structure constants*. Thus for constraints that compose the generators of a compact Lie group, the conditions are well fulfilled since the compact group requirement ensures that the spectrum of the generators is purely discrete and includes zero. Systems for which the constraints are generators of a noncompact Lie group satisfy condition (2), but they may not have normalizable eigenvectors, thus requiring special techniques. Systems whose constraints are not the generators of a Lie algebra will typically fail to satisfy the Dirac conditions. Special exceptions can be found as follows. Suppose, instead of the constraints satisfying the properties of a Lie algebra, they instead satisfy the commutation relation

$$[\Phi_\alpha(P,Q),\Phi_\beta(P,Q)]=iC_{\alpha\beta}{}^{\gamma}(P,Q)\,\Phi_\gamma(P,Q)\ ,$$

involving this time *structure functions* as opposed to structure constants. Note further that the operators on the right side stand so that the constraint operator lies to the right of the structure functions; if this situation holds, then the constraints satisfy condition (2) of Dirac.

As examples of the last two topics, we first recall the three well-known rotation generators (summation understood)

$$J_j(P,Q)=\epsilon_{jkl}[Q_k\,P_l-Q_l\,P_k]\ ,$$

which fulfill the Lie algebra for SO(3),

$$[J_j(P,Q),J_k(P,Q)]=i\epsilon_{jkl}\,J_l(P,Q)\ ,$$

and thus fulfill condition (2). Second, consider the operator

$$R(P,Q)\equiv 1+(P_1^2+Q_1^2)+(P_2^2+Q_2^2)+(P_3^2+Q_3^2)\ .$$

Due to rotational invariance of R, it follows that

$$[R(P,Q),J_j(P,Q)]=0$$

for all j. Therefore, if we introduce

$$T_j(P,Q)\equiv\tfrac{1}{2}[R(P,Q)\,J_j(P,Q)+J_j(P,Q)\,R(P,Q)]=R(P,Q)\,J_j(P,Q)\ ,$$

it follows that

$$[T_j(P,Q),T_k(P,Q)]=i\,\epsilon_{jkl}\,R(P,Q)\,T_l(P,Q)\ ,$$

which still satisfies condition (2) even though the T_j do not generate a Lie algebra.

It is noteworthy that if we define

$$S(P,Q) \equiv 1 + (P_1^2 + Q_1^2) + \sqrt{2}(P_2^2 + Q_2^2) + \sqrt{3}(P_3^2 + Q_3^2) \,,$$

and then [in place of $T_j(P,Q)$] introduce

$$U_j(P,Q) \equiv \tfrac{1}{2}[S(P,Q)\,J_j(P,Q) + J_j(P,Q)\,S(P,Q)] \,,$$

it follows that the operators U_j will *not* satisfy condition (2) of Dirac because the rotational symmetry enjoyed by R is no longer respected by S. This latter example is one for which the procedure of Dirac does not work and alternative methods are called for. It is an example of a system with an *anomaly*, namely, one for which the open first-class set of classical constraints become partially second-class when quantized. Such systems, along with straightforward second-class constraints, are among those that cannot be treated by the Dirac procedure.

An example of a straightforward second-class example would be the classical constraints $\phi_1 = p$ and $\phi_2 = q$ for a single degree of freedom. As noted before, the Poisson bracket of the two constraints $\{p,q\} = -1$ fails to vanish on the constraint hypersurface $p = q = 0$, which is the hallmark of a second-class constraint. Classically, the nonvanishing of the constraint Poisson bracket is not serious. The real difficulty lies after quantization, which requires that

$$P\,|\psi\rangle_{phys} = 0 \,, \qquad Q\,|\psi\rangle_{phys} = 0 \,,$$

two relations which imply that

$$[P,Q]\,|\psi\rangle_{phys} = -i\hbar|\psi\rangle_{phys} = 0 \,,$$

namely, that $|\psi\rangle_{phys} = 0$ itself, i.e., the physical Hilbert space is empty. This is not entirely a satisfactory result.

To handle such situations, Dirac proposed that second-class constraint variables should be eliminated classically before quantization. This can be done either directly or by using so-called Dirac brackets, a procedure that does not require actually solving for the constraint variables that need to be eliminated. In either case, these procedures raise the potential difficulty that the remaining classical variables to be quantized may not admit Cartesian coordinates, in which case the procedure Dirac advocates could very well yield an incorrect result.

9.3 The Projection Operator Method

The set of general classical constraint functions, namely, $\{\phi_\alpha(p,q)\}_{\alpha=1}^{A} = 0$, can be recast as a *single* classical constraint simply by considering

$$\sum_{\alpha=1}^{A} \phi_\alpha(p,q)^2 = 0 \ ,$$

which to be true requires that every separate element must vanish. Likewise, the Dirac condition on the quantum constraints, namely, that

$$\Phi_\alpha(P,Q)\,|\psi\rangle_{phys} = 0$$

for all α, $1 \leq \alpha \leq A$, can also be rephrased as the single quantum constraint condition

$$\sum_{\alpha=1}^{A} \Phi_\alpha(P,Q)^2\,|\psi\rangle_{phys} = 0$$

without any loss of generality. This condition holds since the latter equation implies that

$$\sum_{\alpha=1}^{A} {}_{phys}\langle\psi|\Phi_\alpha(P,Q)^2\,|\psi\rangle_{phys} = 0 \ ,$$

and as vectors in Hilbert space,

$$0 = {}_{phys}\langle\psi|\Phi_\alpha(P,Q)^2\,|\psi\rangle_{phys} \iff \Phi_\alpha(P,Q)\,|\psi\rangle_{phys} = 0$$

for all α. Being equivalent to the Dirac condition means that the single constraint condition also fails when the Dirac procedure fails.

The *projection operator method* [Kla97, Kla01] modifies slightly the single constraint formulation of the Dirac procedure. As a self-adjoint operator

$$X \equiv \sum_{\alpha=1}^{A} \Phi_\alpha(P,Q)^2 = \int_0^\infty u\, d\mathbf{E}(\mathbf{u}) \ ,$$

where $\mathbf{E}(u)$ is a *spectral family of projection operators* [BEH] associated with the nonnegative, self-adjoint operator X. The integral over a limited range given by

$$\int_0^{\delta(\hbar)^2} d\mathbf{E}(u) \equiv \mathbb{E}(X \leq \delta(\hbar)^2)$$

defines a projection operator on the spectral range of the operator X that lies in the closed interval $[0, \delta(\hbar)^2]$.

In more familiar terms, for an operator X with a discrete spectrum $\{x_k, |k\rangle\}_{k=0}^{\infty}$, where $X\,|k\rangle = x_k|k\rangle$ with $x_{k+1} \geq x_k$, for example, it follows that

$$\mathbb{E}(X \leq \delta^2) = \sum_{k=0}^{K} |k\rangle\langle k| \ ,$$

where K is chosen so that $x_K \leq \delta^2$ and $x_{K+1} > \delta^2$. If the operator X has a continuous (absolutely continuous) spectrum, $X\,|y\rangle = y\,|y\rangle$, then

$$\mathbb{E}(X \leq \delta^2) = \int_0^{\delta^2} |y\rangle\langle y|\, dy = \mathbb{E}(X < \delta^2) \ .$$

In these relations, $\delta(\hbar) > 0$ denotes a *regularization parameter* that is fundamental to the projection operator method. (N.B. The symbol $\delta(\hbar)$ as used here does *not* mean a Dirac delta function!) When $\delta(\hbar)$ is a function of $\hbar$, as indicated, it has the additional feature that

$$\lim_{\hbar \to 0} \delta(\hbar) = 0 \, ,$$

implying that in the true classical limit in which $\hbar \to 0$ the quantum constraints $\Phi_\alpha(P, Q)$ pass in some appropriate sense to the classical constraints $\phi_\alpha(p, q)$. However, the quantum constraints are initially discussed with a finite, nonzero regularization parameter $\delta(\hbar) > 0$, as will be illustrated in the next section.

On the other hand, there are other cases where δ is independent of $\hbar$ and in those cases we need to take a suitable limit in which $\delta \to 0$ before we declare that the quantum constraints are fulfilled. Such limits will take many forms, and they are best described by illustrative examples.

In all cases, the regularized physical Hilbert space is defined by

$$\mathfrak{H}_{phys} = \mathbb{E}\mathfrak{H} \, .$$

Sometimes, when the true physical Hilbert space is a valid subspace of the original Hilbert space, this already becomes the final form of the physical Hilbert space. However, as examples below will illustrate, that is not always the case.

9.3.1 Observables and the classical limit

Any operator $\mathcal{O}$ for which

$$[\mathbb{E}, \mathcal{O}] = 0$$

qualifies as an observable in the (regularized) physical Hilbert space. Likewise, the observable component of any operator $\mathcal{G}$ is given by

$$\mathcal{G}^E \equiv \mathbb{E}\mathcal{G}\mathbb{E} \, .$$

The connection between the quantum theory and its associated classical theory is given by the diagonal coherent state matrix elements

$$G(p, q) \equiv \langle p, q | \mathcal{G} | p, q \rangle \, .$$

This expression still contains $\hbar$, and what we shall call the true classical limit is given by

$$G_c(p, q) \equiv \lim_{\hbar \to 0} G(p, q) = \lim_{\hbar \to 0} \langle p, q | \mathcal{G} | p, q \rangle \, ,$$

an expression which is best used when $\mathcal{G}$ is some form of generator.

Instead, if one deals with the (regularized) physical Hilbert space and observable operators, it follows that

$$G^E(p,q) \equiv \frac{\langle p,q|\mathcal{G}^E|p,q\rangle}{\langle p,q|\mathbb{E}|p,q\rangle} = \frac{\langle p,q|\mathbb{E}\mathcal{G}\mathbb{E}|p,q\rangle}{\langle p,q|\mathbb{E}|p,q\rangle} \ ,$$

and

$$G^E_c(p,q) \equiv \lim_{\hbar\to 0} \frac{\langle p,q|\mathcal{G}^E|p,q\rangle}{\langle p,q|\mathbb{E}|p,q\rangle} = \lim_{\hbar\to 0} \frac{\langle p,q|\mathbb{E}\mathcal{G}\mathbb{E}|p,q\rangle}{\langle p,q|\mathbb{E}|p,q\rangle} \ .$$

9.3.2 Basic examples of the projection operator method

Let us first illustrate three basic examples of the use of the projection operator method. These examples are presented in some detail to make sure the principles are fully clear.

Example 1. Consider the case of three constraints that are the Hermitian generators of the three-dimensional rotation group (summation understood):

$$J_j = \epsilon_{jkl}(Q_k P_l - Q_l P_k) \ ,$$

which satisfy

$$[J_j, J_k] = i\hbar\epsilon_{jkl} J_l \ ,$$

where $j,k,l \in \{1,2,3\}$. The constraints $\Phi_j = J_j$ are generators of a compact Lie group, SO(3), for which the original Dirac procedure works well. As is well known, the irreducible representations of this group involve angular momentum values $\ell \in \{0,1,2,3,\ldots\}$, or if we consider the covering group SU(2), then the angular momentum values are expanded to include all spin values such that now $\ell \in \{0,1/2,1,3/2,2,\ldots\}$. The three conditions $J_j|\psi\rangle_{phys} = 0$ ensure that only the angular momentum value $\ell = 0$ is included. This follows because for each of the irreducible representations

$$J_1^2 + J_2^2 + J_3^2 = \ell(\ell+1)\hbar^2 \, \mathbb{1}_\ell \ ,$$

where $\mathbb{1}_\ell$ is the unit operator in a $(2\ell+1)$-dimensional (sub)space of the particular irreducible representation in question. Thus if we select $\delta(\hbar)^2 = \hbar^2/2$ (or any other value strictly less than $3\hbar^2/4$), it follows that

$$\mathbb{E}(\Sigma_{j=1}^3 J_j^2 \le \hbar^2/2) = \mathbb{E}(\Sigma_{j=1}^3 J_j^2 = 0) \ .$$

It is important to appreciate that Example 1 is one for which $X = \Sigma_{j=1}^3 J_j^2$ is an operator with a *discrete spectrum* that *includes* the value zero in that spectrum. This characterization is for a set of *first-class constraints*.

Before leaving this example it is important to see it within a representation. For convenience and generality, the choice of representation we make is a

coherent state one. For this example we need coherent states for three degrees of freedom $\mathbf{p} = \{p_j\}_{j=1}^3$ and $\mathbf{q} = \{q_j\}_{j=1}^3$ given (with $\hbar = 1$) by

$$|\mathbf{p}, \mathbf{q}\rangle = e^{-i\mathbf{q}\cdot\mathbf{P}}\, e^{i\mathbf{p}\cdot\mathbf{Q}}\, |0\rangle \ ,$$

and an overlap

$$\langle \mathbf{p}', \mathbf{q}' | \mathbf{p}, \mathbf{q}\rangle = e^{i\frac{1}{2}(\mathbf{p}'+\mathbf{p})\cdot(\mathbf{q}'-\mathbf{q})}\, e^{-\frac{1}{4}[(\mathbf{p}'-\mathbf{p})^2 + (\mathbf{q}'-\mathbf{q})^2]} \ .$$

This function provides a reproducing kernel to define a coherent state representation of the original Hilbert space $\mathfrak{H}$. We observe (for $\hbar = 1$) that the function

$$K_E(\mathbf{p}', \mathbf{q}'; \mathbf{p}, \mathbf{q}) \equiv \langle \mathbf{p}', \mathbf{q}' | \mathbb{E} | \mathbf{p}, \mathbf{q}\rangle = e^{i(\mathbf{p}'\cdot\mathbf{q}' - \mathbf{p}\cdot\mathbf{q})/2} \times e^{-\frac{1}{4}(\mathbf{p}'^2 + \mathbf{q}'^2 + \mathbf{p}^2 + \mathbf{q}^2)} \frac{\sinh[\sqrt{(\mathbf{q}'-i\mathbf{p})^2(\mathbf{q}+i\mathbf{p})^2}/2]}{[\sqrt{(\mathbf{q}'-i\mathbf{p})^2(\mathbf{q}+i\mathbf{p})^2}/2]} \ ,$$

a proof of which we leave as an exercise for the reader. We note that $K_E(\mathbf{p}', \mathbf{q}'; \mathbf{p}, \mathbf{q})$ serves as a reproducing kernel for a coherent state representation of the physical Hilbert space $\mathfrak{H}_{phys} = \mathbb{E}\mathfrak{H}$.

Example 2. In this example we consider the two constraints P and Q, which as observed above is an example of second-class constraints. In this case $X = P^2 + Q^2$, which is just like the Hamiltonian for a harmonic oscillator, and therefore is an operator, like that in Example 1, with a *discrete spectrum*, but, unlike Example 1, it is an operator that does *not include* zero in its spectrum. Indeed, the spectrum of X is $X/\hbar \in \{1, 3, 5, 7, \ldots\}$. Thus if we choose

$$\mathbb{E}(P^2 + Q^2 \leq \hbar) \ ,$$

then $\mathbb{E}$ is a nonvanishing projection operator onto a space where the otherwise incompatible constraints P and Q are satisfied to the highest degree possible; the value $\delta(\hbar)^2 = \hbar$ can be increased to any value strictly less than $3\hbar$; on the other hand, if $\delta^2(\hbar)$ is made any smaller than $\hbar$, then the projection operator vanishes.

As already noted, it is important to appreciate that Example 2 is one for which $X = P^2 + Q^2$ is an operator with a discrete spectrum that does not include the value zero in the spectrum. Thus, it is fundamentally important that δ remain separated from zero. This characterization is for a set of *second-class constraints*. Note well the similarities and the differences involved in Examples 1 and 2.

To discuss a representation of the present example, we choose coherent states

$$|p, q\rangle = e^{-iqP/\hbar}\, e^{ipQ/\hbar}\, |0\rangle \ ,$$

with an overlap function given by

$$\langle p', q' | p, q \rangle = e^{i(p'+p)(q'-q)/2\hbar - [(p'-p)^2 + (q'-q)^2]/4\hbar} ,$$

a function which serves as the reproducing kernel for a reproducing kernel Hilbert space coherent state representation of the original Hilbert space $\mathfrak{H}$. The function

$$\begin{aligned} K_E(p', q'; p, q) &\equiv \langle p', q' | \mathbb{E} | p, q \rangle \\ &= e^{ip'q'/2\hbar - (p'^2 + q'^2)/4\hbar} \, e^{-ipq/2\hbar - (p^2+q^2)/4\hbar} \end{aligned}$$

is a function which serves as a reproducing kernel for the one-dimensional physical Hilbert space $\mathfrak{H}_{phys} = \mathbb{E}\mathfrak{H}$.

Next we consider an example of a very different kind.

Example 3. In this example there is just one constraint Q. The operator $X = Q^2$ is a nonnegative operator with a *continuous spectrum*. In this case δ is not a function of $\hbar$. The projection operator

$$\mathbb{E}(Q^2 \leq \delta^2) = \int_0^{\delta^2} d\mathbf{E}(u)$$

has the feature that

$$\mathbb{E}(Q^2 \leq \delta^2) = \mathbb{E}(-\delta \leq Q \leq \delta) = \mathbb{E}(-\delta < Q < \delta) ,$$

which clearly vanishes as $\delta \to 0$, i.e.,

$$\lim_{\delta \to 0} \mathbb{E}(Q^2 \leq \delta^2) = 0 ,$$

in the strong sense, for example. To analyze this situation more closely, let us take expectation values of this projection operator and diagonalize the constraint, leading to

$$\langle \phi | \mathbb{E}(Q^2 \leq \delta^2) | \phi \rangle = \int \int_{-\delta}^{\delta} \phi(x, y)^* \, \phi(x, y) \, dx \, dy ;$$

here the variable y accounts for any degeneracy in this system. Again, it is clear that all expectation values vanish as $\delta \to 0$; but there is also a clue how to get something out of this expression that does not vanish, namely, we need to first *rescale* this expression before we take the limit $\delta \to 0$. As a trial case we might adopt

$$\frac{1}{2\delta} \int \int_{-\delta}^{\delta} \phi(x, y)^* \, \phi(x, y) \, dx \, dy ,$$

which *for certain functions* leads to

$$\lim_{\delta \to 0} \frac{1}{2\delta} \int \int_{-\delta}^{\delta} \phi(x, y)^* \, \phi(x, y) \, dx \, dy = \int \phi(0, y)^* \, \phi(0, y) \, dy .$$

This limit would apply, for example, to functions of the form

$$\phi(x,y) \equiv \textstyle\sum_{m=0}^{M} x^m\, e^{-x^2}\, f_m(y)\ ,$$

where $M < \infty$ and the coefficients $f_m(y)$ are general, square-integrable functions. The resultant function would then be $\phi(0,y) = f_0(y)$. However, there are many vectors $|\phi\rangle$ and associated functions $\phi(x,y)$ that would fail to give a meaningful limit. In particular, let us consider square-integrable functions of the form

$$\phi^\epsilon(x,y) \equiv |x|^{-\epsilon} \textstyle\sum_{m=0}^{M} x^m\, e^{-x^2}\, f_m(y)\ ,$$

where $0 < \epsilon < 1/2$ and $M < \infty$. If we choose $\epsilon = 1/4$, for example, then we could consider

$$\lim_{\delta\to 0}\ \frac{(1-2/4)}{2\delta^{1-2/4}} \int\int_{-\delta}^{\delta} \phi^\epsilon(x,y)^*\, \phi^\epsilon(x,y)\, dx\, dy = \int \phi(0,y)^*\, \phi(0,y)\, dy\ ,$$

which informs us that this form of scaling works for $\epsilon = 1/4$ but clearly not for all values. The message is clear: we cannot expect the rescaled limit to hold for *all* vectors in the original Hilbert space. Instead, we choose a *select set of bras and kets*, which forms a total set of vectors in the original Hilbert space, and restrict limits to such vectors; recall that a total set spans the space. In particular, it suffices to choose a set of coherent states $|p,q\rangle$ as our select total set, and so we ask that a rescaled form of $\langle p',q'|\mathbb{E}|p,q\rangle$ is nonzero in the limit where $\delta \to 0$. For the case at hand, it follows (with $\hbar = 1$) that

$$\begin{aligned} K(p',q';p,q) &= \frac{1}{2\delta}\langle p',q'|\mathbb{E}|p,q\rangle \\ &= \frac{1}{2\delta\,\pi^{1/2}}\, e^{i(p'q'-pq)} \int_{-\delta}^{\delta} e^{-(x-q')^2/2} e^{-i(p'-p)x}\, e^{-(x-q')^2/2}\, dx\ . \end{aligned}$$

For $\delta \ll 1$ this expression may be approximately evaluated as

$$K(p',q';p,q) = e^{i(p'q'-pq)}\, e^{-(q'^2+q^2)/2}\, \frac{\sin[(p'-p)\,\delta]}{[\pi^{1/2}\,(p'-p)\,\delta]} + O(\delta)\ .$$

For very small δ it is noteworthy that the leading factor in this expression serves as a perfectly acceptable reproducing kernel, but it suffers from the fact that it is only an approximation since $\delta > 0$ still holds. It is clear that the limit exists as $\delta \to 0$ in which case

$$\lim_{\delta\to 0}\ \frac{1}{2\delta}\langle p',q'|\mathbb{E}|p,q\rangle = \frac{1}{\pi^{1/2}}\, e^{i(p'q'-pq)}\, e^{-(q'^2+q^2)/2}\ ,$$

but this limit lacks a natural, self-consistent definition and is needlessly complicated by an irrelevant factor ($\pi^{-1/2}$). If we take advantage of the fact that finite, positive multiples of a reproducing kernel do not change the space of

functions that make up the reproducing kernel Hilbert space, then a more satisfactory, and clearly self-consistent, rescaling procedure is to define

$$K_E(p', q'; p, q) \equiv \lim_{\delta \to 0} \frac{\langle p', q' | \mathbb{E} | p, q \rangle}{\langle \eta | \mathbb{E} | \eta \rangle} ,$$

for some choice of $|\eta\rangle$, which, in the present case if we choose $|\eta\rangle = |0\rangle$ (and restore $\hbar$), leads to

$$K_E(p', q'; p, q) = e^{ip'q'/\hbar - q'^2/2\hbar} \, e^{-ipq/\hbar - q^2/2\hbar} .$$

Apart from trivial phase factors that could be absorbed into the defintion of the coherent states if desired, it is noteworthy that this expression no longer depends on the parameters p' and p, which, since those variables enter the coherent states via the forms $p'Q$ and pQ, is the real signal that we have reached the space where "$Q = 0$." But this is only part of the story since we have focused on a select total set of bras and kets. We extend this expression by interpreting $K_E(p', q'; p, q)$ as a reproducing kernel for a functional representation of a reproducing kernel representation of the physical Hilbert space, which strictly speaking is no longer a subspace of the original Hilbert space. The resultant functional representation typically involves a new version of the inner product. It may not involve any form of a local integral representation of the inner product in this realization of the physical Hilbert space, but it always admits the usual inner product of traditional reproducing kernel Hilbert space formulations.

Alternatively, we can simplify the resulting reproducing kernel by fixing irrelevant parameters and leave an expression that can still serve as a suitable reproducing kernel. For example, in the final expression of Example 3, one could fix $p' = p = 1$ to yield

$$K'_E(q'; q) = e^{iq'/\hbar - q'^2/2\hbar} \, e^{-iq/\hbar - q^2/2\hbar} ,$$

or even simpler by setting $p' = p = 0$ giving

$$K''_E(q'; q) = e^{-q'^2/2\hbar} \, e^{-q^2/2\hbar} ,$$

both of which give completely satisfactory reproducing kernels for the same physical Hilbert space. The results of such processes lead to *reduced reproducing kernels*, and they can clearly be useful.

In summary, if the operator X representing the sum of the squared constraints has its zero in the continuous part of its spectrum, then a rescaled form limit is a convenient way for coherent state matrix elements to generate a function of positive type that can serve as a reproducing kernel for the physical Hilbert space; recall that a form refers to an entity, like an operator, but which requires a restriction on both kets *and* bras, while operators involve restrictions on kets alone.

9.3.3 Additional examples of the projection operator method

In the strong belief that examples are a great way to learn, we continue our discussion of the projection operator method through the study of additional examples. However, in contrast to the previous section, we will proceed more rapidly, building on the familiarity gained through a study of the initial three examples.

Example 4. We noted that reducible constraints—in which linear dependencies among the constraints exist—may cause some trouble, and thus our next example focuses on an elementary example of such behavior. In particular, we focus on an example with two constraints $\phi_1 = q = 0$ and $\phi_2 = q = 0$, namely, two *identical* constraints. In promoting this situation to the quantum level we encounter the two quantum constraints $\Phi_1 = Q$ and $\Phi_2 = Q$. This means that we focus on the projection operator

$$\mathbb{E} = \mathbb{E}(Q^2 + Q^2 \leq \delta^2)$$

where we have dropped any $\hbar$ dependence since the total constraint operator X has its zero in the continuous spectrum. In fact,

$$\mathbb{E} = \mathbb{E}(Q^2 + Q^2 \leq \delta^2) = \mathbb{E}(Q^2 \leq \delta^2/2) = \mathbb{E}(Q^2 \leq \delta'^2) \ ,$$

which shows that this reducible pair of constraints is mapped onto Example 3. Thus it is clear that

$$\lim_{\delta \to 0} \langle p', q' | \mathbb{E} | p, q \rangle / \langle 0 | \mathbb{E} | 0 \rangle = K_E(p', q'; p, q) \ ,$$

i.e., the very same reduced reproducing kernel arrived at in Example 3. Hence, two identical constraints are equivalent to one constraint. It is clear by this argument that M, $M < \infty$, identical constraints are equivalent to one as well. Procedures to deal with the situation where $M = \infty$ have been treated in the literature [Kla01].

Example 5. The second troublesome category of constraints was that of the irregular constraint, such as $\phi = q^3$, as opposed to the regular constraint $\phi' = q$. Quantum mechanically we must consider the constraint $\Phi = Q^3$ for which it follows that

$$\mathbb{E} = \mathbb{E}(Q^6 \leq \delta^2) = \mathbb{E}(Q^2 \leq (\delta^2)^{1/3}) = \mathbb{E}(Q^2 \leq \delta''^2) \ ,$$

which shows, just as above, that this example of an irreducible constraint is mapped onto Example 3. In other words,

$$\lim_{\delta \to 0} \langle p', q' | \mathbb{E} | p, q \rangle / \langle 0 | \mathbb{E} | 0 \rangle = K_E(p', q'; p, q) \ ,$$

the same reproducing kernel as in Example 3. Here the virtue of the self-consistent rescaling is very clear to see. Indeed, the same result holds for any

constraint of the form $\Phi = Q^\beta$, $\beta > 1$. One can even have two reducible, irregular constraints, such as $\Phi_1 = Q^5$ and $\Phi_2 = Q^5$, and the result is the same as in Example 3. Less obvious, but still true, is the case of two reducible but distinct constraints such as $\Phi_1 = Q$ and $\Phi_2 = Q^3$, which then leads to the projection operator

$$\mathbb{E} = \mathbb{E}(Q^2 + Q^6 \leq \delta^2) = \mathbb{E}(Q^2 \leq \tilde{\delta}^2) \, ,$$

where $\delta^2 = \tilde{\delta}^2 + \tilde{\delta}^6$. With this connection even this example leads to a reduced reproducing kernel given by that of Example 3.

Example 6. The two previous examples dealt with reducible and irregular constraints with their zeros in their continuous spectrum. Let us next examine examples that deal with discrete spectra. We suppose that the two constraints are $\phi_1 = p$ and $\phi_2 = q^2$. Thus we must consider $X = P^2 + Q^4$. This too is like some kind of Hamiltonian with a discrete spectrum of the form $X/\hbar^{4/3} \in \{c_0, c_1, c_2, \ldots\}$, where the set of pure numbers $\{c_j\}$ is an ordered set, $0 < c_0 < c_1 < c_2 < \ldots$. While the numbers $\{c_j\}$ are not known, it is clear that they exist. Armed with this information, we choose

$$\mathbb{E} = \mathbb{E}(P^2 + Q^4 \leq c_0 \hbar^{4/3})$$

which ensures that we project onto the ground state of this kind of anharmonic oscillator. The answer is similar to the case of the two constraints P and Q in that it is a projection operator onto a one-dimensional space for the physical Hilbert space. It is different from that case because it is not the same state. Does that matter? It may lead to some technical differences, but they can only be quantum-type corrections. After all, we could expect some difference to show up between regular second-class constraints and irregular second-class constraints, but both cases are well-defined. The case of two irregular second-class constraints given by $\Phi_1 = P^3$ and $\Phi_2 = Q^4$ leads to a projection operator given by

$$\mathbb{E} = \mathbb{E}(P^6 + Q^8 \leq c_0' \hbar^{24/7})$$

for a suitable choice of $c_0' > 0$, which then projects onto a one-dimensional space.

Example 7. We next study a constraint with multiple zeros such as $\phi = q(2 - q)$ which vanishes in a regular way at $q = 0$ and $q = 2$. The projection operator in this case reads

$$\mathbb{E} = \mathbb{E}(Q^2(2 - Q)^2 \leq \delta^2) \, .$$

For small enough δ, it follows that

$$\mathbb{E}(Q^2(2 - Q)^2 \leq \delta^2) = \mathbb{E}_0(4Q^2 \leq \delta^2) + \mathbb{E}_2(4(Q - 2)^2 \leq \delta^2) \, ,$$

where $\mathbb{E}_0$ and $\mathbb{E}_2$ are two orthogonal projection operators, the first centered about $Q = 0$ and the second centered about $Q = 2$. Using coherent states and rescaling the matrix elements leads us to consider

$$\langle p', q' | \mathbb{E}_0 | p, q \rangle / \langle 0 | \mathbb{E} | 0 \rangle + \langle p', q' | \mathbb{E}_2 | p, q \rangle / \langle 0 | \mathbb{E} | 0 \rangle \ ,$$

which converges as $\delta \to 0$ to

$$K_{E_0}(p', q'; p, q) + K_{E_2}(p', q'; p, q) \ ,$$

which is the reproducing kernel for a direct sum of two one-dimensional spaces, one for each separate kernel. It is possible to use the same rescaling factor in this case since each zero in the constraint is of the same degree, i.e., each are first order zeros.

Example 8. We modify the previous example to study the single constraint $\phi = q(2-q)^3$, which has a regular zero at $q = 0$ but an irregular zero at $q = 2$. For very small δ,

$$\mathbb{E} = \mathbb{E}(Q^2 (2 - Q)^6 \leq \delta^2) = \mathbb{E}_0(64 Q^2 \leq \delta^2) + \mathbb{E}_2(4(Q-2)^6 \leq \delta^2) \ .$$

Here we encounter a difficulty, namely, that the two projection operators vanish at different rates as $\delta \to 0$. Therefore, a single scaling factor will not catch both contributions at the same time. However, when we treated this example from a classical standpoint, we argued that both contributions, the one at $q = 0$ and the other at $q = 2$, should be part of the constraint hypersurface (such as it is). We seek to ensure that this happens at the quantum level as well.

Again let us introduce coherent state matrix elements leading to

$$\begin{aligned} K(p', q'; p, q) &\equiv \langle p', q' | \mathbb{E} | p, q \rangle \\ &= \langle p', q' | \mathbb{E}_0 | p, q \rangle + \langle p', q' | \mathbb{E}_2 | p, q \rangle \\ &\equiv K_0(p', q'; p, q) + K_2(p', q'; p, q) \ , \end{aligned}$$

and observe that the two kernels are orthogonal,

$$\int K_0(p'', q''; p, q)\, K_2(p, q; p', q')\, d\mu(p, q) = 0 \ ,$$

due to the orthogonality, $\mathbb{E}_0 \mathbb{E}_2 = 0$, of the two projection operators for sufficiently small δ. That means we can craft an *amplification factor*

$$A(p', q'; p, q) \equiv K_0(p', q'; p, q) / \sqrt{\langle \eta_0 | \mathbb{E}_0 | \eta_0 \rangle} + K_2(p', q'; p, q) / \sqrt{\langle \eta_2 | \mathbb{E}_2 | \eta_2 \rangle}$$

that has the virtue that the new reproducing kernel on the original Hilbert space given by

$$\begin{aligned}\tilde{K}(p'',q'';p',q') &\equiv \int A(p'',q'';\overline{p},\overline{q})\,K(\overline{p},\overline{q};p,q)\,A(p,q;p',q')\,d\mu(\overline{p},\overline{q})\,d\mu(p,q)\\ &= K_0(p',q';p,q)/\langle\eta_0|\,\mathbb{E}_0\,|\eta_0\rangle + K_2(p',q';p,q)/\langle\eta_2|\,\mathbb{E}_2\,|\eta_2\rangle\\ &\equiv C_0\,K_0(p',q';p,q) + C_2\,K_2(p',q';p,q)\end{aligned}$$

has been rescaled in such a way that *both* factors will contribute to the limit in which $\delta \to 0$. It is clear, now, that

$$K'_E(p',q';p,q) \equiv \lim_{\delta\to 0}\,\tilde{K}(p',q';p,q) \equiv K'_{E_0}(p',q';p,q) + K'_{E_2}(p',q';p,q)$$

includes suitable contributions from both sectors $Q = 0$ and $Q = 2$. For future reference, we note that the *diagonal elements* for this example are given by

$$K'_{E_0}(p,q;p,q) = e^{-q^2/\hbar}\,, \qquad K'_{E_2}(p,q;p,q) = e^{-(q\,-\,2)^2/\hbar}\,;$$

in determining these expressions, we have assumed that $|\eta_0\rangle = |0\rangle = |p = 0, q = 0\rangle$ and $|\eta_2\rangle = |p = 0, q = 2\rangle$.

Here, in this example, we see the power of the projection operator method. The procedure is the same for all constraints in the regularized part of the process where δ is positive and large enough. In the second part of the process, the details of the constraint, as embodied in the spectral properties of $X = \Sigma_\alpha \Phi_\alpha^2$, are examined critically. In the most recent example, for instance, there are two regions where X vanishes, one near $Q = 0$, the other near $Q = 2$. We could select just one of these regions for the physical Hilbert space, or, as we actually did, we could choose both regions on an equal footing to contribute to the physical Hilbert space, by specifying the manner in which the limit $\delta \to 0$ is taken. The use of the amplification factor enabled us to put the discordant constraint features onto common ground. This could have been done in many ways. For example, we could have chosen a somewhat different amplification factor such as

$$\begin{aligned}A^\star(p',q';p,q) \equiv{}& \sqrt{a}\,K_0(p',q';p,q)/\sqrt{\langle\eta_0|\,\mathbb{E}_0\,|\eta_0\rangle}\\ &+\sqrt{b}\,K_2(p',q';p,q)/\sqrt{\langle\eta_2|\,\mathbb{E}_2\,|\eta_2\rangle}\,,\end{aligned}$$

where $a > 0$ and $b > 0$. The final result would have been changed to read

$$K'^\star_E(p',q';p,q) \equiv \lim_{\delta\to 0}\,\tilde{K}^\star(p',q';p,q) \equiv a\,K'_{E_0}(p',q';p,q) + b\,K'_{E_2}(p',q';p,q)\,,$$

namely, we have introduced arbitrary multipliers of each sector of the physical Hilbert space. These could also arise by choosing different terms in the original amplification factor by making different choices for $|\eta_0\rangle$ and $|\eta_2\rangle$. In any case, it is important to understand what differences such factors as a and b introduce to the story.

As an analogy, consider the case of the direct sum of two finite-dimensional Hilbert spaces $\mathfrak{H} = \mathfrak{H}_1 \oplus \mathfrak{H}_2$. Any vector $V \in \mathfrak{H}$ is given by $V = V_1 \oplus$

V_2. However, the inner product $\mathfrak{H}$ is yet to be defined, and for that we can arbitrarily choose

$$(V, W)_{\mathfrak{H}} \equiv a(V_1, W_1)_{\mathfrak{H}_1} + b(V_2, W_2)_{\mathfrak{H}_2} \ .$$

This freedom is that of a *similarity transformation*; it changes the nature of the *representation*, but it does not change the *physical meaning*. Such is also the interpretation of the parameters a and b in the modified reduced reproducing kernel above.

To see that no physical objects are changed by the presence of these factors, let us study the allowed reduced phase space that arises in the classical limit of the proposed quantization of the present system under discussion. Recall that the observable aspect of the coordinate Q in the (regularized) physical Hilbert space is given by $Q^E = \mathbb{E}Q\mathbb{E}$. Furthermore, with $q_{reduced}(p,q) \equiv Q^E_c(p,q)$, the true classical limit is given by the expression

$$\begin{aligned} q_{reduced}(p,q) &\equiv \lim_{\hbar\to 0}\lim_{\delta\to 0} \frac{\langle p,q|\,\mathbb{E}Q\mathbb{E}\,|p,q\rangle}{\langle p,q|\,\mathbb{E}\,|p,q\rangle} \\ &= \lim_{\hbar\to 0}\lim_{\delta\to 0} \frac{aC_0\langle p,q|\,\mathbb{E}_0Q\mathbb{E}_0\,|p,q\rangle + bC_2\langle p,q|\,\mathbb{E}_2Q\mathbb{E}_2\,|p,q\,\rangle}{aC_0\langle p,q|\,\mathbb{E}_0\,|p,q\rangle + bC_2\langle p,q|\,\mathbb{E}_2\,|p,q\,\rangle} \\ &= \lim_{\hbar\to 0}\lim_{\delta\to 0} \frac{2\,bC_2\,\langle p,q|\,\mathbb{E}_2\,|p,q\,\rangle}{aC_0\langle p,q|\,\mathbb{E}_0\,|p,q\rangle + bC_2\langle p,q|\,\mathbb{E}_2\,|p,q\,\rangle} \\ &= \lim_{\hbar\to 0} \frac{2\,be^{-(q-2)^2/\hbar}}{ae^{-q^2/\hbar} + be^{-(q-2)^2/\hbar}} \ . \end{aligned}$$

Before $\hbar \to 0$ one observes that this expression lies between 0 and 2, with a transition region of width $O(\sqrt{\hbar})$ at $q = 1$. In the limit that $\hbar \to 0$, it follows that

$$\begin{aligned} q_{reduced}(p,q) &= 0 \qquad\qquad \text{when } \ q < 1 \ , \\ q_{reduced}(p,q) &= 2b/(a+b) \quad \text{when } \ q = 1 \ , \\ q_{reduced}(p,q) &= 2 \qquad\qquad \text{when } \ q > 1 \ . \end{aligned}$$

Observe that the classical reduced phase space contains exactly the two expected values: $q = 0$ and $q = 2$. The value at $q = 1$ is a residue of the quantum representation, but it is classically unobservable since it is occurs on a set of measure zero in the original phase space.

Further discussion of examples of this sort may be found in [KL06].

9.3.4 Representation of the projection operator

Thus far in our discussion we have focused on how the projection operator can be used in various constraint situations. Now it is time to see how to make the projection operator itself from more elementary ingredients. In certain cases,

there will turn out to be several distinct procedures to generate the desired projection operator. Some of these different ways have the property that the method of construction depends heavily on the set of constraints involved, implying that for each set of constraints there is most likely a different rule of construction of the projection operator. This fact is not intrinsically bad, but it also turns out that there is a *universal* way to make the projection operator that is completely independent of the nature of the constraints themselves. In the author's judgment, this universal way is preferable and will be the mode almost exclusively used. The only exception to that rule would be for certain situations for which an alternative scheme of construction was far simpler to implement.

Let us first suppose we deal with a set of constraints that form the generators of a compact Lie group. In this case, we can represent a unitary representation of the group by so-called canonical group coordinates of the first kind for which (summation understood)

$$U(g) \equiv e^{-ig^k \Phi_k} \ ,$$

where $g = \{g_k\}_{k=1}^K$. The usual group multiplication rule is written as

$$U(g')\, U(g) = U(g' \circ g) \ ,$$

with $g' \circ g$ denoting an appropriate combination rule among group coordinates. Lie groups admit group invariant measures, and for compact groups there is one unique form of measure (up to a multiple, since the left-invariant and right-invariant group invariant measures coincide). We denote the group-invariant measure by $d\rho(g)$, and without loss of generality we assume it is normalized so that $\int d\rho(g) = 1$. This measure has several invariances such as

$$d\rho(g) = d\rho(g_0 \circ g) = d\rho(g \circ g_0) = d\rho(g^{-1}) \ ,$$

where the last expression involves the inverse group elements. These facts can be used to show that

$$\mathbb{E}(\Sigma_k \Phi_k^2 = 0) = \int U(g)\, d\rho(g) \ .$$

The argument goes as follows: let $A \equiv \int U(g)\, d\rho(g)$, then

$$\begin{aligned} AA &= \int U(g')\, U(g)\, d\rho(g')\, d\rho(g) = \int d\rho(g) \int U(g' \circ g)\, d\rho(g') \\ &= \int d\rho(g) \int U(g')\, d\rho(g') = A \ , \end{aligned}$$

and

$$A^\dagger = \int U(g)^\dagger\, d\rho(g) = \int U(g^{-1})\, d\rho(g^{-1}) = A \ .$$

Thus A is a projection operator. The fact that

$$U(g')\,A = \int U(g' \circ g)\,d\rho(g) = A$$

proves that $A = \mathbb{E}(\Sigma_k \Phi_k^2 = 0)$ as was to be shown. This construction directly yields the limiting case where $\delta = 0$, and it should be kept in mind that for different Lie groups, even those with the same number of parameters, the form of the measure will typically change for each new set of constraints. An early version of the projection operator method, as described in this paragraph and suitable for constraints that form a compact Lie algebra, was introduced by Shabanov [Sha89].

Noncompact Lie groups have invariant measures such that $\int d\rho(g) = \infty$, and since such measures cannot be normalized they cannot fit into the previous scheme. Such groups can also cause other problems because some of the generators (i.e., constraints) may have their zero in the continuous spectrum.

A variety of examples for the realization of the projection operators $\mathbb{E}$ of interest is presented in [Kla97].

9.3.5 A universal representation for the projection operator

We now present a universal scheme that applies to *any* set of constraints $\{\Phi_\alpha\}_{\alpha=1}^{A}$, and which does not rely on any group property that may or may not hold among the constraint elements. To win this much generality with a universal scheme, it is not surprising that the procedure we offer is more complex than that described above. We begin by discussing the first part of the integrand given by

$$\mathsf{T}\, e^{-i\int_{t_1}^{t_2} \lambda^\alpha(t)\,\Phi_\alpha\,dt}\,.$$

Here, T denotes *time ordering*, which refers to the ordering of the integral in the exponential; in particular,

$$\begin{aligned}
&\mathsf{T}[\textstyle\int_{t_1}^{t_2} \lambda^\alpha(t)\,\Phi_\alpha\,dt\,]^p \\
&\quad = \textstyle\int_{t_1}^{t_2}\int_{t_1}^{t_2}\cdots\int_{t_1}^{t_2} \mathsf{T}[\lambda^{\alpha_p}(s_p)\,\Phi_{\alpha_p}]\cdots[\lambda^{\alpha_1}(s_1)\Phi_{\alpha_1}]\,ds_p\cdots ds_1 \\
&\quad \equiv p!\textstyle\int_{t_1}^{t_2} ds_p \int_{t_1}^{s_p} ds_{p-1}\cdots\int_{t_1}^{s_2} ds_1\,[\lambda^{\alpha_p}(s_p)\,\Phi_{\alpha_p}]\,[\lambda^{\alpha_{p-1}}(s_{p-1})\,\Phi_{\alpha_{p-1}}] \\
&\qquad\quad \times\cdots\times\,[\lambda^{\alpha_1}(s_1)\,\Phi_{\alpha_1}]\,,
\end{aligned}$$

which shows how each term in a power series expansion of the exponential is to be treated. Under a time-ordering operation, as we have, one can freely move the operators past one another as if they commuted with each other since the proper order will always be controlled in the end by the magic of the time-ordering symbol T. Thus we are next free to integrate the basic integrand over the set $\{\lambda^\alpha(t)\}$ of ordinary real functions. We choose a Gaussian weighting factor for this integration and are led to

$$\mathcal{K} \int \mathsf{T}\, e^{-i\int_{t_1}^{t_2} \lambda^\alpha(t)\, \Phi_\alpha\, dt}\; e^{\frac{1}{4} i\gamma^{-1} \int_{t_1}^{t_2} \Sigma_{\alpha=1}^{A} \lambda_\alpha(t)^2\, dt}\; \Pi_t\, \Pi_{\alpha=1}^{A}\, d\lambda^\alpha(t)\ ,$$

where γ is a real parameter. The proper answer is given by

$$e^{-i\gamma \int_{t_1}^{t_2} \Sigma_{\alpha=1}^{A} \Phi_\alpha^2\, dt} = e^{-i\gamma\, \Delta\, \Sigma_{\alpha=1}^{A} \Phi_\alpha^2}\ ,$$

where $\Delta \equiv (t_2 - t_1) > 0$ and the initial factor $\mathcal{K}$ has been chosen to achieve the final normalization as shown. If necessary to ensure convergence, one may let γ have a tiny imaginary component, which is then set to zero after the integration; we shall regard any such modification as implicit. This is the first part of the story.

Next we recall the Fourier transform of the function

$$f(u) = 1\ , \quad |u| \le \Delta\delta^2\ ; \qquad\qquad f(u) = 0\ , \quad |u| > \Delta\delta^2\ ,$$

given by

$$\int f(u)\, e^{i\, u\gamma}\, du = \int_{-\Delta\,\delta^2}^{\Delta\,\delta^2} e^{i\, u\gamma}\, du = 2\, \frac{\sin[\gamma\, \Delta\delta^2]}{\gamma}\ .$$

Consequently,

$$f(u) = \frac{1}{2\pi} \int 2\, \frac{\sin[\gamma\, \Delta\delta^2]}{\gamma}\, e^{-i\, u\gamma}\, d\gamma\ ;$$

in fact, since we are dealing with conditionally convergent integrals, it may be necessary to adopt

$$f(u) = \lim_{\zeta \to 0^+} \frac{1}{2\,\pi} \int 2\, \frac{\sin[\gamma\, \Delta(\delta^2 + \zeta)]}{\gamma}\, e^{-i\, u\gamma}\, d\gamma$$

to ensure we secure the desired equality sign in $f(u)$, i.e., $f(\Delta\delta^2) = 1$. In any case, we shall regard any such operation as implicit.

Putting all the ingredients together, we are led directly to our final relation

$$\mathbb{E}(\Sigma_\alpha \Phi_\alpha^2 \le \delta^2) = \mathbb{E}(\Delta\, \Sigma_\alpha \Phi_\alpha^2 \le \Delta\delta^2) = \int \frac{\sin[\gamma\, \Delta\delta^2]}{\pi\, \gamma}\, e^{-i\, \Delta\gamma\, \Sigma_\alpha \Phi_\alpha^2}\, d\gamma\ .$$

Thus, with this twofold procedure, first a Gaussian time-ordered integral, followed by a $\sin(x)/x$ type of Fourier transform, we have achieved our goal of a single form of integral representation, valid for any set of constraints; note that the combined operations depend on: (1) the number of constraints A, (2) the time interval involved $\Delta = (t_2 - t_1)$, and (3) the parameter $\delta = \delta(\hbar)$. *The procedure to construct the projection operator does not depend on the nature of the constraints in any fashion.*

We combine these operations, leaving the integration over the variance parameter γ as implicit, and hereafter we simply write

$$\mathbb{E}(\Sigma_\alpha \Phi_\alpha^2 \leq \delta(\hbar)^2) = \int \mathsf{T}\, e^{-i\int \lambda^\alpha(t)\Phi_\alpha\, dt}\, \mathcal{D}R(\lambda)\ .$$

To emphasize its universal validity, we observe that this formula is correct even if $\Phi_\alpha = 0$ for all α. In that case it follows that

$$\mathbb{E}(0 \leq \delta(\hbar)^2) = \mathbb{1} = \int \mathbb{1}\, \mathcal{D}R(\lambda)\ ,$$

which can be read as the normalization condition

$$\int \mathcal{D}R(\lambda) = 1\ .$$

This is an important property of the projection operator method.

9.4 Constrained Dynamics in Operator Form

Let us turn our attention from kinematics to dynamics. The essence of the main idea is very simple indeed. Unconstrained temporal evolution (with $\hbar = 1$) for a time-independent Hamiltonian operator $\mathcal{H}$ is expressed by the operator formula

$$e^{-iT\mathcal{H}} = e^{-i\epsilon\mathcal{H}}\, e^{-i\epsilon\mathcal{H}} \cdots e^{-i\epsilon\mathcal{H}}\ ,$$

where $\epsilon \equiv T/(N+1)$ as we have used previously. When constraints are present, we must respect the fact that all behavior must take place in the physical Hilbert space, or at least initially in the regularized physical Hilbert space. Thus the factorized evolution equation presented above should be replaced with the expression

$$\lim_{N\to\infty} \mathbb{E}\, e^{-i\epsilon\mathcal{H}}\, \mathbb{E}\, e^{-i\epsilon\mathcal{H}}\, \mathbb{E} \cdots \mathbb{E}\, e^{-i\epsilon\mathcal{H}}\, \mathbb{E}\ .$$

Such an equation ensures that we start in the physical Hilbert space, and after each evolution by a small time step we ensure that we retain only those changes that reside in the physical Hilbert space. Finally, we must take the limit as $N \to \infty$ to ensure that such a construction is done as smoothly as possible. A theorem of Chernoff [Che68] implies that the result of this limit is given by

$$\mathbb{E}\, e^{-iT(\mathbb{E}\mathcal{H}\mathbb{E})}\, \mathbb{E} = \mathbb{E}\, e^{-iT\mathcal{H}^E}\, \mathbb{E}\ ,$$

which is easy to understand based on a hypothetical expansion in powers of $\mathcal{H}$. This is a very satisfactory result. The Hamiltonian $\mathcal{H}$ has been replaced by its observable component $\mathcal{H}^E \equiv \mathbb{E}\mathcal{H}\mathbb{E}$. Since $\mathcal{H}$ was originally self-adjoint, it follows that $\mathcal{H}^E$ is always a symmetric operator. If $\mathcal{H}$ is bounded below, i.e., $\mathcal{H} \geq c\mathbb{1}$, then $\mathcal{H}^E$ is also bounded below and we are assured that it admits a

self-adjoint extension. The point of this remark is to argue that in most cases of interest the induced evolution in the physical Hilbert space is unitary just as it was in the original Hilbert space before the constraints were imposed.

The evolution described in the previous paragraph is the most involved that can arise, one where the temporal evolution tends to leave the physical Hilbert space and thus requires to be forced back into the physical Hilbert space at each moment of time. This scenario is very much like our classical discussion regarding *second-class constraints*, which also required that the Lagrange multipliers assume particular values to ensure that the system remained on the constraint hypersurface. Besides those situations, there were also the *first-class constraint* situations for which once the system was put on the constraint hypersurface it remained there whatever choice was made for the Lagrange multipliers, i.e., whatever gauge choice was made. Such systems have their quantum analogs as well, and they correspond to situations where the Hamiltonian is *already an observable* to begin with. If that is the case, then

$$[\mathbb{E}, \mathcal{H}] = 0$$

and, consequently, the dynamical evolution simplifies dramatically so that

$$\mathbb{E}\, e^{-iT\,(\mathbb{E}\mathcal{H}\mathbb{E})}\, \mathbb{E} = e^{-iT\mathcal{H}}\, \mathbb{E}\,.$$

The appearance of the initial projection operator is analogous to the initial value equation of the classical theory, namely, putting the system initially in compliance with the constraints, which is all that is necessary for a system with first-class constraints.

To summarize, the general formula derived initially always applies and is necessary to deal with the analogs of second-class classical systems; the simpler formula, with only an initial projection onto the physical Hilbert space, applies whenever the original Hamiltonian is already an observable and therefore is appropriate for first-class constraint systems.

As a modest generalization of the foregoing story, we can consider evolution in the presence of a time-dependent Hamiltonian $\mathcal{H}(t)$. Time development without the application of any constraints leads to the unitary evolution operator (still with $\hbar = 1$)

$$\mathsf{T}\, e^{-i\int_0^T \mathcal{H}(t)\,dt} = \lim_{N\to\infty} e^{-i\epsilon\mathcal{H}_N}\, e^{-i\epsilon\mathcal{H}_{N-1}}\, \cdots\, e^{-i\epsilon\mathcal{H}_0}$$

where $\mathcal{H}_k \equiv \mathcal{H}(k\epsilon)$. Introduction of the projection operator $\mathbb{E}$ leads to

$$\lim_{N\to\infty} \mathbb{E}\, e^{-i\epsilon\mathcal{H}_N}\, \mathbb{E}\, e^{-i\epsilon\mathcal{H}_{N-1}}\, \mathbb{E}\, \cdots\, \mathbb{E}\, e^{-i\epsilon\mathcal{H}_0}\, \mathbb{E}\,,$$

which becomes

$$\mathbb{E}\, \mathsf{T}\, e^{-i\int_0^T (\mathbb{E}\mathcal{H}(t)\mathbb{E})\,dt}\, \mathbb{E} = \mathbb{E}\, \mathsf{T}\, e^{-i\int_0^T \mathcal{H}^E(t)\,dt}\, \mathbb{E}\,.$$

If $\mathcal{H}(t)$ is an observable, i.e., if $[\mathbb{E}, \mathcal{H}(t)] = 0$ for all t, $0 \le t \le T$, then

$$\mathbb{E}\,\mathsf{T}\, e^{-i\int_0^T (\mathbb{E}\mathcal{H}(t)\mathbb{E})\,dt}\,\mathbb{E} = \mathsf{T}\, e^{-i\int_0^T \mathcal{H}(t)\,dt}\,\mathbb{E}\,.$$

9.5 Coherent State Path Integrals for Systems with Constraints

With the foregoing extensive discussion about the projection operator method and temporal evolution, the realization of these expressions as coherent state path integrals is readily obtained. For simplicity, we deal with a single degree of freedom, although the extension to finitely many degrees of freedom is straightforward. As usual we adopt as a basis the coherent states (with $\hbar = 1$)

$$|p, q\rangle = e^{-iqP}\, e^{ipQ}\, |0\rangle\,,$$

where, again, as usual, $(Q + iP)|0\rangle = 0$. We will make heavy use of the resolution of unity

$$\mathbb{1} = \int |p, q\rangle\langle p, q|\, d\mu(p, q)\,,$$

where, as before, $d\mu(p, q) = dp\,dq/2\pi$ and integration extends over the whole phase space $\mathbb{R}^2$. Focusing on the case of a time-independent Hamiltonian, we are led to

$$\begin{aligned}
&\langle p'', q''|\,\mathbb{E}\, e^{-iT(\mathbb{E}\mathcal{H}\mathbb{E})}\,\mathbb{E}\,|p', q'\rangle \\
&\quad = \lim_{N\to\infty} \int\cdots\int \textstyle\prod_{n=0}^{N} \langle p_{n+1}, q_{n+1}|\mathbb{E}\, e^{-i\epsilon\mathcal{H}}\,\mathbb{E}|p_n, q_n\rangle \prod_{n=1}^{N} d\mu(p_n, q_n) \\
&\quad = \lim_{N\to\infty} M''M' \int\cdots\int \textstyle\prod_{n=0}^{N} \langle\!\langle p_{n+1}, q_{n+1}|e^{-i\epsilon\mathcal{H}}|p_n, q_n\rangle\!\rangle \\
&\qquad\qquad \times \textstyle\prod_{n=1}^{N} \langle p_n, q_n|\mathbb{E}|p_n, q_n\rangle\, d\mu(p_n, q_n)\,,
\end{aligned}$$

where $|p, q\rangle\!\rangle \equiv \mathbb{E}|p, q\rangle / \|\,\mathbb{E}|p, q\rangle\|$ and $M(p, q) \equiv \|\,\mathbb{E}|p, q\rangle\| = \langle p, q|\mathbb{E}|p, q\rangle^{1/2}$. As usual, $p_{N+1}, q_{N+1} = p'', q''$ and $p_0, q_0 = p', q'$, and thus $M'' = M(p'', q'')$ and $M' = M(p', q')$. Since the states $|p, q\rangle\!\rangle$ are normalized, the previous expression may be approximated by

$$\begin{aligned}
\lim_{N\to\infty} M''M' \int\cdots\int \exp\Big(i\textstyle\sum_{n=0}^{N} \{ i\,\langle\!\langle p_{n+1}, q_{n+1}|\,[\,|p_{n+1}, q_{n+1}\rangle\!\rangle - |p_n, q_n\rangle\!\rangle\,] \\
-\epsilon H(p_{n+1}, q_{n+1}; p_n, q_n)\}\Big) \textstyle\prod_{n=1}^{N} \langle p_n, q_n|\mathbb{E}|p_n, q_n\rangle\, d\mu(p_n, q_n)\,,
\end{aligned}$$

where

$$H(p_{n+1}, q_{n+1}; p_n, q_n) = \frac{\langle\!\langle p_{n+1}, q_{n+1}|\mathcal{H}|p_n, q_n\rangle\!\rangle}{\langle\!\langle p_{n+1}, q_{n+1}|p_n, q_n\rangle\!\rangle} = \frac{\langle p_{n+1}, q_{n+1}|\,\mathbb{E}\mathcal{H}\mathbb{E}|p_n, q_n\rangle}{\langle p_{n+1}, q_{n+1}|\,\mathbb{E}|p_n, q_n\rangle}\,.$$

If we interchange the limit and integrations, we are led to the formal expression

$$M'' M' \int \exp\{i \int [\, i \langle\!\langle p,q|\, (d/dt) |p,q\rangle\!\rangle - \langle\!\langle p,q|\mathcal{H}|p,q\rangle\!\rangle \,]\, dt\}\, \mathcal{D}\mu_E(p,q)\ ,$$

where the measure μ_E implicitly incorporates the proper reweighting. We can reexpress this equation in terms of the original coherent states as follows:

$$M'' M' \int \exp\{i \int [\, i \langle p,q|\mathbb{E}(d/dt)\mathbb{E}|p,q\rangle / \langle p,q|\mathbb{E}|p,q\rangle - \langle p,q|\mathbb{E}\mathcal{H}\mathbb{E}|p,q\rangle / \langle p,q|\mathbb{E}|p,q\rangle \,]\, dt\}\, \mathcal{D}\mu_E(p,q)\ .$$

The formulas for temporal evolution of constrained systems discussed above have been based on a path integral in which the quantum constraints have already been enforced. As a variation on the foregoing formulation, let us illustrate how the universal integral representation for the projection operator can be used to introduce the constraints as part of the path integral expression. In the following (rather lengthy!) equation chain, the weak measure $R(\lambda)$ is made explicit as we choose an expression that achieves our goal, namely:

$$\begin{aligned}
&\mathcal{M} \int e^{i \int_0^T [p_j \dot{q}^j - H(p,q) - \lambda^a \phi_\alpha(p,q)]\, dt}\, \mathcal{D}p\, \mathcal{D}q\, \mathcal{D}R(\lambda) \equiv \lim_{N\to\infty} \int \prod_{n=0}^{N} \\
&\times \Bigg\{ \lim_{M\to\infty} \int \prod_{m=1}^{M} \Bigg[\Bigg(\langle p_{n+m/M}, q_{n+m/M} | p_{n+(m-1)/M}, q_{n+(m-1)/M} \rangle \\
&+ \delta_{m,M} \langle p_{n+m/M}, q_{n+m/M} | (-i\epsilon\mathcal{H}) | p_{n+(m-1)/M}, q_{n+(m-1)/M} \rangle \\
&+ \langle p_{n+m/M}, q_{n+m/M} | [-i(\epsilon/M)\lambda^\alpha_{n,m} \Phi_\alpha - (\epsilon^2/2M^2)\lambda^\alpha_{n,m} \lambda^\beta_{n,m} \Phi_\alpha \Phi_\beta] \\
&\times | p_{n+(m-1)/M}, q_{n+(m-1)/M} \rangle \Bigg) (c\gamma_n)^{-A/2}\, e^{i\epsilon/(4M\gamma_n)\, \Sigma_\alpha \lambda^{\alpha\,2}_{n,m}}\, \Pi_\alpha\, d\lambda^\alpha_{n,m} \Bigg] \Bigg\} \\
&\times \prod_{n=1}^{N} \Bigg[\Bigg(\prod_{m=1}^{M} dp_{n+m/M}\, dq_{n+m/M}/(2\pi) \Bigg) \frac{\sin[\gamma_n \epsilon \delta(\hbar)^2]}{\pi\gamma_n}\, d\gamma_n \Bigg] \\
&= \lim_{N\to\infty} \int \prod_{n=0}^{N} \Bigg\{ \langle p_{n+1}, q_{n+1} | (1 - i\epsilon\mathcal{H})\, e^{-i\gamma_n \epsilon \Sigma_\alpha \Phi^2_\alpha} | p_n, q_n \rangle \\
&\times \frac{\sin[\gamma_n \epsilon \delta(\hbar)^2]}{\pi\gamma_n}\, d\gamma_n \Bigg\} \prod_{n=1}^{N} dp_n\, dq_n/(2\pi) \\
&= \lim_{N\to\infty} \int \prod_{n=0}^{N} \langle p_{n+1}, q_{n+1} | e^{-i\epsilon\mathcal{H}} \mathbb{E} | p_n, q_n \rangle \prod_{n=1}^{N} dp_n\, dq_n/(2\pi) \\
&= \lim_{N\to\infty} \langle p'', q'' | e^{-i\epsilon\mathcal{H}} \mathbb{E} \cdots e^{-i\epsilon\mathcal{H}} \mathbb{E} e^{-i\epsilon\mathcal{H}} \mathbb{E} | p', q' \rangle \\
&= \langle p'', q'' | \mathbb{E} e^{-i(\mathbb{E}\mathcal{H}\mathbb{E})T}\, \mathbb{E} | p', q' \rangle\ .
\end{aligned}$$

Here, as usual, $p'', q'' = p_{N+1}, q_{N+1}$ as well as $p', q' = p_0, q_0$, and the constant $c = 4\pi i M/\epsilon$, that appears part way through the equation chain, is a normalization chosen to ensure the form of the equation which follows the one in which c appears.

It is important to observe that, unlike the Hamiltonian $\mathcal{H}$, it was necessary to expand the expression involving the constraints Φ_α to *second* order in the small parameter ϵ. In addition, it was necessary to introduce an additional refinement (M) of each small time step (ϵ) in order to construct a projection operator ($\mathbb{E}$) to go along with each of the large number (N) of small time step evolutions for the Hamiltonian ($\mathcal{H}$) [Kla05].

This concludes our study of classical and quantum constraints, and how they appear within coherent state path integrals.

Exercises

9-1 We impose the constraints that the three rotation generators $J_j = \epsilon_{jkl}(Q_k P_l - Q_l P_k)$, which satisfy $[J_j, J_k] = i\hbar\epsilon_{jkl} J_l$, should vanish so that $\mathbb{E} = \mathbb{E}(\Sigma_{j=1}^3 J_j^2 = 0)$ denotes the projection operator of interest.

Choose coherent states given (for $\hbar = 1$) by $|\mathbf{p}, \mathbf{q}\rangle = e^{-i\,\mathbf{q}\cdot\mathbf{P}}\, e^{i\,\mathbf{p}\cdot\mathbf{Q}}\,|0\rangle$, with an overlap

$$\langle \mathbf{p}', \mathbf{q}' | \mathbf{p}, \mathbf{q}\rangle = e^{i\,\frac{1}{2}(\mathbf{p}'+\mathbf{p})\cdot(\mathbf{q}'-\mathbf{q})}\, e^{-\frac{1}{4}[(\mathbf{p}'-\mathbf{p})^2 + (\mathbf{q}'-\mathbf{q})^2]}\ ,$$

where $\mathbf{p} = \{p_j\}_{j=1}^3$ and $\mathbf{q} = \{q_j\}_{j=1}^3$.

Show that the coherent state matrix elements of the desired projection operator are given by

$$K_E(\mathbf{p}', \mathbf{q}'; \mathbf{p}, \mathbf{q}) \equiv \langle \mathbf{p}', \mathbf{q}' | \mathbb{E} | \mathbf{p}, \mathbf{q}\rangle = e^{i\,(\mathbf{p}'\cdot\mathbf{q}' - \mathbf{p}\cdot\mathbf{q})/2}$$
$$\times e^{-\frac{1}{4}(\mathbf{p}'^2 + \mathbf{q}'^2 + \mathbf{p}^2 + \mathbf{q}^2)}\, \frac{\sinh[\sqrt{(\mathbf{q}'-i\mathbf{p})^2\,(\mathbf{q}+i\mathbf{p})^2}/2]}{[\sqrt{(\mathbf{q}'-i\mathbf{p})^2\,(\mathbf{q}+i\mathbf{p})^2}/2]}\ .$$

9-2 For two irregular constraints $\Phi_1 = P^3$ and $\Phi_2 = Q^4$, show that the appropriate projection operator is given by

$$\mathbb{E} = \mathbb{E}(P^6 + Q^8 \leq c_0'\,\hbar^{24/7})\ ,$$

with the given dependence on $\hbar$, and for a suitable choice of $c_0' > 0$.

9-3 Consider the reproducing kernel (RK) defined by $\langle p', q' | \mathbb{E} | p, q\rangle / \langle 0 | \mathbb{E} | 0\rangle$. Evaluate the RK for $\mathbb{E} = \mathbb{E}(P^2 + \alpha^2 Q^2 \leq \alpha\hbar)$ as a function of α. Take the limit as $\alpha \to 0$, and show that the RK for the *second-class* constraint system of two constraints $P = 0$ and $\alpha Q = 0$ passes smoothly to the RK for the *first-class* constraint system for the single constraint $P = 0$.

Part III

Quantum Field Theory

10

Application to Quantum Field Theory

10.1 Introduction and Overview

In the present chapter we extend the basic ideas of quantum mechanics and elementary path integral representations of that quantum mechanics to a limited discussion of contemporary quantum field theory, namely, scalar fields. Initial emphasis is placed on understanding the free relativistic quantum field theory in Minkowski and Euclidean spacetimes with dimensions $n \geq 2$. Interacting theories are examined from the point of view of perturbation theory about the free theory, and a brief discussion of their nonperturbative analysis by lattice regularized Euclidean spacetime is offered.

10.1.1 Classical preliminaries

Classical field theory is a theory of infinitely many degrees of freedom, and thus it stands to reason that quantum field theory is also a theory of infinitely many degrees of freedom. The subject of quantum field theory is vast and varied, and this chapter does not pretend to cover the subject in detail; from a conventional viewpoint, there are many excellent books that can serve that purpose [IZ80]. Instead we focus on scalar fields, which are in some way the least complicated examples, and which still have a good number of mathematically interesting issues that need to be carefully dealt with. Thus our discussion is not intended to prepare the reader to cope with the issues that arise in particle physics due to the lack of any discussion of higher spin fields, including both bosons and fermions. Nevertheless, apart from sometimes rather involved technical issues, the principal problems that one encounters in general quantum field theory are mostly already present in a discussion of scalar fields.

For continuum field theory, by far the most important examples are relativistically covariant, and we will focus on just two such models: the free field theory and the model with a quartic self-interaction. We will discuss the

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9_10,

perturbative analysis of this model as a function of the spacetime dimension n, and show how that analysis dramatically changes as a function of n.

10.2 Relativistic Free Fields

10.2.1 A brief survey of classical and quantum properties

We focus on a scalar field $\phi(x)$ where $x = (t, \mathbf{x})$ (in units where the speed of light $c = 1$) in an n-dimensional spacetime $\mathbb{R}^n = \mathbb{R} \times \mathbb{R}^s$, comprising a one-dimensional time and an s-dimensional space ($s = n-1$). To describe the dynamics of this field, we adopt the free field action functional

$$\begin{aligned} I &= \tfrac{1}{2}\int \{\, [\partial_\mu \phi(x)]^2 - m^2 \phi(x)^2 \,\}\, d^n x \ , \\ &= \tfrac{1}{2}\int \{\, [\partial\phi(x)/\partial t]^2 - [\partial\phi(x)/\partial x^k]^2 - m^2 \phi(x)^2 \,\}\, d^n x \ , \end{aligned}$$

with a summation on repeated indices understood. Stationary variation of the action leads to the equation of motion

$$\left[\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^{k\,2}}\right]\phi(x) + m^2\,\phi(x) \equiv \Box\,\phi(x) + m^2\,\phi(x) = 0 \ .$$

In the Hamiltonian formulation one keeps the field $\phi(x)$ and introduces the canonical momentum $\pi(x) = \dot{\phi}(x)$. Classically these fields obey the equal-time Poisson bracket

$$\{\, \phi(t, \mathbf{x}), \pi(t, \mathbf{y}) \,\} = \delta(\mathbf{x} - \mathbf{y}) \ ,$$

and quantum mechanically the corresponding operators obey the equal-time commutation relation

$$[\varphi(t, \mathbf{x}), \pi(t, \mathbf{y})] = i\hbar\delta(\mathbf{x} - \mathbf{y}) \ ,$$

with a factor of the identity operator implicit on the right-hand side. The dynamics of the relativistic free field is described by a proper choice of the Hamiltonian. This is already true in the classical theory and it is no less true in the quantum theory. In the classical theory, the Hamiltonian for the relativistic free field of mass m is given by

$$\begin{aligned} H &= \tfrac{1}{2}\int \{[\pi(\mathbf{x})]^2 + [\boldsymbol{\nabla}\phi(\mathbf{x})]^2 + m^2[\phi(\mathbf{x})]^2\}\, d^s x \\ &\equiv \int H(\mathbf{x})\, d^s x \ . \end{aligned}$$

Correspondingly, the quantum Hamiltonian operator is taken as

$$\begin{aligned} \mathcal{H} &= \tfrac{1}{2}\int :\{[\pi(\mathbf{x})]^2 + [\boldsymbol{\nabla}\varphi(\mathbf{x})]^2 + m^2[\varphi(\mathbf{x})]^2\}: \, d^s x \\ &\equiv \int \mathcal{H}(\mathbf{x})\, d^s x \ , \end{aligned}$$

where $:(\cdot):$ denotes normal ordering, which is further discussed below. The indicated expression for $\mathcal{H}$ is quite formal due to the fact that the field operators at points are not genuine operators and so their square is ill-defined. Nevertheless, the specifically indicated sum of the several ill-defined expressions leads to a meaningful operator that may, alternatively and properly, be defined as

$$\mathcal{H} = \int a(\mathbf{k})^\dagger \sqrt{\mathbf{p}^2 + m^2}\, a(\mathbf{k})\, d^s k \ ,$$

where we have used $\mathbf{p} \equiv \hbar\, \mathbf{k}$. Thus we deal with an operator for which

$$\hbar^2 \omega(\mathbf{k})^2 \equiv \mathbf{p}^2 + m^2 \ .$$

Here we have a dispersion relation appropriate to a relativistic system of mass m; this relation is exactly the familiar relativistic dispersion law $E^2 = \mathbf{p}^2 + m^2$ in units where the velocity of light $c = 1$.

Hereafter, for the remainder of this chapter, we choose units such that $\hbar = 1$.

Observe that the relativistic choice of Hamiltonian is not only translationally invariant but it is rotationally invariant as well. As a consequence, we have the Poisson bracket relation $\{J_{ab}, H\} = 0$ as well as the operator commutation relation $[\mathcal{J}_{ab}, \mathcal{H}] = 0$ for all pairs a, b, $a \neq b$, where the expressions J_{ab} and $\mathcal{J}_{ab}$ represent, respectively, the classical and quantum rotation generators.

In the case of a relativistic theory there are additional generators of interest that relate to *boosts*, namely, the *relativity transformations*. Classically, the generator is a vector quantity defined by

$$K_a = -\int [\pi(\mathbf{x})\, t\, \partial_a\, \phi(\mathbf{x}) + x_a\, H(\mathbf{x})]\, d^s x \ .$$

Correspondingly, the quantum operator is given by

$$\mathcal{K}_a = -\int :[\pi(\mathbf{x})\, t\, \partial_a\, \varphi(\mathbf{x}) + x_a\, \mathcal{H}(\mathbf{x})]: \ d^s x \ .$$

It is straightforward to verify that as Poisson brackets $\{P_l, K_m\} = \delta_{lm} H$ and $\{H, K_l\} = P_l$, as well as operator commutators $[\mathcal{P}_l, \mathcal{K}_m] = i\delta_{lm} \mathcal{H}$ and $[\mathcal{H}, \mathcal{K}_l] = i\mathcal{P}_l$, as required for the Lie algebra generators of the Poincaré group [IZ80]. These relations effectively establish that the example under discussion refers to a relativistic free field.

Based on direct analogies with harmonic oscillators with finitely many degrees of freedom, we are able to immediately offer a variety of expectation values. For example, in terms of the fields φ and π smeared at any sharp time, e.g., $t = 0$, it follows for a relativistic free field of mass m that

$$\begin{aligned} &\langle 0|\ \exp\{i[\varphi(f) - \pi(g)]\}\, |0\rangle \\ &\qquad = \exp\{-(1/4)\textstyle\int [\omega(\mathbf{k})^{-1} |\tilde{f}(\mathbf{k})|^2 + \omega(\mathbf{k}) |\tilde{g}(\mathbf{k})|^2]\, d^s k\} \ , \end{aligned}$$

where $|0\rangle$ denotes the normalized, unique ground state of the free field Hamiltonian operator, $\mathcal{H}|0\rangle = 0$.

As an extension of the foregoing story, let us introduce the spacetime smeared field

$$\varphi(h) \equiv \int h(x)\varphi(x)\, d^n x \ ,$$

for which it follows that

$$\langle 0|\exp[i\varphi(h)]|0\rangle = \exp\{-(1/2)\int h(x)\, \Delta_1(x-y)\, h(y)\, d^n x\, d^n y\} \ ,$$

where

$$\begin{aligned}\Delta_1(x-y) &\equiv \frac{1}{2}\langle 0|[\varphi(x)\varphi(y)+\varphi(y)\varphi(x)]|0\rangle \\ &= \frac{1}{(2\pi)^s}\int e^{i\,\mathbf{k}\,\cdot\,(\mathbf{x}-\mathbf{y})}\ \cos\left[\omega(\mathbf{k})(t_x-t_y)\right]\frac{d^s k}{2\omega(\mathbf{k})} \\ &= \frac{1}{2(2\pi)^s}\int e^{ik(x-y)}\delta(p^2-m^2)\, d^n k \ .\end{aligned}$$

In Fourier space this expectation value reads

$$\langle 0|\exp[i\varphi(h)]|0\rangle = \exp\{-(\pi/2)\int|\tilde{h}(k)|^2\delta(p^2-m^2)\, d^n k\,\} \ ,$$

which implies that unless the test function h contains energy and momentum values "on the mass shell," i.e., that satisfy the relativistic dispersion relation, the expectation value in question is unity.

Another expectation value of interest involves the time-ordering operator T and is given by

$$\langle 0|\mathsf{T}\exp[i\varphi(h)]|0\rangle = \exp\{(i/2)\int h(x)\, \Delta_F(x-y)\, h(y)\, d^n x\, d^n y\} \ ,$$

where F stands for "Feynman," and

$$\begin{aligned}\Delta_F(x-y) &\equiv i\langle 0|\mathsf{T}\varphi(x)\varphi(y)|0\rangle \\ &= -\lim_{\epsilon\to 0}\frac{1}{(2\pi)^n}\int\frac{e^{ik(x-y)}}{p^2-m^2+i\epsilon}\, d^n k \ .\end{aligned}$$

Here ϵ is a small positive parameter that orients the contour of integration properly around the poles in the complex $k_0(=p_0)$ plane, and the limit $\epsilon \to 0^+$ is taken at the end of the calculation.

Expressions involving ϵ in the manner indicated above may be written in an alternative fashion as well. In particular, for $u \in \mathbb{R}$, we observe that

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{1}{u+i\epsilon} &= \lim_{\epsilon\to 0}\frac{u-i\epsilon}{u^2+\epsilon^2} \\ &= \mathcal{P}\frac{1}{u} - i\pi\delta(u) \ .\end{aligned}$$

Here $\mathcal{P}$ denotes the *principal value* integral, while the second term is based on the fact that $\int [\epsilon/(u^2+\epsilon^2)]\, du = \pi$ identically for all $\epsilon > 0$.

If we put this relation to use for us in Fourier space, then the time-ordered vacuum expectation value reads

$$\begin{aligned}\langle 0|\mathsf{T}\exp[i\varphi(h)]\,|0\rangle &= \lim_{\epsilon\to 0}\exp\left\{-\frac{i}{2}\int \frac{|\tilde{h}(k)|^2}{p^2-m^2+i\epsilon}\,d^nk\right\}\\ &= \exp\left\{-\frac{i}{2}\int [\mathcal{P}\frac{1}{p^2-m^2} - i\pi\delta(p^2-m^2)]|\tilde{h}(k)|^2\,d^nk\right\}.\end{aligned}$$

This result contains two terms: (1) a phase term that exists, in general, for any nonvanishing test function h, and (2) a damping term that arises only provided the test function h contains energy and momenta on the mass shell.

It is important to note that what is here referred to from a mathematical point of view as a *test function* may also be called from a physical point of view an *external source*. In particular, if the external source contains the appropriate values of energy and momenta, then it is possible for the source to create particles out of the vacuum. If that is the case, then the probability amplitude for the new state which includes a number of such created particles has an overlap with the vacuum that is less than unity. Indeed, the absolute square of the probability amplitude $\langle 0|\mathsf{T}\exp[i\varphi(h)]|0\rangle$ is the probability to find the vacuum in the new state. On the other hand, if the external source does *not* contain energy and momenta in the right relationship, then the probability amplitude is simply a phase factor with an absolute magnitude squared of unity showing that the probability to make real particles from the vacuum is zero whenever the correct dispersion relation is not satisfied by the external source.

10.3 Functional Integral Formulation

Let us follow the discussion for path integrals presented in Chapter 6, and generalize that argument to the case of scalar free fields. First recall that for finitely many degrees of freedom, the formal path integral in the presence of an external source may be given by

$$\begin{aligned}&\langle 0|\mathsf{T}\exp[i\textstyle\int \Sigma\, h_k(t)Q_k(t)\,dt]|0\rangle\\ &\qquad = \lim_{\epsilon\to 0}\mathcal{N}\int \exp\big(i\textstyle\int\{\Sigma\, h_k(t)q_k(t) + \frac{1}{2}\Sigma\,[\dot{q}_k(t)^2 + i\epsilon q_k(t)^2]\\ &\qquad\qquad -V(q(t))\}\,dt\big)\,\Pi\,\mathcal{D}q_k\,,\end{aligned}$$

where the time integral in the exponent runs from $-\infty$ to ∞ and as boundary conditions we choose $\lim_{|t|\to\infty} q_k(t) = 0$. Thanks to this infinite integration range and to the presence of a small damping factor proportional to ϵ, all final states save for the ground state are projected out. The limit $\epsilon \to 0$ is reserved

for the final step. To reduce the number of symbols, the presence of both the small damping factor as well as the ultimate limit that $\epsilon \to 0$ are frequently left implicit, and we shall adopt such a notational shorthand as well. From a formal point of view the extension of the path integral given above to the case of infinitely many degrees of freedom leads to exactly the same expression as for finitely many degrees of freedom! Of course, a limit is involved, and that limit may exist only under special conditions. For example, it may be enough to ensure that the coordinates of the external source h_k fall to zero sufficiently fast as $k \to \infty$. In that case the existence of the limit characterizes the allowed test sequences for the problem at hand. With this discussion in mind, we may simply reinterpret the formal path integral above as applying to the case of infinitely many degrees of freedom as well. Just as one is able to trade infinitely many coordinates for fields in the classical theory and in the operator formulation of quantum field theory, we may do so again within the path integral; in a certain sense, such a trade amounts to little more than a formal change of variables within an integration. Since the expression in the exponent of the formal path integral represents the classical action integral as the integral of the Lagrangian, we can—maintaining the same level of formal manipulations—pass from one expression of the classical action to another. In the present case we can pass to fields expressed in either configuration space or to fields expressed in momentum space, and for the sake of completeness we shall illustrate both cases. Let us begin with free fields.

Initially, we consider a general dispersion relation $\Omega(\mathbf{k})^2$, and to ease the transition to fields, we initially adopt a particle-like language and concentrate on the expression

$$\begin{aligned}
&\langle 0|\mathsf{T}\exp[i\textstyle\int h(x)\varphi(x)\,d^n x]|0\rangle \\
&\quad = \langle 0|\mathsf{T}\exp\{i\textstyle\int[\tilde{h}(t,\mathbf{k})^* e^{-i\Omega(\mathbf{k})t}a(\mathbf{k}) \\
&\qquad\qquad +\tilde{h}(t,\mathbf{k})e^{i\Omega(\mathbf{k})t}a(\mathbf{k})^\dagger\,]\,\frac{d^s k\,dt}{\sqrt{2\Omega(\mathbf{k})}}\,\}|0\rangle \\
&\quad = \mathcal{N}\int \exp\{i\textstyle\int[\tilde{h}(t,\mathbf{k})^* q(t,\mathbf{k}) \\
&\qquad\qquad +\tfrac{1}{2}|\dot{q}(t,\mathbf{k})|^2 - \tfrac{1}{2}\Omega(\mathbf{k})^2|q(t,\mathbf{k})|^2\,]\,d^s k\,dt\}\,\mathcal{D}q\,.
\end{aligned}$$

This formal expression is an immediate consequence of the corresponding path integral for a single degree of freedom and the fact that for each $\mathbf{k}$ value, the field at $\mathbf{k}$ is independent of the field for all other $\mathbf{k}$ values. Let us reexpress this final result in the language of space and time. In that case, and using the field notation, a Fourier transformation leads directly to the formal expression

$$\begin{aligned}
&\langle 0|\mathsf{T}\exp[i\textstyle\int h(x)\varphi(x)\,d^n x]\,|0\rangle \\
&\quad = \mathcal{N}\int \exp\{i\textstyle\int[h(t,\mathbf{x})\phi(t,\mathbf{x}) + \tfrac{1}{2}\dot{\phi}(t,\mathbf{x})^2]\,d^s x\,dt \\
&\qquad\qquad -\tfrac{1}{2}i\textstyle\int\phi(t,\mathbf{x})\,W(\mathbf{x}-\mathbf{y})\,\phi(t,\mathbf{y})\,d^s x\,d^s y\,dt\}\,\mathcal{D}\phi\,.
\end{aligned}$$

In particular, the relativistic free particle of mass m is such that $\Omega(\mathbf{k})^2 = \omega(\mathbf{k})^2 = \mathbf{k}^2 + m^2$, and it may be characterized by the functional integral (including the $i\epsilon$ factor but leaving the limit $\epsilon \to 0$ implicit)

$$\langle 0|\mathsf{T}\exp[i\textstyle\int h(x)\varphi(x)\,d^nx]\,|0\rangle$$
$$= \mathcal{N}\int \exp\big(i\textstyle\int\{h(x)\phi(x) + \frac{1}{2}[\partial_\mu\phi(x)]^2 - \frac{1}{2}(m^2 - i\epsilon)\phi(x)^2\}\,d^nx\,\big)\,\mathcal{D}\phi\ ,$$

where the field is integrated over all spacetime with the boundary condition that $\phi(x)$ vanishes at both spatial and temporal infinity. Although formal, this is an expression of fundamental importance.

Recall that we have discussed the general N-dimensional Gaussian integral in Sec. 6.1.3 leading to the result that

$$\int e^{i\Sigma b_k y_k - \frac{1}{2}\Sigma y_k A_{kl} y_l}\ \Pi\, dy_k = \frac{(2\pi)^{N/2}}{\sqrt{\det(A)}} e^{-\frac{1}{2}\Sigma b_k A^{(-1)}_{kl} b_l}$$

provided, of course, that $A = \{A_{kl}\}$ is a symmetric $N \times N$ matrix such that the integral converges, a condition which also ensures that A has an inverse $A^{(-1)}$ defined so that $A^{(-1)}A = 1$. Observe that the final form of the exponent, $-\frac{1}{2}\Sigma b_k A^{(-1)}_{kl} b_l$, may be determined from the exponent of the integrand, $i\Sigma b_k y_k - \frac{1}{2}\Sigma y_k A_{kl} y_l$, evaluated at the value of y_l corresponding to the extremal point, namely, for $ib_k = \Sigma A_{kl} y_l$. This is an extremely useful method to determine the exponent in a Gaussian integral.

We can immediately use this result to evaluate the functional integral for the relativistic free field of mass m. The result will be the exponential of a quadratic form that is given by the extremal value of the exponent of the integrand. That extremal occurs when

$$[\Box + (m^2 - i\epsilon)]\phi(x) = h(x)\ ,$$

where we have retained the necessary factor $(i\epsilon)$ in order for the integral to have converged in the first place. The solution to this equation consistent with its vanishing at infinity is

$$\begin{aligned}\phi(x) &= \int \frac{1}{[\Box + (m^2 - i\epsilon)]}\,\delta(x-y)\,h(y)\,d^ny \\ &= \int\left\{-\frac{1}{(2\pi)^n}\int \frac{e^{ip(x-y)}}{p^2 - m^2 + i\epsilon}\,d^np\right\} h(y)\,d^ny \\ &\equiv \int \Delta_F(x-y)\,h(y)\,d^ny\ .\end{aligned}$$

Substitution of this extremal ϕ into the exponent of the integrand leads to the result that

$$\langle 0|\mathsf{T}\exp[i\textstyle\int h(x)\varphi(x)\,d^nx]\,|0\rangle$$
$$= \exp\{(i/2)\textstyle\int h(x)\,\Delta_F(x-y)\,h(y)\,d^nx\,d^ny\}\ ,$$

where the overall normalization has been fixed by the required value of unity when $h \equiv 0$. This result confirms the one previously obtained for the generating function of the time-ordered vacuum expectation values.

Up to this point our discussion has been a purely formal one and it is important to understand under what circumstances it is permissible to argue in strictly formal terms—and also when it is not. In the present case there is, underlying our formal discussion, a fully satisfactory proper mathematical meaning to the manipulations involved. Let us focus on the example of the relativistic free particle since the discussion in any other case is entirely analogous. To give meaning to a formal functional integral of the kind under consideration it is necessary to employ some form of *regularization* designed to give meaning to an otherwise ill-defined expression. There are a great many ways to do so, but for the present let us focus on the so-called *lattice regularization* wherein one replaces the continuum of spacetime by a (hyper)cubic lattice with a uniform lattice spacing $a > 0$. Initially, we also assume that the size of the entire lattice is finite so that we have in effect replaced the field theory with its infinite number of degrees of freedom by a system with a finite number, say N, of degrees of freedom. For this system with finitely many degrees of freedom, we can formulate the functional integral as an ordinary integral for N variables. Ultimately, the answer of physical interest should be given by a limit in which $N \to \infty$. Let us see how this works out in practice for the relativistic free field.

As our regularized form for the path integral we consider

$$C_N \int \exp\{i[\Sigma_k\, h_k \phi_k a^n + \tfrac{1}{2}\Sigma_k\, (\phi_{k^\#} - \phi_k)^2 a^{n-2} \\ -\tfrac{1}{2}\Sigma_k\, (\phi_{k^*} - \phi_k)^2 a^{n-2} - \tfrac{1}{2}(m^2 - i\epsilon)\Sigma_k\, \phi_k^2 a^n]\,\}\, \Pi_k\, d\phi_k\ .$$

There are a number of notational points to clarify in this expression: C_N denotes a normalization factor adjusted so that the result is unity if $h_k = 0$ for all k; a^n represents the volume of a unit cell on the lattice; k denotes the label of a lattice site, i.e., $k = (k^0, k^1, \ldots, k^s)$, $k^j \in \mathbb{Z} \equiv \{0, \pm 1, \pm 2, \ldots\}$; $k^\# = (k^0 + 1, k^1, \ldots, k^s)$, namely, the next future point in time, while, analogously, k^* denotes each one of the s nearest neighbors in the (positive) spatial directions, $k^1, \ldots, k^s$. The sums run over finitely many lattice sites and (implicitly over) nearest neighbors as needed with (say) periodic boundary conditions. It is clear that the exponent of the integrand is nothing other than a Riemann sum approximation to the integral expression that holds in the continuum, and in which, for example, h_k denotes the average value of the test function h over the lattice cell centered at the lattice site k, and correspondingly for ϕ_k. Clearly, with a lattice regularization present, only a finite number (N) of ordinary integrals are involved, and the desired answer for the generating functional of the time-ordered vacuum expectation values is formally given as the limit $N \to \infty$, at least in principle. For the case of the free field, the result of those N integrations can be readily determined by the general

formula for Gaussian integrals, and so the answer in the continuum limit is obtained by taking the limit $N \to \infty$ of the expression obtained after doing the Gaussian integrations. In turn, the expression that results from this operation is recognized as the same as that obtained by directly solving for the continuum form of the exponent as we have already done, and this observation provides a justification for the procedure outlined above for determining the result of the functional integral.

10.4 Euclidean-Space Functional Integral Formulation

An appropriate analytic continuation in the "time" variable takes one from Minkowski to Euclidean space. In so doing, the integrand of the functional integral changes from an oscillatory to a positive one, but even for this real functional integral it is generally necessary to still introduce a lattice regularized formulation. Thus we may as well go directly from some form of regularization such as the Minkowski lattice formulation to the Euclidean lattice formulation by a simple analytic procedure. Let one of the lattice spacing "a" terms in the lattice formulation be identified as referring to the time direction, and analytically continue *that* a (and not any other) from a to $-ia$, a rotation of $-\pi/2$ in the complex plane. This has the effect of changing the Minkowski lattice formulation for the free theory into the Euclidean expression

$$M_N \int \exp[\Sigma_k h_k \phi_k a^n - \tfrac{1}{2}\Sigma_k (\phi_{k^*} - \phi_k)^2 a^{n-2} - \tfrac{1}{2} m^2 \Sigma_k \phi_k^2 a^n] \, \Pi_k \, d\phi_k \ .$$

Note that for the *time* increment, the (temporal) a in question is in the *denominator*, while for the *space* increments, the (temporal) a in question lies in the *numerator*. Thus the time- and space-increment terms respond differently to the complex rotation of the variable. In the expression for the integrand, the implicit sum over k^* covers all n dimensions, i.e., both the former time direction as well as all s spatial directions, and they all enter with a common sign characteristic of a Euclidean (rather than a Minkowski) metric. This is a very important formulation for the relativistic free field, and we shall soon extend it to interacting field models as well. If we formally take the continuum limit and interchange it with the multidimensional integrations, then we are led to a formal functional integral in the case of the Euclidean free field theory given by

$$S\{h\} = \mathcal{N} \int \exp\Big(\int \{h(x)\phi(x) - \tfrac{1}{2}[(\nabla\phi)(x)^2 + m^2 \phi(x)^2]\} \, d^n x \Big) \, \mathcal{D}\phi \ .$$

Here we have begun to use the symbol $\mathcal{N}$ as a formal normalization constant; although its specific value may actually change from one equation to another, the manner of its definition will generally be clear from the context. The formal expression $S\{h\}$ may be regarded in either of two ways. On the one

hand, it may be viewed as just a formal shorthand notation for the lattice regularized expression followed by the continuum limit. On the other hand, there is in fact a probability measure μ_F (F for "free") on distribution-valued fields such that

$$S\{h\} = \int e^{\phi(h)}\, d\mu_F(\phi)$$

holds in a well-defined mathematical sense [GV64, Sko74]. In the Euclidean case under discussion, the measure μ_F is a proper measure (i.e., countably additive) unlike the case for the Minkowski functional integral where the "measure" that one often assumes is not a proper measure (namely, it is only finitely additive). As such, it is important to observe that the Euclidean situation offers a level of mathematical characterization that is unavailable in the Minkowski formulation, namely, that of a genuine *stochastic process*. Let us. denote the ensemble average of some quantity $(\cdot)$ by $\langle(\cdot)\rangle \equiv \int(\cdot)\, d\mu_F$, and let us use the same symbol ϕ to denote the stochastic variable. In that case the Euclidean-space moment generating functional may be written in the form

$$\begin{aligned} S(h) &= \langle\, e^{\phi(h)}\,\rangle = \exp\{\tfrac{1}{2}\langle\,[\phi(h)]^2\,\rangle\} \\ &= \exp\{\tfrac{1}{2}\textstyle\int h(x)\,\langle\phi(x)\phi(y)\rangle\, h(y)\, d^n\!x\, d^n\!y\,\}\ . \end{aligned}$$

Consequently, the entire expression is determined completely by the two-point correlation function $\langle\phi(x)\phi(y)\rangle$, just as the case for any mean-zero Gaussian random variable.

The form of the Euclidean-space, free field, two-point function is given by the integral expression

$$C(x-y) \equiv \langle\,\phi(x)\phi(y)\,\rangle = \frac{1}{(2\pi)^n}\int \frac{e^{ip\,\cdot\,(x\,-\,y)}}{p^2+m^2}\, d^n\!p\ .$$

This integral may be explicitly evaluated in terms of modified Bessel functions [JR80]. However, we are not especially interested in the exact functional form of this correlation function; rather we are interested in *the behavior of* $C(x-y)$ *when* x *and* y *are close to one another*. In our units the distance $|x-y|$ is small when $m|x-y| \ll 1$, in which case, when $n \geq 3$, we may use the approximation that

$$\begin{aligned} C(x-y) \simeq C_0(x-y) \ &\equiv \frac{1}{(2\pi)^n}\int \frac{e^{ip\,\cdot\,(x\,-\,y)}}{p^2}\, d^n\!p \\ &\propto |x-y|^{-(n-2)}\,, \qquad n \geq 3\ , \end{aligned}$$

where the indicated functional dependence for small $|x-y|$ is essentially determined simply on dimensional grounds, or equivalently, by a simple change of variables. This short-distance behavior is characteristic of all relativistic free fields in dimensions $n \geq 3$ and is independent of the mass m. In two spacetime dimensions ($n = 2$) the mass parameter cannot be eliminated from

the denominator for otherwise the integral would diverge at zero momentum as well as infinite momentum. The short-distance behavior in the case of two spacetime dimensions involves the mass m and (with $p \equiv mr$) is given by

$$\begin{aligned} C(x-y) &= \frac{1}{(2\pi)^2} \int \frac{e^{imr\cdot(x-y)}}{r^2+1}\, d^2r \\ &\propto -\ln(m|x-y|)\,, \qquad n=2\,. \end{aligned}$$

The short-distance behavior for the two-point correlation functions that we have arrived at will be of great significance as we turn our attention to the construction of local powers of the field. We shall find the proper proportionality constant for each of the cases later.

10.4.1 Local products

As a preliminary to the study of additional scalar field models we take up the question of *local powers* of the stochastic variable $\phi(x)$. A brief argument will convince us that we cannot use the conventional notion of local power. First of all, we note that for any $n \geq 2$ it already follows that $\langle \phi(x)^2 \rangle = C(0) = \infty$. For some random variables this condition may simply imply the divergence of a moment, but for a mean-zero Gaussian variable it actually implies that $\phi(x)^2 = \infty$ almost everywhere. A proof of that statement runs as follows: let X denote a mean-zero Gaussian random variable. Then

$$\begin{aligned} \langle e^{-\frac{1}{2}X^2} \rangle &= \frac{1}{(2\pi)^{1/2}} \int e^{-\frac{1}{2}t^2} \langle e^{itX} \rangle\, dt \\ &= \frac{1}{(2\pi)^{1/2}} \int e^{-\frac{1}{2}(1+\langle X^2\rangle)t^2}\, dt = \frac{1}{\sqrt{1+\langle X^2\rangle}}\,. \end{aligned}$$

Thus, if $\langle X^2 \rangle = \infty$, it follows that $\langle \exp[-\frac{1}{2}X^2] \rangle = 0$, which can only hold provided $X^2 = \infty$ almost everywhere.

Let us instead focus on *normally-ordered local products*, which, because of the similarity with the definition in the case of operator fields, we denote by $:\phi(x)^p:$, $p \in \{1,2,3,\ldots\}$, and which we formally define by the generating function

$$\begin{aligned} :e^{c\phi(x)}: &\equiv e^{c\phi(x)} / \langle e^{c\phi(x)} \rangle \\ &= e^{c\phi(x) - \frac{1}{2}c^2\langle\phi(x)^2\rangle}\,. \end{aligned}$$

Normal-ordered field powers formally follow by expansion of this expression in a power series in the parameter c. For example, it formally follows that

$$\begin{aligned} :\phi(x): &= \phi(x)\,, \\ :\phi(x)^2: &= \phi(x)^2 - \langle\phi(x)^2\rangle\,, \\ :\phi(x)^3: &= \phi(x)^3 - 3\phi(x)\langle\phi(x)^2\rangle\,, \\ :\phi(x)^4: &= \phi(x)^4 - 6\phi(x)^2\langle\phi(x)^2\rangle + 3\langle\phi(x)^2\rangle^2\,, \end{aligned}$$

etc. In these expressions observe that $\langle \phi(x)^2 \rangle = C(0) = \langle \phi(0)^2 \rangle$ is a divergent constant that is independent of the point x. To see that such expressions make sense as distributions we first observe—without proof for the moment—that

$$\langle [\int h(x) : \phi(x)^p : \, d^n x]^2 \rangle = p! \int h(x) \, \langle \phi(x)\phi(y) \rangle^p \, h(y) \, d^n x \, d^n y \; .$$

This expression converges at infinity because we are dealing with test functions that fall off at infinity as fast as we like. On the other hand, convergence (or lack of it) when $x \approx y$ is quite another matter. For $n = 2$ the short-distance behavior has a logarithmic divergence, $-\ln(m|x-y|)$, and so we may conclude that

$$\int_{m|x-y|<1} [-\ln(m|x-y|)]^p \, d^n x < \infty$$

holds for all $p \geq 0$. For $n \geq 3$, the convergence depends on a specific relationship of p and n. Since the short-distance behavior of the correlation function behaves as $|x-y|^{-(n-2)}$, we will have convergence when $x \approx y$ provided that

$$\int_{m|x-y|<1} \frac{1}{|x-y|^{p(n-2)}} \, d^n x < \infty \; ,$$

and this criterion demands that $p(n-2) < n$. That is, we have convergence whenever

$$p < \frac{n}{n-2} \; ,$$

a relation that holds for all $n \geq 2$, assuming the proper interpretation in the case $n = 2$. The finiteness of such a moment implies that the stochastic variable

$$: \phi^p : (h) \equiv \int h(x) : \phi(x)^p : \, d^n x \; , \qquad p < n/(n-2) \; ,$$

is almost everywhere finite. Therefore, $: \phi^p : (h)$ is necessarily well-defined almost everywhere. These normal-ordered powers will serve as the local powers, especially when we seek to go beyond free theories.

We now return to the proof that the second moment of the stochastic variable $: \phi^p : (h)$ has the indicated form. The basic arguments in the proof are already present in the simplest case where $p = 2$. Because the stochastic process under consideration is Gaussian with a zero mean, it follows that

$$\begin{aligned} \langle \phi(u)\phi(v)\phi(x)\phi(y) \rangle \, &= \langle \phi(u)\phi(v) \rangle \langle \phi(x)\phi(y) \rangle \\ &\quad + \langle \phi(u)\phi(x) \rangle \langle \phi(v)\phi(y) \rangle + \langle \phi(u)\phi(y) \rangle \langle \phi(v)\phi(x) \rangle \; . \end{aligned}$$

Now we specialize this expression by setting $u = x$ and $v = y$. Thus we are led to the relation that

$$\langle \phi(x)^2 \phi(y)^2 \rangle = 2 \langle \phi(x)\phi(y) \rangle^2 + \langle \phi(x)\phi(x) \rangle^2 \; .$$

The term on the left-hand side as well as the last term on the right-hand side are divergent, and we now move the two divergent terms to the same (left-hand) side of the equation to learn that

$$\langle : \phi(x)^2 :: \phi(y)^2 : \rangle = \langle \phi(x)^2 \phi(y)^2 \rangle - \langle \phi(x)\phi(x) \rangle^2 = 2 \langle \phi(x)\phi(y) \rangle^2 \ .$$

In like manner it follows that all divergences from coinciding points are subtracted off in forming the higher order, normally ordered local powers, thus confirming the quoted relation for the moments.

10.5 Interacting Scalar Fields

As a typical example of an interacting scalar quantum field theory let us focus on the formal functional integral given by

$$\mathcal{N} \int \exp\Big(\int \{ h(x)\phi(x) - \tfrac{1}{2}[\nabla \phi(x)]^2 - \tfrac{1}{2} m^2 \phi(x)^2 - g : \phi(x)^4 : \} \, d^n x \Big) \, \mathcal{D}\phi \ .$$

Let us set about trying to give some mathematical meaning to this formal expression. The constant term that is part of the normal ordering is already included in the overall normalization factor $\mathcal{N}$, which may always be determined at the end of the calculation by an appropriate normalization condition.

If we wish to give better meaning to the formal functional integral we may invoke the stochastic process that describes the free field of mass m. Once again we denote by μ_F the measure on fields that describes the free field. Thus a somewhat better formulation of the formal path integral is given by

$$\mathcal{N} \int \exp\{ \int [h(x)\phi(x) - g : \phi(x)^4 :] \, d^n x \} \, d\mu_F(\phi) \ .$$

This expression is still formal and thus unsatisfactory, however, because the local density $: \phi(x)^4 :$ leads to a stochastic variable only when integrated against a test function and, in particular, not the function "one" for all spacetime. A closer look at the test function criterion reveals that the function "one" defined over a compact region, and zero elsewhere, would be an acceptable test function in the present case. Thus we modify our formal path integral once again to read

$$S(h) = \lim_{\Lambda \to \infty} C_\Lambda \int \exp\{ \textstyle\int_\Lambda [h(x)\phi(x) - g : \phi(x)^4 :] \, d^n x \} \, d\mu_F(\phi) \ ,$$

where by Λ we mean a finite volume of spacetime which we can take to be a (hyper)cube centered on the origin of coordinates and by $\Lambda \to \infty$ we mean that the volume approaches infinity in a natural and symmetric way. Actually, this formulation is entirely satisfactory whenever the locally integrated normal-ordered power is a proper stochastic variable, and that situation occurs, as

derived above, whenever $p < n/(n-2)$, or specifically for $p = 4$ when $n < 8/3$, i.e., $n = 2$. Thus the formulation of the Euclidean φ_2^4 field theory offered above is a perfectly acceptable one, and the theory has been successfully analyzed from this starting point [GJ87, FFS92]. Unfortunately, it is not a suitable starting point for higher spacetime dimensions.

10.5.1 Perturbation theory

Let us take up the important question of a perturbation analysis in a power series in the coupling constant g. The perturbation theoretic approach to study the question of the existence and the (approximate) calculation of $S(h)$ is an important one and can be used for any spacetime dimension n that one chooses. To a certain extent some features of the calculation are common to any dimension, and so let us discuss those initially. For the sake of convenience, we employ a formal notation suppressing the volume Λ and its associated limit. Thus, first of all, let us consider the formal power series given by

$$\begin{aligned} S(h) &= \sum_{q=0}^{\infty} \frac{(-g)^q}{q!}\, C' \langle\, e^{\phi(h)} [\textstyle\int :\phi(x)^4: \, d^n x]^q \,\rangle \\ &= \sum_{q=0,\, p=0}^{\infty} \frac{(-g)^q}{q!p!}\, C' \langle\, [\phi(h)]^p \, [\textstyle\int :\phi(x)^4: \, d^n x]^q \,\rangle \,. \end{aligned}$$

In the last line the problem has been reduced, in effect, to a sum of correlation functions of the field in the Gaussian ensemble of the relativistic free field with mass m. In turn, this leads to a number of contributions each of which involves sums of products of the two-point correlation function of the Gaussian ensemble. The factor C' is to be chosen so that the overall normalization $S(0) = 1$. In practice, the effect of this normalization can be most simply incorporated within perturbation theory by confining attention to those contributions that are entirely connected to the external terms $\phi(h)$ [IZ80, Kak93, PS95].

As indicated earlier in this chapter, the correlation functions of a mean-zero Gaussian ensemble are entirely determined by the two-point function, which for the relativistic free field is given by

$$\begin{aligned} C(x-y) \equiv \langle \phi(x)\phi(y) \rangle &= \frac{1}{(2\pi)^n} \int \frac{e^{ip\,\cdot\,(x-y)}}{p^2 + m^2}\, d^n p \\ &\simeq \frac{k_n}{|x-y|^{n-2}}\,, \qquad\qquad m|x-y| \ll 1\,, \quad n \geq 3\,, \\ &\simeq -k_2 \ln(m|x-y|)\,, \qquad m|x-y| \ll 1\,, \quad n = 2\,, \end{aligned}$$

where the vector square and vector inner product are taken here in the Euclidean metric. The last two lines are devoted to the short-distance behavior in the two cases $n \geq 3$ and $n = 2$. Here the constants k_n may be determined by the following relatively simple argument. For $n \geq 3$ we consider the distributional relation

$$\frac{k_n}{|x|^{n-2}} \equiv \frac{1}{(2\pi)^n} \int \frac{e^{ip\,\cdot\,x}}{p^2}\, d^n p \;,$$

which holds simply on the grounds of a change of variables in the integration. We now integrate both sides of this expression with a Gaussian test function, leading to

$$\begin{aligned} \int \frac{k_n}{|x|^{n-2}} e^{-x^2/2}\, d^n x \;&= \frac{1}{(2\pi)^n} \int \frac{e^{ip\,\cdot\,x}}{p^2} e^{-x^2/2}\, d^n p\, d^n x \\ &= \frac{1}{(2\pi)^{n/2}} \int \frac{1}{p^2} e^{-p^2/2}\, d^n p \;. \end{aligned}$$

Let us introduce $u^2 \equiv x^2/2$ and $v^2 \equiv p^2/2$, where $0 < u < \infty$ as well as $0 < v < \infty$. Since the angular integrations on both sides cancel we are left with just one-dimensional radial integrals, leading to

$$k_n \int_0^\infty e^{-u^2} u\, du = \frac{1}{4(\pi)^{n/2}} \int_0^\infty e^{-v^2} v^{n-3}\, dv \;,$$

from which we conclude that

$$k_n = \frac{\Gamma(n/2-1)}{4(\pi)^{n/2}} \;, \qquad n \geq 3 \;.$$

For $n = 2$ we can use the fact that as $m \to 0$ the natural logarithm holds over an ever-increasing range of spacetime. Let us again integrate with a Gaussian test function, leading to

$$\int e^{-x^2/2} \left[-k_2 \ln(m|x|)\right] d^2 x \simeq \frac{1}{(2\pi)} \int \frac{e^{-p^2/2}}{p^2+m^2}\, d^2 p \;.$$

As m becomes smaller and smaller, we may divide the right-hand side so that

$$\begin{aligned} -k_2 \ln(m) \int e^{-x^2/2}\, d^2 x + \text{l.o.t.} \\ = \frac{1}{(2\pi)} \int_{|p|^2<1} \frac{1}{p^2+m^2}\, d^2 p \\ + \frac{1}{(2\pi)} \int_{|p|^2<1} \frac{[e^{-p^2/2}-1]}{p^2+m^2}\, d^2 p + \text{l.o.t.} \;, \end{aligned}$$

where l.o.t. means lower order terms. It follows directly from the leading divergence as $m \to 0$ that

$$k_2 = \frac{1}{2\pi} \;.$$

10.5.2 Spacetime dimension $n = 3$

For $n = 3$, as we shall observe below, the naturally defined formal theory has one ultraviolet divergence which requires an infinite mass renormalization. This means that besides normal ordering it is necessary to add an additional, divergent multiple of $: \phi(x)^2 :$. The need for such a modification may be readily demonstrated. For the present we proceed formally, and consider the two-point function given by

$$\langle \phi(h)^2 \exp\{-g \int : \phi(x)^4 : \, d^3x\} \rangle \, .$$

We expand the exponent in a power series in g and focus in particular on the second-order term

$$\tfrac{1}{2} g^2 \int \langle \phi(h)^2 : \phi(x)^4 :: \phi(y)^4 : \rangle \, d^3x \, d^3y \, .$$

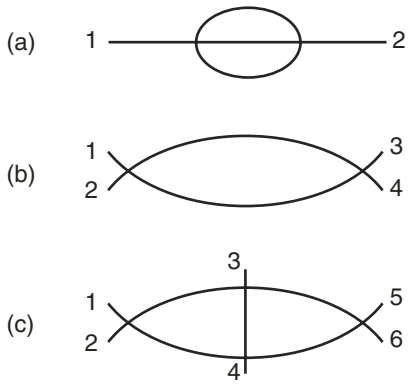

Fig. 10.1. Several contributions to the generating functional represented graphically. The free ends labeled 1, 2, 3, etc., each denote a separate factor $\phi(h)$. The crosses resulting from the intersection of two lines represent terms of the form $: \phi(\cdot)^4 :$ arising from the interaction. Each intersection is accompanied by one power of the coupling constant g.

Among the several contributions this integral leads to, we single out a particular term, conveniently represented by Fig. 10.1, part (a), for further consideration. The term in question is analytically given by[1]

$$96 \, g^2 \int \langle \phi(h)\phi(x) \rangle \langle \phi(x)\phi(y) \rangle^3 \langle \phi(y)\phi(h) \rangle \, d^3x \, d^3y$$

[1] The prefactor $96 = (4!)^2/3!$ is a combinatorial factor representing the number of similar terms arising from the previous expression. For rules on such factors see, e.g., [IZ80].

$$\equiv 96\, g^2 \int \overline{h}(x)\, C(x-y)^3\, \overline{h}(y)\, d^3x\, d^3y$$

$$\simeq [96/(4\pi)^3]\, g^2 \int \overline{h}(x)\, |x-y|^{-3}\, \overline{h}(y)\, d^3x\, d^3y$$

$$= (3/2\pi^3)\; g^2 \int \overline{h}(u+v/2)\, |v|^{-3}\, \overline{h}(u-v/2)\, d^3u\, d^3v\ ,$$

where $\overline{h}(x) \equiv \langle \phi(h)\phi(x)\rangle$. Due to the rapid falloff of $\overline{h}(x)$ for large $|x|$, the integrals over u and v may be considered to be restricted to a finite domain of $\mathbb{R}^3$, say $|v| \leq 1$. On the other hand, there is a divergence in the integral that arises at $v = 0$. Observe that if a Taylor expansion is made about u of either term $\overline{h}$, then the extra v terms that appear render the integral over v convergent at the origin. Thus the divergence is equivalent to that of

$$(3/2\pi^3)\, g^2 \int \overline{h}(u)\, |v|^{-3}\, \overline{h}(u)\, d^3u\, d^3v\ .$$

This term is first made finite by regularization, say by replacing $|v|^{-3}$ by $(|v|+\epsilon)^{-3}$ for a very small $\epsilon > 0$. However, we also may note that

$$[4\pi|\ln(\epsilon)|]^{-1}(|v|+\epsilon)^{-3}$$

converges to the three-dimensional delta function as $\epsilon \to 0$. Thus we observe that we can renormalize our regularized expression by introducing

$$\lim_{\epsilon\to 0} (3/2\pi^3)\, g^2 \int \overline{h}(x)\{(|x-y|+\epsilon)^{-3} - [4\pi|\ln(\epsilon)|]\, \delta(x-y)\}\overline{h}(y)\, d^3x\, d^3y$$

which now has the virtue of being finite. Note that the effect of the $\delta(x-y)$ term is identical to that which would follow from an additional interaction proportional to $\frac{1}{2} : \phi(x)^2 :$. Therefore, if we instead reconsider the problem stated in the form

$$\lim_{\epsilon\to 0} \langle \phi(h)^2 \exp\{-\textstyle\int [g : \phi_\epsilon(x)^4 : + c_\epsilon\, g^2 : \phi(x)^2 :]\, d^3x\}\rangle\ ,$$

where $\phi_\epsilon(x)$ is a field with the regularized two-point function,

$$\langle \phi_\epsilon(x)\phi_\epsilon(0)\rangle = (1/4\pi)(|x|+\epsilon)^{-1}\ , \qquad |x| \ll 1\ ,$$

and

$$c_\epsilon \equiv\ (3/\pi^2)|\ln(\epsilon)|\ ,$$

then we would have eliminated the indicated divergence to order g^2. It is a triumph of combinatorics and renormalization theory that in this case this is the *only* counterterm needed to render the expansion finite to all orders of the coupling constant g. Although by itself this fact is suggestive, it no way proves that this regularization and renormalization provide a meaningful regularization and renormalization procedure *outside* of perturbation theory.

Nevertheless, this is the case. In a triumph of mathematical physics (see, e.g., [FFS92]) it has been shown that the expression

$$\lim_{\Lambda\to\infty}\lim_{\epsilon\to 0} C_\Lambda \langle \exp\{\textstyle\int_\Lambda [h(x)\phi(x) - g : \phi_\epsilon(x)^4 : - c_\epsilon\, g^2 : \phi(x)^2 :]\, d^3x\}\rangle$$

converges to a nontrivial result and that a perturbation theory in the coupling constant g is asymptotic to that result.

10.5.3 Spacetime dimension $n = 4$

For $n = 4$ the naturally defined formal theory has three distinct categories of ultraviolet divergences that require a mass renormalization, a field strength renormalization, and a coupling constant renormalization. In fact, it is necessary to add counterterms for each of the three basic contributions in the formal action. Let us initially argue for the first two types of counterterms. The start of our analysis follows the discussion in the case $n = 3$, and we again consider the two-point function given by

$$\langle \phi(h)^2 \exp\{-g\textstyle\int : \phi(x)^4 :\, d^4x\}\rangle \ .$$

Let us expand the exponent in a power series in g and focus again, in particular, on the second-order term

$$\tfrac{1}{2} g^2 \textstyle\int \langle \phi(h)^2 : \phi(x)^4 :: \phi(y)^4 : \rangle\, d^4x\, d^4y \ .$$

Among the several contributions this integral leads to, we single out the following term [see Fig. 10.1, part (a), again]:

$$\begin{aligned} 96\ g^2 &\int \langle\phi(h)\phi(x)\rangle\langle\phi(x)\phi(y)\rangle^3\langle\phi(y)\phi(h)\rangle\, d^4x\, d^4y \\ &\equiv 96\ g^2 \int \overline{h}(x)\, C(x-y)^3\, \overline{h}(y)\, d^4x\, d^4y \\ &\simeq [96/(4\pi^2)^3]\, g^2 \int \overline{h}(x)\, |x-y|^{-6}\, \overline{h}(y)\, d^4x\, d^4y \\ &= (3/2\pi^6)\, g^2 \int \overline{h}\,(u + v/2)|v|^{-6}\, \overline{h}(u - v/2)\, d^4u\, d^4v \ , \end{aligned}$$

where the integrals over u and v may be considered to be restricted to a finite domain of $\mathbb{R}^4$, say $|v| \le 1$. On the other hand, there is a divergence in the integral that arises at $v = 0$. Observe that if a Taylor expansion is made of either term $\overline{h}$, then after two such terms, the extra v terms that appear render the integral over v convergent at the origin. Thus the divergence is equivalent to that for

$$\begin{aligned} &(3/2\pi^6)\, g^2 [\int \overline{h}(u)\, |v|^{-6}\, \overline{h}(u)\, d^4u\, d^4v \\ &-(1/4) \int \overline{h}_{,j}\,(u)\, v^j\, |v|^{-6}\, v^k\, \overline{h}_{,k}\,(u)\, d^4u\, d^4v \\ &+(1/4) \int \overline{h}_{,jk}\,(u)\, v^j\, |v|^{-6}\, v^k\, \overline{h}(u)\, d^4u\, d^4v] \ ; \end{aligned}$$

a possible term involving only one derivative vanishes, on regularization, by symmetry. We shall regularize the divergences by replacing the factor $|v|$ in the denominator by $(|v| + \epsilon)$ just as before. Let us consider the two contributions separately. The first term becomes $\int (|v| + \epsilon)^{-6}\, d^4v = \epsilon^{-2}\pi^2$, which may be renormalized by the introduction of a mass counterterm of the form $(3/4\pi^4)\, g^2\epsilon^{-2} : \phi(x)^2 :$. In the two other factors only those terms for which $j = k$ contribute due to symmetry, each such term has the same divergence, and therefore the regularized form of such terms is also given, with summation on j implied, by

$$-(3/16\pi^6)\, g^2 \int \overline{h}_{,j}\,(u)(|v| + \epsilon)^{-4}\, \overline{h}_{,j}\,(u)\, d^4u\, d^4v\ .$$

The divergence here is given by $\int_0^1 (|v| + \epsilon)^{-4}\, d^4v = 2\pi^2\, |\ln(\epsilon)|$, which in turn may be canceled by the counterterm $(3/16\pi^4)\, g^2\, |\ln(\epsilon)|\ :\ (\nabla\phi)(x)^2\ :$. This kind of renormalization is referred to as field strength renormalization.

To deal with the coupling constant renormalization we focus next on another expression of the form

$$\tfrac{1}{2} g^2 \int \langle \phi(h)^4 : \phi(x)^4 :: \phi(y)^4 : \rangle\, d^4x\, d^4y\ .$$

Of all the terms to which this gives rise we consider further the term [see Fig. 10.1, part (b)] [note $288 = (4!)^2/2!$]

$$\begin{aligned} &288\, g^2 \int \overline{H}(x)\, C(x-y)^2\, \overline{H}(y)\, d^4x\, d^4y \\ &\qquad \simeq [288/(4\pi^2)^2]\, g^2 \int \overline{H}(x)\, |x-y|^{-4}\, \overline{H}(y)\, d^4x\, d^4y\ . \end{aligned}$$

Here $\overline{H}(x) \equiv \langle \phi(h)\phi(x) \rangle^2 = \overline{h}(x)^2$ which is a smooth function that we may assume falls to zero rapidly as $|x|$ becomes large. Thus, as before, the divergence arises when $x = y$. The term in question, therefore, is equivalent to

$$(18/\pi^4)\, g^2 \int \overline{H}(u + v/2)\, |v|^{-4}\, \overline{H}(u - v/2)\, d^4u\, d^4v\ ,$$

and the divergence is given by

$$(18/\pi^4)\, g^2 \int \overline{H}(u)\, |v|^{-4}\, \overline{H}(u)\, d^4u\, d^4v\ .$$

Regularization leads to the integral $\int_0^1 (|v| + \epsilon)^{-4}\, d^4v = (2\pi^2)|\ln(\epsilon)|$, which may be compensated by the counterterm $(3/2\pi^2)\, g^2 |\ln(\epsilon)|\ :\ \phi(x)^4\ :$. Such a modification amounts to a renormalization of the quartic interaction coupling constant.

If we had included the three renormalization counterterms, then we would have eliminated the indicated divergence to order g^2. However, and unlike the

case of $n = 3$, when higher-order terms in the perturbation series are considered, additional divergences arise, but, fortunately, they all may be subsumed into modifications of the mass, coupling constant, and the coefficient of the gradient-squared term in the formal Lagrangian. Thus the coefficients of these factors become power series in the coupling constant, all terms of which have a common characteristic divergence behavior. This behavior is typical of theories that are called strictly renormalizable. Although by itself this fact is suggestive, it no way proves that this kind of regularization and renormalization provide a meaningful regularization and renormalization procedure *outside* of perturbation theory. In fact, it is now widely believed that this is *not* the case, namely, that the full theory treated outside of perturbation theory yields a qualitatively different answer than is given by the perturbation solution. Indeed, outside of perturbation theory, the quartic self-coupled scalar quantum field theory is believed to be a *free* theory, i.e., to lead to a free field theory or perhaps to a generalized free field theory [FFS92, Cal88]. It is of considerable interest that the results of a perturbation treatment, on the one hand, and of a nonperturbative analysis, on the other hand, can have such diverse answers.

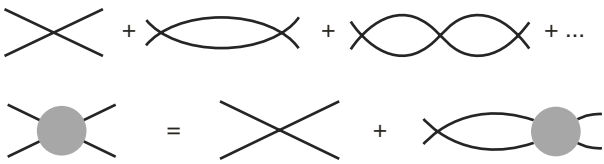

Fig. 10.2. A graphical representation for the sum of an infinite number of terms of a special family of contributions. The top line has separate graphs each of which has the interpretation given for those in Fig. 10.1. The bottom line graphically represents an equation, in fact, an integral equation, whose solution leads, simply by iteration, to the complete sum of the contributions represented in the top line.

Fortunately, this dilemma seems to have a resolution. A careful examination of the perturbation series, especially including a partial resummation of certain selected contributions, leads one to believe that a resummed perturbation series may well also lead to a free theory result. Let us give a qualitative picture of such an argument. First consider Fig. 10.2. This figure represents an integral equation for the summation of an infinite number of contributions, thereby leading to an expression beyond any finite order of perturbation theory. The qualitative structure of such a partial summation is of the form

$$\frac{gC}{(1 + g|\ln(\epsilon)|C)} \,,$$

where C denotes a finite, positive contribution, the details of which are not important for the present discussion. The essential fact to note is that as the regularization is removed, i.e., as $\epsilon \to 0$, the denominator diverges, sending the overall term to zero. In other words, although order by order the contributions to the interaction involve divergences, when selected contributions are resummed it follows that the overall contribution of an infinite number of terms actually forces the nonlinear coupling constant to vanish; this kind of argument is said to be due to a "Landau pole" (or "Landau ghost") [IZ80]. Here, then, in the context of perturbation theory is an argument, heuristic to be sure, that the "complete" theory including the nonlinear interaction corresponds to a noninteracting theory. This result offers some support for the widely held belief that the conventionally formulated ϕ_4^4 is trivial [FFS92].

10.5.4 Spacetime dimension $n \geq 5$

For $n = 5$ the naturally defined formal theory has an unlimited number of distinct categories of ultraviolet divergences which require not only a mass renormalization, a field strength renormalization, and a coupling constant renormalization, but also require the introduction and renormalization of an infinite number of additional self-interactions. In other words, the formal starting Lagrangian that contains a quartic interaction ϕ^4 must be supplemented by an unending series of interactions of the form $:\phi^{2p}:$ where $p = 3, 4, 5 \ldots$ (as well as other counterterms involving derivatives). The divergent part of the coefficients of these auxiliary interactions is determined so as to cancel those divergences that arise from lower-order interactions. However, that leaves open the question of any finite part remaining for these interactions, and the arbitrariness in these factors, as the argument is generally made, renders such theories useless for any predictive purposes inasmuch as an infinite number of measurements need to be made first so as to fix the finite parts of these needed counterterm interactions. Such theories are called nonrenormalizable, and it is clear from this qualitative account that they lead to essentially insurmountable difficulties. Our purpose in this chapter is not to offer any miracle cure for such "sick" theories, but rather to indicate how the need for such an endless chain of higher-order interactions arises. (However, in the next chapter we shall revisit these theories and indeed propose a "miracle cure"!) For convenience, we shall confine our discussion to the case $n = 5$.

To assess the specific need for renormalizations and various counterterms for such a theory we accept the lessons we have learned from the analysis in spacetime dimensions $n = 3$ and $n = 4$. In particular, for $n = 4$ we found the need for a coupling constant renormalization. In the present case let us examine the term

$$(1/4!)g^4\langle\phi(h)^6\,[\int :\phi(x)^4: d^5x]^4\,\rangle\ .$$

Observe that this term is higher-order in g than we have previously considered. Among the several contributions to which this factor gives rise, we focus on

[see Fig. 10.1, part (c)] [note $144 = (4!)^2/(2!)^2$]

$$144\, g^4 \int \overline{H}(w)\, \overline{H}(x)\, \overline{h}(y)\, \overline{h}(z)\, C(x-y)\, C(x-z)\, C(y-z) \\ \times C(w-y)\, C(w-z)\, d^5w\, d^5x\, d^5y\, d^5z\ ,$$

where $\overline{h}$ and $\overline{H}$ have their previous meanings. As was the case before, the integration domains are effectively restricted for large arguments. For short distances $C(x-y) \simeq (1/8\pi^2)|x-y|^{-3}$, and therefore we are faced with estimating

$$144(1/8\pi^2)^5\, g^4 \int \overline{H}(w)\, \overline{H}(x)\, \overline{h}(y)\, \overline{h}(z)\, |x-y|^{-3}\, |x-z|^{-3}\, |y-z|^{-3} \\ \times |w-y|^{-3}\, |w-z|^{-3}\, d^5w\, d^5x\, d^5y\, d^5z\ .$$

We may estimate this expression by carrying out the integrations over both w and x, which leads to a result proportional to

$$g^4 \int \overline{h}(y)\, |y-z|^{-5}\, \overline{h}(z)\, d^5y\, d^5z\ .$$

A change of variables leads to

$$g^4 \int \overline{h}\,(u+v/2)|v|^{-5}\, \overline{h}(u-v/2)\, d^5u\, d^5v\ ,$$

and much as before the divergent part is given by

$$g^4 \int \overline{h}(u)\, |v|^{-5}\, \overline{h}(u)\, d^5u\, d^5v\ ,$$

which is logarithmically divergent, i.e., proportional to $|\ln(\epsilon)|$. For present purposes it is sufficient to determine that the integral in question is divergent, and we do not need to explicitly evaluate the coefficient. The counterterm for this divergence is proportional to $g^4|\ln(\epsilon)|\ :\phi(x)^6:$, which is seen to be an entirely new local interaction. In turn, this nonvanishing interaction may be used in the evaluation of even further contributions. For example, we may next examine

$$(g^7/3!)\, |\ln(\epsilon)|\, \langle \phi(h)^8 \textstyle\int :\phi(w)^6:\, d^5w\, [\int :\phi(x)^4:\, d^5x]^3 \rangle$$

which, among others, contributes a term proportional to

$$g^7\, |\ln(\epsilon)| \int \overline{H}(w)\, \overline{H}(x)\, \overline{h}(y)^3\, \overline{h}(z)\, C(x-y)\, C(x-z)\, C(y-z) \\ \times C(w-y)\, C(w-z)\, d^5w\, d^5x\, d^5y\, d^5z\ ,$$

with an integral which is just as divergent as was the previously considered case. This particular divergence requires the introduction of a counterterm

proportional to $g^7\,|\ln(\epsilon)|^2 : \phi(x)^8 :$. And so it goes *ad infinitum* with the need to introduce counterterms of an ever higher order.

One may ask the question of how these conclusions obtained on the basis of perturbation theory relate to conclusions that can be drawn from nonperturbative analyses. In fact, a nonperturbative analysis of such models has been made on the basis of a Euclidean lattice model formulation (see next section), and it has been shown [FFS92] that models such as ϕ_5^4 rigorously tend to free or generalized free theories in the limit that the regularization is removed. In other words, while the perturbation analysis requires an unending set of higher-order counterterms for its definition order by order in the coupling constant, the nonperturbative analysis leads to a theory in which all nonlinear interactions effectively vanish. One may again argue—as was the case for $n = 4$—that a partial summation of suitable contributions leads to the suppression of any nonlinear interaction, and so in this heuristic sense the results of the nonperturbative study may be qualitatively understood.

We note without proof that for the scalar quantum field models ϕ_n^4, with $n \geq 6$, the results are qualitatively the same as for $n = 5$ and all such theories are nonrenormalizable.

10.6 Euclidean-Space Lattice Regularization

An alternative regularization provides another approach to self-interacting scalar quantum field theories, and it is a natural generalization of the lattice formulation already studied in the case of the relativistic free theories. In particular, let us consider

$$S(h) = C \int \exp\{\Sigma_k [h_k\,\phi_k\,a^n - \tfrac{1}{2}(\phi_{k^*} - \phi_k)^2 a^{n-2} \\ -\tfrac{1}{2} m_0^2\,\phi_k^2\,a^n - g : \phi_k^4 : a^n\,]\}\,\Pi_k\,d\phi_k\ .$$

In this expression we assume that the total number of lattice sites $N \equiv L^n$, say, is finite, and recall that the continuum limit means not only that $a \to 0$, but that $N \to \infty$ in such a way that eventually the integration covers all of $\mathbb{R}^n$. Guided by the continuum theory, we have included normal ordering for the quartic interaction term. For $n = 2$ and $n = 3$, and allowing for an appropriate definition of the bare parameters, it has been demonstrated that the continuum limit of this lattice formulation generates a nontrivial Euclidean-space quantum field theory that in addition satisfies the appropriate axioms to permit a transformation back to a Minkowski theory, and which yields a theory that is equivalent to the one defined asymptotically by its perturbation series in the coupling constant g. On the other hand, for $n \geq 5$, and quite probably for $n = 4$, the so-defined theory leads to a Gaussian—often called trivial—result [FFS92].

We will reexamine these results for $n \geq 5$ from an entirely different viewpoint in the following chapter and will find potentially very different conclusions. These new conclusions involve a new renormalization procedure called "mashing the measure"; keep an eye out for it in the next chapter.

Exercises

10-1 Verify the equality implied in the equation

$$\begin{aligned}\langle 0|\mathsf{T}\exp[i\textstyle\int \Sigma\, h_k(t)Q_k(t)\,dt]|0\rangle \\ = \lim_{\epsilon\to 0}\mathcal{N}\int \exp\Big(i\textstyle\int\{\Sigma\, h_k(t)q_k(t) + \frac{1}{2}\Sigma\,[\dot{q}_k(t)^2 + i\epsilon q_k(t)^2] \\ -V(q(t))\}\,dt\Big)\,\Pi\,\mathcal{D}q_k\ .\end{aligned}$$

What kind of time dependence is meant by $Q_k(t)$ on the left-hand side?

10-2 Complete the argument that

$$k_2 = \frac{1}{2\pi}$$

for the short-distance behavior of the function

$$\frac{1}{(2\pi)}\int \frac{e^{ip(x-y)}}{p^2+m^2}\,d^2p \simeq k_2\,[-\ln(m\,|x-y|)]$$

when $m|x-y| \ll 1$.

10-3 Using the formal fact that

$$:e^{\phi(x)}: \equiv e^{\phi(x)}/\langle e^{\phi(x)}\rangle\ ,$$

show that in general

$$\langle :\phi(x)^p::\phi(y)^p:\rangle = p!\,\langle\phi(x)\,\phi(y)\rangle^p\ .$$

10-4 Set up the several integral expressions represented by the expression

$$\langle \phi(h)^2\,:\phi(x)^4::\phi(y)^4:\phi(h)^2\rangle\ .$$

11

A Modern Approach to Nonrenormalizable Models

11.1 Introduction

In the preceding chapter we discussed a sequence of quartic self-interacting relativistic scalar fields that range from super renormalizable models (for spacetime dimension $n = 2, 3$), to strictly renormalizable models ($n = 4$), and finally to nonrenormalizable models ($n \geq 5$). In the super renormalizable cases, the perturbative solutions we outlined agree completely with the rigorous results obtained by more sophisticated, nonperturbative techniques. Thus these cases are unambiguous and are considered solved (from an existence point of view, at least). The strictly renormalizable case exhibits all the expected features of an acceptable perturbation analysis, but there is good reason to believe that it does *not* agree with the nonperturbative solution. That is, the nonperturbative solution is not equal to the one derived by a perturbation theory, and as renormalization group and Monte Carlo studies have indicated, the nonperturbative strictly renormalizable model leads to a free theory in the continuum limit. By summing a suitable set of infinitely many perturbation terms, even the perturbation theory can be made to suggest that the continuum limit is also a free theory. Moreover, when we examine the nonrenormalizable cases from a nonperturbative viewpoint and limit the set of counterterms to include only mass and coupling constant modifications, they too can be rigorously shown to approach free theories in the continuum limit. This situation means that for the quartic relativistic scalar field theories, a perturbation analysis for spacetime dimension $n = 4$ and $n \geq 5$, namely, the perturbative solution discussed in the preceding chapter, fails to properly represent the quantization of the corresponding classical field theory. This situation represents a genuine gap in our understanding of quantum field theory.

Recently, a general philosophy and an associated approach to nonrenormalizable models has arisen that we wish to develop in the present chapter. It is not yet clear whether these same ideas may offer a viable, alternative approach for strictly renormalizable cases that are also asymptotically nonfree,

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9_11,

but that possibility is not yet excluded. Therefore, because we seem to be on safer ground, we shall focus our attention solely on the nonrenormalizable cases in what follows.

11.2 Nonrenormalizable Classical Models

11.2.1 Relativistic models

The Euclidean action for a quartic, self-interacting, relativistic model is given by

$$I = \int \{ \tfrac{1}{2} [\nabla \phi(x)]^2 + \tfrac{1}{2} m_0^2 \phi(x)^2 + g_0 \phi(x)^4 \} \, d^n x \ ,$$

where n denotes the spacetime dimension and $g_0 \geq 0$. The case where $g_0 = 0$ defines the all-important free model, the Euclidean action for which is given by

$$I_0 = \tfrac{1}{2} \int \{ [\nabla \phi(x)]^2 + m_0^2 \phi(x)^2 \} \, d^n x \ .$$

As outlined in the preceding chapter, the quartic scalar field theory is nonrenormalizable for $n \geq 5$ since a perturbation series of the interacting theory about the free theory leads to an infinite sequence of distinct, ever more divergent counterterms.

In this chapter, we will argue that: (i) there is a clear and natural reason for such models to be nonrenormalizable, (ii) this reason suggests that a perturbation analysis about an alternative model—to be called the *pseudofree model*—leads to a perturbation series which is term-by-term finite, and very importantly, (iii) the pseudofree model can be reached by an extremely elementary procedure. However, to motivate the analysis for the relativistic models it is pedagogically useful to first discuss two (nonrelativistic) nonrenormalizable models, both of which are even more singular than relativistic models, yet with the great virtue of being completely soluble!

11.2.2 Ultralocal models

The Euclidean action for ultralocal models is obtained from the relativistic models by dropping the spatial gradient terms. Thus the quartic Euclidean action functional for ultralocal models reads

$$I = \int \{ \tfrac{1}{2} \dot{\phi}(t, \mathbf{x})^2 + \tfrac{1}{2} m_0^2 \phi(t, \mathbf{x})^2 + g_0 \phi(t, \mathbf{x})^4 \} \, dt \, d^s \mathbf{x} \ .$$

In this expression, $s = n - 1$, $x = (t, \mathbf{x})$ in which we have singled out the Euclidean time t, and the spatial variable $\mathbf{x} \in \mathbb{R}^s$. Only the time derivative—represented by $\dot{(\ )}$—is present in the ultralocal model.

Besides the quartic self-interacting ultralocal model, we also focus on the free ultralocal model for which the Euclidean action reads

$$I_0 = \tfrac{1}{2}\int\{\,\dot{\phi}(t,\mathbf{x})^2 + m_0^2\phi(t,\mathbf{x})^2\,\}\,dt\,d^s\mathbf{x}\ .$$

Observe that both the free and interacting ultralocal models lack spatial derivatives, implying that the time development at one point $\mathbf{x}$ is completely independent of the time development at any other spatial point. It is as if the forward light cone of influence relevant for relativistic models has been collapsed to a forward *line* of influence. These models do not represent the physical situation in the real world—our motivation for studying ultralocal models lies in their complete solubility, not their physicality.

11.2.3 Independent value models

A small modification of the ultralocal models leads us to the independent value models. That modification is to drop from the Euclidean action the remaining (time) derivative term, leading to the Euclidean action for the independent value models given by

$$I = \int\{\,\tfrac{1}{2}m_0^2\phi(x)^2 + g_0\,\phi(x)^4\,\}\,d^nx\ .$$

The free independent value models are in turn given by the Euclidean action

$$I_0 = \tfrac{1}{2}m_0^2\int\phi(x)^2\,d^nx\ .$$

Unlike the relativistic and ultralocal models, the independent value models exhibit no dynamics whatsoever; the role of these latter models lies in their functional integral study which, as we shall see, renders them completely soluble.

For the three sets of models that we have introduced, we have confined our attention only to quartic nonlinear interactions. The reason for this is that quartic nonlinear interactions are the easiest to deal with, and they also have the virtue that higher-order interactions can be treated by very similar techniques. In point of fact, as a concrete example of what is implied by the previous remark, we note that our treatment of the three models in their nonrenormalizable regimes can be carried out not only for ϕ^4 nonlinear interactions but for ϕ^{44} or ϕ^{444}, etc., equally well! This should indicate to the knowledgeable reader that our approach will be vastly different from that which is conventionally used.

We next explore the classical behavior of our chosen models in some detail.

11.2.4 Classical pseudofree models

Although we have presented the Euclidean (imaginary time) action for the relativistic and ultralocal models above, it is straightforward to imagine dealing with the real-time classical actions and the dynamical equations to which they lead upon stationary variation. For example, the set of classical solutions

W_f for the *free* ultralocal model is limited to the class of functions $\phi(t,\mathbf{x})$ that satisfy

$$W_f \equiv \{\phi : \int [\dot{\phi}(t,\mathbf{x})^2 + m_0^2 \phi(t,\mathbf{x})^2]\, dt\, d^s\mathbf{x} < \infty\}$$

because such a limitation is imposed on the variation of the functions in the classical action in order to derive the classical equations of motion.

Now consider the solutions of the dynamical equations for the interacting ultralocal model for which $g_0 > 0$. The set of such solutions is given by

$$W_{int,g_0} \equiv \{\phi : \int [\dot{\phi}(t,\mathbf{x})^2 + m_0^2 \phi(t,\mathbf{x})^2 + 2g_0 \phi(t,\mathbf{x})^4]\, dt\, d^s\mathbf{x} < \infty\}\,,$$

again because this is the maximal class of functions that are to be varied. Lastly, we consider the set W_{pf} defined as the set of interacting solutions in which the coupling constant has been reduced to zero and which we call the *pseudofree* model. In particular,

$$W_{pf} \equiv \lim_{g_0 \to 0} \{\phi : \int [\dot{\phi}(t,\mathbf{x})^2 + m_0^2 \phi(t,\mathbf{x})^2 + 2g_0 \phi(t,\mathbf{x})^4]\, dt\, d^s\mathbf{x} < \infty\}\,.$$

The question immediately arises whether $W_{pf} = W_f$ or, instead, $W_{pf} \subset W_f$, i.e., a proper subset in the latter case. This is a question that could be asked about any particular theory and the answer is clearly model dependent. For the ultralocal case at hand, it is clear that $W_{pf} \subset W_f$. This is because functions, e.g., like $\phi_{singular}(t,\mathbf{x}) = g(t)|\mathbf{x}|^{-\alpha} \exp(-\mathbf{x}^2)$, where $g(t)$ is a smooth function and $2\alpha < s \leq 4\alpha$, are such that the free term is finite but the quartic interaction term diverges, leading to the fact that $\phi_{singular} \in W_f$ while $\phi_{singular} \notin W_{pf}$. Thus, for the ultralocal model, the classical pseudofree model is *not* equivalent to the classical free model. This fact will play an important role in our discussion of the functional integral approach to quantization of the ultralocal models, as will the fact that the dynamical evolution at any spatial point $\mathbf{x} \in \mathbb{R}^s$ is completely independent of that at any other point $\mathbf{x}' \neq \mathbf{x}$.

A similar discussion applies to the case of the independent value models even though there is no dynamical story to go along with it in this case. In particular, the issue of a finite free Euclidean action requires that

$$W_f \equiv \{\phi : \int \phi(x)^2\, d^n x < \infty\} \equiv \{\phi : \phi \in L^2(\mathbb{R}^n)\}\,,$$

while the corresponding pseudofree theory is characterized by

$$W_{pf} \equiv \{\phi : \phi \in L^2(\mathbb{R}^n) \cap L^4(\mathbb{R}^n)\}\,.$$

It is clear in this case as well that $W_{pf} \subset W_f$ as exemplified by such functions as $\phi_{singular}(x) = |x|^{-\beta} |\exp(-x^2)$, where $2\beta < n \leq 4\beta$.

Most interesting, however, is the relativistic case. For this case we recall a classical Sobolev-type inequality [LSU, Kla00] which states that

$$[\int \phi(x)^4 \, d^n x]^{1/2} \leq C \int \{[\nabla \phi(x)]^2 + m_0^2 \phi(x)^2\} \, d^n x \ ,$$

where $C = 4/3$ suffices whenever $n \leq 4$, while $C = \infty$ holds whenever $n \geq 5$. When $C = \infty$, it means that there are fields ϕ such that the free term is finite while the interaction term is infinite; for example, such a field is given by $\phi_{singular}(x) = |x|^{-p} \exp(-x^2)$, where $n/4 \leq p < n/2 - 1$. It is noteworthy, in the case of relativistic models, that the class of perturbatively renormalizable models, for which $n \leq 4$, is such that $W_{pf} = W_f$. On the other hand, the class of perturbatively nonrenormalizable models, for which $n \geq 5$, is such that $W_{pf} \subset W_f$.

Thus, in the three cases of (1) independent value models, (2) ultralocal models, and (3) relativistic models (for $n \geq 5$), we deal with perturbatively nonrenormalizable models for which the space of solutions of the classical free theory, which we have called W_f, is *strictly larger* than the space of solutions of the classical pseudofree theory, which we have called W_{pf}. This fact will play a fundamental role when we quantize these respective models by functional integrals, at which time we will find ourselves led naturally to alternative approaches to those that are normally used.

11.3 Euclidean Space Functional Integrals—Preliminary Remarks

The formal functional integrals of interest in this section all involve the Euclidean action functionals given above. We focus on the Euclidean formulations for their better mathematical properties, and even for their ultimate relevance for possible Monte Carlo calculational studies. Nevertheless, although we are in reality considering stochastic theories such as those discussed in Part I of this monograph, we are at the same time studying properties of quantum systems as studied in Part II.

11.3.1 Independent value models

For the independent value (IV) models we focus attention on evaluating the formal functional integral given by

$$S_{IV}(h) \equiv \mathcal{N} \int e^{(1/\hbar)\int [h\phi - \frac{1}{2} m_0^2 \phi^2 - g_0 \phi^4] \, d^n x} \, \mathcal{D}\phi \ ,$$

as well as the associated free independent value model given by

$$S_{IV,\,f}(h) \equiv \mathcal{N} \int e^{(1/\hbar)\int [h\phi - \frac{1}{2} m_0^2 \phi^2] \, d^n x} \, \mathcal{D}\phi \ .$$

Observe that $\mathcal{N}$ denotes a formal normalization that can vary from one expression to another.

Note well the appearance of $\hbar$ in these integrals. Although there is not much quantum mechanics in the IV model, the introduction of Planck's constant is in keeping with its presence when we quantize the ultralocal and relativistic models. Moreover, the very process of quantization entails $\hbar$ ambiguities, namely, the appearance of additional interaction terms that involve positive powers of $\hbar$ which in the "classical limit" (such as it is for the IV models) would disappear in the limit in which $\hbar \to 0$. We make these remarks because—unlike well-defined and fully specified functional integrals—the quantization process is *inherently ambiguous* and the appearance of additional $\hbar$ "corrections" in that process is oftentimes necessary and welcome. As a consequence, the very formulation of the functional integrals above is intrinsically incomplete, which is both a *bad thing* because it is not completely clear just how it should be defined, and also a *good thing* because this very freedom may permit $\hbar$ corrections that lead to a well-defined, possibly nontrivial, functional integral after all. Since we do not know or even have a proposal for what the $\hbar$ corrections should be at this point, they are left out of the formal functional integral in this preliminary discussion. Such an omission is absolutely standard at the present stage.

An important feature regarding the IV models is the vast symmetry in the formal functional integral, which implies that the field distribution at any one point $x \in \mathbb{R}^n$ is independent of the field distribution at any other point $x' \neq x$. This very property is embodied within the name of the models and will be strongly used to find suitable $\hbar$-dependent counterterms that both respect this basic symmetry as well as provide a fully acceptable resolution of how the evaluation of such a functional integral should be carried out.

11.3.2 Ultralocal models

For the ultralocal (UL) models we introduce the formal functional integrals of interest given by

$$S_{UL}(h) \equiv \mathcal{N} \int e^{(1/\hbar)\int \{h\phi - [\frac{1}{2}\dot{\phi}^2 + \frac{1}{2}m_0^2\phi^2 + g_0\phi^4]\}\, dt\, d^s\mathbf{x}}\, \mathcal{D}\phi \,,$$

along with the corresponding free ultralocal model given by

$$S_{UL,\,f}(h) \equiv \mathcal{N} \int e^{(1/\hbar)\int \{h\phi - \frac{1}{2}[\dot{\phi}^2 + m_0^2\phi^2]\}\, dt\, d^s\mathbf{x}}\, \mathcal{D}\phi \,.$$

The functional integrals for the UL models are clearly designed to create a Euclidean version of the quantization of the classical, ultralocal, scalar field theory. As a proper quantization of such a model, it is natural to expect that $\hbar$ ambiguities exist and may well be ultimately necessary to give these formal expressions a genuine meaning. It is conceivable that these quantum "corrections" may consist only of modifications to the constant factors m_0 and g_0; alternatively, it may be necessary to introduce a variety of counterterms with $\hbar$-dependent coefficients as dictated by a perturbation analysis

such as that outlined in the previous chapter; or, if a traditional perturbation approach proves unsatisfactory, it may be necessary to introduce nontypical, $\hbar$-dependent counterterms, and what form those counterterms would take may be hard to predict at the outset. As it will turn out, normal counterterms suggested by perturbation theory can be shown to be inadequate, and we will be forced to seek a solution within the essentially unlimited family of general $\hbar$-dependent functions. At face value, the prospect of finding within this vast family of $\hbar$-dependent functions a suitable counterterm—or perhaps a few satisfactory counterterms—seems all but hopeless. The prospect of success for the ultralocal models lies in the fact that such models have an enormous amount of symmetry that should, in the end, be sufficient to fully determine just the right form of $\hbar$-dependent counterterms to solve this problem completely.

All the remarks given above regarding the appearance of $\hbar$ in these functional integrals apply equally well to the ultralocal models. In particular, these integrals are intrinsically undefined based partially on the usual $\hbar$ ambiguity inherent in the quantization process.

Having sung the song regarding $\hbar$ ambiguity loudly enough, let us give this poor symbol a rest by setting $\hbar = 1$ for the time being.

11.3.3 Solution of independent value models

In this section we outline the indirect arguments that enable us to evaluate the functional integral (with $\hbar = 1$) given by

$$S_{IV}(h) = \mathcal{N} \int e^{\int [h(x)\,\phi(x) - \frac{1}{2}\,m_0^2\,\phi(x)^2 - g_0\,\phi(x)^4]\,d^n x}\,\mathcal{D}\phi\ .$$

The answer we will find is given by

$$S_{IV}(h) = \exp\{\,b \int d^n x \int [\cosh(h(x)\,u) - 1]\,e^{-\frac{1}{2}bm^2\,u^2 - g b^3\,u^4}\,du/|u|\,\}\ ,$$

where the bare parameters $m_0^2 \equiv [b/\delta(0)]\,m^2$ and $g_0 \equiv [b/\delta(0)]^3 g$ are related to the finite, physical parameters m^2 and g. Observe that $b > 0$ is an arbitrary parameter with dimension L^{-n} chosen so as to cancel the dimensions of the n-dimensional delta function $\delta(0)$.

Since the derivation of this answer is indirect, we do not present the full details here; the complete story can be found in [Kla00]. Instead, we simply sketch the central steps involved:

(1) On the basis of symmetry, the answer must be of the form

$$S(h) = \exp\{\,\int L[h(x)]\,d^n x\,\}\ .$$

The discussion of infinitely divisible probability distributions presented in Sec. 2.3.2 applies, and since $L[-p] = L[p]$, we conclude that

$$L[p] = \tilde{b}p^2 + \int [\cosh(pu) - 1]\,C(u)^2\,du\ ,$$

where $p \in \mathbb{R}$, $\tilde{b} \geq 0$, and we have already assumed that the measure $C(u)^2\,du$ is absolutely continuous. The function $C(u)$ is called the "model function." It follows that

$$\int [u^2/(1+u^2)]\, C(u)^2\, du < \infty\ ,$$

but it is possible that

$$\int C(u)^2\, du = \infty\ ,$$

which in fact will be true in our case. The term $\tilde{b}p^2$ applies to the (Gaussian) free theory, and hereafter we focus on the second (Poisson) factor assuming that $\tilde{b} = 0$.

(2) The stochastic field $\phi(x)$ for the second factor may be realized as

$$\phi(x) = \sum_{n=1}^{\infty} r_n\, \delta(x - y_n)\ ,$$

where the r_n variables have a distribution determined by $C(u)^2$ and the variables $\{y_n\}$ have a Poisson distribution in $\mathbb{R}^n$ with a unit density. The procedure by which such a realization leads to $S(h)$ is completely analogous to how a similar expression is derived in Sec. 4.9.

(3) Local field products may be defined by the operator product expansion. In particular, we note that

$$\begin{aligned}\phi(x)\,\phi(x') &= \sum_{n,m=1}^{\infty} r_n r_m\, \delta(x - y_n)\, \delta(x' - y_m) \\ &= \delta(x - x') \sum_{n=1}^{\infty} r_n^2\, \delta(x - y_n) + \sum_{n \neq m = 1}^{\infty} r_n r_m\, \delta(x - y_n)\, \delta(x' - y_m)\ .\end{aligned}$$

As $x' \to x$ the first term dominates and thus we define the renormalized (R) local product as

$$\begin{aligned}\phi(x)_R^2 &= [b/\delta(0)]\, \phi(x)^2 \\ &= b \sum_{n=1}^{\infty} r_n^2\, \delta(x - y_n)\ ,\end{aligned}$$

where we have introduced the positive factor b with the dimension L^{-n} to cancel the dimensions of the delta function. Observe that division by $\delta(0)$ has eliminated any contribution to $\phi(x)_R^2$ from the terms where $n \neq m$. (The limit involved here is better understood after smearing the operators with test functions, but the result is the same.) In like fashion,

$$\phi(x)_R^p = b^{p-1} \sum_{n=1}^{\infty} r_n^p\, \delta(x - y_n)\ ,$$

which represents the pth renormalized field power, where we have used the same factor b for each power. This convention ensures that

$$[\phi(x)_R^p \phi(x)_R^q]_R = \phi(x)_R^{p+q} \ .$$

(4) The addition of an interaction term begins with the expression

$$\begin{aligned} \mathcal{N} \int \exp\{\textstyle\int [h(x)\,\phi(x) - \frac{1}{2}\,m_0^2\,\phi(x)^2 - g_0\,\phi(x)^4 - q_0\,w(x)\,\phi(x)^p]\,d^n x\}\,\mathcal{D}\phi \\ = \exp\{\textstyle\int d^n x \int [e^{h(x)\,u - q\,w(x)\,b^{p-1}\,u^p} - 1]\,C(u)^2\,du\} \ , \end{aligned}$$

where $q_0 > 0$, $w(x) \geq 0$ and compactly supported, and p is even. To add the new term as an interaction over all spacetime requires that we first rescale the normalization factor so that $S(0) = 1$, and this leads to

$$\begin{aligned} S(h) \ &= \frac{\exp\{\int d^n x \int [e^{h(x)\,u - q\,w(x)\,b^{p-1}\,u^p} - 1]\,C(u)^2\,du\}}{\exp\{\int d^n x \int [e^{-q\,w(x)\,b^{p-1}\,u^p} - 1]\,C(u)^2\,du\}} \\ &= \exp\{\textstyle\int d^n x \int [\cosh(h(x)\,u) - 1]\,e^{-q\,w(x)\,b^{p-1}\,u^p}\,C(u)^2\,du\} \ . \end{aligned}$$

At this point we can let $w(x) \to 1$, leading to the evaluation given by

$$\begin{aligned} S(h) \ &= \mathcal{N} \int \exp\{\textstyle\int [h(x)\,\phi - \frac{1}{2}\,m_0^2\phi^2 - g_0\,\phi^4 - q_0\phi^p]\,d^n x\}\,\mathcal{D}\phi \\ &= \exp\{\textstyle\int d^n x \int [\cosh(h(x)\,u) - 1]\,e^{-q\,b^{p-1}\,u^p}\,C(u)^2\,du\} \ . \end{aligned}$$

Hence, we can include the original quartic interaction by setting $p = 4$, which leads to

$$C^2(u) = e^{-g\,b^3\,u^4}\,C_0^2(u) \ ,$$

where $C_0^2(u)$ applies to the case when only the mass term survives.

The next expression, although superficially appearing to be a Gaussian integral because any $\hbar$-dependent counterterms are not included, requires the reader to keep an open mind!

To determine the expression $C_0^2(u)$ we make a mass insertion such that

$$\begin{aligned} \mathcal{N} \int \exp\{\textstyle\int [h(x)\,u - \frac{1}{2}\,m_0^2\,\phi^2 - \frac{1}{2}\,q_o\phi^2]\,d^n x\}\,\mathcal{D}\phi \\ = \exp\{\textstyle\int d^n x \int [\cosh(h(x)\,u) - 1]\,e^{-\frac{1}{2}\,q\,b u^2}\,C_0(u)^2\,du\} \ . \end{aligned}$$

Since the interaction and mass terms are now similar, a rescaling of the field variables returns us to the original expression when $q_0 = 0$. This implies the functional identity that

$$\begin{aligned}&\int[\cosh(zu)-1]e^{-\frac{1}{2}qbu^2}\,C_0^2(u)\,du\\&\quad=\int[\cosh(zu/\sqrt{1+q/m^2}\,)-1]\,C_0^2(u)\,du\\&\quad=\int[\cosh(zu)-1]\,\sqrt{1+q/m^2}\,C_0^2(\sqrt{1+q/m^2}\,u)\,du\;.\end{aligned}$$

Consistency of this expression for all z values requires that

$$e^{-\frac{1}{2}qbu^2}\,C_0^2(u)=\sqrt{1+q/m^2}\,C_0^2(\sqrt{1+q/m^2}\,u)$$

for all arguments. As a consequence we determine that

$$C_0^2(u)=b\,\frac{e^{-\frac{1}{2}m^2bu^2}}{|u|}\;,$$

up to an overall constant that we have arbitrarily chosen to be b which has the correct dimensions.

Putting all this together, we learn that

$$\begin{aligned}S(h)\;&=\mathcal{N}\int\exp\{\textstyle\int[h(x)\phi-\frac{1}{2}m_0^2\phi^2-g_0\phi^4]\,d^nx\}\,\mathcal{D}\phi\\&=\exp\{b\textstyle\int d^nx\int[\cosh(h(x)u)-1]\,e^{-\frac{1}{2}bm^2u^2}-gb^3u^4\,du/|u|\}\;,\end{aligned}$$

where, as noted previously, $m_0^2=[b/\delta(0)]m^2$ and $g_0=[b/\delta(0)]^3g$.

The first remark to be made about this result is to point out that the pseudofree model, which arises when $g\to 0$, is given by

$$S_{pf}(h)=\exp\{b\textstyle\int d^nx\int[\cosh(h(x)u)-1]\,e^{-\frac{1}{2}bm^2u^2}\,du/|u|\}\;,$$

and this expression is decidedly *not* the (Gaussian) free model. Moreover, the interacting model solution has an expansion in the coupling constant g about the pseudofree model that is divergence free at each order, as seen in the expansion series

$$\begin{aligned}&S(h)\\&=\exp\{\textstyle\sum_{l=0}^{\infty}(l!)^{-1}(-gb^3)^lb\int d^nx\int[\cosh(h(x)u)-1]\,u^{4l}e^{-\frac{1}{2}bm^2u^2}\,du/|u|\}\,.\end{aligned}$$

11.3.4 Solution of ultralocal models

Although the ultralocal models have dynamics and thus appear to be very different from the independent value models, they are much closer than might otherwise be thought. The spacetime functional integral that formally describes the UL models is (again with $\hbar=1$) given by

$$S_{UL}(h)=\mathcal{N}\int e^{\int\{h(t,\mathbf{x})\phi-\frac{1}{2}[\dot\phi^2+m_0^2\phi^2]-g_0\phi^4\}\,dt\,d^s\mathbf{x}}\,\mathcal{D}\phi\;.$$

This expression is formal and perturbatively nonrenormalizable; yet, when viewed nonperturbatively, this functional integral is apparently doomed to describe a free theory as a consequence of the Central Limit Theorem.

Based on a rigorous analysis, we can find a different interpretation of this formal functional integral allowing for more general $\hbar$-dependent counterterms, and thereby determine a fully satisfactory solution that is described by the following set of equations:

$$S_{UL}(h) = \exp\{b\textstyle\int d^s\mathbf{x} \int c(\lambda)[\mathsf{T}\cosh(\int h(t,\mathbf{x})\lambda(t)dt) - 1]\, c(\lambda)\, d\lambda\}\ ,$$

where b is a new positive parameter chosen, in the present case, with dimensions L^{-s}, T denotes time ordering,

$$\lambda(t) \equiv e^{\hbar t}\lambda e^{-\hbar t}\ ,$$

and finally the Hamiltonian $\hbar$ for each point is given by

$$\hbar \equiv -\tfrac{1}{2}b^{-1}\hbar^2(\partial^2/\partial\lambda^2) + \tfrac{1}{2}m^2 b\lambda^2 + gb^3\lambda^4 + \tfrac{3}{8}b^{-1}\hbar^2\lambda^{-2}\ ;$$

the connection of m_0 with m and of g_0 with g is given later.

Rather than working with the full spacetime path integral, we shall focus on the *ground state* for the ultralocal models. It is basically clear that any system can be equally well defined by means of the lattice regularized ground state, or the lattice regularized Hamiltonian, or, finally, by the lattice regularized Euclidean action function; for the IV model we took the last choice, while for the UL model, we will pick the first choice. The rigorous derivation of the solution given above is indirect and for a thorough discussion of this model, the reader is directed to the literature [Kla00]. Here, we simply present an outline of that derivation:

(1) As noted when discussing the classical properties of the UL models we recall that the dynamics is such that the temporal evolution at one point $\mathbf{x} \in \mathbb{R}^s$ is completely independent of the temporal evolution at any other spatially separated point. As a consequence, the ground state distribution $\Psi_0(\phi)^2$ must formally be of the form $\mathcal{M}\exp[-\int W[\phi(\mathbf{x})]\, d^s\mathbf{x}]$, in units where $\hbar = 1$. Stated otherwise, the characteristic functional for the ground state distribution must be given by

$$\begin{aligned} C(f) &= \mathcal{M}\int e^{\int\{if(\mathbf{x})\phi(\mathbf{x}) - W[\phi(\mathbf{x})]\}d^s\mathbf{x}} \\ &= \exp\{-\textstyle\int L[f(\mathbf{x})]\, d^s\mathbf{x}\}\ . \end{aligned}$$

At this point we begin to see a close connection between the IV and UL models! In particular, the *characteristic functional* of the UL ground state distribution $C(f)$ is analogous to the full spacetime *Schwinger functional* $S(h)$ for the IV model. This similarity is not compromised by the fact that we have used a real test (i.e., source) function h for the Schwinger functional $S(h)$

and a pure imaginary test function f for the characteristic functional of the ground state distribution $C(f)$. Furthermore, this similarity means that we can borrow from the analysis of the IV model to claim that the function $L[p]\,(= L[-p])$ that may arise in $C(f)$ is given by

$$L[p] = \tilde{b}p^2 + \int [1 - \cos(p\lambda)]\, c(\lambda)^2\, d\lambda\ ;$$

if we choose $\tilde{b} > 0$ and $c(\lambda) = 0$, we recover the free ultralocal model, so hereafter we limit our discussion to those cases where $\tilde{b} = 0$ and $c(\lambda) > 0$. There are restrictions on the new "model function" $c(\lambda)$ similar to those on the previous model function $C(u)$. In particular, it follows that

$$\int [\lambda^2/(1+\lambda^2)]\, c(\lambda)^2\, d\lambda < \infty\ ,$$

while it may be (and for our cases will be) true that

$$\int c(\lambda)^2\, d\lambda = \infty\ .$$

(2) The time zero stochastic field may be realized by

$$\phi(\mathbf{x}) = \sum_{l=1}^{\infty} \lambda_l\, \delta(\mathbf{x} - \mathbf{y}_l)\ ,$$

where the distribution of the independent variables λ_l is determined by the function $c(\lambda)^2$, and the variables $\mathbf{y}_l$ have a Poisson distribution in $\mathbb{R}^s$ with a unit density.

(3) Local field products are defined by the operator product expansion and lead to the renormalized square as given by

$$\begin{aligned} \phi(\mathbf{x})^2_R &= [b/\delta(\mathbf{0})]\, \phi(\mathbf{x})^2 \\ &= b \sum_{l=1}^{\infty} \lambda_l^2\, \delta(\mathbf{x} - \mathbf{y}_l)\ , \end{aligned}$$

analogous to the story for the IV models. Likewise higher-order local powers are given by

$$\phi(\mathbf{x})^p_R = b^{p-1} \sum_{l=1}^{\infty} \lambda_l^p\, \delta(\mathbf{x} - \mathbf{y}_l)\ .$$

(4) At this point we begin to deviate from the IV model. In particular, the ground state $\Psi_0(\phi)$ is the lowest eigenstate of the Hamiltonian operator $\mathcal{H} = \int \mathcal{H}(\mathbf{x})\, d^s\mathbf{x}$. We also assume that $\mathcal{H}(\mathbf{x}) = \mathcal{K}(\mathbf{x}) + \mathcal{V}(\mathbf{x})$, i.e., a decomposition into local kinetic and potential operators. In the representation in which the field operator is diagonalized, it follows that a very general form for an even potential is given by

$$\mathcal{V}(\mathbf{x}) = \sum_{k=1}^{K} v_k \phi(\mathbf{x})_R^{2k} = \sum_{k,l=1}^{K,\infty} v_k b^{2k-1} \lambda_l^{2k} \delta(\mathbf{x} - \mathbf{y}_l) \ .$$

In an analogous manner, the local kinetic energy is given by

$$\begin{aligned} \mathcal{K}(\mathbf{x}) &= \tfrac{1}{2} \pi_R^2(\mathbf{x}) + \tfrac{1}{2} \sigma \hbar^2 \phi_R^{-2}(\mathbf{x}) \\ &= -\tfrac{1}{2} b^{-1} \hbar^2 \sum_{l=1}^{\infty} \frac{\partial^2}{\partial \lambda_l^2} \delta(\mathbf{x} - \mathbf{y}_l) + \tfrac{1}{2} \sigma b^{-1} \hbar^2 \sum_{l=1}^{\infty} \lambda_l^{-2} \delta(\mathbf{x} - \mathbf{y}_l) \ , \end{aligned}$$

where we have introduced an unusual term in the kinetic energy which scales in the same fashion that the derivative term scales, hence the same factor of $b^{-1}\hbar^2$ applies to both such terms. It also has a dimensionless coefficient σ, which, for the present—if the reader finds this term offensive—could be chosen to be zero. Later, we shall offer the reader compelling arguments for choosing $\sigma = \frac{3}{4}$.

Observe that the Hamiltonian density $\mathcal{H}(\mathbf{x})$ involves two quite separate components, namely, the sum over delta functions involving the spatial point $\mathbf{x}$ and a Poisson-distributed set of spatial points $\{\mathbf{y}_n\}$, along with a differential operator (augmented now by a constant e_0)

$$\boldsymbol{h} \equiv -\tfrac{1}{2} b^{-1} \hbar^2 \frac{\partial^2}{\partial \lambda^2} + \tfrac{1}{2} \sigma b^{-1} \hbar^2 \lambda^{-2} + \sum_{k=1}^{K} v_k b^{2k-1} \lambda^{2k} - e_0 \ .$$

The operator $\boldsymbol{h}$ acts as a Hamiltonian operator at each point $\mathbf{x}$, and correspondingly, the model function $c(\lambda)$ acts as a ground state for the Hamiltonian $\boldsymbol{h}$ at each point $\mathbf{x}$, namely,

$$[-\tfrac{1}{2} b^{-1} \hbar^2 \frac{\partial^2}{\partial \lambda^2} + \tfrac{1}{2} \sigma b^{-1} \hbar^2 \lambda^{-2} + \sum_{k=1}^{K} v_k b^{2k-1} \lambda^{2k} - e_0] \, c(\lambda) = 0 \ .$$

This relation applies for a wide class of Hamiltonians, i.e., a wide class of potentials. To see a simple example, we shall focus on a model that we shall soon identify as the pseudofree model for the ultralocal models. In this case we choose only $v_1 = \frac{1}{2} m^2$ to be nonzero, leading to

$$[-\tfrac{1}{2} b^{-1} \hbar^2 \frac{\partial^2}{\partial \lambda^2} + \tfrac{1}{2} \sigma b^{-1} \hbar^2 \lambda^{-2} + \tfrac{1}{2} b \, m^2 \lambda^2 - e_0] \, \bar{c}(\lambda) = 0 \ .$$

In fact, the desired analytic solution to this equation leads to

$$\bar{c}(\lambda) = \sqrt{b} \, \frac{e^{-bm\lambda^2/2\hbar}}{|\lambda|^{\gamma}} \ ,$$

where $\sigma \equiv \gamma(\gamma + 1)$ and $e_0 = \hbar m(\gamma - \frac{1}{2})$. Hence, if $\gamma = \frac{1}{2}$, then $e_0 = 0$.

If we identify $\bar{c}(\lambda)$ as a model function, it follows that we must choose $\gamma \geq \frac{1}{2}$ in order that our field operator has a purely continuous spectrum. In this work we shall focus on the case that $\gamma = \frac{1}{2}$. [**Remark:** In [Kla00] the entire acceptable range where $1 \leq 2\gamma < 3$ is analyzed, but in fact all such cases may be obtained by invertible coordinate transformations from just the case

$\gamma = \frac{1}{2}$.] With attention focused on $\gamma = \frac{1}{2}$, it then follows that $\sigma = \frac{1}{2}(1 + \frac{1}{2}) = \frac{3}{4}$, as promised. The resultant model described by the model function $\bar{c}(\lambda)$ can be identified as the pseudofree model for the general ultralocal model $\boldsymbol{\hbar}$. This choice leads to the pseudofree characteristic function for the ground state distribution given (with $\hbar = 1$) by

$$\langle 0| e^{i\int f(\mathbf{x})\,\hat{\phi}(\mathbf{x})\,d^s x}\,|0\rangle = \exp(-b\int d^s x \int \{1 - \cos[f(\mathbf{x})\,\lambda]\}\, e^{-bm\lambda^2}\, d\lambda/|\lambda|\,) \ .$$

Interacting models, such as a quartic ultralocal self-interaction, involve model functions that are solutions of the ground state differential equation when an additional term in the potential, i.e., $v_2 b^3 \lambda^4$, is included. Unfortunately, the analytic form of such a model function is unknown. Thus, let us choose as an alternative interacting model the case where

$$\begin{aligned}&\langle 0| e^{i\int f(\mathbf{x})\,\hat{\phi}(\mathbf{x})\,d^s x}\,|0\rangle \\ &\quad = \exp(-b\int d^s x \int \{1 - \cos[f(\mathbf{x})\,\lambda]\}\, e^{-bm\lambda^2 - gb^3\lambda^4}\, d\lambda/|\lambda|\,) \ .\end{aligned}$$

Just as was the case for the independent model, it is possible to exhibit a divergence-free, term-by-term expansion of the interaction about the pseudofree model in this case, as given by

$$\begin{aligned}&\langle 0| e^{i\int f(\mathbf{x})\,\hat{\phi}(\mathbf{x})\,d^s x}\,|0\rangle \\ &= \exp(-\textstyle\sum_{l=0}^{\infty}(l!)^{-1}(-gb^3)^l b\int d^s x \int \{1 - \cos[f(\mathbf{x})\,\lambda]\}\,\lambda^{4l}\, e^{-bm\lambda^2}\, d\lambda/|\lambda|\,) \ .\end{aligned}$$

This relation is important because, as argued below, finite moments evaluated at a sharp time imply a corresponding finiteness of the full spacetime moments.

This concludes our discussion of both the independent models and the ultralocal models. Many more details about these models may be found in [Kla00].

11.3.5 An alternative approach to both the IV and UL models—an overview with details to follow

The solutions found for the IV and UL models discussed above were not obtained by actually doing the integration implicit in the functional integral expression, but they were in fact obtained—as with the evaluation of most integrals—by *indirect means*. An additional explanation of the phrase *indirect means* is offered by the following example of a one-dimensional integral given by

$$T(s) = \frac{1}{2\pi}\int_0^{2\pi} e^{is\cos(\theta)}\, d\theta \ .$$

There are several different ways to evaluate this integral: One is to expand the integrand in a power series in s, evaluate the integrals term by term,

and sum up the result; a second approach is to take derivatives of $T(s)$ with respect to s and show that $T(s)$ solves a second-order differential equation with respect to s subject to the elementary boundary conditions that $T(0) = 1$ and $T'(0) = 0$. Either of these methods leads to the conclusion that the integral in question is given by $T(s) = J_0(s)$, namely, the zeroth-order Bessel function of the first kind. Note well that this result for the value of the integral has *not* been obtained by summing up the area under the curve. Instead, it has been obtained *indirectly*, i.e., by completely independent means. The point to be emphasized is that the integrals for the IV and UL models as discussed above have also been obtained *indirectly*. A complete evaluation in these two field theoretic examples has been possible thanks to the vast symmetry inherent in these very special models.

When we come to analyze the relativistic models of real interest we will not have anything like the necessary symmetry to carry out the functional integral for such models by similar indirect means. In the face of that remark, we would have to admit that we would not know how to proceed. The story of trying to find an alternative quantization procedure for the relativistic models would seem to be at a dead end—*except for a remarkable fact that has only recently emerged!*

The remarkable fact is that a new and very simple argument has been found to quantize both the IV and the UL models within the context of a regularized version of the functional integral itself, and to do so in such a way that the very same answer that was obtained without regularization, but which was based entirely on special symmetry properties, is regained. The beauty of this new procedure is that it does not significantly rely on the existence of very special symmetry properties, therefore it can be applied more generally to additional models, including, we believe, the class of self-interacting, nonrenormalizable, relativistic scalar quantum fields, such as the φ_n^4 models with $n \geq 5$. It is the outline of this new argument that is offered below when we discuss the quantization of the nonrenormalizable relativistic models.

11.4 An Alternative Method to Solve the IV and UL Models

11.4.1 A reexamination of IV models

We first discuss the IV models. To regularize the formal functional integral, we introduce an n-dimensional, hypercubic spacetime lattice with lattice spacing $a > 0$ and with $L < \infty$ sites on an edge. For the free IV model, the regularized functional integral for the generating functional becomes

$$\begin{aligned} S_f(h) &= K \int e^{\Sigma_k h_k \phi_k a^n - \frac{1}{2} \Sigma_k m_0^2 \phi_k^2 a^n} \Pi_k d\phi_k \\ &= e^{\frac{1}{2} \Sigma_k m_0^{-2} h_k^2 a^n} , \end{aligned}$$

where $k = \{k_0, k_1, \ldots, k_s\} \in \mathbb{Z}^n$.

Moments of mass-like perturbations in the associated probability distribution are given by

$$I_p(m_0) \equiv K \int [\Sigma_k \phi_k^2 \, a^n]^p \, e^{-\frac{1}{2} \Sigma_k m_0^2 \phi_k^2 \, a^n} \, \Pi_k \, d\phi_k \; .$$

These integrals can be evaluated exactly, but let us examine them by steepest descent arguments. For that purpose we pass to *hyperspherical coordinates* given by

$$\phi_k \equiv \kappa \eta_k \; , \qquad \Sigma_k \phi_k^2 = \kappa^2 \; , \qquad \Sigma_k \eta_k^2 = 1 \; ,$$
$$\kappa \geq 0 \; , \qquad -1 \leq \eta_k \leq 1 \; ,$$

which leads to

$$I_p(m_0) = 2K \int [\kappa^2 \, a^n]^p \, e^{-\frac{1}{2} m_0^2 \kappa^2 \, a^n} \, \kappa^{N-1} \, d\kappa \, \delta(1 - \Sigma_k \eta_k^2) \, \Pi_k \, d\eta_k \; .$$

Based in part on a steepest descent argument, it follows that

$$I_p(m_0) = O(\, (N/m_0^2)^p \,) \, I_0(m_0) \; ,$$

which leads to a divergence in the continuum limit [$a \to 0$ and $L \to \infty$ (and $N \equiv L^n \to \infty$ as well), with La fixed and large] that becomes larger as p increases. A perturbation calculation of the value $I_1(M_0)$ expanded about $I_1(m_0)$ is given by

$$I_1(M_0) = I_1(m_0) - \tilde{\delta} \, I_2(m_0) + \tfrac{1}{2} \tilde{\delta}^2 \, I_3(m_0) + \cdots \; ,$$

where $\tilde{\delta} \equiv M_0^2 - m_0^2$, and this series is seen to have ever-increasing divergences as $N \to \infty$. **Note well: The source of those divergences is clearly identified as the factor N in the term κ^{N-1} arising from the measure. If we could somehow replace the factor N in that part of the measure by a positive, finite number—call it R—then all the divergences would vanish**. Of course, the factor κ^{N-1} in the measure is fixed and it cannot be changed directly. But it can be changed *indirectly* by choosing an alternative weighting for the integrand that differs from the original one by a factor that involves $\hbar$. Roughly speaking, let us assume that the "correct" form of the integrand, as modified by the appropriate counterterm, leads to the substitution of the original weighting proportional to

$$e^{-\frac{1}{2} m_0^2 \kappa^2 \, a^n}$$

by an alternative weighting factor proportional to

$$\kappa^{-(N-R)} \, e^{-\frac{1}{2} m_0^2 \kappa^2 \, a^n} \; ,$$

where R is a positive, finite factor to be determined. Respecting the symmetry of the IV model, this change is really taken to be given by

$$e^{-\frac{1}{2}m_0^2\kappa^2\,a^n} \to [\Pi_k\,|\phi_k|^{(1-R/N)}]^{-1}\,e^{-\frac{1}{2}m_0^2\kappa^2\,a^n}\ ,$$

which then leads to a new form for the lattice version of the modified generating function given by

$$\int \Pi_k\,K'\,e^{h_k\phi_k a^n}\,\frac{e^{-\frac{1}{2}m_0^2\phi_k^2 a^n}}{|\phi_k|^{(1-R/N)}}\,d\phi_k$$
$$= \Pi_k\Big\{\,1 + K'\textstyle\int[e^{h_k\phi_k a^n} - 1]\,\dfrac{e^{-\frac{1}{2}m_0^2\phi_k^2 a^n}}{|\phi_k|^{(1-R/N)}}\,d\phi_k\Big\}\ .$$

Next we must choose the parameter R in order to proceed. The principle we adopt is that the given expression must have a meaningful continuum limit as $a \to 0$ and $L \to \infty$ such that the spacetime volume $V = N a^n$ remains finite. It can be as large as required, but it should be finite. Afterwards, if it is needed, a limit in which $V \to \infty$ may be introduced, but we shall not find that necessary. The key to choosing R so as to ensure we have a meaningful continuum limit is connected with the relation between the normalization constant K' and the parameter R. As written it should be reasonably clear that we would like $K' = ba^n$ in order that we obtain a proper integration measure in the limit. To arrange that K' has this value it suffices to consider

$$1 = (ba^n)\int \frac{-\frac{1}{2}m_0^2\phi^2\,a^n}{|\phi|^{(1-R/N)}}\,d\phi\ ,$$

which, after choosing $m_0^2 = ba^n\,m^2$ and a change of variables $a^n\,\phi \to u$, leads effectively to

$$1 = (ba^n)\int \frac{e^{-\frac{1}{2}m^2 b u^2}}{|u|^{(1-R/N)}}\,du\ .$$

In turn we can roughly evaluate this as

$$1 = 2(ba^n)\int_0^U [u^{(1-R/N)}]^{-1}\,du = (2N/R)(ba^n)U^{R/N}\ ,$$

which, to leading order, and independently of $U < \infty$, gives $R = 2N(ba^n)$. We accept that choice, and it follows that the continuum limit of the generating functional becomes

$$S_{pf}(h) = \exp(b\textstyle\int d^n x\int\{\cosh[h(x)\,u] - 1\}\,e^{-\frac{1}{2}m^2 b u^2}\,du/|u|)\ ,$$

which is exactly the expression derived before for the pseudofree IV model! With this expression in hand, and assuming for simplicity that the function

$h(x)$ has compact support, this result applies equally well for an infinite space-time volume.

Let us briefly summarize the principal steps that so quickly and painlessly led us to the desired result for the pseudofree IV model. Starting with the free IV model, we observed that mass-like moments of the free model distribution led to divergences in the continuum limit. With the help of symmetry and the use of hyperspherical coordinates, we were able to uniquely identify the very parameter that was the cause of those divergences, namely, the factor N in the measure of the hyperspherical radius: $\kappa^{N-1}\, d\kappa$. While such a measure term is sacred, we can nevertheless design a distribution for a suitable pseudofree IV model so that it cancels that factor N before it tends to infinity and replaces it with a positive, finite factor R. There are many ways that such a cancellation can take place, and we found that the underlying symmetry of the IV models dictated a very specific form for the entire pseudofree IV distribution apart from fixing the new factor R. As a final step, we chose R in order to ensure that the continuum limit had a meaningful form—and even that choice still left us the freedom of an additional, positive, dimensionless multiple of R that we arbitrarily chose to be unity in the present analysis.

Note well that replacing the diverging factor N by the fixed, finite factor R in the integration measure does *not* mean that we have dropped or eliminated any of the integration variables in the problem. In fact, the change mentioned has arisen from the introduction of a special factor as a modifier of the weight function of the appropriate integrand: nothing more and nothing less.

It is important to emphasize that the *source of the divergences for the IV model was identified within the functional integral formulation*—particularly by using hyperspherical coordinates—which then exposed the factor N in the measure of the hyperspherical radius as the source; isolating the source would be much less evident, or perhaps not evident at all, in other coordinate choices. After removing that source by invoking a possible pseudofree model, evaluation of the resultant Schwinger functional for the pseudofree model was found, somewhat indirectly, by insisting that the continuum limit had a nontrivial behavior. Although the functional integral evaluation has indeed again been indirect, the initial formulation of the basic problem as a lattice space integral in special coordinates was, most likely, *indispensable* in identifying the source of the divergences!

We shall see below that a similar story can be used to choose the ground state distribution for the pseudofree ultralocal model in an equally rapid and painless fashion. This should not be surprising because the IV and UL models have a great deal in common, but it does suggest that this same procedure—which was referred to in Chapter 1 as the "simple surgical procedure"—may have a wider application than just to models of high symmetry. Indeed, modulo the step in choosing the explicit form of the pseudofree ground state for the model in question, we shall argue that this procedure—which we unceremoniously call *measure mashing*—suggests a possible procedure to eliminate the divergences in relativistic, nonrenormalizable scalar quantum field theory.

But we get ahead of ourselves. Let us first reexamine the UL models.

11.4.2 A reexamination of UL models

In light of the previous discussion regarding the IV models, we can significantly compress the usual analysis of the UL models. On a hypercubic spatial lattice, the characteristic function for the lattice regularized ground state distribution of the free UL model is given by

$$\begin{aligned} C(f) &= M \int e^{i \Sigma'_k f_k \phi_k \, a^s - m_0 \Sigma'_k \phi_k^2 \, a^s} \; \Pi'_k d\phi_k \\ &= e^{-\frac{1}{4} m_0^{-1} \Sigma'_k f_k^2 a^s} \; , \end{aligned}$$

and moments of mass-like terms in this lattice regularized ground state distribution are given by

$$I'_p(m_0) = M \int [\Sigma'_k \phi_k^2 \, a^s]^p \, e^{-m_0 \Sigma'_k \phi_k^2 \, a^s} \; \Pi'_k d\phi_k \; ;$$

note that primes on the sums and products signify that only those lattice sites on a spatial slice at a fixed Euclidean time are involved. Reexpressing this latter expression in terms of hyperspherical coordinates on a spatial slice of the lattice at fixed Euclidean time, given now by

$$\begin{gathered} \phi_k \equiv \kappa \eta_k \quad , \quad \Sigma'_k \phi_k^2 = \kappa^2 \quad , \quad \Sigma'_k \eta_k^2 = 1 \quad , \\ 0 \le \kappa < \infty \quad , \quad 1 \le \eta_k \le 1 \, , \end{gathered}$$

leads to

$$I'_p(m_0) = 2 M \int [\kappa^2 a^s]^p \, e^{-m_0 \kappa^2 a^s} \; \kappa^{N'-1} d\kappa \, \delta(1 - \Sigma'_k \eta_k^2) \, \Pi'_k d\eta_k \; .$$

A steepest descent argument rapidly leads to

$$I'_p(m_0) = O(\, (N'/m_0)^p) \, I'_0(m_0) \; ,$$

which for all $p \ge 1$ exhibits a divergence in the continuum limit as $N' = L^s \to \infty$. A suitable pseudofree UL ground state distribution should replace

$$e^{-m_0 \kappa^2 a^s}$$

by

$$\kappa^{-(N'-R')} \, e^{-m_0 \kappa^2 a^s} \; .$$

Among the many ways to do so, UL symmetry points to

$$\Psi_{pf}(\phi)^2 = \prod_k{}' M' \frac{e^{-m_0\,\phi_k^2\,a^s}}{|\phi_k|^{(1-R'/N')}}$$

as the right choice for the pseudofree ground state distribution.

In the present case, we choose $m_0 = ba^s m$, and pick R' so that a meaningful continuum limit emerges. In particular, we set $R' = 2ba^s N'$ so that the pseudofree characteristic function becomes

$$\begin{aligned} C(f) &= \lim_{a\to 0} \prod_k{}' \int \left\{ 1 - (ba^s) \int [1 - \cos(f_k\,\lambda)]\, e^{-bm\lambda^2}\, d\lambda/|\lambda| \right\} \\ &= \exp(-b \textstyle\int d^s x \int \{1 - \cos[f(x)\,\lambda]\}\, e^{-bm\,\lambda^2}\, d\lambda/|\lambda|) \ . \end{aligned}$$

Once again, for compactly supported functions $f(x)$, this relation also applies for an infinite spatial volume.

After having determined the pseudofree UL ground state in this fashion, to complete the description of the UL models it suffices to use this ground state to define the pseudofree Hamiltonian operator, and from there to introduce the lattice action for this model.

Additional terms added to the potential of the Hamiltonian operator serve to define interacting UL models, and although we cannot explicitly solve the Hamiltonian operator for a quartic interaction, we can make a quasi-solution that can be useful. In particular the ground state for an interacting model with a quartic interaction is "qualitatively" of the form

$$c(\lambda) \simeq |\lambda|^{-1/2}\, e^{-\frac{1}{2}bm\lambda^2 - \sqrt{2g}\,b^2|\lambda|^3/3} \ ,$$

valid for large $|\lambda|$. It follows from this analysis that $g_0 = (ba^s)^4 g$. In summary, for the UL models, it follows that the original parameters satisfy the relations

$$m_0^2 = [b/\delta(\mathbf{0})]^2 m^2 \ , \qquad\qquad g_0 = [b/\delta(\mathbf{0})]^4 g \ .$$

These values for the bare parameters in terms of more physical quantities differ from those of the IV model due, in part, to contributions from a field strength renormalization factor that our analysis has implicitly ignored. In our study of the relativistic models below, the field strength renormalization factor will play a prominent and important role.

As already noted for the IV models, the practice of "measure mashing" has once again rapidly led to a divergence-free formulation of the UL models.

11.5 Relativistic Nonrenormalizable Scalar Models

At long last, we have arrived at the discussion of the relativistic nonrenormalizable models, namely, the so-called φ_n^4 models for $n \geq 5$. Just as with the IV and UL models, these relativistic models have the property that their classical

pseudofree solution space W_{pf} is a proper subset of the classical free solution space W_f. Just as with the IV and UL models, this fact suggests that the nonlinear interaction term acts partially as a hard core classically and it may do so quantum mechanically as well. For the IV and UL models it was shown above that dealing with the quantum hard core seems to be no more complicated than replacing the factor N or N', respectively, by a finite, positive parameter R or R' in the hyperspherical radius measure, a process we called *measure mashing*. While the classical nature of the hard core seems unduly complicated, namely, deleting fields of the form $\phi_{singular}$ as given above, how is it possible that the changes involved in the quantum analysis seem so different? Our answer to this puzzle is first to note that the fields involved in the classical action terms are generally but a set of measure zero in the quantum measure and second to observe that the simple change of the diverging factor N, N' to the finite factors R, R' actually accomplishes the miracle of rendering the associated IV and UL models divergence free. In brief, from a quantum perspective, it seems that the presence of a classical hard core merely serves to put one on notice that a quantum hard core can be expected, but the details of the classical hard core action, in such complex field problems, appear to get essentially lost when it comes to implementing the quantum hard core. An implementation of the quantum hard core which works is one that renders the theory divergence free, and for the IV and UL models, that implementation is primarily—and indisputably—the replacement of the diverging factor N, N' in the measure by a well-chosen, finite factor R, R'. Accepting this fact as a robust principle that deserves to have a wider application than just the IV and UL models will be our guiding principle in a proposed analysis of the relativistic, nonrenormalizable, scalar models. Only further investigation—which is beyond the scope of this monograph—will demonstrate whether this hypothesis is valid or not.

11.5.1 Motivation for alternative studies of relativistic nonrenormalizable models

Nonrenormalizable quantum field models, such as φ_n^4, with a spacetime dimension $n \geq 5$, require the introduction of nonclassical, nontrivial counterterms to avoid triviality [Aiz81, Fro82]. A conventional regularized perturbation theory is designed to cancel divergences one-by-one as they are encountered in a perturbation analysis, and as such they lead to an infinite set of higher-order contributions that have an ever more divergent behavior. However, a suitable summation of perturbative Feynman graphs can be shown to support the triviality of such models. In hope of attaining a meaningful and nontrivial formulation of such examples, it is necessary to go beyond the class of counterterms suggested by perturbation theory about the free theory. Our approach to this huge task is to design a suitable pseudofree theory on the basis of measure mashing as was shown to be successful in the IV and UL

cases. Let us first examine the free theory and try to isolate the cause of the divergences in the ground state distribution.

11.5.2 The free ground state distribution

For reasons of simplicity and clarity, we focus on the ground state distribution for a free, *massless*, relativistic scalar model on a hypercubic lattice such as we have used before, but our first task is to recall the analysis of massless lattice models. On the one hand, in the continuum, a massless theory in spatial dimension $s \geq 2$ defines a proper spatial kernel since it involves the Fourier transform of the measure $d^s p/|p|$, which near the origin is integrable (despite the fact that the integrand diverges at $p = 0$). Our spatial dimensions of interest will certainly qualify, but, on the other hand, when such a system is put on a lattice space with a lattice spacing $a > 0$, there is a "zero mode" to contend with. The zero mode is where all the fields are translated by an equal amount, such as $\phi_k \rightarrow \phi_k + c$. The ground state of a free lattice massless theory is invariant under such a transformation, and consequently the zero mode is normally omitted from the calculation. Instead of omitting the zero mode, we prefer to add a small artificial mass-like term. In a discrete Fourier series representation of the field values that is dual to the spatial lattice representation, the addition of this small artificial term is straightforward; see below. The added mass-like term is chosen so it will automatically disappear in the continuum limit.

The characteristic functional for the ground state distribution in this case is therefore given (for $\hbar = 1$) by

$$C(f) = M \int e^{i\Sigma'_k f_k \phi_k a^s \,-\, \Sigma'_{k,l} \phi_k A_{k-l} \phi_l a^{2s}} \, \Pi'_k d\phi_k \ ,$$

where A_{k-l} is a positive definite matrix designed to yield the spatial gradient terms plus a small artificial mass-like term. In particular, two derivatives of the lattice ground state by the lattice Hamiltonian, which has a derivative term of the form

$$\mathcal{H} = -\tfrac{1}{2} a^{-s} \textstyle\sum'_k \dfrac{\partial^2}{\partial \phi_k^2} + \cdots \ ,$$

lead to a contribution to the lattice Hamiltonian of the form

$$\tfrac{1}{2} \Sigma'_{k,l} \, \phi_k \, [\Sigma'_j A_{k-j} \, A_{j-l}] \, \phi_l \, a^{3s} \ .$$

In turn, this expression must equal the conventional spatial lattice gradient squared term plus our small artificial mass-like term, i.e.,

$$\tfrac{1}{2} \Sigma'_k \, \Sigma'_{k^*} (\phi_{k^*} - \phi_k)^2 \, a^{s-2} + \tfrac{1}{2} s (L^{-2s} a^{-2}) \, \Sigma'_k \phi_k^2 \, a^s \ ,$$

where the sum over k^* (here made explicit!) includes each of the s nearest neighbors in a positive sense in all spatial directions away from the lattice point k. In order for this relation to be true, it is necessary that the matrix A_{k-l} be proportional to $a^{-(s+1)}$, a fact which will be of importance subsequently.

11.5.3 Properties of the matrix A_{k-l}

It is of some interest to see exactly just what the matrix elements A_{k-l} are given by. We introduce a discrete Fourier series representation given by the general transformations

$$\begin{aligned} y_{(l)} &\equiv \textstyle\sum_k' e^{-i2\pi k\cdot l/L}\, Y_k \ , \\ Y_k &\equiv L^{-s} \textstyle\sum_l' e^{i2\pi k\cdot l/L}\, y_{(l)} \ , \end{aligned}$$

and we say that these expressions constitute a transform pair, $Y_k \Longleftrightarrow y_{(l)}$. The range of the sums in these expressions is given by $0 \le k_j, l_j \le L-1$ for each $j = 1, \ldots, s$. Additionally, we extend the range of the labels k on Y_k in a periodic fashion compatible with the second line of the transform pair above. We also use the convolution rule

$$\textstyle\sum_m' X_{k-m} Y_m \Longleftrightarrow x_{(l)}\, y_{(l)} \ .$$

Using a periodic extension where necessary, it follows that

$$C_{k-l} \equiv sL^{-2s}\,\delta_{k-l,0} + \Sigma_{j=1}^{s}[2\delta_{k-l,0} - \delta_{k-l,1_j} - \delta_{k-l,-1_j}]$$

corresponds to the gradient squared plus the small auxiliary mass matrix, less an overall factor of a^{s-2}. We define

$$\begin{aligned} c_{(l)} &= \textstyle\sum_k' e^{-i2\pi k\cdot l/L}\, C_k \\ &= sL^{-2s} + 2\,\Sigma_{j=1}^{s}[1-\cos(2\pi l_j/L)] \ . \end{aligned}$$

Thus, with $\overline{A}_{k-l} \equiv a^{s+1} A_{k-l}$, we have

$$\begin{aligned} \overline{A}_k &= L^{-s}\textstyle\sum_l' e^{i2\pi k\cdot l/L}\,\{c_{(l)}\}^{1/2} \\ &= L^{-s}\textstyle\sum_l'\{sL^{-2s} + 2\,\Sigma_{j=1}^{s}[1-\cos(2\pi l_j/L)]\}^{1/2}\,\cos(2\pi k\cdot l/L) \ . \end{aligned}$$

In the limit of very large L, this expression may be replaced by an integral, where we let $l_j/L \to y_j$, as $L\to\infty$, with $0 \le y_j < 1$, for each j. Thus,

$$\overline{A}_k = \sqrt{2}\textstyle\int d^s y\{\Sigma_{j=1}^{s}[1-\cos(2\pi y_j)]\}^{1/2}\,\cos(2\pi k\cdot y) \ ,$$

and using the relation $\{a+b+c+\cdots\}^{1/2} \le a^{1/2}+b^{1/2}+c^{1/2}+\cdots$, valid for nonnegative a, b, c, etc., it follows that

$$|\overline{A}_k| \le s\sqrt{2}\textstyle\int_0^1[1-\cos(2\pi u)]^{1/2}\,du \le s\sqrt{2} \ .$$

For future reference, we introduce

$$f(L) \equiv \max_k\,|\overline{A}_k| = \max_k\, a^{s+1}\,|A_k| \ ,$$

which we expect to be $O(1)$, a value consistent with the large L behavior derived above.

11.5.4 Mass-like moments in the ground state distribution

It follows that mass-like moments in the free, almost massless ground state distribution are given by expressions of the form

$$J_p(A) \equiv M \int [\Sigma'_k \phi_k^2 \, a^s]^p \, e^{-\Sigma'_{k,l} \phi_k A_{k-l} \phi_l \, a^{2s}} \, \Pi'_k d\phi_k \ .$$

This relation is similar in structure to an analogous relation for the UL model save for the fact that the exponent is no longer of the form $\Sigma'_k \phi_k^2 \, a^s$. In fact, we can change variables from $\{\phi_k\}$ to $\{\xi_k\}$, where $\phi_k = \Sigma'_l B_{k,l} \, \xi_l$, and the matrix B represents a similarity transformation such that

$$\Sigma'_{k,l} \phi_k A_{k-l} \phi_l \, a^{2s} = \mathsf{m} \Sigma'_l \xi_l^2 \, a^s \ .$$

Here m is an appropriate function of the lattice spacing and the lattice size. Reexpressing the integral for the moments of the almost massless ground state distribution in the ξ variables leads to

$$J_p(A) = \tilde{M} \int [\Sigma'_{k,l} \xi_k B^{(2)}_{k,l} \xi_l \, a^s]^p \, e^{-\mathsf{m} \Sigma'_l \xi_l^2 \, a^s} \, \Pi'_l d\xi_l \ ,$$

an expression that more nearly resembles the relations studied previously. In either form of the integral, it follows for nonrenormalizable field theories that these integrals diverge, much like $O(N'^p)$ as before. In terms of hyperspherical variables, the new integrals do not factor into strictly hyperspherical radius integrals times spectator $\{\eta_k\}$ integrals; nevertheless, we still wish to identify the factor N' in the hyperspherical radius measure as the primary cause of the divergences. To remove divergences caused by integration over κ, we will eliminate the same factor N' in the measure as before, and for that we follow the lead of the IV and UL models. Other factors will need to be changed as well, but they will be discussed in due course.

11.5.5 Mashing the relativistic measure

We return to our original set of ϕ_k variables and write once again the free lattice ground state distribution given by

$$M \, e^{-\Sigma'_{k,l} \phi_k A_{k-l} \phi_l \, a^{2s}}$$

and recognize that we wish to choose a pseudofree lattice ground state distribution qualitatively of the form

$$M' \, \kappa^{-(N'-1)} \, e^{-\Sigma'_{k,l} \phi_k A_{k-l} \phi_l \, a^{2s}} \ ,$$

where we have chosen the factor R (or R') previously used to be simply unity. To make a concrete proposal for the pseudofree ground state, we might first try

$$M' \frac{e^{-\frac{1}{2}\Sigma'_{k,l}\,\phi_k\,A_{k-l}\,\phi_l\,a^{2s}}}{\Pi'_k\,|\phi_k|^{(1-1/N')/2}} , \qquad wrong! ,$$

that was successful for the UL models, but this suggestion fails on several accounts. Although this expression has the desired measure mashing property, it also has incipient divergences for each k at $\phi_k = 0$ as $N' \to \infty$. This feature was required by symmetry for the UL model and the divergence was welcome in constructing the eventual generalized Poisson ground state distribution, but for the relativistic case, we do expect neighboring spatial sites to "talk" to each other and we do *not* want to construct a generalized Poisson distribution for such models. The way out of this situation is to propose another form for the pseudofree ground state given by

$$\Psi_{pf-prov}(\phi) = M' \frac{e^{-\frac{1}{2}\Sigma'_{k,l}\,\phi_k\,A_{k-l}\,\phi_l\,a^{2s}}}{\Pi'_k\,[\Sigma'_l\,J_{k,l}\phi_l^2]^{(1-1/N')/4}} , \qquad provisional! ,$$

where the constants $J_{k,l}$ will be chosen momentarily. Observe that the modified denominator still has the feature that it contributes the desired measure mashing factor $\kappa^{-(N'-1)}$ to the pseudofree ground state distribution. But more than that, the new version makes a certain *average* controlled by the constants $J_{k,l}$ over several different field values. If there are enough terms in such an average, this will ensure that there is *no* incipient divergence of the integral should all the values in a given denominator factor vanish simultaneously. To ensure this property holds throughout the lattice it is necessary that the field at point k gets a contribution from other fields in every spatial direction. It would be adequate to choose just one extra field in each spatial direction, but then the natural lattice symmetry would be unnecessarily broken. Instead, we choose to include every nearest neighbor in a spatial sense, within every average. Specifically, we choose

$$J_{k,l} \equiv \frac{1}{2s+1}\,\delta_{k,l\in\{k\cup k_{nn}\}} ,$$

where $\delta_{k,l}$ is a Kronecker delta. This notation means that an equal weight of $1/(2s+1)$ is given to the $2s+1$ points in the set composed of k and its $2s$ nearest neighbors in the spatial sense only; $J_{k,l} = 0$ for all other points in that spatial slice. [Specifically, we define $J_{k,l} = 1/(2s+1)$ for the points $l = k = (k_0, k_1, k_2, \ldots, k_s)$, $l = (k_0, k_1 \pm 1, k_2, \ldots, k_s)$, $l = (k_0, k_1, k_2 \pm 1, \ldots, k_s)$,..., $l = (k_0, k_1, k_2, \ldots, k_s \pm 1)$.] This definition implies that $\Sigma'_l J_{k,l} = 1$.

With the provisional pseudofree ground state chosen it follows that the provisional pseudofree ground state distribution is given by

$$\Psi_{pf-prov}(\phi)^2 = M \frac{e^{-\Sigma'_{k,l}\,\phi_k\,A_{k-l}\,\phi_l\,a^{2s}}}{\Pi'_k\,[\Sigma'_l\,J_{k,l}\phi_l^2]^{(1-1/N')/2}} , \qquad provisional! ,$$

where M denotes a normalization constant appropriate to each case in which it appears.

Of course, there are many alternative choices that could be made for the denominator contribution. To spread out the reach of $J_{k,l}$ further does not seem to offer an improvement, and it would seem that in the continuum limit this distinction would most likely disappear. One could replace the factor $[\Sigma'_l J_{k,l}\,\phi_l^2]$ by $[\Sigma'_l J_{k,l}\,\phi_l^4]^{1/2}$, but this change may or may not add anything new to the mix; it seems prudent to choose the first version and see to what it might lead before diluting the study further. Moreover, we have to see what counterterm in the Hamiltonian we are led to on the basis of our favoring the provisional pseudofree ground state.

11.5.6 Lattice Hamiltonian for the free and pseudofree models

Given any ground state wave function—let us call it $\Psi_0(\phi)$—it is straightforward to derive the Hamiltonian operator from the general expression

$$\mathcal{H} = -\tfrac{1}{2}\hbar^2\, a^{-s} \textstyle\sum'_k \dfrac{\partial^2}{\partial\phi_k^2} + \tfrac{1}{2}\hbar^2 \dfrac{a^{-s}}{\Psi_0(\phi)} \textstyle\sum'_k \dfrac{\partial^2 \Psi_0(\phi)}{\partial\phi_k^2}\ .$$

For the free relativistic model, the ground state is given by

$$\Psi_f(\phi) = M\, e^{-(1/2\hbar)\Sigma'_{k,l}\,\phi_k\, A_{k-l}\,\phi_l\, a^{2s}}\ ,$$

and the associated Hamiltonian reads

$$\begin{aligned}\mathcal{H}_f = &-\tfrac{1}{2}\hbar^2\, a^{-s} \textstyle\sum'_k \dfrac{\partial^2}{\partial\phi_k^2} + \tfrac{1}{2}\textstyle\sum'_k \sum'_{k^*} (\phi_{k^*} - \phi_k)^2\, a^{s-2} \\ &+\tfrac{1}{2}\, s\,(L^{-2s} a^{-2}) \textstyle\sum'_k \phi_k^2\, a^s - \tfrac{1}{2}\hbar N' A_0\, a^s\ ,\end{aligned}$$

as expected. This Hamiltonian has all the appearance of a covariant expression as it would appear on a lattice space.

We next analyze the provisional pseudofree ground state given by

$$\Psi_{pf-prov}(\phi) = M' \frac{e^{-\frac{1}{2}\Sigma'_{k,l}\,\phi_k\, A_{k-l}\,\phi_l\, a^{2s}}}{\Pi'_k\, [\Sigma'_l J_{k,l}\phi_l^2]^{(1-1/N')/4}}\ ,$$

and if we accept this expression, the associated—and provisional—pseudofree Hamiltonian is given by

$$\begin{aligned}\mathcal{H}_{pf-prov} = &-\tfrac{1}{2}\hbar^2\, a^{-s} \textstyle\sum'_k \dfrac{\partial^2}{\partial\phi_k^2} + \tfrac{1}{2}\textstyle\sum'_k \sum'_{k^*} (\phi_{k^*} - \phi_k)^2\, a^{s-2} \\ &+\tfrac{1}{2}\, s\,(L^{-2s} a^{-2}) \textstyle\sum'_k \phi_k^2\, a^s \\ &+\tfrac{1}{2}\hbar^2 \textstyle\sum'_k \mathcal{F}_k(\phi)\, a^s + \tfrac{1}{2}\hbar \textstyle\sum'_k w_k(\phi)\, a^s\ .\end{aligned}$$

In this expression, the factor $\mathcal{F}_k$ arises from two derivatives acting on the denominator factor in $\Psi_{pf-prov}(\phi)$, which leads to

$$\mathcal{F}_k(\phi) \equiv \frac{1}{4}\left(\frac{N'-1}{N'}\right)^2 a^{-2s}\sum\nolimits'_{r,t}\frac{J_{r,k}\,J_{t,k}\,\phi_k^2}{[\Sigma'_l\,J_{r,l}\,\phi_l^2][\Sigma'_m\,J_{t,m}\,\phi_m^2]}$$
$$-\frac{1}{2}\left(\frac{N'-1}{N'}\right)a^{-2s}\sum\nolimits'_t\frac{J_{t,k}}{[\Sigma'_m\,J_{t,m}\,\phi_m^2]}$$
$$+\left(\frac{N'-1}{N'}\right)a^{-2s}\sum\nolimits'_t\frac{J_{t,k}^2\,\phi_k^2}{[\Sigma'_m\,J_{t,m}\,\phi_m^2]^2}\ .$$

This rather cumbersome expression represents the "counterterm" added to the Hamiltonian that generates the measure mashing term. This cumbersome term is exactly what we want to have in our pseudofree Hamiltonian. However, there is another term to be dealt with denoted by $w_k(\phi)$. This term arises from one derivative acting on the factor in the exponential and the other derivative acting on the denominator; *this field-dependent cross term is decidedly not wanted in the Hamiltonian!* How do we get rid of it?

To resolve the cross term issue, we now need to make the pseudofree Hamiltonian the primary quantity, and we deliberately choose it to have the form

$$\mathcal{H}_{pf} = -\tfrac{1}{2}\hbar^2\,a^{-s}\sum\nolimits'_k\frac{\partial^2}{\partial\phi_k^2} + \tfrac{1}{2}\sum\nolimits'_k\sum\nolimits'_{k^*}(\phi_{k^*}-\phi_k)^2\,a^{s-2}$$
$$+\tfrac{1}{2}\,s\,(L^{-2s}a^{-2})\sum\nolimits'_k\phi_k^2\,a^s$$
$$+\tfrac{1}{2}\hbar^2\sum\nolimits'_k\mathcal{F}_k(\phi)\,a^s - E_0\ .$$

In this expression we have retained the usual terms that make up the free Hamiltonian plus the nonclassical (i.e., $\hbar$ dependent) counterterm $\mathcal{F}_k(\phi)$ that has the feature that it is homogeneous in ϕ^{-2}, which is exactly the property that arose in our analysis of the UL model. We insist that the unwanted term based on $w_k(\phi)$ does *not* appear and that the only other term allowed is a constant, E_0, that represents the ground state energy subtracted from the Hamiltonian. The pseudofree ground state $\Psi_{pf}(\phi)$ satisfies the equation

$$\mathcal{H}_{pf}\,\Psi_{pf}(\phi) = 0\ ,$$

and leads to a solution of the form

$$\Psi_{pf}(\phi) = M\,\frac{e^{-(1/2\hbar)\Sigma'_{k,l}\,\phi_k\,A_{k-l}\,\phi_l\,a^{2s} - W(\phi\,a^{(s-1)/2}/\hbar^{1/2})/2}}{\Pi'_k\,[\Sigma'_l\,J_{k,l}\phi_l^2]^{(1-1/N')/4}}\ .$$

In this expression we have kept the same form of the quadratic term in the exponent, which is the most relevant term for very large field values, and we have kept the terms in the denominator that are the most relevant for very small field values. We have added the term $W(\phi\,a^{(s-1)/2}/\hbar^{1/2})$, which modifies the mixed derivative terms so that in the Hamiltonian, they become a simple constant E_0. The dependence of $\mathcal{H}_{pf}$ on the parameters a and $\hbar$ is such that we can determine the dependence of the function W on these parameters. Unfortunately, we do not know the form of the function W itself, so we are

unable to analytically specify the full form of $\Psi_{pf}(\phi)$, but we have nonetheless been able to identify what we want as the counterterm, i.e., $\mathcal{F}_k(\phi)$, which has the desired effect of mashing the relativistic measure by just the right factor!

In the continuum limit, it is important to observe that the form of the counterterm $\mathcal{F}_k(\phi)$ leads to a local covariant potential, albeit an unconventional one.

Next, let us make a brief detour to learn why we have devoted so much attention to the ground state distribution.

11.5.7 The importance of sharp time moments

We have focused on the ground state distribution for the relativistic models for a very good reason. In this section we wish to argue that moments of certain expressions of interest in the full spacetime distribution can be bounded by suitable averages of related quantities in the ground state distribution. In particular, let the full spacetime average on a lattice be given by

$$\langle\, [\Sigma_{k_0} F(\phi, a)\, a]^p \,\rangle \equiv M \int [\Sigma_{k_0} F(\phi, a)\, a]^p \; e^{-I(\phi, a, \hbar)} \; \Pi_k \, d\phi_k \;,$$

where I is the lattice action, Σ_{k_0} denotes a summation over the temporal direction k_0 only, and $F(\phi, a)$ is an expression that depends only on fields ϕ_k at a fixed value of k_0. For example, one may consider $F(\phi, a) = \Sigma'_k \phi_k^4 \, a^s$ or $F(\phi, a) = \Sigma'_{k,l} \, \Omega_{k,l} \, \phi_k \, \phi_l \, a^{2s}$, for some c-number kernel $\Omega_{k,l}$, etc., where the primed sum implies summation over a spatial slice at fixed k_0. It follows that

$$\langle\, [\Sigma_{k_0} F(\phi, a)\, a]^p \,\rangle = \Sigma_{k_0, \ldots, k_0} \, a^p \, \langle\, F(\phi_1, a) \cdots F(\phi_p, a) \,\rangle \;,$$

where each ϕ_j refers to the fields at Euclidean time $k_0 = j$. A straightforward inequality shows that

$$|\langle\, F(\phi_1, a) \cdots F(\phi_p, a) \,\rangle| \leq |\langle\, F(\phi_1, a)^p \,\rangle \cdots \langle\, F(\phi_p, a)^p \,\rangle|^{1/p} \;.$$

Finally, for sufficiently large $N'(ba^s)$, we note that

$$\langle\, F(\phi, a)^p \,\rangle = \int F(\phi, a)^p \, \Psi_0(\phi)^2 \, \Pi'_k \, d\phi_k \;,$$

namely, an average in the ground state distribution. The argument behind the previous equation is as follows. Quite generally,

$$\langle\, F(\phi, a)^p \,\rangle = M_0 \textstyle\sum_l \displaystyle\int \langle \phi | l \rangle \, e^{-E_l \, T} \, \langle l | \, \phi \rangle \, F(\phi, a)^p \; \Pi'_k \, d\phi_k \;,$$

where we have used the resolution of unity $1 = \int |\, \phi \rangle \langle \phi \,| \, \Pi'_k \, d\phi_k$ for states for which $\hat{\phi}(x) \,|\, \phi \rangle = \phi(x) \,|\, \phi \rangle$, as well as the eigenvectors $|\, l \rangle$ and eigenvalues E_l for which $\mathcal{H} \,|\, l \rangle = E_l \,|\, l \rangle$. At this point in the argument it may be helpful to let

the period of the lattice in the temporal direction be larger than the period in the several spatial directions. This can be done by putting L_0 ($L_0 > L$) sites in the temporal direction, and we can then examine this problem for ever larger values of $T = L_0 a$ while keeping the spatial length La fixed. For asymptotically large T, it follows that only the (unique) ground state contributes, and the former expression becomes

$$\langle F(\phi, a)^p \rangle = \int F(\phi, a)^p \, |\langle \phi | 0 \rangle|^2 \, \Pi'_k \, d\phi_k \ ,$$

now with $M_0 = 1$, which is just the expression given above.

In summary, for a finite, hypercubic lattice with periodic boundary conditions, we have derived an important result: If the sharp time average of $[F(\phi, a)]^p$ is finite, then it follows that the spacetime average of $[\Sigma_{k_0} F(\phi, a)\, a]^p$ is also finite.

11.5.8 Statement of the fundamental problem

Given the pseudofree Hamiltonian, it is straightforward to define the lattice Hamiltonian operator for the full theory including the mass term and the quartic coupling constant. The full Hamiltonian is given by

$$\begin{aligned} \mathcal{H} = & -\tfrac{1}{2}\hbar^2 a^{-s} \textstyle\sum'_k \dfrac{\partial^2}{\partial \phi_k^2} + \tfrac{1}{2} \textstyle\sum'_k \sum'_{k^*} (\phi_{k^*} - \phi_k)^2 \, a^{s-2} \\ & + \tfrac{1}{2}\, s\, L^{-2s} a^{-2} \textstyle\sum'_k \phi_k^2 \, a^s \\ & + \tfrac{1}{2}\, m_0^2 \textstyle\sum'_k \phi_k^2 \, a^s + \lambda_0 \textstyle\sum'_k \phi_k^4 \, a^s \\ & + \tfrac{1}{2}\hbar^2 \textstyle\sum'_k \mathcal{F}_k(\phi)\, a^s - E \ . \end{aligned}$$

In turn, the full lattice action arises from the full Hamiltonian, and reads

$$\begin{aligned} I(\phi, a, N) \equiv & \tfrac{1}{2} \textstyle\sum_k \sum_{k^*} (\phi_{k^*} - \phi_k)^2 \, a^{n-2} + \tfrac{1}{2}(m_0^2 + L^{-2s} a^{-2}) \textstyle\sum_k \phi_k^2 \, a^n \\ & + \lambda_0 \textstyle\sum_k \phi_k^4 \, a^n + \tfrac{1}{2}\hbar^2 \textstyle\sum_k \mathcal{F}_k(\phi)\, a^n \ , \end{aligned}$$

in which we have dropped the energy term E as it plays no role in the classical action. Moreover, when the mass term is restored, it becomes less necessary to maintain the small, auxiliary mass-like factor $L^{-2s} a^{-2}$. The lattice action is the principal term in the lattice version of the Schwinger generating functional, which takes the form

$$\begin{aligned} S(h) & \equiv M \int \exp[Z^{-1/2} \, \Sigma_k h_k \, \phi_k \, a^n / \hbar - I(\phi, a, N)/\hbar] \ \Pi_k \, d\phi_k \\ & \equiv \langle e^{Z^{-1/2} \, \Sigma_k h_k \, \phi_k \, a^n / \hbar} \rangle \ , \end{aligned}$$

where $\{h_k\}$ determines an appropriate test sequence, and the normalization factor M ensures that $S(0) = 1$. In the expression for $S(h)$, Z denotes the

field strength renormalization factor, which along with the bare mass m_0^2 and the bare coupling constant λ_0, need to be determined in order to complete the formulation of the model.

Our procedure for determining the dependence of Z, m_0^2, and λ_0 on the lattice parameters—the cutoff-dependent parameters in this lattice formulation—will be to show that moments of the three different factors that contain these terms are finite in the full spacetime distribution, and we do that by studying the corresponding question for the sharp time moments. Insisting on a meaningful continuum limit will fix the dependence of the three unknown constants on the lattice parameters as required.

In what follows, we closely follow reference [Kl08a].

11.6 The Continuum Limit, and Term-by-Term Finiteness of a Perturbation Analysis

Before focusing on the limit $a \to 0$ and $L \to \infty$, we note several important facts about ground state averages of the direction field variables $\{\eta_k\}$. First, we assume that such averages have two important symmetries: (i) averages of an odd number of η_k variables vanish, i.e.,

$$\langle \eta_{k_1} \cdots \eta_{k_{2p+1}} \rangle = 0 \ ,$$

and (ii) such averages are invariant under any spacetime translation, i.e.,

$$\langle \eta_{k_1} \cdots \eta_{k_{2p}} \rangle = \langle \eta_{k_1+l} \cdots \eta_{k_{2p}+l} \rangle$$

for any $l \in \mathbb{Z}^n$ due to a similar translational invariance of the lattice action. Second, we note that for any ground state distribution, it is necessary that $\langle \eta_k^2 \rangle = 1/N'$ for the simple reason that $\Sigma_k' \eta_k^2 = 1$. Hence, $|\langle \eta_k \eta_l \rangle| \leq 1/N'$ as follows from the Schwarz inequality. Since $\langle [\Sigma_k' \eta_k^2]^2 \rangle = 1$, it follows that $\langle \eta_k^2 \eta_l^2 \rangle = O(1/N'^2)$. Similar arguments show that for any ground state distribution

$$\langle \eta_{k_1} \cdots \eta_{k_{2p}} \rangle = O(1/N'^p) \ ,$$

which will be useful in the sequel.

11.6.1 Field strength renormalization

For a suitable spatial test sequence $\{h_k\}$, we insist that expressions such as

$$\int Z^{-p} \, [\Sigma_k' h_k \, \phi_k \, a^s]^{2p} \, \Psi_{pf}(\phi)^2 \, \Pi_k' d\phi_k$$

are finite in the continuum limit. Due to the intermediate field relevance of the factor W in the pseudofree ground state, an approximate evaluation of the integral will be adequate for our purposes. Thus, we are led to consider

$$K \int Z^{-p} \, [\Sigma'_k h_k \, \phi_k \, a^s]^{2p} \, \frac{e^{-\Sigma'_{k,l} \, \phi_k \, A_{k-l} \, \phi_l \, a^{2s}/\hbar} - W}{\Pi'_k [\, \Sigma'_l J_{k,l} \phi_l^2 \,]^{(N'-1)/2N'}} \, \Pi'_k \, d\phi_k$$

$$\simeq 2 K_0 \int Z^{-p} \, \kappa^{2p} \, [\Sigma'_k h_k \, \eta_k \, a^s]^{2p}$$

$$\times \frac{e^{-\kappa^2 \, \Sigma'_{k,l} \, \eta_k \, A_{k-l} \, \eta_l \, a^{2s}/\hbar}}{\Pi'_k [\, \Sigma'_l J_{k,l} \eta_l^2 \,]^{(N'-1)/2N'}} \, d\kappa \, \delta(1 - \Sigma'_k \eta_k^2) \, \Pi'_k \, d\eta_k \ ,$$

where K_0 is the normalization factor when W is dropped. Our goal is to use this integral to determine a value for the field strength renormalization constant Z. To estimate this integral we first replace two factors with η variables by their appropriate averages. In particular, the quadratic expression in the exponent is estimated by

$$\kappa^2 \, \Sigma'_{k,l} \, \eta_k \, A_{k-l} \, \eta_l \, a^{2s} \simeq \kappa^2 \, \Sigma'_{k,l} \, N'^{\,-1} A_{k-l} \, a^{2s} \propto \kappa^2 \, N' \, a^{2s} \, a^{-(s+1)} \, f(L) \ ,$$

where $f(L)$ expresses an order of magnitude of the matrix $A_{k-l} \, a^{(s+1)}$, as defined earlier in Sec. 11.5.3. Recall that $f(L)$ has no dependence on the lattice spacing a, and is $O(1)$. [**Remark:** In prior publications on this subject, it was the rule to ignore $f(L)$ and simply treat it as unity; this practice has turned out to be quite acceptable.] Next, the expression in the integrand is estimated by

$$[\Sigma'_k h_k \, \eta_k \, a^s]^{2p} \simeq N'^{\,-p} \, [\Sigma'_k h_k \, a^s]^{2p} \ .$$

The integral over κ is then estimated by first rescaling the variable $\kappa^2 \to \kappa^2/(N' a^{s-1} f(L)/\hbar)$, which then leads to an overall integral estimate proportional to

$$Z^{-p} \, N'^{-p} \, [\Sigma'_k h_k \, a^s]^{2p} / [N' \, a^{s-1} \, f(L)]^p \ .$$

At this point, all factors of a are now outside the integral. [We note that had we kept the term $W = W(\phi a^{(s-1)/2}/\hbar^{1/2})$ in our calculations, it too would no longer depend on the lattice spacing a after the latest change of variables.] For this final result to be meaningful in the continuum limit, we are led to choose $Z = N'^{\,-2} a^{-(s-1)} \, f(L)^{-1}$. However, Z must be dimensionless, so we introduce a fixed positive quantity q with dimensions of an inverse length, which allows us to set

$$Z = N'^{\,-2} \, (qa)^{-(s-1)} \, f(L)^{-1} \ .$$

11.6.2 Mass and coupling constant renormalization

A power series expansion of the mass and coupling constant terms leads to the expressions $\langle [m_0^2 \, \Sigma_k \phi_k^2 a^n]^p \rangle$ and $\langle [\lambda_0 \, \Sigma_k \phi_k^4 a^n]^p \rangle$ for $p \geq 1$, which we treat together as part of the larger family governed by $\langle [g_{0,r} \, \Sigma_k \phi_k^{2r} a^n]^p \rangle$ for integral $r \geq 1$. Thus we consider

$$K \int [g_{0,r} \Sigma'_k \phi_k^{2r}\, a^s]^p \, \frac{e^{-\Sigma'_{k,l}\, \phi_k\, A_{k-l}\, \phi_l\, a^{2s}/\hbar} - W}{\Pi'_k [\Sigma'_l J_{k,l} \phi_l^2]^{(N'-1)/2N'}} \, \Pi'_k \, d\phi_k$$

$$\simeq 2K_0 \int g_{0,r}^p \, \kappa^{2rp} \, [\Sigma'_k \eta_k^{2r}\, a^s]^p$$

$$\times \frac{e^{-\kappa^2\, \Sigma'_{k,l}\, \eta_k\, A_{k-l}\, \eta_l\, a^{2s}/\hbar}}{\Pi'_k [\Sigma'_l J_{k,l} \eta_l^2]^{(N'-1)/2N'}} \, d\kappa \, \delta(1 - \Sigma'_k \eta_k^2) \, \Pi'_k \, d\eta_k \ .$$

The quadratic exponent is again estimated as

$$\kappa^2\, \Sigma'_{k,l}\, \eta_k\, A_{k-l}\, \eta_l\, a^{2s} \propto \kappa^2\, N'\, a^{2s}\, a^{-(s+1)}\, f(L) \ ,$$

while the integrand factor

$$[\Sigma'_k \eta_k^{2r}]^p \simeq N'^p\, N'^{-rp} \ .$$

The same transformation of variables used above precedes the integral over κ, and the result is an integral, no longer depending on a, that is proportional to

$$g_{0,r}^p\, N'^{-(r-1)p}\, a^{sp} / N'^{rp}\, a^{(s-1)rp}\, f(L)^{rp} \ .$$

To have an acceptable continuum limit, it suffices that

$$g_{0,r} = N'^{(2r-1)}\, (qa)^{(s-1)r-s}\, f(L)^r\, g_r \ ,$$

where g_r may be called the physical coupling factor. Moreover, it is noteworthy that $Z^r\, g_{0,r} = [N'(qa)^s]^{-1}\, g_r$, for all values of r—independently of the expression $f(L)$—and which for a finite spatial volume $V' = N' a^s$ leads to a finite nonzero result for $Z^r\, g_{0,r}$. It should not be a surprise that there are no divergences for all such interactions because the source of all divergences has been neutralized!

We may specialize the general result established above to the two cases of interest to us. Namely, when $r = 1$ this last relation implies that $m_0^2 = N'(qa)^{-1}\, f(L)\, m^2$, while when $r = 2$, it follows that $\lambda_0 = N'^3 (qa)^{s-2}\, f(L)^2\, \lambda$. In these cases it also follows that $Z\, m_0^2 = [N'(qa)^s]^{-1} m^2$ and $Z^2\, \lambda_0 = [N'(qa)^s]^{-1}\, \lambda$, which for a finite spatial volume $V' = N' a^s$ leads to a finite nonzero result for $Z\, m_0^2$ and $Z^2\, \lambda_0$, respectively. Expressions for Z, m_0^2, and λ_0 provide sufficient information to fully define the model.

11.7 Conclusion

For scalar nonrenormalizable quantum field models, we have shown that the choice of a nonconventional counterterm, but one that is still nonclassical, leads to a formulation for which a perturbation analysis of both the mass term

and the nonlinear interaction term, expanded about the appropriate pseudo-free model, are term-by-term finite. Alternative insight into such models may possibly be obtained by Monte Carlo studies of the full, nonperturbative model including the special counterterm; for a preliminary discussion of such an approach, see [Kla08].

****

Between These Covers—and Beyond

We have taken the reader on a long journey, beginning with random and stochastic variables, through infinite-dimensional integrals, and real functional integrals often having probabilistic significance. The subject matter then shifted toward quantum mechanics and complex functional integrals that encoded important properties of the quantum story. These complex functional integrals had complications of their own and some effort was expended in trying to advance reasonable formulations of those important quantities. This effort eventually led us to continuous-time regularized phase space path integrals whose interpretation was intimately linked with coherent states, which also had the virtue of providing a relatively smooth bridge between the quantum and classical formalisms. Extending such stories to infinitely many degrees of freedom, as befits a quantum field theory, led to divergences that appear whenever a quantum field theory is analyzed. The most troublesome of those cases—the so-called nonrenormalizable models—entail an unending series of new and ever-more divergent terms that need cancellation, even though the corresponding classical theory is well-behaved. By studying soluble nonrenormalizable models and dissecting them to find the "trick" that is involved in their well-behaved solution, we were led to conjecture how relativistic nonrenormalizable models may possibly find a reformulation that is divergence free. Although our argument is not yet as complete as we would like it to be, it would appear sufficiently plausible that Monte Carlo computations should be able to lend credible support—or the contrary, of course—when they are performed.

If the new proposals do pass the proper test and thus can serve as acceptable quantum solutions of hitherto insoluble quantum field theory problems, then the whole formalism of functional integration will have once again shown its extraordinary power as a tool to approach and occasionally help solve very difficult problems.

References

[Aiz81] Aizenman, M.: Proof of the Triviality of φ_d^4 Field Theory and Some Mean-Field Features of Ising Models for $d > 4$, Phys. Rev. Lett. **47**, 1–4, E-886 (1981)

[AKL] Alicki, R., Klauder, J.R., and Lewandowski, J.: Landau-Level Ground State Degeneracy and its Relevance for a General Quantization Procedure, Phys. Rev. A **48**, 2538–2548 (1993)

[Aro43] Aronszajn, N.: La théorie des noyaux reproduisants et ses applications Premire Partie, Proc. Cambridge Philos. Soc. **39**, 133–153 (1943); Note additionnelle l'article La théorie des noyaux reproduisants et ses applications, Proc. Cambridge Phil. Soc. **39**, 205–205 (1943); Theory of Reproducing Kernels, Trans. Am. Math. Soc. **68**, 337–404 (1950)

[AK69] Aslaksen, E., and Klauder, J.R.: Continuous Representation Theory Using the Affine Group, J. Math. Phys. **10**, 2267–2275 (1969)

[Bar61] Bargmann, V.: On a Hilbert Space of Analytic Functions and an Associated Integral Transform, Commun. Pure Appl. Math. **14**, 187–214 (1961)

[Ber72] Berezin, F.A.: Covariant and Contravariant Symbols of Operators, (Russian), Izv. Akad. Nauk SSSR Ser. Mat. **36**, 1134–1167 (1972)

[Ber75] Berezin, F.A.: General Concept of Quantization, Commun. Math. Phys. **40**, 153–174 (1975)

[BEH] Blank, J., Exner, P., and Havlicek, M.: Hilbert Space Operators in Quantum Physics, AIP Press, New York (1994)

[Bod03] Bodmann, B.G.: Construction of Self-Adjoint Berezin–Toeplitz Operators on Kähler Manifolds and a Probabilistic Representation of the Associated Semigroups, J. Geom. Phys. **47**, 128–160 (2003)

[Cal88] Callaway, D.J.E.: Triviality Pursuit: Can Elementary Scalar Particles Exist?, Phys. Rep. **167**, 241–320 (1988)

[Cam62] Cameron, R.H.: The Ilstow and Feynman Integrals, J. Anal. Math. **10**, 287–361 (1962)

[Che68] Chernoff, P.: Note on Product Formulas for Operator Semigroups, J. Funct. Anal. **2**, 238–242 (1968)

J.R. Klauder, *A Modern Approach to Functional Integration*, Applied and Numerical Harmonic Analysis, DOI 10.1007/978-0-8176-4791-9,

[Che46] Chevalley, C.: Theory of Lie Groups I, Princeton University Press, Princeton, New Jersey, Chap. V (1946)

[CKW] Coffey, W.T., Kalmykov, Yu. P., and Waldron, J.T.: The Langevin Equation, with Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering, Second Edition, World Scientific, Singapore (2004)

[Coh61] Cohn, P.M.: Lie Groups, Cambridge University Press, London, Chaps. IV and V (1961)

[DK85] Daubechies, I., and Klauder, J.R.: Quantum Mechanical Path Integrals with Wiener Measures for All Polynomial Hamiltonians, II, J. Math. Phys. **26**, 2239–2256 (1985)

[DKP] Daubechies, I., Klauder, J.R., and Paul, T.: Wiener Measures for Path Integrals with Affine Kinematic Variables, J. Math. Phys. **28**, 85–102 (1987)

[DMS] De Finetti, B., Machi, A, and Smith, A.: Theory of Probability: A Critical Introductory Treatment, Wiley Series in Probability & Mathematical Statistics, New York, Vols. 1 and 2 (1992)

[Dir33] Dirac, P.A.M.: The Lagrangian in Quantum Mechanics, Phys. Z. Sov. Union **3**, 64–72 (1933)

[Dir64] Dirac, P.A.M.: Lectures on Quantum Mechanics, Belfer Graduate School of Science, Yeshiva University, New York (1964)

[Dir74] Dirac, P.A.M.: The Principles of Quantum Mechanics, Oxford University Press, Oxford, Fourth Edition (1974)

[Emc74] Emch, G.G.: Algebraic Methods in Statistical Mechanics and Quantum Field Theory, Wiley–Interscience, New York (1974)

[FFS92] Fernández, R., Fröhlich, J., and Sokal, A.: Random Walks, Critical Phenomena, and Triviality in Quantum Field Theory, Springer Verlag, New York (1992)

[Fey48] Feynman, R.P.: Space-time Approach to Non-relativistic Quantum Mechanics, Rev. Mod. Phys. **20**, 367–387 (1948); Feynman, R.P., and Hibbs, A.: Quantum Mechanics and Path Integrals, McGraw–Hill, New York (1965)

[Fey51] Feynman, R.P.: An Operator Calculus Having Applications in Quantum Electrodynamics, Phys. Rev. **84**, 108–128 (1951)

[Fro82] Fröhlich, J.: On the Triviality of $\lambda\varphi_4^4$ Theories and the Approach to the Critical Point in $d \geq 4$ Dimensions, Nuclear Phys. B **200**, 281–296 (1982)

[GY60] Gel'fand, I.M., and Yaglom, A.M.: Integration in Functional Spaces and its Applications in Quantum Physics, J. Math. Phys. **1**, 48–69 (1960)

[GV64] Gel'fand, I.M., and Vilenkin, N.Ya.: Generalized Functions, Vol. 4: Applications of Harmonic Analysis, Academic Press, New York (1964)

[Gil06] Gilmore, R.: Lie Groups, Lie Algebras, and Some of Their Applications, Dover, New York (2006)

[Gla63] Glauber, R.J.: The Quantum Theory of Optical Coherence, Phys. Rev. **130**, 2529–2539 (1963); Coherent and Incoherent States of the Radiation Field, Phys. Rev. **131**, 2766–2788 (1963)

[GJ87] Glimm, J., and Jaffe, A.: Quantum Physics, Springer Verlag, New York, Second Edition (1987)

[GS98] Grosche, C., and Steiner, F.: Handbook of Feynman Path Integrals, Springer Verlag, Berlin (1998)

[HK04] Hartmann, L., and Klauder, J.R.: Weak Coherent State Path Integrals, J. Math. Phys. **45**, 87–99 (2004)

[Hid70] Hida, T.: Stationary Stochastic Processes, Princeton University Press, Princeton, New Jersey (1970)

[Hus40] Husimi, K.: Some Format Properties of the Density Matrix, Proc. Phys. Math. Soc. Japan **22**, 264–314 (1940)

[Itô61] Itô, K.: Wiener Integral and Feynman Integral, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Univerity of California Press, Berkeley, Vol. 2, 227–238 (1961); Generalized Uniform Complex Measures in the Hilbertian Metric Space with their Application to the Feynman Integral, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, Vol. 2, Part 1, 145–161 (1967)

[IMc65] Itô, K., and McKean, H.: Diffusion Processes and Their Sample Paths, Springer Verlag, New York (1965)

[IZ80] Itzykson, C., and Zuber, J.-B.: Quantum Field Theory, McGraw–Hill, New York (1980)

[JR80] Jauch, J.M., and Rohrlich, F.: The Theory of Photons and Electrons, Springer Verlag, Berlin, Second Expanded Edition, Second Corrected Printing (1980)

[Kak93] Kaku, M.: Quantum Field Theory, Oxford University Press, New York (1993)

[Kla60] Klauder, J.R.: The Action Option and the Feynman Quantization of Spinor Fields in Terms of Ordinary *c*-Numbers, Ann. Phys. **11**, 123–168 (1960)

[Kla64] Klauder, J.R.: Continuous-Representation Theory. III. On Functional Quantization of Classical Systems, J. Math. Phys. **5**, 177–187 (1964)

[KMC] Klauder, J.R., McKenna, J., and Currie, D.G.: On 'Diagonal' Coherent-State Representations for Quantum-Mechanical Density Matrices, J. Math. Phys. **6**, 734–739 (1965)

[KS68] Klauder, J.R., and Sudarshan, E.C.G.: Fundamentals of Quantum Optics, W.A. Benjamin, New York (1968) and Dover Publications, New York (2006)

[KlS85] Klauder, J.R., and Skagerstam, B.-S.: Coherent States: Applications in Physics and Mathematical Physics, World Scientific, Singapore (1985)

[Kla88] Klauder, J.R.: Quantization *Is* Geometry, After All, Ann. Phys. **188**, 120–141 (1988)

[Kla97] Klauder, J.R.: Coherent State Quantization of Constraint Systems, Ann. Phys. **254**, 419–453 (1997)

[Kla99] Klauder, J.R.: Universal Procedure for Enforcing Quantum Constraints, Nuclear Phys. B **547**, 397–412 (1999)

[Kl99a] Klauder, J.R.: Product Representations and the Quantization of Constrained Systems, Proc. V. A. Steklov Inst. Math. **226**, 212–222 (1999)

[Kla00] Klauder, J.R.: Beyond Conventional Quantization, Cambridge University Press, Cambridge (2000 & 2005)

[Kla01] Klauder, J.R.: Quantization of Constrained Systems, Lect. Notes Phys. **572**, 143–182 (2001)

[Kl01a] Klauder, J.R.: Coherent State Path Integrals *without* Resolutions of Unity, Found. Phys. **31**, 57–67 (2001)

[Kla05] Klauder, J.R.: Path Integrals, and Classical and Quantum Constraints, arXiv:quant-ph/0507222 (2005)

[KL06] Klauder, J.R., and Little, J.S.: Highly Irregular Quantum Constraints, Class. Quant. Grav. **23**, 3641–3654 (2006)

[Kla08] Klauder, J.R.: Divergence-free Nonrenormalizable Models, J. Phys. A **41**, 335206 (15pp) (2008)

[Kl08a] Klauder, J.R.: Taming Nonrenormalizability, J. Phys. A **42**, 335208 (7pp) (2009)

[Kle03] Kleinert, H.: Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics, 3rd Edition, World Scientific, Singapore (2003)

[LSU] Ladyzenskaja, O.A., Solonnikov, V., and Ural'ceva, N.N.: Linear and Quasi-linear Equations of Parabolic Type, Am. Math. Soc., Providence, Vol. 23 (1968)

[Leb07] Lebl, J.: Cantor Function, http://planetmath.org/encyclopedia/CantorFunction.html (2007)

[Lie73] Lieb, E.: The Classical Limit of Quantum Spin Systems, Commun. Math. Phys. **31**, 327–340 (1973)

[Luk70] Lukacs, E.: Characteristic Functions, Hafner, New York, Second Edition (1970)

[Mar92] Maraner, P.: Landau Ground State on Riemannian Surfaces, Mod. Phys. Lett. A **7**, 2555–2558 (1992)

[MS74] McShane, E.J.: Stochastic Calculus and Stochastic Models, Academic Press, New York (1974)

[Nel64] Nelson, E.: Feynman Integrals and the Schrödinger Equation, J. Math. Phys. **5**, 332–343 (1964)

[On80] Onofri, E.: Path Integrals Over Coherent States, in Functional Integration, Eds. J.-P. Antoine and E. Tirapegui, Plenum, New York, 121–124 (1980)

[PS95] Peskin, M.E., and Schroeder, D.V.: An Introduction to Quantum Field Theory, Addison–Wesley, Reading, Massachusetts (1995)

[Roe96] Roepstorff, G.: Path Integral Approach to Quantum Physics, Springer Verlag, Berlin (1996)

[Seg63] Segal, I.: Mathematical Problems of Relativistic Physics, Am. Math. Soc., Providence (1963)

[Sha89] Shabanov, S.V.: Path Integral in Holomorphic Representation without Gauge Fixation, JINR Preprint, E2-89-687, JINR, Dubna (1989),

reprinted in: Proceeding of Path Intgerals 96 (Eds. V.S. Yarunin and M.A. Smondyrev) 133–138, JINR, Dubna (1996)

[Sim79] Simon, B.: Functional Integration and Quantum Physics, Academic Press, New York (1979)

[Sko65] Skorohod, A.V.: Studies in the Theory of Random Processes, Addison-Wesley, Reading, Massachusetts(1965)

[Sko74] Skorohod, A.V.: Integration in Hilbert Space, Springer Verlag, Berlin (1974)

[Sni80] Sniatycki, J.: Geometric Quantization and Quantum Mechanics, Springer Verlag, New York (1980)

[Sud63] Sudarshan, E.C.G.: Equivalence of Semiclassical and Quantum Mechanical Descriptions of Statistical Light Beams, Phys. Rev. Lett. **10**, 277–279 (1963)

[Tro58] Trotter, H.F.: Approximation of Semi-Groups of Operators, Pac. J. Math. **8**, 887–919 (1958)

[WK02] Watson, G., and Klauder, J.R.: Metric and Curvature in Gravitational Phase Space, Class. Quant. Grav. **19**, 3617–3623 (2002)

[ZK93] Zhu, C.-J., and Klauder, J.R.: Classical Symptoms of Quantum Illnesses, Am. J. Phys. **61**, 605–611 (1993)

Index

Applied and Numerical Harmonic Analysis

J.M. Cooper: *Introduction to Partial Differential Equations with MATLAB* (ISBN 978-0-8176-3967-9)

C.E. D'Attellis and E.M. Fernández-Berdaguer: *Wavelet Theory and Harmonic Analysis in Applied Sciences* (ISBN 978-0-8176-3953-2)

H.G. Feichtinger and T. Strohmer: *Gabor Analysis and Algorithms* (ISBN 978-0-8176-3959-4)

T.M. Peters, J.H.T. Bates, G.B. Pike, P. Munger, and J.C. Williams: *The Fourier Transform in Biomedical Engineering* (ISBN 978-0-8176-3941-9)

A.I. Saichev and W.A. Woyczyński: *Distributions in the Physical and Engineering Sciences* (ISBN 978-0-8176-3924-2)

R. Tolimieri and M. An: *Time-Frequency Representations* (ISBN 978-0-8176-3918-1)

G.T. Herman: *Geometry of Digital Spaces* (ISBN 978-0-8176-3897-9)

A. Procházka, J. Uhlíř, P.J.W. Rayner, and N.G. Kingsbury: *Signal Analysis and Prediction* (ISBN 978-0-8176-4042-2)

J. Ramanathan: *Methods of Applied Fourier Analysis* (ISBN 978-0-8176-3963-1)

A. Teolis: *Computational Signal Processing with Wavelets* (ISBN 978-0-8176-3909-9)

W.O. Bray and C.V. Stanojević: *Analysis of Divergence* (ISBN 978-0-8176-4058-3)

G.T Herman and A. Kuba: *Discrete Tomography* (ISBN 978-0-8176-4101-6)

J.J. Benedetto and P.J.S.G. Ferreira: *Modern Sampling Theory* (ISBN 978-0-8176-4023-1)

A. Abbate, C.M. DeCusatis, and P.K. Das: *Wavelets and Subbands* (ISBN 978-0-8176-4136-8)

L. Debnath: *Wavelet Transforms and Time-Frequency Signal Analysis* (ISBN 978-0-8176-4104-7)

K. Gröchenig: *Foundations of Time-Frequency Analysis* (ISBN 978-0-8176-4022-4)

D.F. Walnut: *An Introduction to Wavelet Analysis* (ISBN 978-0-8176-3962-4)

O. Bratteli and P. Jorgensen: *Wavelets through a Looking Glass* (ISBN 978-0-8176-4280-8)

H.G. Feichtinger and T. Strohmer: *Advances in Gabor Analysis* (ISBN 978-0-8176-4239-6)

O. Christensen: *An Introduction to Frames and Riesz Bases* (ISBN 978-0-8176-4295-2)

L. Debnath: *Wavelets and Signal Processing* (ISBN 978-0-8176-4235-8)

J. Davis: *Methods of Applied Mathematics with a MATLAB Overview* (ISBN 978-0-8176-4331-7)

G. Bi and Y. Zeng: *Transforms and Fast Algorithms for Signal Analysis and Representations* (ISBN 978-0-8176-4279-2)

J.J. Benedetto and A. Zayed: *Sampling, Wavelets, and Tomography* (ISBN 978-0-8176-4304-1)

E. Prestini: *The Evolution of Applied Harmonic Analysis* (ISBN 978-0-8176-4125-2)

O. Christensen and K.L. Christensen: *Approximation Theory* (ISBN 978-0-8176-3600-5)

L. Brandolini, L. Colzani, A. Iosevich, and G. Travaglini: *Fourier Analysis and Convexity* (ISBN 978-0-8176-3263-2)

W. Freeden and V. Michel: *Multiscale Potential Theory* (ISBN 978-0-8176-4105-4)

O. Calin and D.-C. Chang: *Geometric Mechanics on Riemannian Manifolds* (ISBN 978-0-8176-4354-6)

Applied and Numerical Harmonic Analysis (Cont'd)

J.A. Hogan and J.D. Lakey: *Time-Frequency and Time-Scale Methods*
(ISBN 978-0-8176-4276-1)

C. Heil: *Harmonic Analysis and Applications* (ISBN 978-0-8176-3778-1)

K. Borre, D.M. Akos, N. Bertelsen, P. Rinder, and S.H. Jensen: *A Software-Defined GPS and Galileo Receiver* (ISBN 978-0-8176-4390-4)

T. Qian, V. Mang I, and Y. Xu: *Wavelet Analysis and Applications* (ISBN 978-3-7643-7777-9)

G.T. Herman and A. Kuba: *Advances in Discrete Tomography and Its Applications*
(ISBN 978-0-8176-3614-2)

M.C. Fu, R.A. Jarrow, J.-Y. J. Yen, and R.J. Elliott: *Advances in Mathematical Finance*
(ISBN 978-0-8176-4544-1)

O. Christensen: *Frames and Bases* (ISBN 978-0-8176-4677-6)

P.E.T. Jorgensen, K.D. Merrill, and J.A. Packer: *Representations, Wavelets, and Frames*
(ISBN 978-0-8176-4682-0)

M. An, A.K. Brodzik, and R. Tolimieri: *Ideal Sequence Design in Time-Frequency Space*
(ISBN 978-0-8176-4737-7)

S.G. Krantz: *Explorations in Harmonic Analysis* (ISBN 978-0-8176-4668-4)

G.S. Chirikjian: *Stochastic Models, Information Theory, and Lie Groups, Volume 1*
(ISBN 978-0-8176-4802-2)

C. Cabrelli and J.L. Torrea: *Recent Developments in Real and Harmonic Analysis*
(ISBN 978-0-8176-4531-1)

B. Luong: *Fourier Analysis on Finite Abelian Groups* (ISBN 978-0-8176-4915-9)

M.V. Wickerhauser: *Mathematics for Multimedia* (ISBN 978-0-8176-4879-4)

P. Massopust and B. Forster: *Four Short Courses on Harmonic Analysis*
(ISBN 978-0-8176-4890-9)

O. Christensen: *Functions, Spaces, and Expansions*
(ISBN 978-0-8176-4979-1)

J.R. Klauder: *A Modern Approach to Functional Integration*
(ISBN 978-0-8176-4790-2)