Michael E. Taylor

APPLIED MATHEMATICAL SCIENCES 116

Partial Differential Equations II

Qualitative Studies of Linear Equations

2nd Edition

Springer

Applied Mathematical Sciences

Volume 116

For further volumes:
http://www.springer.com/series/34

Michael E. Taylor

Partial Differential Equations II

Qualitative Studies of Linear Equations

Second Edition

Springer

Michael E. Taylor
Department of Mathematics
University of North Carolina
Chapel Hill, NC 27599
USA
met@math.unc.edu

ISSN 0066-5452
ISBN 978-1-4419-7051-0 e-ISBN 978-1-4419-7052-7
DOI 10.1007/978-1-4419-7052-7
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010937758

Mathematics Subject Classification (2010): 35A01, 35A02, 35J05, 35J25, 35K05, 35L05, 35Q30, 35Q35, 35S05

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my wife and daughter, Jane Hawkins and Diane Taylor

Contents

Contents of Volumes I and III

Volume I: Basic Theory

Volume III: Nonlinear Equations

Preface

Partial differential equations is a many-faceted subject. Created to describe the mechanical behavior of objects such as vibrating strings and blowing winds, it has developed into a body of material that interacts with many branches of mathematics, such as differential geometry, complex analysis, and harmonic analysis, as well as a ubiquitous factor in the description and elucidation of problems in mathematical physics.

This work is intended to provide a course of study of some of the major aspects of PDE. It is addressed to readers with a background in the basic introductory graduate mathematics courses in American universities: elementary real and complex analysis, differential geometry, and measure theory.

Chapter 1 provides background material on the theory of ordinary differential equations (ODE). This includes both very basic material – on topics such as the existence and uniqueness of solutions to ODE and explicit solutions to equations with constant coefficients and relations to linear algebra – and more sophisticated results – on flows generated by vector fields, connections with differential geometry, the calculus of differential forms, stationary action principles in mechanics, and their relation to Hamiltonian systems. We discuss equations of relativistic motion as well as equations of classical Newtonian mechanics. There are also applications to topological results, such as degree theory, the Brouwer fixed-point theorem, and the Jordan–Brouwer separation theorem. In this chapter we also treat scalar first-order PDE, via Hamilton–Jacobi theory.

Chapters 2–6 constitute a survey of basic linear PDE. Chapter 2 begins with the derivation of some equations of continuum mechanics in a fashion similar to the derivation of ODE in mechanics in Chap. 1, via variational principles. We obtain equations for vibrating strings and membranes; these equations are not necessarily linear, and hence they will also provide sources of problems later, when nonlinear PDE is taken up. Further material in Chap. 2 centers around the Laplace operator, which on Euclidean space $\mathbb{R}^n$ is

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}, \tag{1}$$

and the linear wave equation,

$$\frac{\partial^2 u}{\partial t^2} - \Delta u = 0. \tag{2}$$

We also consider the Laplace operator on a general Riemannian manifold and the wave equation on a general Lorentz manifold. We discuss basic consequences of Green's formula, including energy conservation and finite propagation speed for solutions to linear wave equations. We also discuss Maxwell's equations for electromagnetic fields and their relation with special relativity. Before we can establish general results on the solvability of these equations, it is necessary to develop some analytical techniques. This is done in the next couple of chapters.

Chapter 3 is devoted to Fourier analysis and the theory of distributions. These topics are crucial for the study of linear PDE. We give a number of basic applications to the study of linear PDE with constant coefficients. Among these applications are results on harmonic and holomorphic functions in the plane, including a short treatment of elementary complex function theory. We derive explicit formulas for solutions to Laplace and wave equations on Euclidean space, and also the heat equation,

$$\frac{\partial u}{\partial t} - \Delta u = 0. \tag{3}$$

We also produce solutions on certain subsets, such as rectangular regions, using the method of images. We include material on the discrete Fourier transform, germane to the discrete approximation of PDE, and on the fast evaluation of this transform, the FFT. Chapter 3 is the first chapter to make extensive use of functional analysis. Basic results on this topic are compiled in Appendix A, Outline of Functional Analysis.

Sobolev spaces have proven to be a very effective tool in the existence theory of PDE, and in the study of regularity of solutions. In Chap. 4 we introduce Sobolev spaces and study some of their basic properties. We restrict attention to L^2-Sobolev spaces, such as $H^k(\mathbb{R}^n)$, which consists of L^2 functions whose derivatives of order $\le k$ (defined in a distributional sense, in Chap. 3) belong to $L^2(\mathbb{R}^n)$, when k is a positive integer. We also replace k by a general real number s. The L^p-Sobolev spaces, which are very useful for nonlinear PDE, are treated later, in Chap. 13.

Chapter 5 is devoted to the study of the existence and regularity of solutions to linear elliptic PDE, on bounded regions. We begin with the Dirichlet problem for the Laplace operator,

$$\Delta u = f \text{ on } \Omega, \qquad u = g \text{ on } \partial\Omega, \tag{4}$$

and then treat the Neumann problem and various other boundary problems, including some that apply to electromagnetic fields. We also study general boundary problems for linear elliptic operators, giving a condition that guarantees regularity and solvability (perhaps given a finite number of linear conditions on the data). Also in Chap. 5 are some applications to other areas, such as a proof of the Riemann mapping theorem, first for smooth simply connected domains in the complex plane $\mathbb{C}$, then, after a treatment of the Dirichlet problem for the Laplace operator on domains with rough boundary, for general simply connected domains in $\mathbb{C}$. We also develop Hodge theory and apply it to DeRham cohomology, extending the study of topological applications of differential forms begun in Chap. 1.

In Chap. 6 we study linear evolution equations, in which there is a "time" variable t, and initial data are given at $t = 0$. We discuss the heat and wave equations. We also treat Maxwell's equations, for an electromagnetic field, and more general hyperbolic systems. We prove the Cauchy–Kowalewsky theorem, in the linear case, establishing local solvability of the Cauchy initial value problem for general linear PDE with analytic coefficients, and analytic data, as long as the initial surface is "noncharacteristic." The nonlinear case is treated in Chap. 16. Also in Chap. 6 we treat geometrical optics, providing approximations to solutions of wave equations whose initial data either are highly oscillatory or possess simple singularities, such as a jump across a smooth hypersurface.

Chapters 1–6, together with Appendix A and B, Manifolds, Vector Bundles, and Lie Groups, make up the first volume of this work. The second volume consists of Chaps. 7–12, covering a selection of more advanced topics in linear PDE, together with Appendix C, Connections and Curvature.

Chapter 7 deals with pseudodifferential operators (ψDOs). This class of operators includes both differential operators and parametrices of elliptic operators, that is, inverses modulo smoothing operators. There is a "symbol calculus" allowing one to analyze products of ψDOs, useful for such a parametrix construction. The L^2-boundedness of operators of order zero and the Gårding inequality for elliptic ψDOs with positive symbol provide very useful tools in linear PDE, which will be used in many subsequent chapters.

Chapter 8 is devoted to spectral theory, particularly for self-adjoint elliptic operators. First we give a proof of the spectral theorem for general self-adjoint operators on Hilbert space. Then we discuss conditions under which a differential operator yields a self-adjoint operator. We then discuss the asymptotic distribution of eigenvalues of the Laplace operator on a bounded domain, making use of a construction of a parametrix for the heat equation from Chap. 7. In the next four sections of Chap. 8 we consider the spectral behavior of various specific differential operators: the Laplace operator on a sphere, and on hyperbolic space, the "harmonic oscillator"

$$-\Delta + |x|^2, \tag{5}$$

and the operator

$$-\Delta - \frac{K}{|x|}, \tag{6}$$

which arises in the simplest quantum mechanical model of the hydrogen atom. Finally, we consider the Laplace operator on cones.

In Chap. 9 we study the scattering of waves by a compact obstacle K in $\mathbb{R}^3$. This scattering theory is to some degree an extension of the spectral theory of the

Laplace operator on $\mathbb{R}^3 \setminus K$, with the Dirichlet boundary condition. In addition to studying how a given obstacle scatters waves, we consider the *inverse* problem: how to determine an obstacle given data on how it scatters waves.

Chapter 10 is devoted to the Atiyah–Singer index theorem. This gives a formula for the index of an elliptic operator D on a compact manifold M, defined by

$$\text{Index}\, D = \dim \ker D - \dim \ker D^*. \tag{7}$$

We establish this formula, which is an integral over M of a certain differential form defined by a pair of "curvatures," when D is a first order differential operator of "Dirac type," a class that contains many important operators arising from differential geometry and complex analysis. Special cases of such a formula include the Chern–Gauss–Bonnet formula and the Riemann–Roch formula. We also discuss the significance of the latter formula in the study of Riemann surfaces.

In Chap. 11 we study Brownian motion, described mathematically by Wiener measure on the space of continuous paths in $\mathbb{R}^n$. This provides a probabilistic approach to diffusion and it both uses and provides new tools for the analysis of the heat equation and variants, such as

$$\frac{\partial u}{\partial t} = -\Delta u + Vu, \tag{8}$$

where V is a real-valued function. There is an integral formula for solutions to (8), known as the Feynman–Kac formula; it is an integral over path space with respect to Wiener measure, of a fairly explicit integrand. We also derive an analogous integral formula for solutions to

$$\frac{\partial u}{\partial t} = -\Delta u + Xu, \tag{9}$$

where X is a vector field. In this case, another tool is involved in constructing the integrand, the stochastic integral. We also study stochastic differential equations and applications to more general diffusion equations.

In Chap. 12 we tackle the $\overline{\partial}$-Neumann problem, a boundary problem for an elliptic operator (essentially the Laplace operator) on a domain $\Omega \subset \mathbb{C}^n$, which is very important in the theory of functions of several complex variables. From a technical point of view, it is of particular interest that this boundary problem does not satisfy the regularity criteria investigated in Chap. 5. If Ω is "strongly pseudoconvex," one has instead certain "subelliptic estimates," which are established in Chap. 12.

The third and final volume of this work contains Chaps. 13–18. It is here that we study nonlinear PDE.

We prepare the way in Chap. 13 with a further development of function space and operator theory, for use in nonlinear analysis. This includes the theory of L^p-Sobolev spaces and Hölder spaces. We derive estimates in these spaces on nonlinear functions $F(u)$, known as "Moser estimates," which are very useful. We extend the theory of pseudodifferential operators to cases where the symbols have limited smoothness, and also develop a variant of ψDO theory, the theory of "paradifferential operators," which has had a significant impact on nonlinear PDE since about 1980. We also estimate these operators, acting on the function spaces mentioned above. Other topics treated in Chap. 13 include Hardy spaces, compensated compactness, and "fuzzy functions."

Chapter 14 is devoted to nonlinear elliptic PDE, with an emphasis on second order equations. There are three successive degrees of nonlinearity: semilinear equations, such as

$$\Delta u = F(x, u, \nabla u), \tag{10}$$

quasi-linear equations, such as

$$\sum a^{jk}(x, u, \nabla u)\partial_j \partial_k u = F(x, u, \nabla u), \tag{11}$$

and completely nonlinear equations, of the form

$$G(x, D^2 u) = 0. \tag{12}$$

Differential geometry provides a rich source of such PDE, and Chap. 14 contains a number of geometrical applications. For example, to deform conformally a metric on a surface so its Gauss curvature changes from $k(x)$ to $K(x)$, one needs to solve the semilinear equation

$$\Delta u = k(x) - K(x)e^{2u}. \tag{13}$$

As another example, the graph of a function $y = u(x)$ is a minimal submanifold of Euclidean space provided u solves the quasilinear equation

$$\big(1 + |\nabla u|^2\big)\Delta u + (\nabla u) \cdot H(u)(\nabla u) = 0, \tag{14}$$

called the minimal surface equation. Here, $H(u) = (\partial_j \partial_k u)$ is the Hessian matrix of u. On the other hand, this graph has Gauss curvature $K(x)$ provided u solves the completely nonlinear equation

$$\det H(u) = K(x)\big(1 + |\nabla u|^2\big)^{(n+2)/2}, \tag{15}$$

a Monge–Ampère equation. Equations (13)–(15) are all scalar, and the maximum principle plays a useful role in the analysis, together with a number of other tools. Chapter 14 also treats nonlinear systems. Important physical examples arise in studies of elastic bodies, as well as in other areas, such as the theory of liquid crystals. Geometric examples of systems considered in Chap. 14 include equations for harmonic maps and equations for isometric imbeddings of a Riemannian manifold in Euclidean space.

In Chap. 15, we treat nonlinear parabolic equations. Partly echoing Chap. 14, we progress from a treatment of semilinear equations,

$$\frac{\partial u}{\partial t} = Lu + F(x, u, \nabla u), \tag{16}$$

where L is a linear operator, such as $L = \Delta$, to a treatment of quasi-linear equations, such as

$$\frac{\partial u}{\partial t} = \sum \partial_j a^{jk}(t, x, u)\partial_k u + X(u). \tag{17}$$

(We do very little with completely nonlinear equations in this chapter.) We study systems as well as scalar equations. The first application of (16) we consider is to the parabolic equation method of constructing harmonic maps. We also consider "reaction-diffusion" equations, $\ell \times \ell$ systems of the form (16), in which $F(x, u, \nabla u) = X(u)$, where X is a vector field on $\mathbb{R}^\ell$, and L is a diagonal operator, with diagonal elements $a_j \Delta,\ a_j \geq 0$. These equations arise in mathematical models in biology and in chemistry. For example, $u = (u_1, \dots, u_\ell)$ might represent the population densities of each of ℓ species of living creatures, distributed over an area of land, interacting in a manner described by X and diffusing in a manner described by $a_j \Delta$. If there is a nonlinear (density-dependent) diffusion, one might have a system of the form (17).

Another problem considered in Chap. 15 models the melting of ice; one has a linear heat equation in a region (filled with water) whose boundary (where the water touches the ice) is moving (as the ice melts). The nonlinearity in the problem involves the description of the boundary. We confine our analysis to a relatively simple one-dimensional case.

Nonlinear hyperbolic equations are studied in Chap. 16. Here continuum mechanics is the major source of examples, and most of them are systems, rather than scalar equations. We establish local existence for solutions to first order hyperbolic systems, which are either "symmetric" or "symmetrizable." An example of the latter class is the following system describing compressible fluid flow:

$$\frac{\partial v}{\partial t} + \nabla_v v + \frac{1}{\rho}\,\mathrm{grad}\, p = 0, \quad \frac{\partial \rho}{\partial t} + \nabla_v \rho + \rho\, \mathrm{div}\, v = 0, \tag{18}$$

for a fluid with velocity v, density ρ, and pressure p, assumed to satisfy a relation $p = p(\rho)$, called an "equation of state." Solutions to such nonlinear systems tend to break down, due to shock formation. We devote a bit of attention to the study of weak solutions to nonlinear hyperbolic systems, with shocks.

We also study second-order hyperbolic systems, such as systems for a k-dimensional membrane vibrating in $\mathbb{R}^n$, derived in Chap. 2. Another topic covered in Chap. 16 is the Cauchy–Kowalewsky theorem, in the nonlinear case. We use a method introduced by P. Garabedian to transform the Cauchy problem for an analytic equation into a symmetric hyperbolic system.

In Chap. 17 we study incompressible fluid flow. This is governed by the Euler equation

$$\frac{\partial v}{\partial t} + \nabla_v v = -\operatorname{grad} p, \qquad \operatorname{div} v = 0, \tag{19}$$

in the absence of viscosity, and by the Navier–Stokes equation

$$\frac{\partial v}{\partial t} + \nabla_v v = \nu \mathcal{L} v - \operatorname{grad} p, \qquad \operatorname{div} v = 0, \tag{20}$$

in the presence of viscosity. Here $\mathcal{L}$ is a second-order operator, the Laplace operator for a flow on flat space; the "viscosity" ν is a positive quantity. Equation (19) shares some features with quasilinear hyperbolic systems, though there are also significant differences. Similarly, (20) has a lot in common with semilinear parabolic systems.

Chapter 18, the last chapter in this work, is devoted to Einstein's gravitational equations:

$$G_{jk} = 8\pi\kappa T_{jk}. \tag{21}$$

Here G_{jk} is the Einstein tensor, given by $G_{jk} = \operatorname{Ric}_{jk} - (1/2)S g_{jk}$, where Ric_{jk} is the Ricci tensor and S the scalar curvature, of a Lorentz manifold (or "spacetime") with metric tensor g_{jk}. On the right side of (21), T_{jk} is the stress-energy tensor of the matter in the spacetime, and κ is a positive constant, which can be identified with the gravitational constant of the Newtonian theory of gravity. In local coordinates, G_{jk} has a nonlinear expression in terms of g_{jk} and its second order derivatives. In the empty-space case, where $T_{jk} = 0$, (21) is a quasilinear second order system for g_{jk}. The freedom to change coordinates provides an obstruction to this equation being hyperbolic, but one can impose the use of "harmonic" coordinates as a constraint and transform (21) into a hyperbolic system. In the presence of matter one couples (21) to other systems, obtaining more elaborate PDE. We treat this in two cases, in the presence of an electromagnetic field, and in the presence of a relativistic fluid.

In addition to the 18 chapters just described, there are three appendices, already mentioned above. Appendix A gives definitions and basic properties of Banach and Hilbert spaces (of which L^p-spaces and Sobolev spaces are examples), Fréchet spaces (such as $C^\infty(\mathbb{R}^n)$), and other locally convex spaces (such as spaces of distributions). It discusses some basic facts about bounded linear operators, including some special properties of compact operators, and also considers certain classes of unbounded linear operators. This functional analytic material plays a major role in the development of PDE from Chap. 3 onward.

Appendix B gives definitions and basic properties of manifolds and vector bundles. It also discusses some elementary properties of Lie groups, including a little representation theory, useful in Chap. 8, on spectral theory, as well as in the Chern–Weil construction.

Appendix C, Connections and Curvature, contains material of a differential geometric nature, crucial for understanding many things done in Chaps. 10–18. We consider connections on general vector bundles, and their curvature. We discuss in detail special properties of the primary case: the Levi–Civita connection and Riemann curvature tensor on a Riemannian manifold. We discuss basic properties of the geometry of submanifolds, relating the second fundamental form to curvature via the Gauss–Codazzi equations. We describe how vector bundles arise from principal bundles, which themselves carry various connections and curvature forms. We then discuss the Chern–Weil construction, yielding certain closed differential forms associated to curvatures of connections on principal bundles. We give several proofs of the classical Gauss–Bonnet theorem and some related results on two-dimensional surfaces, which are useful particularly in Chaps. 10 and 14. We also give a geometrical proof of the Chern–Gauss–Bonnet theorem, which can be contrasted with the proof in Chap. 10, as a consequence of the Atiyah–Singer index theorem.

We mention that, in addition to these "global" appendices, there are appendices to some chapters. For example, Chap. 3 has an appendix on the gamma function. Chapter 6 has two appendices; Appendix A has some results on Banach spaces of harmonic functions useful for the proof of the linear Cauchy–Kowalewsky theorem, and Appendix B deals with the stationary phase formula, useful for the study of geometrical optics in Chap. 6 and also for results later, in Chap. 9. There are other chapters with such "local" appendices. Furthermore, there are two *sections*, both in Chap. 14, with appendices. Section 6, on minimal surfaces, has a companion, Sect. 6B, on the second variation of area and consequences, and Sect. 13, on nonlinear elliptic systems, has a companion, Sect. 12B, with complementary material.

Having described the scope of this work, we find it necessary to mention a number of topics in PDE that are not covered here, or are touched on only very briefly.

For example, we devote little attention to the real analytic theory of PDE. We note that harmonic functions on domains in $\mathbb{R}^n$ are real analytic, but we do not discuss analyticity of solutions to more general elliptic equations. We do prove the Cauchy–Kowalewsky theorem, on analytic PDE with analytic Cauchy data. We derive some simple results on unique continuation from these few analyticity results, but there is a large body of lore on unique continuation, for solutions to nonanalytic PDE, neglected here.

There is little material on numerical methods. There are a few references to applications of the FFT and of "splitting methods." Difference schemes for PDE are mentioned just once, in a set of exercises on scalar conservation laws. Finite element methods are neglected, as are many other numerical techniques.

There is a large body of work on free boundary problems, but the only one considered here is a simple one space dimensional problem, in Chap. 15.

While we have considered a variety of equations arising from classical physics and from relativity, we have devoted relatively little attention to quantum mechanics. We have considered one quantum mechanical operator, given in formula (6) above. Also, there are some exercises on potential scattering mentioned in Chap. 9. However, the physical theories behind these equations are not discussed here.

There are a number of nonlinear evolution equations, such as the Korteweg–deVries equation, that have been perceived to provide infinite dimensional analogues of completely integrable Hamiltonian systems, and to arise "universally" in asymptotic analyses of solutions to various nonlinear wave equations. They are not here. Nor is there a treatment of the Yang–Mills equations for gauge fields, with their wonderful applications to the geometry and topology of four dimensional manifolds.

Of course, this is not a complete list of omitted material. One can go on and on listing important topics in this vast subject. The author can at best hope that the reader will find it easier to understand many of these topics with this book, than without it.

Acknowledgments

I have had the good fortune to teach at least one course relevant to the material of this book, almost every year since 1976. These courses led to many course notes, and I am grateful to many colleagues at Rice University, SUNY at Stony Brook, the California Institute of Technology, and the University of North Carolina, for the supportive atmospheres at these institutions. Also, a number of individuals provided valuable advice on various portions of the manuscript, as it grew over the years. I particularly want to thank Florin David, David Ebin, Frank Jones, Anna Mazzucato, Richard Melrose, James Ralston, Jeffrey Rauch, Santiago Simanca, and James York. The final touches were put on the manuscript while I was visiting the Institute for Mathematics and its Applications, at the University of Minnesota, which I thank for its hospitality and excellent facilities.

Finally, I would like to acknowledge the impact on my studies of my senior thesis and Ph.D. thesis advisors, Edward Nelson and Heinz Cordes.

Introduction to the Second Edition

In addition to making numerous small corrections to this work, collected over the past dozen years, I have taken the opportunity to make some very significant changes, some of which broaden the scope of the work, some of which clarify previous presentations, and a few of which correct errors that have come to my attention.

There are seven additional sections in this edition, two in Volume 1, two in Volume 2, and three in Volume 3. Chapter 4 has a new section, “Sobolev spaces on rough domains,” which serves to clarify the treatment of the Dirichlet problem on rough domains in Chap. 5. Chapter 6 has a new section, “Boundary layer phenomena for the heat equation,” which will prove useful in one of the new sections in Chap. 17. Chapter 7 has a new section, “Operators of harmonic oscillator type,” and Chap. 10 has a section that presents an index formula for elliptic systems of operators of harmonic oscillator type. Chapter 13 has a new appendix, “Variations on complex interpolation,” which has material that is useful in the study of Zygmund spaces. Finally, Chap. 17 has two new sections, “Vanishing viscosity limits” and “From velocity convergence to flow convergence.”

In addition, several other sections have been substantially rewritten, and numerous others polished to reflect insights gained through the use of these books over time.

7

Pseudodifferential Operators

Introduction

In this chapter we discuss the basic theory of pseudodifferential operators as it has been developed to treat problems in linear PDE. We define pseudodifferential operators with symbols in classes denoted $S^m_{\rho,\delta}$, introduced by L. Hörmander. In §2 we derive some useful properties of their Schwartz kernels. In §3 we discuss adjoints and products of pseudodifferential operators. In §4 we show how the algebraic properties can be used to establish the regularity of solutions to elliptic PDE with smooth coefficients. In §5 we discuss mapping properties on L^2 and on the Sobolev spaces H^s. In §6 we establish Gårding's inequality.

In §7 we apply some of the previous material to establish the existence of solutions to hyperbolic equations. In §8 we show that certain important classes of pseudodifferential operators are preserved under the action of conjugation by solution operators to (scalar) hyperbolic equations, a result of Y. Egorov. We introduce the notion of wave front set in §9 and discuss the microlocal regularity of solutions to elliptic equations. We also discuss how solution operators to a class of hyperbolic equations propagate wave front sets. In §10 there is a brief discussion of pseudodifferential operators on manifolds.

We give some further applications of pseudodifferential operators in the next three sections. In §11 we discuss, from the perspective of the pseudodifferential operator calculus, the classical method of layer potentials, applied particularly to the Dirichlet and Neumann boundary problems for the Laplace operator. Historically, this sort of application was one of the earliest stimuli for the development of the theory of singular integral equations. One function of §11 is to provide a warm-up for the use of similar integral equations to tackle problems in scattering theory, in §7 of Chap. 9. Section 12 looks at general regular elliptic boundary problems and includes material complementary to that developed in §11 of Chap. 5. In §13 we construct a parametrix for the heat equation and apply this to obtain an asymptotic expansion of the trace of the solution operator. This expansion will be useful in studies of the spectrum in Chap. 8 and in index theory in Chap. 10.

In §14 we introduce the Weyl calculus. This can provide a powerful alternative to the operator calculus developed in §§1–6, as can be seen in [Ho4] and in Vol. 3

M.E. Taylor, *Partial Differential Equations II: Qualitative Studies of Linear Equations*, Applied Mathematical Sciences 116, DOI 10.1007/978-1-4419-7052-7_1,

of [Ho5]. Here we concentrate on identities, tied to symmetries in the Weyl calculus. We show how this leads to a quicker construction of a parametrix for the heat equation than the method used in §13. We will make use of this in §10 of Chap. 10, on a direct attack on the index theorem for elliptic differential operators on two-dimensional manifolds.

In §15, we study a class of pseudodifferential operators of "Harmonic oscillator type." This class contains the Harmonic oscillator,

$$H = -\Delta + |x|^2,$$

with symbol $|x|^2 + |\xi|^2$, and results on these operators are interesting variants on those with symbols in $S^m_{1,0}$.

Material in §§1–10 is taken from Chap. 0 of [T4], and the author thanks Birkhäuser Boston for permission to use this material. We also mention some books that take the theory of pseudodifferential operators farther than is done here: [Ho5, Kg, T1], and [Tre].

1. The Fourier integral representation and symbol classes

Using a slightly different convention from that established in Chap. 3, we write the Fourier inversion formula as

$$f(x) = \int \hat{f}(\xi)\, e^{ix\cdot\xi}\, d\xi, \tag{1.1}$$

where $\hat{f}(\xi) = (2\pi)^{-n} \int f(x) e^{-ix\cdot\xi}\, dx$ is the Fourier transform of a function on $\mathbb{R}^n$. If one differentiates (1.1), one obtains

$$D^\alpha f(x) = \int \xi^\alpha \hat{f}(\xi) e^{ix\cdot\xi}\, d\xi, \tag{1.2}$$

where $D^\alpha = D_1^{\alpha_1} \cdots D_n^{\alpha_n}$, $D_j = (1/i)\, \partial/\partial x_j$. Hence, if

$$p(x, D) = \sum_{|\alpha| \le k} a_\alpha(x) D^\alpha$$

is a differential operator, we have

$$p(x, D) f(x) = \int p(x, \xi) \hat{f}(\xi) e^{ix\cdot\xi}\, d\xi \tag{1.3}$$

where

$$p(x, \xi) = \sum_{|\alpha| \le k} a_\alpha(x) \xi^\alpha.$$

One uses the Fourier integral representation (1.3) to define pseudodifferential operators, taking the function $p(x, \xi)$ to belong to one of a number of different classes of symbols. In this chapter we consider the following symbol classes, first defined by Hörmander [Ho2].

Assuming $\rho, \delta \in [0, 1]$, $m \in \mathbb{R}$, we define $S^m_{\rho,\delta}$ to consist of C^∞-functions $p(x, \xi)$ satisfying

$$|D^\beta_x D^\alpha_\xi p(x, \xi)| \le C_{\alpha\beta} \langle \xi \rangle^{m-\rho|\alpha|+\delta|\beta|}, \tag{1.4}$$

for all α, β, where $\langle \xi \rangle = (1 + |\xi|^2)^{1/2}$. In such a case we say the associated operator defined by (1.3) belongs to $OPS^m_{\rho,\delta}$. We say that $p(x, \xi)$ is the symbol of $p(x, D)$. The case of principal interest is $\rho = 1$, $\delta = 0$. This class is defined by [KN].

Recall that in Chap. 3, §8, we defined $P(\xi) \in S^m_1(\mathbb{R}^n)$ to satisfy (1.4), with $\rho = 1$, and with no x-derivatives involved. Thus $S^m_{1,0}$ contains $S^m_1(\mathbb{R}^n)$.

If there are smooth $p_{m-j}(x, \xi)$, homogeneous in ξ of degree $m-j$ for $|\xi| \ge 1$, that is, $p_{m-j}(x, r\xi) = r^{m-j} p_{m-j}(x, \xi)$ for $r, |\xi| \ge 1$, and if

$$p(x, \xi) \sim \sum_{j \ge 0} p_{m-j}(x, \xi) \tag{1.5}$$

in the sense that

$$p(x, \xi) - \sum_{j=0}^{N} p_{m-j}(x, \xi) \in S^{m-N-1}_{1,0}, \tag{1.6}$$

for all N, then we say $p(x, \xi) \in S^m_{cl}$, or just $p(x, \xi) \in S^m$. We call $p_m(x, \xi)$ the *principal symbol* of $p(x, D)$. We will give a more general definition of the principal symbol in §10.

It is easy to see that if $p(x, \xi) \in S^m_{\rho,\delta}$ and $\rho, \delta \in [0, 1]$, then $p(x, D) : \mathcal{S}(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$. In fact, multiplying (1.3) by x^α, writing $x^\alpha e^{ix\cdot\xi} = (-D_\xi)^\alpha e^{ix\cdot\xi}$, and integrating by parts yield

$$p(x, D) : \mathcal{S}(\mathbb{R}^n) \longrightarrow \mathcal{S}(\mathbb{R}^n). \tag{1.7}$$

Under one restriction, $p(x, D)$ also acts on tempered distributions:

Lemma 1.1. *If* $\delta < 1$, *then*

$$p(x, D) : \mathcal{S}'(\mathbb{R}^n) \longrightarrow \mathcal{S}'(\mathbb{R}^n). \tag{1.8}$$

Proof. Given $u \in \mathcal{S}'$, $v \in \mathcal{S}$, we have (formally)

$$\langle v, p(x, D)u \rangle = \langle p_v, \hat{u} \rangle, \tag{1.9}$$

where

$$p_v(\xi) = (2\pi)^{-n} \int v(x) p(x,\xi) e^{ix\cdot\xi}\, dx.$$

Now integration by parts gives

$$\xi^\alpha p_v(\xi) = (2\pi)^{-n} \int D_x^\alpha \big(v(x) p(x,\xi)\big) e^{ix\cdot\xi}\, dx,$$

so

$$|p_v(\xi)| \le C_\alpha \langle\xi\rangle^{m+\delta|\alpha|-|\alpha|}.$$

Thus if $\delta < 1$, we have rapid decrease of $p_v(\xi)$. Similarly, we get rapid decrease of derivatives of $p_v(\xi)$, so it belongs to $\mathcal{S}$. Thus the right side of (1.9) is well defined.

In §5 we will analyze the action of pseudodifferential operators on Sobolev spaces.

Classes of symbols more general than $S^m_{\rho,\delta}$ have been introduced by R. Beals and C. Fefferman [BF, Be], and still more general classes were studied by Hörmander [Ho4]. These classes have some deep applications, but they will not be used in this book.

Exercises

1. Show that, for $a(x,\xi) \in \mathcal{S}(\mathbb{R}^{2n})$,

(1.10)
$$a(x,D)u = \int \hat{a}(q,p)\, e^{iq\cdot X} e^{ip\cdot D} u(x) \quad dq \quad dp,$$

where $\hat{a}(q,p)$ is the Fourier transform of $a(x,\xi)$, and the operators $e^{iq\cdot X}$ and $e^{ip\cdot D}$ are defined by

$$e^{iq\cdot X}u(x) = e^{iq\cdot x}u(x), \quad e^{ip\cdot D}u(x) = u(x+p).$$

2. Establish the identity

(1.11)
$$e^{ip\cdot D} e^{iq\cdot X} = e^{iq\cdot p}\, e^{iq\cdot X} e^{ip\cdot D}.$$

Deduce that, for $(t,q,p) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n = \mathcal{H}^n$, the binary operation

(1.12)
$$(t,q,p) \circ (t',q',p') = (t+t'+p\cdot q', q+q', p+p')$$

gives a group and that

(1.13)
$$\tilde{\pi}(t,q,p) = e^{it} e^{iq\cdot X} e^{ip\cdot X}$$

defines a unitary representation of $\mathcal{H}^n$ on $L^2(\mathbb{R}^n)$; in particular, it is a group homomorphism: $\tilde{\pi}(z\circ z') = \tilde{\pi}(z)\tilde{\pi}(z')$. $\mathcal{H}^n$ is called the Heisenberg group.

3. Give a definition of $a(x-q, D-p)$, acting on $u(x)$. Show that

$$a(x-q, D-p) = \tilde{\pi}(0,q,p)\, a(x,D)\, \tilde{\pi}(0,q,p)^{-1}.$$

4. Assume $a(x,\xi) \in S^m_{\rho,\delta}$ and $b(x,\xi) \in \mathcal{S}(\mathbb{R}^n \times \mathbb{R}^n)$. Show that $c(x,\xi) = (b * a)(x,\xi)$ belongs to $S^m_{\rho,\delta}$ ($*$ being convolution on $\mathbb{R}^{2n}$). Show that

$$c(x,D)u = \int b(y,\eta)\, a(x-y, D-\eta)\, dy\, d\eta.$$

5. Show that the map $\Psi(p,u) = p(x,D)u$ has a unique, continuous, bilinear extension from $S^m_{\rho,\delta} \times \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ to

$$\Psi : \mathcal{S}'(\mathbb{R}^{2n}) \times \mathcal{S}(\mathbb{R}^n) \longrightarrow \mathcal{S}'(\mathbb{R}^n),$$

so that $p(x,D)$ is "well defined" for any $p \in \mathcal{S}'(\mathbb{R}^n \times \mathbb{R}^n)$.

6. Let $\chi(\xi) \in C_0^\infty(\mathbb{R}^n)$ be 1 for $|\xi| \le 1$, $\chi_\epsilon(\xi) = \chi(\epsilon\xi)$. Given $p(x,\xi) \in S^m_{\rho,\delta}$, let $p_\epsilon(x,\xi) = \chi_\epsilon(\xi)p(x,\xi)$. Show that if $\rho, \delta \in [0,1]$, then

$$u \in \mathcal{S}(\mathbb{R}^n) \Longrightarrow p_\epsilon(x,D)u \to p(x,D)u \;\text{ in } \mathcal{S}(\mathbb{R}^n). \tag{1.14}$$

If also $\delta < 1$, show that

$$u \in \mathcal{S}'(\mathbb{R}^n) \Longrightarrow p_\epsilon(x,D)u \to p(x,D)u \;\text{ in } \mathcal{S}'(\mathbb{R}^n), \tag{1.15}$$

where we give $\mathcal{S}'(\mathbb{R}^n)$ the weak* topology.

7. For $s \in \mathbb{R}$, define $\Lambda^s : \mathcal{S}'(\mathbb{R}^n) \to \mathcal{S}'(\mathbb{R}^n)$ by

$$\Lambda^s u(x) = \int \langle\xi\rangle^s \hat{u}(\xi)\, e^{ix\cdot\xi}\, d\xi, \tag{1.16}$$

where $\langle\xi\rangle = (1+|\xi|^2)^{1/2}$. Show that $\Lambda^s \in OPS^s$.

8. Given $p_j(x,\xi) \in S^{m_j}_{\rho,\delta}$, for $j \ge 0$, with $\rho,\delta \in [0,1]$ and $m_j \searrow -\infty$, show that there exists $p(x,\xi) \in S^{m_0}_{\rho,\delta}$ such that

$$p(x,\xi) \sim \sum_{j\ge 0} p_j(x,\xi),$$

in the sense that, for all k,

$$p(x,\xi) - \sum_{j=0}^{k-1} p_j(x,\xi) \in S^{m_k}_{\rho,\delta}.$$

2. Schwartz kernels of pseudodifferential operators

To an operator $p(x,D) \in OPS^m_{\rho,\delta}$ defined by (1.3) there corresponds a Schwartz kernel $K \in \mathcal{D}'(\mathbb{R}^n \times \mathbb{R}^n)$, satisfying

$$\begin{aligned}\langle u(x)v(y), K\rangle &= \iint u(x)p(x,\xi)\hat{v}(\xi)e^{ix\cdot\xi}\, d\xi\, dx \\ &= (2\pi)^{-n} \iiint u(x)p(x,\xi)e^{i(x-y)\cdot\xi} v(y)\, dy\, d\xi\, dx.\end{aligned} \tag{2.1}$$

Thus, K is given as an "oscillatory integral"

$$K = (2\pi)^{-n} \int p(x,\xi) e^{i(x-y)\cdot\xi}\, d\xi. \tag{2.2}$$

We have the following basic result.

Proposition 2.1. *If $\rho > 0$, then K is C^∞ off the diagonal in $\mathbb{R}^n \times \mathbb{R}^n$.*

Proof. For given $\alpha \geq 0$,

$$(x-y)^\alpha K = \int e^{i(x-y)\cdot\xi}\, D_\xi^\alpha p(x,\xi)\, d\xi. \tag{2.3}$$

This integral is clearly absolutely convergent for $|\alpha|$ so large that $m - \rho|\alpha| < -n$. Similarly, it is seen that applying j derivatives to (2.3) yields an absolutely convergent integral provided $m + j - \rho|\alpha| < -n$, so in that case $(x-y)^\alpha K \in C^j(\mathbb{R}^n \times \mathbb{R}^n)$. This gives the proof.

Generally, if T has the mapping properties

$$T : C_0^\infty(\mathbb{R}^n) \longrightarrow C^\infty(\mathbb{R}^n), \quad T : \mathcal{E}'(\mathbb{R}^n) \longrightarrow \mathcal{D}'(\mathbb{R}^n),$$

and its Schwartz kernel K is C^∞ off the diagonal, it follows easily that

$$\text{sing supp}\, Tu \subset \text{sing supp}\, u, \quad \text{for } u \in \mathcal{E}'(\mathbb{R}^n).$$

This is called the *pseudolocal property*. By (1.7)– (1.8) it holds for $T \in OPS^m_{\rho,\delta}$ if $\rho > 0$ and $\delta < 1$.

We remark that the proof of Proposition 2.1 leads to the estimate

$$|D^\beta_{x,y} K| \leq C|x-y|^{-k}, \tag{2.4}$$

where $k \geq 0$ is any integer strictly greater than $(1/\rho)(m + n + |\beta|)$. In fact, this estimate is rather crude. It is of interest to record a more precise estimate that holds when $p(x,\xi) \in S^m_{1,\delta}$.

Proposition 2.2. *If $p(x,\xi) \in S^m_{1,\delta}$, then the Schwartz kernel K of $p(x,D)$ satisfies estimates*

$$|D^\beta_{x,y} K| \leq C|x-y|^{-n-m-|\beta|} \tag{2.5}$$

provided $m + |\beta| > -n$.

The result is easily reduced to the case $p(x,\xi) = p(\xi)$, satisfying $|D^\alpha p(\xi)| \leq C_\alpha \langle\xi\rangle^{m-|\alpha|}$, for which $p(D)$ has Schwartz kernel $K = \hat{p}(y-x)$. It suffices to prove (2.5) for such a case, for $\beta = 0$ and $m > -n$. We make use of the following simple but important characterization of such symbols.

Lemma 2.3. *Given $p(\xi) \in C^\infty(\mathbb{R}^n)$, it belongs to $S^m_{1,0}$ if and only if*

$$p_r(\xi) = r^{-m} p(r\xi) \text{ is bounded in } C^\infty(1 \le |\xi| \le 2), \ \text{for } r \in [1,\infty). \tag{2.6}$$

Given this, we can write $p(\xi) = p_0(\xi) + \int_0^\infty q_\tau(e^{-\tau}\xi)\,d\tau$ with $p_0(\xi) \in C_0^\infty(\mathbb{R}^n)$ and $e^{-m\tau}q_\tau(\xi)$ bounded in the Schwartz space $\mathcal{S}(\mathbb{R}^n)$, for $\tau \in [0,\infty)$. Hence $e^{-m\tau}\hat{q}_\tau(z)$ is bounded in $\mathcal{S}(\mathbb{R}^n)$. In particular, $e^{-m\tau}|\hat{q}_\tau(z)| \le C_N\langle z\rangle^{-N}$, so

$$\begin{aligned} |\hat{p}(z)| &\le |\hat{p}_0(z)| + C_N \int_0^\infty e^{(n+m)\tau}\big(1+|e^\tau z|\big)^{-N}\,d\tau \\ &\le C + C_N|z|^{-n-m}\int_{\log|z|}^\infty e^{(n+m)\tau}(1+e^\tau)^{-N}\,d\tau, \end{aligned} \tag{2.7}$$

which implies (2.5). We also see that in the case $m + |\beta| = -n$, we obtain a result upon replacing the right side of (2.5) by $C \log|x-y|^{-1}$, (provided $|x-y| < 1/2$).

We can get a complete characterization of $\hat{P}(x) \in \mathcal{S}'(\mathbb{R}^n)$, given $P(\xi) \in S^m_1(\mathbb{R}^n)$, provided $-n < m < 0$.

Proposition 2.4. *Assume $-n < m < 0$. Let $q \in \mathcal{S}'(\mathbb{R}^n)$ be smooth outside the origin and rapidly decreasing as $|x| \to \infty$. Then $q = \hat{P}$ for some $P(\xi) \in S^m_1(\mathbb{R}^n)$ if and only if $q \in L^1_{loc}(\mathbb{R}^n)$ and, for $x \ne 0$,*

$$|D_x^\beta q(x)| \le C_\beta\,|x|^{-n-m-|\beta|}. \tag{2.8}$$

Proof. That $P \in S^m_1(\mathbb{R}^n)$ implies (2.8) has been established above. For the converse, write $q = q_0(x) + \sum_{j\ge0}\psi_j(x)q(x)$, where $\psi_0 \in C_0^\infty(\mathbb{R}^n)$ is supported in $1/2 < |x| < 2$, $\psi_j(x) = \psi_0(2^j x)$, $\sum_{j\ge0}\psi_j(x) = 1$ on $|x| \le 1$. Since $|q(x)| \le C|x|^{-n-m}$, $m < 0$, it follows that $\sum \psi_j(x)q(x)$ converges in L^1-norm. Then $q_0 \in \mathcal{S}(\mathbb{R}^n)$. The hypothesis (2.8) implies that $2^{-nj-mj}\psi_j(2^{-j}x)q(2^{-j}x)$ is bounded in $\mathcal{S}(\mathbb{R}^n)$, and an argument similar to that used for Proposition 2.2 implies $\hat{q}_0(\xi) + \sum_{j=0}^\infty(\psi_j q)\hat{\ }(\xi) \in S^m_1(\mathbb{R}^n)$.

We will deal further with the space of elements of $\mathcal{S}'(\mathbb{R}^n)$ that are smooth outside the origin and rapidly decreasing (with all their derivatives) at infinity. We will denote this space by $\mathcal{S}'_0(\mathbb{R}^n)$.

If $m \le -n$, the argument above extends to show that (2.8) is a sufficient condition for $q = \hat{P}$ with $P \in S^m_1(\mathbb{R}^n)$, but, as noted above, there exist symbols $P \in S^m_1(\mathbb{R}^n)$ for which $q = \hat{P}$ does not satisfy (2.8). Now, given that $q \in \mathcal{S}'_0(\mathbb{R}^n)$, it is easy to see that

$$\nabla q \in \mathcal{F}\big(S^{m+1}_1(\mathbb{R}^n)\big) \Longleftrightarrow q \in \mathcal{F}\big(S^m_1(\mathbb{R}^n)\big). \tag{2.9}$$

Thus, if $-n-1 < m \le -n$, then Proposition 2.4 is almost applicable to ∇q, for $n \ge 2$.

Proposition 2.5. *Assume $n \geq 2$ and $-n-1 < m \leq -n$. If $q \in \mathcal{S}_0'(\mathbb{R}^n) \cap L^1_{loc}$, then $q = \hat{P}$ for some $P \in S^m(\mathbb{R}^n)$ if and only if (2.8) holds for $|\beta| \geq 1$.*

Proof. First note that the hypotheses imply $q \in L^1(\mathbb{R}^n)$; thus $\tilde{q}(\xi)$ is continuous and vanishes as $|\xi| \to \infty$. In the proposition, we need to prove the "if" part. To use the reasoning behind Proposition 2.4, we need only deal with the fact that ∇q is not assumed to be in L^1_{loc}. The sum $\sum \psi_j(x)\nabla q(x)$ still converges in $L^1(\mathbb{R}^n)$, and so $\nabla q - \sum \psi_j(x)\nabla q$ is a sum of an element of $\mathcal{S}(\mathbb{R}^n)$ and possibly a distribution (call it ν) supported at 0. Thus $\hat{\nu}(\xi)$ is a polynomial. But as noted, $\hat{q}(\xi)$ is bounded, so $\hat{\nu}(\xi)$ can have at most linear growth. Hence

$$\xi_j \tilde{q}(\xi) = P_j(\xi) + \ell_j(\xi),$$

where $P_j \in S_1^{m+1}(\mathbb{R}^n)$ and $\ell_j(\xi)$ is a first-order polynomial in ξ. Since $\tilde{q}(\xi) \to 0$ as $|\xi| \to \infty$ and $m+1 \leq -n+1 < 0$, we deduce that $\ell_j(\xi) = c_j$, a constant, that is,

$$\xi_j \tilde{q}(\xi) = P_j(\xi) + c_j, \quad P_j \in S_1^{m+1}(\mathbb{R}^n), \ m+1 < 0. \tag{2.10}$$

Now the left side vanishes on the hyperplane $\xi_j = 0$, which is unbounded if $n \geq 2$. This forces $c_j = 0$, and the proof of the proposition is then easily completed.

If we take $n = 1$ and assume $-2 < m < -1$, the rest of the hypotheses of Proposition 2.5 still yield (2.10), so

$$\frac{dq}{dx} = \hat{P}_1 + c_1\delta.$$

If we also assume q is continuous on $\mathbb{R}$, then $c_1 = 0$ and we again conclude that $q = \hat{P}$ with $P \in S_1^m(\mathbb{R})$. But if q has a simple jump at $x = 0$, then this conclusion fails.

Proposition 2.4 can be given other extensions, which we leave to the reader. We give a few examples that indicate ways in which the result does not extend, making use of results from §8 of Chap. 3. As shown in (8.31) of that chapter, on $\mathbb{R}^n$,

$$v = PF\,|x|^{-n} \Longrightarrow \hat{v}(\xi) = C_n \ \log|\xi|. \tag{2.11}$$

Now v is not rapidly decreasing at infinity, but if $\varphi(x)$ is a cut-off, belonging to $C_0^\infty(\mathbb{R}^n)$ and equal to 1 near $x = 0$, then $f = \varphi v$ belongs to $\mathcal{S}_0'(\mathbb{R}^n)$ and $\hat{f} = c\hat{\varphi} * \hat{v}$ behaves like $\log|\xi|$ as $|\xi| \to \infty$. One can then deduce that, for $n = 1$,

$$f(x) = \varphi(x)\ \log|x| \operatorname{sgn}|x| \Longrightarrow \hat{f}(\xi) \sim C\ \xi^{-1}\ \log|\xi|, \quad |\xi| \to \infty. \tag{2.12}$$

Thus Proposition 2.5 does not extend to the case $n = 1$, $m = -1$. However, we note that, in this case, $\hat{f}$ belongs to $S_1^{-1+\varepsilon}(\mathbb{R})$, for all $\varepsilon > 0$. In contrast to (2.12), note that, again for $n = 1$,

$$g(x) = \varphi(x)\ \log|x| \Longrightarrow \hat{g}(\xi) \sim C\,|\xi|^{-1}, \quad |\xi| \to \infty. \tag{2.13}$$

In this case, $(d/dx)\log|x| = PV(1/x)$.

Of considerable utility is the classification of $\mathcal{F}\big(S^m_{cl}(\mathbb{R}^n)\big)$. When $m = -j$ is a negative integer, this was effectively solved in §§8 and 9 of Chap. 3. The following result is what follows from the proof of Proposition 9.2 in Chap. 3.

Proposition 2.6. *Assume $q \in \mathcal{S}'_0(\mathbb{R}^n) \cap L^1_{loc}(\mathbb{R}^n)$. Let $j = 1, 2, 3, \dots$. Then $q = \hat{P}$ for some $P \in S^{-j}_{cl}(\mathbb{R}^n)$ if and only if*

$$q \sim \sum_{\ell \ge 0} \big(q_\ell + p_\ell(x)\log|x|\big), \tag{2.14}$$

where

$$q_\ell \in \mathcal{H}^{\#}_{j+\ell-n}(\mathbb{R}^n), \tag{2.15}$$

and $p_\ell(x)$ is a polynomial homogeneous of degree $j+\ell-n$; these log coefficients appear only for $\ell \ge n-j$.

We recall that $\mathcal{H}^{\#}_\mu(\mathbb{R}^n)$ is the space of distributions on $\mathbb{R}^n$, homogeneous of degree μ, which are smooth on $\mathbb{R}^n \setminus 0$. For $\mu > -n$, $\mathcal{H}^{\#}_\mu(\mathbb{R}^n) \subset L^1_{\text{loc}}(\mathbb{R}^n)$. The meaning of the expansion (2.14) is that, for any $k \in \mathbb{Z}^+$, there is an $N < \infty$ such that the difference between q and the sum over $\ell < N$ belongs to $C^k(\mathbb{R}^n)$. Note that, for $n = 1$, the function $g(x)$ in (2.13) is of the form (2.14), but the function $f(x)$ in (2.12) is not.

To go from the proof of Proposition 9.2 of Chap. 3 to the result stated above, it suffices to note explicitly that

$$\varphi(x)x^\alpha \log|x| \in \mathcal{F}\big(S_1^{-n-|\alpha|}(\mathbb{R}^n)\big), \tag{2.16}$$

where φ is the cut-off used before. Since $\mathcal{F}$ intertwines D^α_ξ and multiplication by x^α, it suffices to verify the case $\alpha = 0$, and this follows from the formula (2.11), with x and ξ interchanged.

We can also classify Schwartz kernels of operators in $OPS^m_{1,0}$ and OPS^m_{cl}, if we write the kernel K of (2.2) in the form

$$K(x,y) = L(x, x-y), \tag{2.17}$$

with

$$L(x,z) = (2\pi)^{-n} \int p(x,\xi) e^{iz\cdot\xi}\, d\xi. \tag{2.18}$$

The following two results follow from the arguments given above.

Proposition 2.7. *Assume* $-n < m < 0$. *Let* $L \in \mathcal{S}'(\mathbb{R}^n \times \mathbb{R}^n)$ *be a smooth function of* x *with values in* $\mathcal{S}_0'(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$. *Then (2.17) defines the Schwartz kernel of an operator in* $OPS^m_{1,0}$ *if and only if, for* $z \neq 0$,

$$|D_x^\beta D_z^\gamma L(x,z)| \le C_{\beta\gamma}\, |z|^{-n-m-|\gamma|}. \tag{2.19}$$

Proposition 2.8. *Assume* $L \in \mathcal{S}'(\mathbb{R}^n \times \mathbb{R}^n)$ *is a smooth function of* x *with values in* $\mathcal{S}_0'(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$. *Let* $j = 1, 2, 3, \dots$. *Then (2.17) defines the Schwartz kernel of an operator in* OPS^{-j}_{cl} *if and only if*

$$L(x,z) \sim \sum_{\ell \ge 0} \bigl(q_\ell(x,z) + p_\ell(x,z) \log |z|\bigr), \tag{2.20}$$

where each $D_x^\beta q_\ell(x,\cdot)$ *is a bounded continuous function of* x *with values in* $\mathcal{H}^\#_{j+\ell-n}$, *and* $p_\ell(x,z)$ *is a polynomial homogeneous of degree* $j + \ell - n$ *in* z, *with coefficients that are bounded, together with all their* x*-derivatives.*

Exercises

1. Using the proof of Proposition 2.2, show that, given $p(x,\xi)$ defined on $\mathbb{R}^n \times \mathbb{R}^n$, then

$$|D_x^\beta D_\xi^\alpha p(x,\xi)| \le C' \langle \xi \rangle^{-|\alpha|+|\beta|}, \ \text{ for } |\beta| \le 1,\ |\alpha| \le n+1+|\beta|,$$

   implies

$$|K(x,y)| \le C|x-y|^{-n} \ \text{ and } \ |\nabla_{x,y} K(x,y)| \le C|x-y|^{-n-1}.$$

2. If the map κ is given by (2.2) (i.e., $\kappa(p) = K$) show that we get an isomorphism $\kappa : \mathcal{S}'(\mathbb{R}^{2n}) \to \mathcal{S}'(\mathbb{R}^{2n})$. Reconsider Exercise 3 of §1.
3. Show that κ, defined in Exercise 2, gives an isomorphism (isometric up to a scalar factor) $\kappa : L^2(\mathbb{R}^{2n}) \to L^2(\mathbb{R}^{2n})$. Deduce that $p(x,D)$ is a Hilbert-Schmidt operator on $L^2(\mathbb{R}^n)$, precisely when $p(x,\xi) \in L^2(\mathbb{R}^{2n})$.

3. Adjoints and products

Given $p(x,\xi) \in S^m_{\rho,\delta}$, we obtain readily from the definition that the adjoint is given by

$$p(x,D)^* v = (2\pi)^{-n} \int p(y,\xi)^* e^{i(x-y)\cdot\xi} v(y)\, dy\, d\xi. \tag{3.1}$$

This is not quite in the form (1.3), as the amplitude $p(y,\xi)^*$ is not a function of (x,ξ). We need to transform (3.1) into such a form.

Before continuing the analysis of (3.1), we are motivated to look at a general class of operators

$$Au(x) = (2\pi)^{-n} \int a(x,y,\xi) e^{i(x-y)\cdot\xi} u(y)\, dy\, d\xi. \tag{3.2}$$

We assume

$$|D_y^\gamma D_x^\beta D_\xi^\alpha a(x,y,\xi)| \le C_{\alpha\beta\gamma} \langle\xi\rangle^{m-\rho|\alpha|+\delta_1|\beta|+\delta_2|\gamma|} \tag{3.3}$$

and then say $a(x,y,\xi) \in S^m_{\rho,\delta_1,\delta_2}$. A brief calculation transforms (3.2) into

$$(2\pi)^{-n} \int q(x,\xi) e^{i(x-y)\cdot\xi} u(y)\, dy\, d\xi, \tag{3.4}$$

with

$$\begin{aligned} q(x,\xi) &= (2\pi)^{-n} \int a(x,y,\eta) e^{i(x-y)\cdot(\eta-\xi)}\, dy\, d\eta \\ &= e^{iD_\xi\cdot D_y} a(x,y,\xi)\big|_{y=x}. \end{aligned} \tag{3.5}$$

Note that a formal expansion $e^{iD_\xi\cdot D_y} = I + iD_\xi \cdot D_y - (1/2)(D_\xi \cdot D_y)^2 + \cdots$ gives

$$q(x,\xi) \sim \sum_{\alpha\ge 0} \frac{i^{|\alpha|}}{\alpha!} D_\xi^\alpha D_y^\alpha a(x,y,\xi)\big|_{y=x}. \tag{3.6}$$

If $a(x,y,\xi) \in S^m_{\rho,\delta_1,\delta_2}$, with $0 \le \delta_2 < \rho \le 1$, then the general term in (3.6) belongs to $S^{m-(\rho-\delta_2)|\alpha|}_{\rho,\delta}$, where $\delta = \max(\delta_1,\delta_2)$, so the sum on the right is formally asymptotic. This suggests the following result:

Proposition 3.1. *If $a(x,y,\xi) \in S^m_{\rho,\delta_1,\delta_2}$, with $0 \le \delta_2 < \rho \le 1$, then (3.2) defines an operator*

$$A \in OPS^m_{\rho,\delta}, \quad \delta = \max(\delta_1,\delta_2).$$

Furthermore, $A = q(x,D)$, where $q(x,\xi)$ has the asymptotic expansion (3.6), in the sense that

$$q(x,\xi) - \sum_{|\alpha|<N} \frac{i^{|\alpha|}}{\alpha!} D_\xi^\alpha D_y^\alpha a(x,y,\xi)\big|_{y=x} = r_N(x,\xi) \in S^{m-N(\rho-\delta_2)}_{\rho,\delta}.$$

To prove this proposition, one can first show that the Schwartz kernel

$$K(x,y) = (2\pi)^{-n} \int a(x,y,\xi) e^{i(x-y)\cdot\xi}\, d\xi$$

satisfies the same estimates as established in Proposition 2.1, and hence, altering A only by an operator in $OPS^{-\infty}$, we can assume $a(x, y, \xi)$ is supported on $|x - y| \le 1$. Let

$$\hat{b}(x, \eta, \xi) = (2\pi)^{-n} \int a(x, x + y, \xi) e^{-iy\cdot\eta}\, dy, \tag{3.7}$$

so

$$q(x, \xi) = \int \hat{b}(x, \eta, \xi + \eta)\, d\eta. \tag{3.8}$$

The hypotheses on $a(x, y, \xi)$ imply

$$|D_x^\beta D_\xi^\alpha \hat{b}(x, \eta, \xi)| \le C_{\nu\alpha\beta} \langle\xi\rangle^{m+\delta|\beta|+\delta_2\nu-\rho|\alpha|} \langle\eta\rangle^{-\nu}, \tag{3.9}$$

where $\delta = \max\,(\delta_1, \delta_2)$. Since $\delta_2 < 1$, it follows that $q(x, \xi)$ and any of its derivatives can be bounded by some power of $\langle\xi\rangle$.

Now a power-series expansion of $\hat{b}(x, \eta, \xi + \eta)$ in the last argument about ξ gives

$$\begin{aligned} &\Big|\hat{b}(x, \eta, \xi + \eta) - \sum_{|\alpha|<N} \frac{1}{\alpha!} (iD_\xi)^\alpha \hat{b}(x, \eta, \xi)\eta^\alpha\Big| \\ &\qquad\le C_\nu |\eta|^N \langle\eta\rangle^{-\nu} \sup_{0\le t\le 1} \langle\xi + t\eta\rangle^{m+\delta_2\nu-\rho N}. \end{aligned} \tag{3.10}$$

Taking $\nu = N$, we get a bound on the left side of (3.10) by

$$C\langle\xi\rangle^{m-(\rho-\delta_2)N} \quad \text{if } |\eta| \le \frac{1}{2}|\xi|, \tag{3.11}$$

while taking ν large, we get a bound by any power of $\langle\eta\rangle^{-1}$ for $|\xi| \le 2|\eta|$. Hence

$$\Big|q(x, \xi) - \sum_{|\alpha|<N} \frac{1}{\alpha!} (iD_\xi)^\alpha D_y^\alpha a(x, x + y, \xi)\big|_{y=0}\Big| \le C\langle\xi\rangle^{m+n-(\rho-\delta_2)N}. \tag{3.12}$$

The proposition follows from this, plus similar estimates on the difference when derivatives are applied.

If we apply Proposition 3.1 to (3.1), we obtain:

Proposition 3.2. *If* $p(x, D) \in OPS^m_{\rho,\delta}$, $0 \le \delta < \rho \le 1$, *then*

$$p(x, D)^* = p^*(x, D) \in OPS^m_{\rho,\delta}, \tag{3.13}$$

with

$$p^*(x,\xi) \sim \sum_{\alpha\ge 0} \frac{i^{|\alpha|}}{\alpha!} D_\xi^\alpha D_x^\alpha p(x,\xi)^*. \tag{3.14}$$

The result for products of pseudodifferential operators is the following.

Proposition 3.3. *Given* $p_j(x,D) \in OPS^{m_j}_{\rho_j,\delta_j}$, *suppose*

$$0 \le \delta_2 < \rho \le 1, \quad \text{with } \rho = \min(\rho_1,\rho_2). \tag{3.15}$$

Then

$$p_1(x,D)p_2(x,D) = q(x,D) \in OPS^{m_1+m_2}_{\rho,\delta}, \tag{3.16}$$

with $\delta = \max(\delta_1,\delta_2)$, *and*

$$q(x,\xi) \sim \sum_{\alpha\ge 0} \frac{i^{|\alpha|}}{\alpha!} D_\xi^\alpha p_1(x,\xi)\ D_x^\alpha p_2(x,\xi). \tag{3.17}$$

This can be proved by writing

$$p_1(x,D)p_2(x,D)u = p_1(x,D)p_2^*(x,D)^*u = Au, \tag{3.18}$$

for A as in (3.2), with

$$a(x,y,\xi) = p_1(x,\xi)p_2^*(y,\xi)^*, \tag{3.19}$$

and then applying Propositions 3.1 and 3.2, to obtain (3.16), with

$$q(x,\xi) \sim \sum_{\gamma,\sigma\ge 0} \frac{i^{|\sigma|-|\gamma|}}{\sigma!\gamma!} D_\xi^\sigma D_y^\sigma \Big(p_1(x,\xi) D_\xi^\gamma D_x^\gamma p_2(y,\xi)\Big)\Big|_{y=x}. \tag{3.20}$$

The general term in this sum is equal to

$$\frac{i^{|\sigma|-|\gamma|}}{\sigma!\gamma!} D_\xi^\sigma \Big(p_1(x,\xi) D_\xi^\gamma D_x^{\gamma+\sigma} p_2(x,\xi)\Big).$$

Evaluating this by the product rule

$$D_\xi^\sigma(uv) = \sum_{\alpha+\beta=\sigma} \binom{\sigma}{\alpha} D_\xi^\alpha u \cdot D_\xi^\beta v$$

gives

$$(3.21)\qquad q(x,\xi) \sim \sum_{\alpha} \frac{i^{|\alpha|}}{\alpha!} D_\xi^\alpha p_1(x,\xi) \sum_{\beta,\gamma} \frac{i^{|\beta|-|\gamma|}}{\beta!\gamma!} D_\xi^{\beta+\gamma} D_x^{\beta+\gamma+\alpha} p_2(x,\xi).$$

That this yields (3.17) follows from the fact that, whenever $|\mu| > 0$,

$$(3.22)\qquad \sum_{\beta+\gamma=\mu} \frac{i^{|\beta|-|\gamma|}}{\beta!\gamma!} D_\xi^{\beta+\gamma} D_x^{\beta+\gamma+\alpha} p_2(x,\xi) = 0,$$

an identity we leave as an exercise.

An alternative approach to a proof of Proposition 3.3 is to compute directly that $p_1(x,D)p_2(x,D) = q(x,D)$, with

$$(3.23)\qquad \begin{aligned} q(x,\xi) &= (2\pi)^{-n} \int p_1(x,\eta)p_2(y,\xi)e^{i(x-y)\cdot(\eta-\xi)}\, d\eta\, dy \\ &= e^{iD_\eta\cdot D_y} p_1(x,\eta)p_2(y,\xi)\big|_{y=x,\eta=\xi}, \end{aligned}$$

and then apply an analysis such as used to prove Proposition 3.1. Carrying out this latter approach has the advantage that the hypothesis (3.15) can be weakened to

$$0 \le \delta_2 < \rho_1 \le 1,$$

which is quite natural since the right side of (3.17) is formally asymptotic under such a hypothesis. Also, the symbol expansion (3.17) is more easily seen from (3.23).

Note that if $P_j = p_j(x,D) \in OPS^{m_j}_{\rho,\delta}$ are scalar, and $0 \le \delta < \rho \le 1$, then the leading terms in the expansions of the symbols of P_1P_2 and P_2P_1 agree. It follows that the commutator

$$[P_1, P_2] = P_1P_2 - P_2P_1$$

has order lower than $m_1 + m_2$. In fact, the symbol expansion (3.17) implies

$$(3.24)\qquad P_j \in OPS^{m_j}_{\rho,\delta} \text{ scalar} \Longrightarrow [P_1,P_2] \in OPS^{m_1+m_2-(\rho-\delta)}_{\rho,\delta}.$$

Also, looking at the sum over $|\alpha| = 1$ in (3.17), we see that the leading term in the expansion of the symbol of $[P_1, P_2]$ is given in terms of the Poisson bracket:

$$(3.25)\quad [P_1,P_2] = q(x,D), \quad q(x,\xi) = \frac{1}{i}\{p_1,p_2\}(x,\xi) \bmod S^{m_1+m_2-2(\rho-\delta)}_{\rho,\delta}.$$

The Poisson bracket $\{p_1, p_2\}$ is defined by

$$(3.26)\qquad \{p_1,p_2\}(x,\xi) = \sum_j \frac{\partial p_1}{\partial \xi_j}\frac{\partial p_2}{\partial x_j} - \frac{\partial p_1}{\partial x_j}\frac{\partial p_2}{\partial \xi_j},$$

as in §10 of Chap. 1.

The result (3.25) plays an important role in the treatment of Egorov's theorem, in §8.

Exercises

1. Writing $a_j(x, D)$ in the form (1.10), that is,

$$a_j(x, D) = \int \hat{a}_j(q, p) e^{iq\cdot X} e^{ip\cdot D}\, dq\, dp, \tag{3.27}$$

   use the formula (1.11) for $e^{ip\cdot D} e^{iq'\cdot X}$ to express $a_1(x, D)a_2(x, D)$ as a $4n$-fold integral. Show that it gives (3.20).
2. If $Q(x, x)$ is any nondegenerate, symmetric, bilinear form on $\mathbb{R}^n$, calculate the kernel $K_Q(x, y, t)$ for which

$$e^{itQ(D,D)} u(x) = \int_{\mathbb{R}^n} K_Q(x, y, t)\, u(y)\, dy. \tag{3.28}$$

   In case $x \in \mathbb{R}^n$ is replaced by $(x, \xi) \in \mathbb{R}^{2n}$, use this to verify (3.5).
   (*Hint*: Diagonalize Q and recall the treatment of $e^{it\Delta}$ in (6.42) of Chap. 3, giving

$$e^{-it\Delta}\delta(x) = (-4\pi i t)^{-n/2}\, e^{|x|^2/4it}, \quad x \in \mathbb{R}^n.$$

   Compare the treatment of the stationary phase method in Appendix B of Chap. 6.)
3. Establish the identity (3.22), used in the proof of Proposition 3.3.
   (*Hint*: The left side of (3.22) is equal to

$$\left(\sum_{\beta+\gamma=\mu} \frac{i^{|\beta|-|\gamma|}}{\beta!\gamma!} \right) D_\xi^\mu D_x^{\mu+\alpha} p_2(x, \xi),$$

   so one needs to show that the quantity in parentheses here vanishes if $|\mu| > 0$. To see this, make an expansion of $(z + w)^\mu$, and set $z = (i, \dots, i)$, $w = (-i, \dots, -i)$.)

4. Elliptic operators and parametrices

We say $p(x, D) \in OPS^m_{\rho,\delta}$ is elliptic if, for some $r < \infty$,

$$|p(x, \xi)^{-1}| \le C\langle\xi\rangle^{-m}, \quad \text{for } |\xi| \ge r. \tag{4.1}$$

Thus, if $\psi(\xi) \in C^\infty(\mathbb{R}^n)$ is equal to 0 for $|\xi| \le r$, 1 for $|\xi| \ge 2r$, it follows easily from the chain rule that

$$\psi(\xi) p(x, \xi)^{-1} = q_0(x, \xi) \in S^{-m}_{\rho,\delta}. \tag{4.2}$$

As long as $0 \le \delta < \rho \le 1$, we can apply Proposition 3.3 to obtain

$$\begin{aligned} q_0(x, D)p(x, D) &= I + r_0(x, D), \\ p(x, D)q_0(x, D) &= I + \tilde{r}_0(x, D), \end{aligned} \tag{4.3}$$

with

$$r_0(x,\xi),\ \tilde{r}_0(x,\xi) \in S^{-(\rho-\delta)}_{\rho,\delta}. \tag{4.4}$$

Using the formal expansion

$$I - r_0(x,D) + r_0(x,D)^2 - \cdots \sim I + s(x,D) \in OPS^0_{\rho,\delta} \tag{4.5}$$

and setting $q(x,D) = (I + s(x,D))q_0(x,D) \in OPS^{-m}_{\rho,\delta}$, we have

$$q(x,D)p(x,D) = I + r(x,D), \quad r(x,\xi) \in S^{-\infty}. \tag{4.6}$$

Similarly, we obtain $\tilde{q}(x,D) \in OPS^{-m}_{\rho,\delta}$ satisfying

$$p(x,D)\tilde{q}(x,D) = I + \tilde{r}(x,D), \quad \tilde{r}(x,\xi) \in S^{-\infty}. \tag{4.7}$$

But evaluating

$$\big(q(x,D)p(x,D)\big)\tilde{q}(x,D) = q(x,D)\big(p(x,D)\tilde{q}(x,D)\big) \tag{4.8}$$

yields $q(x,D) = \tilde{q}(x,D)$ mod $OPS^{-\infty}$, so in fact

$$\begin{aligned} q(x,D)p(x,D) &= I \text{ mod } OPS^{-\infty}, \\ p(x,D)q(x,D) &= I \text{ mod } OPS^{-\infty}. \end{aligned} \tag{4.9}$$

We say that $q(x,D)$ is a *two-sided parametrix* for $p(x,D)$.

The parametrix can establish the local regularity of a solution to

$$p(x,D)u = f. \tag{4.10}$$

Suppose $u, f \in \mathcal{S}'(\mathbb{R}^n)$ and $p(x,D) \in OPS^m_{\rho,\delta}$ is elliptic, with $0 \le \delta < \rho \le 1$. Constructing $q(x,D) \in OPS^{-m}_{\rho,\delta}$ as in (4.6), we have

$$u = q(x,D)f - r(x,D)u. \tag{4.11}$$

Now a simple analysis parallel to (1.7) implies that

$$R \in OPS^{-\infty} \Longrightarrow R : \mathcal{E}' \longrightarrow \mathcal{S}. \tag{4.12}$$

By duality, since taking adjoints preserves $OPS^{-\infty}$,

$$R \in OPS^{-\infty} \Longrightarrow R : \mathcal{S}' \longrightarrow C^\infty. \tag{4.13}$$

Thus (4.11) implies

$$u = q(x,D)f \text{ mod } C^\infty. \tag{4.14}$$

Applying the pseudolocal property to (4.10) and (4.14), we have the following elliptic regularity result.

Proposition 4.1. *If* $p(x, D) \in OPS^m_{\rho,\delta}$ *is elliptic and* $0 \le \delta < \rho \le 1$*, then, for any* $u \in \mathcal{S}'(\mathbb{R}^n)$*,*

$$\text{sing supp } p(x, D)u = \text{ sing supp } u. \tag{4.15}$$

More refined elliptic regularity involves keeping track of Sobolev space regularity. As we have the parametrix, this will follow simply from mapping properties of pseudodifferential operators, to be established in subsequent sections.

Exercises

1. Give the details of the implication (4.1) ⇒ (4.2) when $p(x,\xi) \in S^m_{\rho,\delta}$, $0 \le \delta < \rho \le 1$. Include the case where $p(x,\xi)$ is a $k \times k$ matrix-valued function, using such identities as
$$\frac{\partial}{\partial x_j}\, p(x,\xi)^{-1} = -p(x,\xi)^{-1}\, \frac{\partial p}{\partial x_j}\, p(x,\xi)^{-1}.$$
2. On $\mathbb{R} \times \mathbb{R}^n$, consider the operator $P = \partial/\partial t - L(x, D_x)$, where
$$L(x, D_x) = \sum a_{jk}(x)\, \partial_j \partial_k u + \sum b_j(x)\, \partial_j u + c(x)u.$$
Assume that the coefficients are smooth and bounded, with all their derivatives, and that L satisfies the strong ellipticity condition
$$-L_2(x,\xi) = \sum a_{jk}(x)\xi_j\xi_k \ge C|\xi|^2, \quad C > 0.$$
Show that
$$\big(i\tau - L_2(x,\xi) + 1\big)^{-1} = E(t,x,\tau,\xi) \in S^{-1}_{1/2,0}.$$
Show that $E(t,x,D)P = A_1(t,x,D)$ and $PE(t,x,D) = A_2(t,x,D)$, where $A_j \in OPS^0_{1/2,0}$ are elliptic. Then, using Proposition 4.1, construct a parametrix for P, belonging to $OPS^{-1}_{1/2,0}$.
3. Assume $-n < m < 0$, and suppose $P = p(x, D) \in OPS^m_{cl}$ has Schwartz kernel $K(x, y) = L(x, x - y)$. Suppose that, at $x_0 \in \mathbb{R}^n$,
$$L(x_0, z) \sim a|z|^{-m-n} + \cdots, \quad z \to 0,$$
with $a \ne 0$, the remainder terms being progressively smoother. Show that
$$p_m(x_0,\xi) = b|\xi|^m, \quad b \ne 0,$$
and hence that P is elliptic near x_0.
4. Let $P = (P_{jk})$ be a $K \times K$ matrix of operators in OPS^*. It is said to be "elliptic in the sense of Douglis and Nirenberg" if there are numbers a_j, b_j, $1 \le j \le K$, such that $P_{jk} \in OPS^{a_j+b_k}$ and the matrix of principal symbols has nonvanishing determinant (homogeneous of order $\sum(a_j + b_j)$), for $\xi \ne 0$. If Λ^s is as in (1,17), let A be a $K \times K$

diagonal matrix with diagonal entries Λ^{-a_j}, and let B be diagonal, with entries Λ^{-b_j}. Show that this "DN-ellipticity" of P is equivalent to the ellipticity of APB in OPS^0.

5. L^2-estimates

Here we want to obtain L^2-estimates for pseudodifferential operators. The following simple basic estimate will get us started.

Proposition 5.1. *Let (X, μ) be a measure space. Suppose $k(x, y)$ is measurable on $X \times X$ and*

$$\int_X |k(x,y)|\, d\mu(x) \le C_1, \quad \int_X |k(x,y)|\, d\mu(y) \le C_2, \tag{5.1}$$

for all y and x, respectively. Then

$$Tu(x) = \int k(x,y)u(y)\, d\mu(y) \tag{5.2}$$

satisfies

$$\|Tu\|_{L^p} \le C_1^{1/p}\, C_2^{1/q}\, \|u\|_{L^p}, \tag{5.3}$$

for $p \in [1, \infty]$, with

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{5.4}$$

This is proved in Appendix A on functional analysis; see Proposition 5.1 there. To apply this result when $X = \mathbb{R}^n$ and $k = K$ is the Schwartz kernel of $p(x, D) \in OPS^m_{\rho,\delta}$, note from the proof of Proposition 2.1 that

$$|K(x,y)| \le C_N |x-y|^{-N}, \quad \text{for } |x-y| \ge 1 \tag{5.5}$$

as long as $\rho > 0$, while

$$|K(x,y)| \le C|x-y|^{-(n-1)}, \quad \text{for } |x-y| \le 1 \tag{5.6}$$

as long as $m < -n + \rho(n-1)$. (Recall that this last estimate is actually rather crude.) Hence we have the following preliminary result.

Lemma 5.2. *If $p(x, D) \in OPS^m_{\rho,\delta}$, $\rho > 0$, and $m < -n + \rho(n-1)$, then*

$$p(x,D) : L^p(\mathbb{R}^n) \longrightarrow L^p(\mathbb{R}^n), \quad 1 \le p \le \infty. \tag{5.7}$$

If $p(x, D) \in OPS^m_{1,\delta}$, then (5.7) holds for $m < 0$.

The last observation follows from the improvement of (5.6) given in (2.5). Our main goal in this section is to prove the following.

Theorem 5.3. *If* $p(x, D) \in OPS^0_{\rho,\delta}$ *and* $0 \le \delta < \rho \le 1$*, then*

$$p(x, D) : L^2(\mathbb{R}^n) \longrightarrow L^2(\mathbb{R}^n). \tag{5.8}$$

The proof we give, following [Ho5], begins with the following result.

Lemma 5.4. *If* $p(x, D) \in OPS^{-a}_{\rho,\delta}$, $0 \le \delta < \rho \le 1$, *and* $a > 0$, *then (5.8) holds.*

Proof. Since $\|Pu\|^2_{L^2} = (P^*Pu, u)$, it suffices to prove that some power of $p(x, D)^*p(x, D) = Q$ is bounded on L^2. But $Q^k \in OPS^{-2ka}_{\rho,\delta}$, so for k large enough this follows from Lemma 5.2.

To proceed with the proof of Theorem 5.3, set $q(x, D) = p(x, D)^*p(x, D) \in OPS^0_{\rho,\delta}$, and suppose $|q(x,\xi)| \le M - b$, $b > 0$, so

$$M - \operatorname{Re} q(x,\xi) \ge b > 0. \tag{5.9}$$

In the matrix case, take $\operatorname{Re} q(x,\xi) = (1/2)\big(q(x,\xi) + q(x,\xi)^*\big)$. It follows that

$$A(x,\xi) = \big(M - \operatorname{Re} q(x,\xi)\big)^{1/2} \in S^0_{\rho,\delta} \tag{5.10}$$

and

$$A(x, D)^*A(x, D) = M - q(x, D) + r(x, D), \quad r(x, D) \in OPS^{-(\rho-\delta)}_{\rho,\delta}. \tag{5.11}$$

Applying Lemma 5.4 to $r(x, D)$, we have

$$M\|u\|^2_{L^2} - \|p(x, D)u\|^2_{L^2} = \|A(x, D)u\|^2_{L^2} - (r(x, D)u, u) \ge -C\|u\|^2_{L^2}, \tag{5.12}$$

or

$$\|p(x, D)u\|^2 \le (M + C)\|u\|^2_{L^2}, \tag{5.13}$$

finishing the proof.

From these L^2-estimates easily follow L^2-Sobolev space estimates. Recall from Chap. 4 that the Sobolev space $H^s(\mathbb{R}^n)$ is defined as

$$H^s(\mathbb{R}^n) = \{u \in \mathcal{S}'(\mathbb{R}^n) : \langle\xi\rangle^s \hat{u}(\xi) \in L^2(\mathbb{R}^n)\}. \tag{5.14}$$

Equivalently, with

$$\Lambda^s u = \int \langle\xi\rangle^s \hat{u}(\xi) e^{ix\cdot\xi}\, d\xi; \quad \Lambda^s \in OPS^s, \tag{5.15}$$

we have

$$H^s(\mathbb{R}^n) = \Lambda^{-s} L^2(\mathbb{R}^n). \tag{5.16}$$

The operator calculus easily gives the next proposition:

Proposition 5.5. *If* $p(x, D) \in OPS^m_{\rho,\delta}$, $0 \le \delta < \rho \le 1$, $m, s \in \mathbb{R}$, *then*

$$p(x, D) : H^s(\mathbb{R}^n) \longrightarrow H^{s-m}(\mathbb{R}^m). \tag{5.17}$$

Given Proposition 5.5, one easily obtains the Sobolev regularity of solutions to the elliptic equations studied in §4.

Calderon and Vaillancourt sharpened Theorem 5.3, showing that

$$p(x, \xi) \in S^0_{\rho,\rho},\ 0 \le \rho < 1 \Longrightarrow p(x, D) : L^2(\mathbb{R}^n) \longrightarrow L^2(\mathbb{R}^n). \tag{5.18}$$

This result, particularly for $\rho = 1/2$, has played an important role in linear PDE, especially in the study of subelliptic operators, but it will not be used in this book. The case $\rho = 0$ is treated in the exercises below.

Another important extension of Theorem 5.3 is that $p(x, D)$ is bounded on $L^p(\mathbb{R}^n)$, for $1 < p < \infty$, when $p(x, \xi) \in S^0_{1,\delta}$. Similarly, Proposition 5.5 extends to a result on L^p-Sobolev spaces, in the case $\rho = 1$. This is important for applications to nonlinear PDE, and will be proved in Chap. 13.

Exercises

Exercises 1–7 present an approach to a proof of the Calderon-Vaillancourt theorem, (5.18), in the case $\rho = 0$. This approach is due to H. O. Cordes [Cor]; see also T. Kato [K] and R. Howe [How]. In these exercises, we assume that $U(y)$ is a (measurable) unitary, operator-valued function on a measure space Y, operating on a Hilbert space $\mathcal{H}$. Assume that, for $f, g \in \mathcal{V}$, a dense subset of $\mathcal{H}$,

$$\int_Y \big|(U(y)f, g)\big|^2\, dm(y) = C_0 \|f\|^2\, \|g\|^2. \tag{5.19}$$

1. Let $\varphi_0 \in \mathcal{H}$ be a unit vector, and set $\varphi_y = U(y)\varphi_0$. Show that, for any $T \in \mathcal{L}(\mathcal{H})$,

$$C_0^2 (Tf_1, f_2) = \int_Y \int_Y L_T(y, y')\, (f_1, \varphi_{y'})\, (\varphi_y, f_2)\, dm(y)\, dm(y'), \tag{5.20}$$

where

$$L_T(y, y') = (T\varphi_{y'}, \varphi_y). \tag{5.21}$$

(*Hint*: Start by showing that $\int (f_1, \varphi_y)(\varphi_y, f_2)\, dm(y) = C_0 (f_1, f_2)$.)

A statement equivalent to (5.20) is

$$T = \iint L_T(y, y')\, U(y)\Phi_0 U(y')\, dm(y)\, dm(y'), \tag{5.22}$$

where Φ_0 is the orthogonal projection of $\mathcal{H}$ onto the span of φ_0.

2. For a partial converse, suppose L is measurable on $Y \times Y$ and

$$\int |L(y, y')|\, dm(y) \le C_1, \quad \int |L(y, y')|\, dm(y') \le C_1. \tag{5.23}$$

Define

$$T_L = \iint L(y, y')\, U(y)\Phi_0 U(y')^*\, dm(y)\, dm(y'). \tag{5.24}$$

Show that the operator norm of T_L on $\mathcal{H}$ has the estimate

$$\|T_L\| \le C_0^2 C_1.$$

3. If G is a trace class operator, and we set

$$T_{L,G} = \iint L(y, y')\, U(y) G U(y')^*\, dm(y)\, dm(y'), \tag{5.25}$$

show that

$$\|T_{L,G}\| \le C_0^2 C_1\, \|G\|_{\mathrm{TR}}. \tag{5.26}$$

(*Hint*: In case $G = G^*$, diagonalize G and use Exercise 2.)

4. Suppose $b \in L^\infty(Y)$ and we set

$$T^\#_{b,G} = \int b(y)\, U(y) G U(y)^*\, dm(y). \tag{5.27}$$

Show that

$$\|T^\#_{b,G}\| \le C_0 \|b\|_{L^\infty}\, \|G\|_{\mathrm{TR}}. \tag{5.28}$$

5. Let $Y = \mathbb{R}^{2n}$, with Lebesgue measure, $y = (q, p)$. Set $U(y) = e^{iq\cdot X} e^{ip\cdot D} = \tilde{\pi}(0, q, p)$, as in Exercises 1 and 2 of §1. Show that the identity (5.19) holds, for $f, g \in L^2(\mathbb{R}^n) = \mathcal{H}$, with $C_0 = (2\pi)^{-n}$. (*Hint*: Make use of the Plancherel theorem.)
6. Deduce that if $a(x, D)$ is a trace class operator,

$$\|(b * a)(x, D)\|_{\mathcal{L}(L^2)} \le C \|b\|_{L^\infty}\, \|a(x, D)\|_{\mathrm{TR}}. \tag{5.29}$$

(*Hint*: Look at Exercises 3–4 of §1.)

7. Suppose $p(x, \xi) \in S^0_{0,0}$. Set

$$a(x, \xi) = \psi(x)\psi(\xi), \quad b(x, \xi) = (1 - \Delta_x)^k (1 - \Delta_\xi)^k p(x, \xi), \tag{5.30}$$

where k is a positive integer, $\hat{\psi}(\xi) = \langle \xi \rangle^{-2k}$. Show that if k is chosen large enough, then $a(x, D)$ is trace class. Note that, for all $k \in \mathbb{Z}^+$, $b \in L^\infty(\mathbb{R}^{2n})$, provided $p \in S^0_{0,0}$. Show that

$$p(x, D) = (b * a)(x, D), \tag{5.31}$$

and deduce the $\rho = 0$ case of the Calderon-Vaillancourt estimate (5.19).

8. Sharpen the results of problems 3–4 above, showing that

$$\|T_{L,G}\|_{\mathcal{L}(\mathcal{H})} \le C_0^2 \, \|L\|_{\mathcal{L}(L^2(Y))} \, \|G\|_{\mathrm{TR}}. \tag{5.32}$$

This is stronger than (5.26) in view of Proposition 5.1.

6. Gårding's inequality

In this section we establish a fundamental estimate, first obtained by L. Gårding in the case of differential operators.

Theorem 6.1. *Assume* $p(x,D) \in OPS^m_{\rho,\delta}$, $0 \le \delta < \rho \le 1$, *and*

$$Re\ p(x,\xi) \ge C|\xi|^m, \ \ for\ |\xi|\ large. \tag{6.1}$$

Then, for any $s \in \mathbb{R}$, *there are* C_0, C_1 *such that, for* $u \in H^{m/2}(\mathbb{R}^n)$,

$$Re\ \big(p(x,D)u,u\big) \ge C_0\|u\|^2_{H^{m/2}} - C_1\|u\|^2_{H^s}. \tag{6.2}$$

Proof. Replacing $p(x,D)$ by $\Lambda^{-m/2}p(x,D)\Lambda^{-m/2}$, we can suppose without loss of generality that $m = 0$. Then, as in the proof of Theorem 5.3, take

$$A(x,\xi) = \Big(\mathrm{Re}\ p(x,\xi) - \frac{1}{2}C\Big)^{1/2} \in S^0_{\rho,\delta}, \tag{6.3}$$

so

$$\begin{gathered} A(x,D)^*A(x,D) = \ \mathrm{Re}\ p(x,D) - \frac{1}{2}C + r(x,D), \\ r(x,D) \in OPS^{-(\rho-\delta)}_{\rho,\delta}. \end{gathered} \tag{6.4}$$

This gives

$$\begin{aligned} \mathrm{Re}\ (p(x,D)u,u) &= \|A(x,D)u\|^2_{L^2} + \frac{1}{2}C\|u\|^2_{L^2} + \big(r(x,D)u,u\big) \\ &\ge \frac{1}{2}C\|u\|^2_{L^2} - C_1\|u\|^2_{H^s} \end{aligned} \tag{6.5}$$

with $s = -(\rho-\delta)/2$, so (6.2) holds in this case. If $s < -(\rho-\delta)/2 = s_0$, use the simple estimate

$$\|u\|^2_{H^{s_0}} \le \varepsilon\|u\|^2_{L^2} + C(\varepsilon)\|u\|^2_{H^s} \tag{6.6}$$

to obtain the desired result in this case.

This Gårding inequality has been improved to a sharp Gårding inequality, of the form

$$\text{Re}\,\big(p(x,D)u,u\big) \ge -C\|u\|_{L^2}^2 \quad \text{when Re } p(x,\xi) \ge 0, \tag{6.7}$$

first for scalar $p(x,\xi) \in S^1_{1,0}$, by Hörmander, then for matrix-valued symbols, with Re $p(x,\xi)$ standing for $(1/2)\big(p(x,\xi)+p(x,\xi)^*\big)$, by P. Lax and L. Nirenberg. Proofs and some implications can be found in Vol. 3 of [Ho5], and in [T1] and [Tre]. A very strong improvement due to C. Fefferman and D. Phong [FP] is that (6.7) holds for scalar $p(x,\xi) \in S^2_{1,0}$. See also [Ho5] and [F] for further discussion.

Exercises

1. Suppose $m > 0$ and $p(x,D) \in OPS^m_{1,0}$ has a symbol satisfying (6.1). Examine the solvability of
$$\frac{\partial u}{\partial t} = p(x,D)u,$$
for $u = u(t,x)$, $u(0,x) = f \in H^s(\mathbb{R}^n)$.
(*Hint*: Look ahead at §7 for some useful techniques. Solve
$$\frac{\partial u_\varepsilon}{\partial t} = J_\varepsilon p(x,D) J_\varepsilon u_\varepsilon$$
and estimate $(d/dt)\|\Lambda^s u_\epsilon(t)\|^2_{L^2}$, making use of Gårding's inequality.)

7. Hyperbolic evolution equations

In this section we examine first-order systems of the form

$$\frac{\partial u}{\partial t} = L(t,x,D_x)u + g(t,x), \quad u(0) = f. \tag{7.1}$$

We assume $L(t,x,\xi) \in S^1_{1,0}$, with smooth dependence on t, so

$$|D_t^j D_x^\beta D_\xi^\alpha L(t,x,\xi)| \le C_{j\alpha\beta}\langle\xi\rangle^{1-|\alpha|}. \tag{7.2}$$

Here $L(t,x,\xi)$ is a $K \times K$ matrix-valued function, and we make the hypothesis of symmetric hyperbolicity:

$$L(t,x,\xi)^* + L(t,x,\xi) \in S^0_{1,0}. \tag{7.3}$$

We suppose $f \in H^s(\mathbb{R}^n)$, $s \in \mathbb{R}$, $g \in C(\mathbb{R}, H^s(\mathbb{R}^n))$.

Our strategy will be to obtain a solution to (7.1) as a limit of solutions u_ε to

$$\frac{\partial u_\varepsilon}{\partial t} = J_\varepsilon L J_\varepsilon u_\varepsilon + g, \quad u_\varepsilon(0) = f, \tag{7.4}$$

where

$$J_\varepsilon = \varphi(\varepsilon D_x), \tag{7.5}$$

for some $\varphi(\xi) \in \mathcal{S}(\mathbb{R}^n)$, $\varphi(0) = 1$. The family of operators J_ε is called a *Friedrichs mollifier*. Note that, for any $\varepsilon > 0$, $J_\varepsilon \in OPS^{-\infty}$, while, for $\varepsilon \in (0,1]$, J_ε is bounded in $OPS^0_{1,0}$.

For any $\varepsilon > 0$, $J_\varepsilon L J_\varepsilon$ is a bounded linear operator on each H^s, and solvability of (7.4) is elementary. Our next task is to obtain estimates on u_ε, independent of $\varepsilon \in (0,1]$. Use the norm $\|u\|_{H^s} = \|\Lambda^s u\|_{L^2}$. We derive an estimate for

$$\frac{d}{dt}\|\Lambda^s u_\varepsilon(t)\|^2_{L^2} = 2\,\mathrm{Re}\,(\Lambda^s J_\varepsilon L J_\varepsilon u_\varepsilon, \Lambda^s u_\varepsilon) + 2\,\mathrm{Re}\,(\Lambda^s g, \Lambda^s u_\varepsilon). \tag{7.6}$$

Write the first two terms on the right as the real part of

$$2(L\Lambda^s J_\varepsilon u_\varepsilon, \Lambda^s J_\varepsilon u_\varepsilon) + 2([\Lambda^s, L] J_\varepsilon u_\varepsilon, \Lambda^s J_\varepsilon u_\varepsilon). \tag{7.7}$$

By (7.3), $L + L^* = B(t,x,D) \in OPS^0_{1,0}$, so the first term in (7.7) is equal to

$$\big(B(t,x,D)\Lambda^s J_\varepsilon u_\varepsilon, \Lambda^s J_\varepsilon u_\varepsilon\big) \le C\,\|J_\varepsilon u_\varepsilon\|^2_{H^s}. \tag{7.8}$$

Meanwhile, $[\Lambda^s, L] \in OPS^s_{1,0}$, so the second term in (7.7) is also bounded by the right side of (7.8). Applying Cauchy's inequality to $2(\Lambda^s g, \Lambda^s u_\varepsilon)$, we obtain

$$\frac{d}{dt}\|\Lambda^s u_\varepsilon(t)\|^2_{L^2} \le C\,\|\Lambda^s u_\varepsilon(t)\|^2_{L^2} + C\,\|g(t)\|^2_{H^s}. \tag{7.9}$$

Thus Gronwall's inequality yields an estimate

$$\|u_\varepsilon(t)\|^2_{H^s} \le C(t)\big[\|f\|^2_{H^s} + \|g\|^2_{C([0,t],H^s)}\big], \tag{7.10}$$

independent of $\varepsilon \in (0,1]$. We are now prepared to establish the following existence result.

Proposition 7.1. *If (7.1) is symmetric hyperbolic and*

$$f \in H^s(\mathbb{R}^n), \quad g \in C(\mathbb{R}, H^s(\mathbb{R}^n)), \quad s \in \mathbb{R},$$

then there is a solution u to (7.1), satisfying

$$u \in L^\infty_{loc}(\mathbb{R}, H^s(\mathbb{R}^n)) \cap \ \mathit{Lip}\ (\mathbb{R}, H^{s-1}(\mathbb{R}^n)). \tag{7.11}$$

Proof. Take $I = [-T, T]$. The bounded family

$$u_\varepsilon \in C(I, H^s) \cap C^1(I, H^{s-1})$$

will have a weak limit point u satisfying (7.11), and it is easy to verify that such u solves (7.1). As for the bound on $[-T, 0]$, this follows from the invariance of the class of hyperbolic equations under time reversal.

Analogous energy estimates can establish the uniqueness of such a solution u and rates of convergence of $u_\varepsilon \to u$ as $\varepsilon \to 0$. Also, (7.11) can be improved to

$$u \in C(\mathbb{R}, H^s(\mathbb{R}^n)) \cap C^1(\mathbb{R}, H^{s-1}(\mathbb{R}^n)). \tag{7.12}$$

To see this, let $f_j \in H^{s+1}$, $f_j \to f$ in H^s, and let u_j solve (7.1) with $u_j(0) = f_j$. Then each u_j belongs to $L^\infty_{\text{loc}}(\mathbb{R}, H^{s+1}) \cap \text{Lip}(\mathbb{R}, H^s)$, so in particular each $u_j \in C(\mathbb{R}, H^s)$. Now $v_j = u - u_j$ solves (7.1) with $v_j(0) = f - f_j$, and $\|f - f_j\|_{H^s} \to 0$ as $j \to \infty$, so estimates arising in the proof of Proposition 7.1 imply that $\|v_j(t)\|_{H^s} \to 0$ locally uniformly in t, giving $u \in C(\mathbb{R}, H^s)$.

There are other notions of hyperbolicity. In particular, (7.1) is said to be *symmetrizable hyperbolic* if there is a $K \times K$ matrix-valued $S(t, x, \xi) \in S^0_{1,0}$ that is positive-definite and such that $S(t, x, \xi)L(t, x, \xi) = \tilde{L}(t, x, \xi)$ satisfies (7.3). Proposition 7.1 extends to the case of symmetrizable hyperbolic systems. Again, one obtains u as a limit of solutions u_ϵ to (7.4). There is one extra ingredient in the energy estimates. In this case, construct $S(t) \in OPS^0_{1,0}$, positive-definite, with symbol equal to $S(t, x, \xi)$ mod $S^{-1}_{1,0}$. For the energy estimates, replace the left side of (7.6) by

$$\frac{d}{dt}\big(\Lambda^s u_\epsilon(t), S(t)\Lambda^s u_\epsilon(t)\big)_{L^2}, \tag{7.13}$$

which can be estimated in a fashion similar to (7.7)–(7.9).

A $K \times K$ system of the form (7.1) with $L(t, x, \xi) \in S^1_{cl}$ is said to be *strictly hyperbolic* if its principal symbol $L_1(t, x, \xi)$, homogeneous of degree 1 in ξ, has K distinct, purely imaginary eigenvalues, for each x and each $\xi \neq 0$. The results above apply in this case, in view of:

Proposition 7.2. *Whenever (7.1) is strictly hyperbolic, it is symmetrizable.*

Proof. If we denote the eigenvalues of $L_1(t, x, \xi)$ by $i\lambda_\nu(t, x, \xi)$, ordered so that $\lambda_1(t, x, \xi) < \cdots < \lambda_K(t, x, \xi)$, then λ_ν are well-defined C^∞-functions of (t, x, ξ), homogeneous of degree 1 in ξ. If $P_\nu(t, x, \xi)$ are the projections onto the $i\lambda_\nu$-eigenspaces of L_1,

$$P_\nu(t, x, \xi) = \frac{1}{2\pi i}\int_{\gamma_\nu} \big(\zeta - L_1(t, x, \xi)\big)^{-1}\, d\zeta, \tag{7.14}$$

where γ_ν is a small circle about $i\lambda_\nu(t,x,\xi)$, then P_ν is smooth and homogeneous of degree 0 in ξ. Then

$$S(t,x,\xi) = \sum_j P_j(t,x,\xi)^* P_j(t,x,\xi) \tag{7.15}$$

gives the desired symmetrizer.

Higher-order, strictly hyperbolic PDE can be reduced to strictly hyperbolic, first-order systems of this nature. Thus one has an analysis of solutions to such higher-order hyperbolic equations.

Exercises

1. Carry out the reduction of a strictly hyperbolic PDE of order m to a first-order system of the form (7.1). Starting with
$$Lu = \frac{\partial^m u}{\partial y^m} + \sum_{j=0}^{m-1} A_j(y,x,D_x)\frac{\partial^j u}{\partial y^j},$$
where $A_j(y,x,D)$ has order $\le m-j$, form $v = (v_1,\dots,v_m)^t$ with
$$v_1 = \Lambda^{m-1}u, \dots, v_j = \partial_y^{j-1}\Lambda^{m-j}u, \dots, v_m = \partial_y^{m-1}u,$$
to pass from $Lu = f$ to
$$\frac{\partial v}{\partial y} = K(y,x,D_x)v + F,$$
with $F = (0,\dots,0,f)^t$. Give an appropriate definition of strict hyperbolicity in this context, and show that this first-order system is strictly hyperbolic provided L is.
2. Fix $r > 0$. Let $\gamma_r \in \mathcal{E}'(\mathbb{R}^2)$ denote the unit mass density on the circle of radius r:
$$\langle u, \gamma_r\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} u(r\cos\theta, r\sin\theta)\,d\theta.$$
Let $\Gamma_r u = \gamma_r * u$. Show that there exist $A_r(\xi) \in S^{-1/2}(\mathbb{R}^2)$ and $B_r(\xi) \in S^{1/2}(\mathbb{R}^2)$, such that
$$\Gamma_r = A_r(D)\cos r\sqrt{-\Delta} + B_r(D)\,\frac{\sin r\sqrt{-\Delta}}{\sqrt{-\Delta}}. \tag{7.16}$$
(*Hint*: See Exercise 1 in §7 of Chap. 6.)

8. Egorov's theorem

We want to examine the behavior of operators obtained by conjugating a pseudodifferential operator $P_0 \in OPS^m_{1,0}$ by the solution operator to a scalar hyperbolic equation of the form

$$\frac{\partial u}{\partial t} = iA(t,x,D_x)u, \tag{8.1}$$

where we assume $A = A_1 + A_0$ with

$$A_1(t,x,\xi) \in S^1_{cl} \text{ real}, \quad A_0(t,x,\xi) \in S^0_{cl}. \tag{8.2}$$

We suppose $A_1(t,x,\xi)$ is homogeneous in ξ, for $|\xi| \geq 1$. Denote by $S(t,s)$ the solution operator to (8.1), taking $u(s)$ to $u(t)$. This is a bounded operator on each Sobolev space H^σ, with inverse $S(s,t)$. Set

$$P(t) = S(t,0)P_0S(0,t). \tag{8.3}$$

We aim to prove the following result of Y. Egorov.

Theorem 8.1. *If $P_0 = p_0(x,D) \in OPS^m_{1,0}$, then for each t, $P(t) \in OPS^m_{1,0}$, modulo a smoothing operator. The principal symbol of $P(t)$ (mod $S^{m-1}_{1,0}$) at a point (x_0,ξ_0) is equal to $p_0(y_0,\eta_0)$, where (y_0,η_0) is obtained from (x_0,ξ_0) by following the flow $\mathcal{C}(t)$ generated by the (time-dependent) Hamiltonian vector field*

$$H_{A_1(t,x,\xi)} = \sum_{j=1}^n \left(\frac{\partial A_1}{\partial \xi_j} \frac{\partial}{\partial x_j} - \frac{\partial A_1}{\partial x_j} \frac{\partial}{\partial \xi_j} \right). \tag{8.4}$$

To start the proof, differentiating (8.3) with respect to t yields

$$P'(t) = i[A(t,x,D), P(t)], \quad P(0) = P_0. \tag{8.5}$$

We will construct an approximate solution $Q(t)$ to (8.5) and then show that $Q(t) - P(t)$ is a smoothing operator.

So we are looking for $Q(t) = q(t,x,D) \in OPS^m_{1,0}$, solving

$$Q'(t) = i[A(t,x,D), Q(t)] + R(t), \quad Q(0) = P_0, \tag{8.6}$$

where $R(t)$ is a smooth family of operators in $OPS^{-\infty}$. We do this by constructing the symbol $q(t,x,\xi)$ in the form

$$q(t,x,\xi) \sim q_0(t,x,\xi) + q_1(t,x,\xi) + \cdots. \tag{8.7}$$

Now the symbol of $i[A, Q(t)]$ is of the form

$$H_{A_1}q + \{A_0, q\} + i \sum_{|\alpha|\geq 2} \frac{i^{|\alpha|}}{\alpha!} \Big(A^{(\alpha)} q_{(\alpha)} - q^{(\alpha)} A_{(\alpha)} \Big), \tag{8.8}$$

where $A^{(\alpha)} = D^\alpha_\xi A$, $A_{(\alpha)} = D^\alpha_x A$, and so on. Since we want the difference between this and $\partial q/\partial t$ to have order $-\infty$, this suggests defining $q_0(t,x,\xi)$ by

$$\Big(\frac{\partial}{\partial t} - H_{A_1} \Big) q_0(t,x,\xi) = 0, \quad q_0(0,x,\xi) = p_0(x,\xi). \tag{8.9}$$

Thus $q_0(t, x_0, \xi_0) = p_0(y_0, \eta_0)$, as in the statement of the theorem; we have $q_0(t, x, \xi) \in S^m_{1,0}$. Equation (8.9) is called a *transport equation*. Recursively, we obtain transport equations

$$\left(\frac{\partial}{\partial t} - H_{A_1}\right) q_j(t, x, \xi) = b_j(t, x, \xi), \quad q_j(0, x, \xi) = 0, \tag{8.10}$$

for $j \geq 1$, with solutions in $S^{m-j}_{1,0}$, leading to a solution to (8.6).

Finally, we show that $P(t) - Q(t)$ is a smoothing operator. Equivalently, we show that, for any $f \in H^\sigma(\mathbb{R}^n)$,

$$v(t) - w(t) = S(t,0)P_0 f - Q(t)S(t,0)f \in H^\infty(\mathbb{R}^n), \tag{8.11}$$

where $H^\infty(\mathbb{R}^n) = \cap_s H^s(\mathbb{R}^n)$. Note that

$$\frac{\partial v}{\partial t} = iA(t, x, D)v, \quad v(0) = P_0 f, \tag{8.12}$$

while use of (8.6) gives

$$\frac{\partial w}{\partial t} = iA(t, x, D)w + g, \quad w(0) = P_0 f, \tag{8.13}$$

where

$$g = R(t)S(t,0)w \in C^\infty(\mathbb{R}, H^\infty(\mathbb{R}^n)). \tag{8.14}$$

Hence

$$\frac{\partial}{\partial t}(v - w) = iA(t, x, D)(v - w) - g, \quad v(0) - w(0) = 0. \tag{8.15}$$

Thus energy estimates for hyperbolic equations yield $v(t) - w(t) \in H^\infty$, for any $f \in H^\sigma(\mathbb{R}^n)$, completing the proof.

A check of the proof shows that

$$P_0 \in OPS^m_{cl} \Longrightarrow P(t) \in OPS^m_{cl}. \tag{8.16}$$

Also, the proof readily extends to yield the following:

Proposition 8.2. *With $A(t, x, D)$ as before,*

$$P_0 \in OPS^m_{\rho,\delta} \Longrightarrow P(t) \in OPS^m_{\rho,\delta} \tag{8.17}$$

provided

$$\rho > \frac{1}{2}, \quad \delta = 1 - \rho. \tag{8.18}$$

One needs $\delta = 1 - \rho$ to ensure that $p(\mathcal{C}(t)(x,\xi)) \in S^m_{\rho,\delta}$, and one needs $\rho > \delta$ to ensure that the transport equations generate $q_j(t,x,\xi)$ of progressively lower order.

Exercises

1. Let $\chi : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism that is a linear map outside some compact set. Define $\chi^* : C^\infty(\mathbb{R}^n) \to C^\infty(\mathbb{R}^n)$ by $\chi^* f(x) = f\big(\chi(x)\big)$. Show that

$$P \in OPS^m_{1,0} \Longrightarrow (\chi*)^{-1} P \chi^* \in OPS^m_{1,0}. \tag{8.19}$$

(*Hint*: Reduce to the case where χ is homotopic to a linear map through diffeomorphisms, and show that the result in that case is a special case of Theorem 8.1, where $A(t,x,D)$ is a t-dependent family of real vector fields on $\mathbb{R}^n$.)

2. Let $a \in C_0^\infty(\mathbb{R}^n)$, $\varphi \in C^\infty(\mathbb{R}^n)$ be real-valued, and $\nabla\varphi \neq 0$ on supp a. If $P \in OPS^m$, show that

$$P\big(a\, e^{i\lambda\varphi}\big) = b(x,\lambda)\, e^{i\lambda\varphi(x)}, \tag{8.20}$$

where

$$b(x,\lambda) \sim \lambda^m\big[b_0^\pm(x) + b_1^\pm(x)\lambda^{-1} + \cdots\big], \quad \lambda \to \pm\infty. \tag{8.21}$$

(*Hint*: Using a partition of unity and Exercise 1, reduce to the case $\varphi(x) = x \cdot \xi$, for some $\xi \in \mathbb{R}^n \setminus 0$.)

3. If a and φ are as in Exercise 2 above and Γ_r is as in Exercise 2 of §7, show that, mod $O(\lambda^{-\infty})$,

$$\Gamma_r\big(a\, e^{i\lambda\varphi}\big) = \cos r\sqrt{-\Delta}\big(A_r(x,\lambda)e^{i\lambda\varphi}\big) + \frac{\sin r\sqrt{-\Delta}}{\sqrt{-\Delta}}\big(B_r(x,\lambda)e^{i\lambda\varphi}\big), \tag{8.22}$$

where

$$A_r(x,\lambda) \sim \lambda^{-1/2}\big[a_{0r}^\pm(x) + a_{1r}^\pm(x)\lambda^{-1} + \cdots\big],$$
$$B_r(x,\lambda) \sim \lambda^{1/2}\big[b_{0r}^\pm(x) + b_{1r}^\pm(x)\lambda^{-1} + \cdots\big],$$

as $\lambda \to \pm\infty$.

9. Microlocal regularity

We define the notion of wave front set of a distribution $u \in H^{-\infty}(\mathbb{R}^n) = \cup_s H^s(\mathbb{R}^n)$, which refines the notion of singular support. If $p(x,\xi) \in S^m$ has principal symbol $p_m(x,\xi)$, homogeneous in ξ, then the characteristic set of $P = p(x,D)$ is given by

$$\text{Char}\, P = \{(x,\xi) \in \mathbb{R}^n \times (\mathbb{R}^n \setminus 0) : p_m(x,\xi) = 0\}. \tag{9.1}$$

If $p_m(x,\xi)$ is a $K \times K$ matrix, take the determinant. Equivalently, (x_0,ξ_0) is noncharacteristic for P, or P is elliptic at (x_0,ξ_0), if $|p(x,\xi)^{-1}| \le C|\xi|^{-m}$, for (x,ξ) in a small conic neighborhood of (x_0,ξ_0) and $|\xi|$ large. By definition, a conic set is invariant under the dilations $(x,\xi) \mapsto (x,r\xi)$, $r \in (0,\infty)$. The wave front set is defined by

$$\mathrm{WF}(u) = \bigcap \{\mathrm{Char}\, P : P \in OPS^0,\ Pu \in C^\infty\}. \tag{9.2}$$

Clearly, $\mathrm{WF}(u)$ is a closed conic subset of $\mathbb{R}^n \times (\mathbb{R}^n \setminus 0)$.

Proposition 9.1. *If π is the projection $(x,\xi) \mapsto x$, then*

$$\pi(WF(u)) = \textit{sing supp}\ u.$$

Proof. If $x_0 \notin$ sing supp u, there is a $\varphi \in C_0^\infty(\mathbb{R}^n)$, $\varphi = 1$ near x_0, such that $\varphi u \in C_0^\infty(\mathbb{R}^n)$. Clearly, $(x_0,\xi) \notin \mathrm{Char}\, \varphi$ for any $\xi \ne 0$, so $\pi(WF(u)) \subset$ sing supp u.

Conversely, if $x_0 \notin \pi(WF(u))$, then for any $\xi \ne 0$ there is a $Q \in OPS^0$ such that $(x_0,\xi) \notin \mathrm{Char}\, Q$ and $Qu \in C^\infty$. Thus we can construct finitely many $Q_j \in OPS^0$ such that $Q_j u \in C^\infty$ and each (x_0,ξ) (with $|\xi| = 1$) is noncharacteristic for some Q_j. Let $Q = \sum Q_j^* Q_j \in OPS^0$. Then Q is elliptic near x_0 and $Qu \in C^\infty$, so u is C^∞ near x_0.

We define the associated notion of $\mathrm{ES}(P)$ for a pseudodifferential operator. Let U be an open conic subset of $\mathbb{R}^n \times (\mathbb{R}^n \setminus 0)$. We say that $p(x,\xi) \in S^m_{\rho,\delta}$ has order $-\infty$ on U if for each closed conic set V of U we have estimates, for each N,

$$|D_x^\beta D_\xi^\alpha p(x,\xi)| \le C_{\alpha\beta N V} \langle\xi\rangle^{-N}, \quad (x,\xi) \in V. \tag{9.3}$$

If $P = p(x,D) \in OPS^m_{\rho,\delta}$, we define the essential support of P (and of $p(x,\xi)$) to be the smallest closed conic set on the complement of which $p(x,\xi)$ has order $-\infty$. We denote this set by $\mathrm{ES}(P)$.

From the symbol calculus of §3, it follows easily that

$$\mathrm{ES}(P_1P_2) \subset \mathrm{ES}(P_1) \cap \mathrm{ES}(P_2) \tag{9.4}$$

provided $P_j \in OPS^{m_j}_{\rho_j,\delta_j}$ and $\rho_1 > \delta_2$. To relate $\mathrm{WF}(Pu)$ to $\mathrm{WF}(u)$ and $\mathrm{ES}(P)$, we begin with the following.

Lemma 9.2. *Let $u \in H^{-\infty}(\mathbb{R}^n)$, and suppose that U is a conic open set satisfying*

$$WF(u) \cap U = \emptyset.$$

If $P \in OPS^m_{\rho,\delta}$, $\rho > 0$, $\delta < 1$, and $ES(P) \subset U$, then $Pu \in C^\infty$.

Proof. Taking $P_0 \in OPS^0$ with symbol identically 1 on a conic neighborhood of $\text{ES}(P)$, so $P = PP_0 \mod OPS^{-\infty}$, it suffices to conclude that $P_0 u \in C^\infty$, so we can specialize the hypothesis to $P \in OPS^0$.

By hypothesis, we can find $Q_j \in OPS^0$ such that $Q_j u \in C^\infty$ and each $(x, \xi) \in \text{ES}(P)$ is noncharacteristic for some Q_j, and if $Q = \sum Q_j^* Q_j$, then $Qu \in C^\infty$ and $\text{Char}\, Q \cap \text{ES}(P) = \emptyset$. We claim there exists an operator $A \in OPS^0$ such that $AQ = P \mod OPS^{-\infty}$. Indeed, let $\tilde{Q}$ be an elliptic operator whose symbol equals that of Q on a conic neighborhood of $\text{ES}(P)$, and let $\tilde{Q}^{-1}$ denote a parametrix for $\tilde{Q}$. Now simply set $A = P\tilde{Q}^{-1}$. Consequently, (mod C^∞) $Pu = AQu \in C^\infty$, so the lemma is proved.

We are ready for the basic result on the preservation of wave front sets by a pseudodifferential operator.

Proposition 9.3. *If $u \in H^{-\infty}$ and $P \in OPS^m_{\rho,\delta}$, with $\rho > 0$, $\delta < 1$, then*

$$WF(Pu) \subset WF(u) \cap ES(P). \tag{9.5}$$

Proof. First we show $\text{WF}(Pu) \subset \text{ES}(P)$. Indeed, if $(x_0, \xi_0) \notin \text{ES}(P)$, choose $Q = q(x, D) \in OPS^0$ such that $q(x, \xi) = 1$ on a conic neighborhood of (x_0, ξ_0) and $\text{ES}(Q) \cap \text{ES}(P) = \emptyset$. Thus $QP \in OPS^{-\infty}$, so $QPu \in C^\infty$. Hence $(x_0, \xi_0) \notin \text{WF}(Pu)$.

In order to show that $\text{WF}(Pu) \subset \text{WF}(u)$, let Γ be any conic neighborhood of $\text{WF}(u)$, and write $P = P_1 + P_2$, $P_j \in OPS^m_{\rho,\delta}$, with $\text{ES}(P_1) \subset \Gamma$ and $\text{ES}(P_2) \cap \text{WF}(u) = \emptyset$. By Lemma 9.2, $P_2 u \in C^\infty$. Thus $\text{WF}(u) = \text{WF}(P_1 u) \subset \Gamma$, which shows $\text{WF}(Pu) \subset \text{WF}(u)$.

One says that a pseudodifferential operator of type (ρ, δ), with $\rho > 0$ and $\delta < 1$, is *microlocal*. As a corollary, we have the following sharper form of local regularity for elliptic operators, called *microlocal regularity*.

Corollary 9.4. *If $P \in OPS^m_{\rho,\delta}$ is elliptic, $0 \le \delta < \rho \le 1$, then*

$$WF(Pu) = WF(u). \tag{9.6}$$

Proof. We have seen that $\text{WF}(Pu) \subset \text{WF}(u)$. On the other hand, if $E \in OPS^{-m}_{\rho,\delta}$ is a parametrix for P, we see that $\text{WF}(u) = \text{WF}(EPu) \subset \text{WF}(Pu)$. In fact, by an argument close to the proof of Lemma 9.2, we have for general P that

$$\text{WF}(u) \subset \text{WF}(Pu) \cup \text{Char}\, P. \tag{9.7}$$

We next discuss how the solution operator e^{itA} to a scalar hyperbolic equation $\partial u/\partial t = iA(x, D)u$ propagates the wave front set. We assume $A(x, \xi) \in S^1_{cl}$, with real principal symbol. Suppose $\text{WF}(u) = \Sigma$. Then there is a countable family of operators $p_j(x, D) \in OPS^0$, each of whose complete symbols vanishes in a neighborhood of Σ, but such that

$$\Sigma = \bigcap_j \{(x,\xi) : p_j(x,\xi) = 0\}. \tag{9.8}$$

We know that $p_j(x,D)u \in C^\infty$ for each j. Using Egorov's theorem, we want to construct a family of pseudodifferential operators $q_j(x,D) \in OPS^0$ such that $q_j(x,D)e^{itA}u \in C^\infty$, this family being rich enough to describe the wave front set of $e^{itA}u$.

Indeed, let $q_j(x,D) = e^{itA}p_j(x,D)e^{-itA}$. Egorov's theorem implies that $q_j(x,D) \in OPS^0$ (modulo a smoothing operator) and gives the principal symbol of $q_j(x,D)$. Since $p_j(x,D)u \in C^\infty$, we have $e^{itA}p_j(x,D)u \in C^\infty$, which in turn implies $q_j(x,D)e^{itA}u \in C^\infty$. From this it follows that $\mathrm{WF}(e^{itA}u)$ is contained in the intersection of the characteristics of the $q_j(x,D)$, which is precisely $C(t)\Sigma$, the image of Σ under the canonical transformation $C(t)$, generated by H_{A_1}. In other words,

$$\mathrm{WF}(e^{itA}u) \subset C(t)\mathrm{WF}(u).$$

However, our argument is reversible; $u = e^{-itA}(e^{itA}u)$. Consequently, we have the following result:

Proposition 9.5. *If $A = A(x,D) \in OPS^1$ is scalar with real principal symbol, then, for $u \in H^{-\infty}$,*

$$WF(e^{itA}u) = C(t)WF(u). \tag{9.9}$$

The same argument works for the solution operator $S(t,0)$ to a time-dependent, scalar, hyperbolic equation.

Exercises

1. If $a \in C_0^\infty(\mathbb{R}^n)$, $\varphi \in C^\infty(\mathbb{R}^n)$ is real-valued, $\nabla\varphi \neq 0$ on supp a, as in Exercise 2 of §8, and $P = p(x,D) \in OPS^m$, so

$$P\big(a\,e^{i\lambda\varphi}\big) = b(x,\lambda)e^{i\lambda\varphi(x)},$$

as in (8.20), show that, mod $O(|\lambda|^{-\infty})$, $b(x,\lambda)$ depends only on the behavior of $p(x,\xi)$ on an arbitrarily small conic neighborhood of

$$C_\varphi = \big\{\big(x,\lambda d\varphi(x)\big) : x \in \text{ supp } a, \lambda \neq 0\big\}.$$

If C_φ^+ is the subset of C_φ on which $\lambda > 0$, show that the asymptotic behavior of $b(x,\lambda)$ as $\lambda \to +\infty$ depends only on the behavior of $p(x,\xi)$ on an arbitrarily small conic neighborhood of C_φ^+.

2. If Γ_r is as in (8.22), show that, given $r > 0$,

$$\big(\cos r\sqrt{-\Delta}\big)(a\,e^{i\lambda\varphi}) = \Gamma_r Q_r(a\,e^{i\lambda\varphi}), \quad \text{mod } O(\lambda^{-\infty}),\ \lambda > 0, \tag{9.10}$$

for some $Q_r \in OPS^{1/2}$. Consequently, analyze the behavior of the left side of (9.10), as $\lambda \to +\infty$, in terms of the behavior of Γ_r analyzed in §7 of Chap. 6.

10. Operators on manifolds

Let M be a smooth manifold. It would be natural to say that a continuous linear operator $P : C_0^\infty(M) \to \mathcal{D}'(M)$ is a pseudodifferential operator in $OPS^m_{\rho,\delta}(M)$ provided its Schwartz kernel is C^∞ off the diagonal in $M \times M$, and there exists an open cover Ω_j of M, a subordinate partition of unity φ_j, and diffeomorphisms $F_j : \Omega_j \to \mathcal{O}_j \subset \mathbb{R}^n$ that transform the operators $\varphi_k P \varphi_j : C^\infty(\Omega_j) \to \mathcal{E}'(\Omega_k)$ into pseudodifferential operators in $OPS^m_{\rho,\delta}$, as defined in §1.

This is a rather "liberal" definition of $OPS^m_{\rho,\delta}(M)$. For example, it poses no growth restrictions on the Schwartz kernel $K \in \mathcal{D}'(M \times M)$ at infinity. Consequently, if M happens to be $\mathbb{R}^n$, the class of operators in $OPS^m_{\rho,\delta}(M)$ as defined above is a bit larger than the class $OPS^m_{\rho,\delta}$ defined in §1. One negative consequence of this definition is that pseudodifferential operators cannot always be composed. One drastic step to fix this would be to insist that the kernel be properly supported, so $P : C_0^\infty(M) \to C_0^\infty(M)$. If M is compact, these problems do not arise. If M is noncompact, it is often of interest to place specific restrictions on K near infinity, but we won't go further into this point here.

Another way in which the definition of $OPS^m_{\rho,\delta}(M)$ given above is liberal is that it requires P to be locally transformed to pseudodifferential operators on $\mathbb{R}^n$ by *some* coordinate cover. One might ask if then P is necessarily so transformed by *every* coordinate cover. This comes down to asking if the class $OPS^m_{\rho,\delta}$ defined in §1 is invariant under a diffeomorphism $F : \mathbb{R}^n \to \mathbb{R}^n$. It would suffice to establish this for the case where F is the identity outside a compact set.

In case $\rho \in (1/2, 1]$ and $\delta = 1 - \rho$, this invariance is a special case of the Egorov theorem established in §8. Indeed, one can find a time-dependent vector field $X(t)$ whose flow at $t = 1$ coincides with F and apply Theorem 8.1 to $iA(t, x, D) = X(t)$. Note that the formula for the principal symbol of the conjugated operator given there implies

$$p(1, F(x), \xi) = p_0(x, F'(x)^t \xi), \tag{10.1}$$

so that the principal symbol is well defined on the cotangent bundle of M.

We will therefore generally insist that $\rho \in (1/2, 1]$ and $\delta = 1 - \rho$ when talking about $OPS^m_{\rho,\delta}(M)$ for a manifold M, without a distinguished coordinate chart. In special situations, it might be natural to use coordinate charts with special structure. For instance, for a Cartesian product $M = \mathbb{R} \times \Omega$, one can stick to product coordinate systems. In such a case, we can construct a parametrix E for the hypoelliptic operator $\partial/\partial t - \Delta_x$, $t \in \mathbb{R}$, $x \in \Omega$, and unambiguously regard E as an operator in $OPS^{-1}_{1/2,0}(\mathbb{R} \times \Omega)$.

We make the following comments on the principal symbol of an operator $P \in OPS^m_{\rho,\delta}(M)$, when $\rho \in (1/2, 1]$, $\delta = 1 - \rho$. By the arguments in §8, the principal symbol is well defined, if it is regarded as an element of the quotient space:

$$p(x, \xi) \in S^m_{\rho,\delta}(T^*M)/S^{m-(2\rho-1)}_{\rho,\delta}(T^*M). \tag{10.2}$$

In particular, by Theorem 8.1, in case $P \in OPS^m_{1,0}(M)$, we have

$$p(x,\xi) \in S^m_{1,0}(T^*M)/S^{m-1}_{1,0}(T^*M). \tag{10.3}$$

If $P \in S^m_{cl}(M)$, then the principal symbol can be taken to be homogeneous in ξ of degree m, by (8.16). Note that the characterizations of the Schwartz kernels of operators in $OPS^m_{1,0}$ and in OPS^m_{cl} given in §2 also make clear the invariance of these classes under coordinate transformations.

We now discuss some properties of an elliptic operator $A \in OPS^m_{1,0}(M)$, when M is a compact Riemannian manifold. Denote by B a parametrix, so we have, for each $s \in \mathbb{R}$,

$$A : H^{s+m}(M) \longrightarrow H^s(M), \quad B : H^s(M) \longrightarrow H^{s+m}(M), \tag{10.4}$$

and $AB = I + K_1$, $BA = I + K_2$, where $K_j : \mathcal{D}'(M) \to C^\infty(M)$. Thus K_j is compact on each Sobolev space $H^s(M)$, so B is a two-sided Fredholm inverse of A in (10.4). In particular, A is a Fredholm operator; $\ker A = \mathcal{K}_{s+m} \subset H^{s+m}(M)$ is finite-dimensional, and $A\big(H^{s+m}(M)\big) \subset H^s(M)$ is closed, of finite codimension, so

$$\mathcal{C}_s = \{v \in H^{-s}(M) : \langle Au, v\rangle = 0 \text{ for all } u \in H^{s+m}(M)\}$$

is finite-dimensional. Note that $\mathcal{C}_s$ is the null space of

$$A^* : H^{-s}(M) \longrightarrow H^{-s-m}(M), \tag{10.5}$$

which is also an elliptic operator in $OPS^m_{1,0}(M)$. Elliptic regularity yields, for all s,

$$\mathcal{K}_{s+m} = \{u \in C^\infty(M) : Au = 0\}, \quad \mathcal{C}_s = \{v \in C^\infty(M) : A^*v = 0\}. \tag{10.6}$$

Thus these spaces are independent of s.

Suppose now that $m > 0$. We will consider A as an unbounded operator on the Hilbert space $L^2(M)$, with domain

$$\mathcal{D}(A) = \{u \in L^2(M) : Au \in L^2(M)\}. \tag{10.7}$$

It is easy to see that A is closed. Also, elliptic regularity implies

$$\mathcal{D}(A) = H^m(M). \tag{10.8}$$

Since A is closed and densely defined, its Hilbert space adjoint is defined, also as a closed, unbounded operator on $L^2(M)$, with a dense domain. The symbol A^* is also our preferred notation for the Hilbert space adjoint. To avoid confusion, we will *temporarily* use A^t to denote the adjoint on $\mathcal{D}'(M)$, so $A^t \in OPS^m(M)$,

$A^t : H^{s+m}(M) \to H^s(M)$, for all s. Now the unbounded operator A^* has domain

$$\mathcal{D}(A^*) = \{u \in L^2(M) : |(u, Av)| \le c(u)\|v\|_{L^2}, \forall\, v \in \mathcal{D}(A)\}, \tag{10.9}$$

and then A^*u is the unique element of $L^2(M)$ such that

$$(A^*u, v) = (u, Av), \quad \text{for all } v \in \mathcal{D}(A). \tag{10.10}$$

Recall that $\mathcal{D}(A) = H^m(M)$. Since, for any $u \in H^m(M)$, $v \in H^m(M)$, we have $(A^t u, v) = (u, Av)$, we see that $\mathcal{D}(A^*) \supset H^m(M)$ and $A^* = A^t$ on $H^m(M)$. On the other hand, $(u, Av) = (A^t u, v)$ holds for all $v \in H^m(M)$, $u \in L^2(M)$, the latter inner product being given by the duality of $H^{-m}(M)$ and $H^m(M)$. Thus it follows that

$$u \in \mathcal{D}(A^*) \Longrightarrow A^*u = A^t u \in L^2(M).$$

But elliptic regularity for $A^t \in OPS^m_{1,0}(M)$ then implies $u \in H^m(M)$. Thus

$$\mathcal{D}(A^*) = H^m(M), \quad A^* = A^t\big|_{H^m(M)}. \tag{10.11}$$

In particular, if A is elliptic in $OPS^m_{1,0}(M)$, $m > 0$, and also symmetric (i.e., $A = A^t$), then the Hilbert space operator is *self-adjoint*; $A = A^*$. For any $\lambda \in \mathbb{C} \setminus \mathbb{R}$, $(\lambda I - A)^{-1} : L^2(M) \to \mathcal{D}(A) = H^m(M)$, so A has compact resolvent. Thus $L^2(M)$ has an orthonormal basis of eigenfunctions of A, $Au_j = \lambda_j u_j$, $|\lambda_j| \to \infty$, and, by elliptic regularity, each u_j belongs to $C^\infty(M)$.

Exercises

In the following exercises, assume that M is a smooth, compact, Riemannian manifold. Let $A \in OPS^m(M)$ be elliptic, positive, and self-adjoint, with $m > 0$. Let u_j be an orthonormal basis of $L^2(M)$ consisting of eigenfunctions of A, $Au_j = \lambda_j u_j$. Given $f \in \mathcal{D}'(M)$, form "Fourier coefficients" $\hat{f}(j) = (f, u_j)$. Thus $f \in L^2(M)$ implies

$$f = \sum_{j=0}^{\infty} \hat{f}(j) u_j, \tag{10.12}$$

with convergence in L^2-norm.

1. Given $s \in \mathbb{R}$, show that $f \in H^s(M)$ if and only if $\sum |\hat{f}(j)|^2 \langle \lambda_j \rangle^{2s/m} < \infty$.
2. Show that, for any $s \in \mathbb{R}$, $f \in H^s(M)$, (10.12) holds, with convergence in H^s-norm. Conclude that if $s > n/2$ and $f \in H^s(M)$, the series converges uniformly to f.
3. If $s > n/2$ and $f \in H^s(M)$, show that (10.12) converges *absolutely*. (*Hint*: Fix $x_0 \in M$ and pick $c_j \in \mathbb{C}$, $|c_j| = 1$, such that $c_j \hat{f}(j) u_j(x_0) \ge 0$. Now consider $\sum c_j \hat{f}(j) u_j$.)
4. Let $-L$ be a second-order, elliptic, positive, self-adjoint differential operator on a compact Riemannian manifold M. Suppose $A \in OPS^1(M)$ is positive, self-adjoint, and

$A^2 = -L + R$, where $R : \mathcal{D}'(M) \to C^\infty(M)$. Show that $A - \sqrt{-L} : \mathcal{D}'(M) \to C^\infty(M)$.

One approach to Exercise 4 is the following.

5. Given $f \in H^s(M)$, form

$$u(y,x) = e^{-y\sqrt{-L}} f(x), \quad v(y,x) = e^{-yA} f(x),$$

for $(y,x) \in [0,\infty) \times M$. Note that

$$\Bigl(\frac{\partial^2}{\partial y^2} + L\Bigr)u = 0, \quad \Bigl(\frac{\partial^2}{\partial y^2} + L\Bigr)v = -Rv(y,x).$$

Use estimates and regularity for the Dirichlet problem for $\partial^2/\partial y^2 + L$ on $[0,\infty) \times M$ to show that $u - v \in C^\infty([0,\infty) \times M)$. Conclude that $\partial u/\partial y - \partial v/\partial y\big|_{y=0} = (A - \sqrt{-L}) f \in C^\infty(M)$.

6. With L as above, use the symbol calculus of §4 to construct a self-adjoint $A \in OPS^1(M)$, with positive principal symbol, such that $A^2 + L \in OPS^{-\infty}(M)$. Conclude that Exercise 4 applies to A.
7. Show that $OPS^0_{1,0}(M)$ has a natural Fréchet space structure.

11. The method of layer potentials

We discuss, in the light of the theory of pseudodifferential operators, the use of "single- and double-layer potentials" to study the Dirichlet and Neumann boundary problems for the Laplace equation. Material developed here will be useful in §7 of Chap. 9, which treats the use of integral equations in scattering theory.

Let $\overline{\Omega}$ be a connected, compact Riemannian manifold with nonempty boundary; $n = \dim \Omega$. Suppose $\overline{\Omega} \subset M$, a Riemannian manifold of dimension n without boundary, on which there is a fundamental solution $E(x,y)$ to the Laplace equation:

$$\Delta_x E(x,y) = \delta_y(x), \tag{11.1}$$

where $E(x,y)$ is the Schwartz kernel of an operator $E(x,D) \in OPS^{-2}(M)$; we have

$$E(x,y) \sim c_n \operatorname{dist}(x,y)^{2-n} + \cdots \tag{11.2}$$

as $x \to y$, if $n \ge 3$, while

$$E(x,y) \sim c_2 \log \operatorname{dist}(x,y) + \cdots \tag{11.3}$$

if $n = 2$. Here, $c_n = -\bigl[(n-2)\mathrm{Area}(S^{n-1})\bigr]^{-1}$ for $n \ge 3$, and $c_2 = 1/2\pi$. The single- and double-layer potentials of a function f on $\partial\Omega$ are defined by

$$\mathcal{S}\ell\, f(x) = \int_{\partial\Omega} f(y) E(x,y)\, dS(y), \tag{11.4}$$

and

$$\mathcal{D}\ell\, f(x) = \int_{\partial\Omega} f(y)\, \frac{\partial E}{\partial \nu_y}(x, y)\, dS(y), \tag{11.5}$$

for $x \in M \setminus \partial\Omega$. Given a function v on $M \setminus \partial\Omega$, for $x \in \partial\Omega$, let $v_+(x)$ and $v_-(x)$ denote the limits of $v(z)$ as $z \to x$, from $z \in \Omega$ and $z \in M \setminus \overline{\Omega} = \mathcal{O}$, respectively, when these limits exist. The following are fundamental properties of these layer potentials.

Proposition 11.1. *For $x \in \partial\Omega$, we have*

$$\mathcal{S}\ell\, f_+(x) = \mathcal{S}\ell\, f_-(x) = Sf(x) \tag{11.6}$$

and

$$\mathcal{D}\ell\, f_\pm(x) = \pm\frac{1}{2} f(x) + \frac{1}{2} Nf(x), \tag{11.7}$$

where, for $x \in \partial\Omega$,

$$Sf(x) = \int_{\partial\Omega} f(y) E(x, y)\, dS(y) \tag{11.8}$$

and

$$Nf(x) = 2\int_{\partial\Omega} f(y) \frac{\partial E}{\partial \nu_y}(x, y)\, dS(y). \tag{11.9}$$

Note that $E(x, \cdot)\big|_{\partial\Omega}$ is integrable, uniformly in x, and that the conclusion in (11.6) is elementary, at least for f continuous; the conclusion in (11.7) is a bit more mysterious. To see what is behind such results, let us look at the more general situation of

$$v = p(x, D)(f\sigma), \tag{11.10}$$

where $\sigma \in \mathcal{E}'(M)$ is surface measure on a hypersurface (here $\partial\Omega$), $f \in \mathcal{D}'(\partial\Omega)$, so $f\sigma \in \mathcal{E}'(M)$. Assume that $p(x, D) \in OPS^m(M)$. Make a local coordinate change, straightening out the surface to $\{x_n = 0\}$. Then, in this coordinate system

$$\begin{aligned} v(x', x_n) &= \int \hat{f}(\xi') e^{ix'\cdot\xi'} p(x, \xi', \xi_n) e^{ix_n\xi_n}\, d\xi_n\, d\xi' \\ &= q(x_n, x', D_{x'}) f, \end{aligned} \tag{11.11}$$

for $x_n \neq 0$, where

$$q(x_n, x', \xi') = \int p(x, \xi', \xi_n) e^{ix_n\xi_n}\, d\xi_n. \tag{11.12}$$

If $p(x, \xi)$ is homogeneous of degree m in ξ, for $|\xi| \geq 1$, then for $|\xi'| \geq 1$ we have

$$q(x_n, x', \xi') = |\xi'|^{m+1} \tilde{p}(x, \omega', x_n|\xi'|), \tag{11.13}$$

where $\omega' = \xi'/|\xi'|$ and

$$\tilde{p}(x, \omega', \tau) = \int p(x, \omega', \zeta) e^{i\zeta\tau}\, d\zeta.$$

Now, if $m < -1$, the integral in (11.12) is absolutely convergent and $q(x_n, x', \xi')$ is continuous in all arguments, even across $x_n = 0$. On the other hand, if $m = -1$, then, temporarily neglecting all the arguments of p but the last, we are looking at the Fourier transform of a smooth function of one variable whose asymptotic behavior as $\xi_n \to \pm\infty$ is of the form $C_1^{\pm}\xi_n^{-1} + C_2^{\pm}\xi_n^{-2} + \cdots$. From the results of Chap. 3 we know that the Fourier transform is smooth except at $x_n = 0$, and if $C_1^+ = C_1^-$, then the Fourier transform has a jump across $x_n = 0$; otherwise there may be a logarithmic singularity.

It follows that if $p(x, D) \in OPS^m(M)$ and $m < -1$, then (11.10) has a limit on $\partial\Omega$, given by

$$v\big|_{\partial\Omega} = Qf, \quad Q \in OPS^{m+1}(\partial\Omega). \tag{11.14}$$

On the other hand, if $m = -1$ and the symbol of $p(x, D)$ has the behavior that, for $x \in \partial\Omega$, ν_x normal to $\partial\Omega$ at x,

$$p(x, \xi \pm \tau\nu_x) = \pm C(x, \xi)\tau^{-1} + O(\tau^{-2}), \quad \tau \to +\infty, \tag{11.15}$$

then (11.10) has a limit from each side of $\partial\Omega$, and

$$v_{\pm} = Q_{\pm} f, \quad Q_{\pm} \in OPS^0(\partial\Omega). \tag{11.16}$$

To specialize these results to the setting of Proposition 11.1, note that

$$\mathcal{S}\ell\, f = E(x, D)(f\sigma) \tag{11.17}$$

and

$$\mathcal{D}\ell\, f = E(x, D)X^*(f\sigma), \tag{11.18}$$

where X is any vector field on M equal to $\partial/\partial\nu$ on $\partial\Omega$, with formal adjoint X^*, given by

$$X^*v = -Xv - (\operatorname{div} X)v. \tag{11.19}$$

The analysis of (11.10) applies directly to (11.17), with $m = -2$. That the boundary value is given by (11.8) is elementary for $f \in C(\partial\Omega)$, as noted before. Given (11.14), it then follows for more general f.

Now (11.18) is also of the form (11.10), with $p(x, D) = E(x, D)X^* \in OPS^{-1}(M)$. Note that the principal symbol at $x \in \partial\Omega$ is given by

$$p_0(x, \xi) = -|\xi|^{-2}\langle \nu(x), \xi\rangle, \tag{11.20}$$

which satisfies the condition (11.15), so the conclusion (11.16) applies. Note that

$$p_0(x, \xi \pm \tau\nu_x) = -|\xi \pm \tau\nu_x|^{-2}\langle \nu_x, \xi \pm \tau\nu_x\rangle,$$

so in this case (11.15) holds with $C(x, \xi) = 1$. Thus the operators $Q_\pm$ in (11.16) have principal symbols $\pm$ const. That the constant is as given in (11.7) follows from keeping careful track of the constants in the calculations (11.11)–(11.13) (cf. Exercise 9 below).

Let us take a closer look at the behavior of $(\partial/\partial\nu_y)E(x, y)$. Note that, for x close to y, if $V_{x,y}$ denotes the unit vector at y in the direction of the geodesic from x to y, then (for $n \geq 3$)

$$\nabla_y E(x, y) \sim (2 - n)c_n \operatorname{dist}(x, y)^{1-n} V_{x,y} + \cdots . \tag{11.21}$$

If $y \in \partial\Omega$ and ν_y is the unit normal to $\partial\Omega$ at y, then

$$\frac{\partial}{\partial\nu_y} E(x, y) \sim (2 - n)c_n \operatorname{dist}(x, y)^{1-n}\langle V_{x,y}, \nu_y\rangle + \cdots . \tag{11.22}$$

Note that $(2-n)c_n = -1/\text{Area}(S^{n-1})$. Clearly, the inner product $\langle V_{x,y}, \nu_y\rangle = \alpha(x, y)$ restricted to $(x, y) \in \partial\Omega \times \partial\Omega$ is Lipschitz and vanishes on the diagonal $x = y$. This vanishing makes $(\partial E/\partial\nu_y)(x, y)$ integrable on $\partial\Omega \times \partial\Omega$. It is clear that in the case (11.7), $Q_\pm$ have Schwartz kernels equal to $(\partial/\partial\nu_y)E(x, y)$ on the complement of the diagonal in $\partial\Omega \times \partial\Omega$. In light of our analysis above of the principal symbol of $Q_\pm$, the proof of (11.7) is complete.

As a check on the evaluation of the constant c in $\mathcal{D}\ell\, f_\pm = \pm cf + (1/2)Nf$, $c = 1/2$, note that applying Green's formula to $\int_\Omega (\Delta 1) \cdot E(x, y)\, dy$ readily gives

$$\int_{\partial\Omega} \frac{\partial E}{\partial\nu_y}(x, y)\, dS(y) = \begin{cases} 1, & \text{for } x \in \Omega, \\ 0, & \text{for } x \in \mathcal{O}, \end{cases}$$

as the value of $\mathcal{D}\ell\, f_\pm$ for $f = 1$. Since $\mathcal{D}\ell\, f_+ - \mathcal{D}\ell\, f_- = 2cf$, this forces $c = 1/2$.

The way in which $\pm(1/2)f(x)$ arises in (11.7) is captured well by the model case of $\partial\Omega$ a hyperplane in $\mathbb{R}^n$, and

$$E\big((x', x_n), (y', 0)\big) = c_n\big[(x'-y')^2 + x_n^2\big]^{(2-n)/2},$$

when (11.22) becomes

$$\frac{\partial}{\partial y_n} E\big((x', x_n), (y', 0)\big) = (2-n)c_n x_n\big[(x'-y')^2 + x_n^2\big]^{-n/2},$$

though in this example $N = 0$.

The following properties of the operators S and N are fundamental.

Proposition 11.2. *We have*

$$S, N \in OPS^{-1}(\partial\Omega), \quad S \text{ elliptic.} \tag{11.23}$$

Proof. That S has this behavior follows immediately from (11.2) and (11.3). The ellipticity at x follows from taking normal coordinates at x and using Exercise 3 of §4, for $n \geq 3$; for $n = 2$, the reader can supply an analogous argument. That N also satisfies (11.23) follows from (11.22) and the vanishing of $\alpha(x, y) = \langle V_{x,y}, \nu_y\rangle$ on the diagonal.

An important result complementary to Proposition 11.1 is the following, on the behavior of the normal derivative at $\partial\Omega$ of single-layer potentials.

Proposition 11.3. *For $x \in \partial\Omega$, we have*

$$\frac{\partial}{\partial\nu}\mathcal{S}\ell\, f_{\pm}(x) = \frac{1}{2}\big(\mp f + N^{\#} f\big), \tag{11.24}$$

where $N^{\#} \in OPS^{-1}(\partial\Omega)$ is given by

$$N^{\#} f(x) = 2\int_{\partial\Omega} f(y)\, \frac{\partial E}{\partial\nu_x}(x, y)\, dS(y). \tag{11.25}$$

Proof. The proof of (11.24) is directly parallel to that of (11.7). To see on general principles why this should be so, use (11.17) to write $(\partial/\partial\nu)\mathcal{S}\ell\, f$ as the restriction to $\partial\Omega$ of

$$X\mathcal{S}\ell\, f = XE(x, D)(f\sigma). \tag{11.26}$$

Using (11.18) and (11.19), we see that

$$\begin{aligned} \mathcal{D}\ell\, f + X\mathcal{S}\ell\, f &= [X, E(x, D)](f\sigma) - E(x, D)(\operatorname{div} X)(f\sigma) \\ &= A(x, D)(f\sigma), \end{aligned} \tag{11.27}$$

with $A(x, D) \in OPS^{-2}(M)$, the same class as $E(x, D)$. Thus the extension of $A(x, D)(f\sigma)$ to $\partial\Omega$ is straightforward, and we have

$$\frac{\partial}{\partial\nu}\mathcal{S}\ell\, f_\pm = -\mathcal{D}\ell\, f_\pm + A(x, D)(f\sigma)\big|_{\partial\Omega}. \tag{11.28}$$

In particular, the jumps across $\partial\Omega$ are related by

$$\frac{\partial}{\partial\nu}\mathcal{S}\ell\, f_+ - \frac{\partial}{\partial\nu}\mathcal{S}\ell\, f_- = \mathcal{D}\ell\, f_- - \mathcal{D}\ell\, f_+, \tag{11.29}$$

consistent with the result implied by formulas (11.7) and (11.24).

It is also useful to understand the boundary behavior of $(\partial/\partial\nu)\mathcal{D}\ell\, f$. This is a bit harder since $\partial^2 E/\partial\nu_x\partial\nu_y$ is more highly singular. From here on, assume $E(x, y) = E(y, x)$, so also $\Delta_y E(x, y) = \delta_x(y)$. We define the *Neumann operator*

$$\mathcal{N} : C^\infty(\partial\Omega) \longrightarrow C^\infty(\partial\Omega) \tag{11.30}$$

as follows. Given $f \in C^\infty(\partial\Omega)$, let $u \in C^\infty(\overline{\Omega})$ be the unique solution to

$$\Delta u = 0 \text{ on } \Omega, \quad u = f \text{ on } \partial\Omega, \tag{11.31}$$

and let

$$\mathcal{N} f = \frac{\partial u}{\partial\nu}\Big|_{\partial\Omega}, \tag{11.32}$$

the limit taken from within Ω. It is a simple consequence of Green's formula that if we form

$$\int_{\partial\Omega} \Big[f(y)\, \frac{\partial E}{\partial\nu_y}(x, y) - \mathcal{N} f(y) E(x, y) \Big] dS(y) = \mathcal{D}\ell\, f(x) - \mathcal{S}\ell\, \mathcal{N} f(x), \tag{11.33}$$

for $x \in M \setminus \partial\Omega$, then

$$\mathcal{D}\ell\, f(x) - \mathcal{S}\ell\, \mathcal{N} f(x) = \begin{cases} u(x), & x \in \Omega, \\ 0, & x \in M \setminus \overline{\Omega}, \end{cases} \tag{11.34}$$

where u is given by (11.31). Note that taking the limit of (11.34) from within Ω, using (11.6) and (11.7), gives $f = (1/2) f + (1/2) N f - S\mathcal{N} f$, which implies the identity

$$S\mathcal{N} = -\frac{1}{2}(I - N). \tag{11.35}$$

Taking the limit in (11.34) from $M \setminus \overline{\Omega}$ gives the same identity. In view of the behavior (11.23), in particular the ellipticity of S, we conclude that

$$\mathcal{N} \in OPS^1(\partial\Omega), \quad \text{elliptic.} \tag{11.36}$$

Now we apply $\partial/\partial\nu$ to the identity (11.34), evaluating on $\partial\Omega$ from both sides. Evaluating from Ω gives

$$\frac{\partial}{\partial\nu}\mathcal{D}\ell\; f_+ - \frac{\partial}{\partial\nu}\mathcal{S}\ell\,\mathcal{N} f_+ = \mathcal{N} f, \tag{11.37}$$

while evaluating from $M \setminus \overline{\Omega}$ gives

$$\frac{\partial}{\partial\nu}\mathcal{D}\ell\; f_- - \frac{\partial}{\partial\nu}\mathcal{S}\ell\,\mathcal{N} f_- = 0. \tag{11.38}$$

In particular, applying $\partial/\partial\nu$ to (11.34) shows that $(\partial/\partial\nu)\mathcal{D}\ell\; f_\pm$ exists, by Proposition 11.3. Furthermore, applying (11.24) to $(\partial/\partial\nu)\mathcal{S}\ell\;\mathcal{N} f_\pm$, we have a proof of the following.

Proposition 11.4. *For $x \in \partial\Omega$, we have*

$$\frac{\partial}{\partial\nu}\mathcal{D}\ell\; f_\pm(x) = \frac{1}{2}(I + N^\#)\mathcal{N} f. \tag{11.39}$$

In particular, there is no jump across $\partial\Omega$ of $(\partial/\partial\nu)\mathcal{D}\ell\; f$.

We have now developed the layer potentials far enough to apply them to the study of the Dirichlet problem. We want an approximate formula for the Poisson integral $u = \text{PI}\; f$, the unique solution to

$$\Delta u = 0 \text{ in } \Omega, \quad u\big|_{\partial\Omega} = f. \tag{11.40}$$

Motivated by the Poisson integral formula on $\mathbb{R}^n_+$, we look for a solution of the form

$$u(x) = \mathcal{D}\ell\; g(x), \quad x \in \Omega, \tag{11.41}$$

and try to relate g to f. In view of Proposition 11.1, letting $x \to z \in \partial\Omega$ in (11.41) yields

$$u(z) = \frac{1}{2}(g + Ng), \quad \text{for } z \in \partial\Omega. \tag{11.42}$$

Thus if we define u by (11.41), then (11.40) is equivalent to

$$f = \frac{1}{2}(I + N)g. \tag{11.43}$$

Alternatively, we can try to solve (11.40) in terms of a single-layer potential:

$$u(x) = \mathcal{S}\ell\, h(x), \quad x \in \Omega. \tag{11.44}$$

If u is defined by (11.44), then (11.40) is equivalent to

$$f = Sh. \tag{11.45}$$

Note that, by (11.23), the operator $(1/2)(I + N)$ in (11.43) is Fredholm, of index zero, on each space $H^s(\partial\Omega)$. It is not hard to verify that S is elliptic of order -1, with real principal symbol, so for each s,

$$S : H^{s-1}(\partial\Omega) \longrightarrow H^s(\partial\Omega)$$

is Fredholm, of index zero.

One basic case when (11.43) and (11.45) can both be solved is the case of bounded Ω in $M = \mathbb{R}^n$, with the standard flat Laplacian.

Proposition 11.5. *If $\overline{\Omega}$ is a smooth, bounded subdomain of $\mathbb{R}^n$, with connected complement, then, for all s,*

$$I + N : H^s(\partial\Omega) \longrightarrow H^s(\partial\Omega) \quad \textit{and} \quad S : H^{s-1}(\partial\Omega) \longrightarrow H^s(\partial\Omega) \tag{11.46}$$

are isomorphisms.

Proof. It suffices to show that $I + N$ and S are injective on $C^\infty(\partial\Omega)$. First, if $g \in C^\infty(\partial\Omega)$ belongs to the null space of $I + N$, then, by (11.42) and the maximum principle, we have $\mathcal{D}\ell\, g = 0$ in Ω. By (11.7), the jump of $\mathcal{D}\ell\, g$ across $\partial\Omega$ is g, so we have for $v = \mathcal{D}\ell\, g|_{\mathcal{O}}$, where $\mathcal{O} = \mathbb{R}^n \setminus \overline{\Omega}$,

$$\Delta v = 0 \text{ on } \mathcal{O}, \quad v\big|_{\partial\Omega} = -g. \tag{11.47}$$

Also, v clearly vanishes at infinity. Now, by (11.39), $(\partial/\partial\nu)\mathcal{D}\ell\, g$ does not jump across $\partial\Omega$, so we have $\partial v/\partial\nu = 0$ on $\partial\Omega$. But at a point on $\partial\Omega$ where $-g$ is maximal, this contradicts Zaremba's principle, unless $g = 0$. This proves that $I + N$ is an isomorphism in this case.

Next, suppose $h \in C^\infty(\partial\Omega)$ belongs to the null space of S. Then, by (11.45) and the maximum principle, we have $\mathcal{S}\ell\, h = 0$ on Ω. By (11.24), the jump of $(\partial/\partial\nu)\mathcal{S}\ell\, h$ across $\partial\Omega$ is $-h$, so we have for $w = \mathcal{S}\ell\, h|_{\mathcal{O}}$ that

$$\Delta w = 0 \text{ on } \mathcal{O}, \quad \frac{\partial w}{\partial\nu}\Big|_{\partial\Omega} = h, \tag{11.48}$$

and w vanishes at infinity. This time, $\mathcal{S}\ell\, h$ does not jump across $\partial\Omega$, so we also have $w = 0$ on $\partial\Omega$. The maximum principle forces $w = 0$ on $\mathcal{O}$, so $h = 0$. This proves that S is an isomorphism in this case.

In view of (11.6), we see that (11.44) and (11.45) also give a solution to $\Delta u = 0$ on the *exterior* region $\mathbb{R}^n \setminus \overline{\Omega}$, satisfying $u = f$ on $\partial\Omega$ and $u(x) \to 0$ as $|x| \to \infty$, if $n \geq 3$. This solution is unique, by the maximum principle.

One can readily extend the proof of Proposition 11.5 and show that $I + N$ and S in (11.46) are isomorphisms in somewhat more general circumstances.

Let us now consider the Neumann problem

$$\Delta u = 0 \text{ on } \Omega, \quad \frac{\partial u}{\partial \nu} = \varphi \text{ on } \partial\Omega. \tag{11.49}$$

We can relate (11.49) to (11.40) via the Neumann operator:

$$\varphi = \mathcal{N} f. \tag{11.50}$$

Let us assume that Ω is connected; then

$$\operatorname{Ker} \mathcal{N} = \{ f = \text{ const. on } \Omega \}, \tag{11.51}$$

so $\dim \operatorname{Ker} \mathcal{N} = 1$. Note that, by Green's theorem,

$$(\mathcal{N} f, g)_{L^2(\partial\Omega)} = -(du, dv)_{L^2(\Omega)} = (f, \mathcal{N} g)_{L^2(\partial\Omega)}, \tag{11.52}$$

where $u = PI\ f$, $v = PI\ g$, so $\mathcal{N}$ is symmetric. In particular,

$$(\mathcal{N} f, f)_{L^2(\partial\Omega)} = -\|du\|^2_{L^2(\Omega)}, \tag{11.53}$$

so $\mathcal{N}$ is negative-semidefinite. The symmetry of $\mathcal{N}$ together with its ellipticity implies that, for each s,

$$\mathcal{N} : H^{s+1}(\partial\Omega) \longrightarrow H^s(\partial\Omega) \tag{11.54}$$

is Fredholm, of index zero, with both $\operatorname{Ker} \mathcal{N}$ and $\mathcal{R}(\mathcal{N})^\perp$ of dimension 1, and so

$$\mathcal{R}(\mathcal{N}) = \Big\{ \varphi \in H^s(\partial\Omega) : \int_{\partial\Omega} \varphi \, dS = 0 \Big\}, \tag{11.55}$$

this integral interpreted in the obvious distributional sense when $s < 0$.

By (11.35), whenever S is an isomorphism in (11.46), we can say that (11.50) is equivalent to

$$(I - N) f = -2S\varphi. \tag{11.56}$$

We can also represent a solution to (11.49) as a single-layer potential, of the form (11.44). Using (11.24), we see that this works provided h satisfies

$$(I - N^\#) h = -2\varphi. \tag{11.57}$$

In view of the fact that (11.44) solves the Dirichlet problem (11.40) with $f = Sh$, we deduce the identity $\varphi = \mathcal{N}Sh$, or

$$\mathcal{N}S = -\frac{1}{2}(I - N^{\#}), \tag{11.58}$$

complementing (11.35). Comparing these identities, representing $S\mathcal{N}S$ in two ways, we obtain the intertwining relation

$$SN^{\#} = NS. \tag{11.59}$$

Also note that, under the symmetry hypothesis $E(x, y) = E(y, x)$, we have $N^{\#} = N^*$.

The method of layer potentials is applicable to other boundary problems. An application to the "Stokes system" will be given in Chap. 17, §A.

We remark that a number of results in this section do not make substantial use of the pseudodifferential operator calculus developed in the early sections; this makes it easy to extend such results to situations where the boundary has limited smoothness. For example, it is fairly straightforward to extend results on the double-layer potential $\mathcal{D}\ell$ to the case where $\partial\Omega$ is a C^{1+r}-hypersurface in $\mathbb{R}^n$, for any $r > 0$, and in particular to extend (partially) the first part of (11.46), obtaining

$$I + N : L^2(\partial\Omega) \longrightarrow L^2(\partial\Omega) \text{ invertible},$$

in such a case, thus obtaining the representation (11.41) for the solution to the Dirichlet problem with boundary data in $L^2(\partial\Omega)$, when $\partial\Omega$ is a C^{1+r}-surface. Results on S in (11.23) and some results on the Neumann operator, such as (11.36), do depend on the pseudodifferential operator calculus, so more work is required to adapt this material to C^{1+r}-surfaces, though that has been done.

In fact, via results of [Ca3] and [CMM], the layer potential approach has been extended to domains in $\mathbb{R}^n$ bounded by C^1-surfaces, in [FJR], and then to domains bounded by Lipschitz surfaces, in [Ver] and [DK]. See also [JK] for nonhomogeneous equations. Extensions to Lipschitz domains in Riemannian manifolds are given in [MT1] and [MT2], and extensions to "uniformly rectifiable" domains in [D, DS], and [HMT]. We mention just one result here; many others can be found in the sources cited above and references they contain.

Proposition 11.6. *If Ω is a Lipschitz domain in a compact Riemannian manifold M, then*

$$\mathrm{PI} : L^2(\partial\Omega) \longrightarrow H^{1/2}(\Omega).$$

Exercises

1. Let M be a compact, connected Riemannian manifold, with Laplace operator L, and let $\overline{\Omega} = [0, 1] \times M$, with Laplace operator $\Delta = \partial^2/\partial y^2 + L$, $y \in [0, 1]$. Show that the Dirichlet problem

$$\Delta u = 0 \text{ on } \Omega, \quad u(0, x) = f_0(x), \; u(1, x) = f_1(x)$$

has the solution

$$u(y,x) = e^{-y\sqrt{-L}}\varphi_0 + e^{-(1-y)\sqrt{-L}}\varphi_1 + \kappa y,$$

where κ is the constant $\kappa = (\text{vol } M)^{-1} \int_M (f_1 - f_0)\, dV$, and

$$\varphi_0 = (1 - e^{-2\sqrt{-L}})^{-1}(f_0 - e^{-\sqrt{-L}} f_1 - \kappa),$$
$$\varphi_1 = (1 - e^{-2\sqrt{-L}})^{-1}(f_1 - \kappa - e^{-\sqrt{-L}} f_0),$$

the operator $(1 - e^{-2\sqrt{-L}})^{-1}$ being well defined on $(\ker L)^\perp$.

2. If $\mathcal{N} f_0(x) = (\partial u/\partial y)(0, x)$, where u is as above, with $f_1 = 0$, show that

$$\mathcal{N} f_0 = -\sqrt{-L}\, f_0 + \mathcal{R} f_0,$$

where $\mathcal{R}$ is a smoothing operator, $\mathcal{R} : \mathcal{D}'(M) \to C^\infty(M)$. Using (11.36), deduce that these calculations imply

$$\sqrt{-L} \in OPS^1(M).$$

Compare Exercises 4–6 of §10.

3. If PI: $C^\infty(\partial\Omega) \to C^\infty(\overline{\Omega})$ is the Poisson integral operator solving (11.40), show that, for $x \in \Omega$,

$$\text{PI } f(x) = \int_{\partial\Omega} k(x,y) f(y)\, dS(y),$$

with

$$|k(x,y)| \le C\big(d(x,y)^2 + \rho(x)^2\big)^{-(n-1)/2},$$

where $n =$ dim Ω, $d(x,y)$ is the distance from x to y, and $\rho(x)$ is the distance from x to $\partial\Omega$.

4. If $\overline{M}$ is an $(n-1)$-dimensional surface with boundary in $\overline{\Omega}$, intersecting $\partial\Omega$ transversally, with $\partial M \subset \partial\Omega$, and $\rho : C^\infty(\overline{\Omega}) \to C^\infty(\overline{M})$ is restriction to $\overline{M}$, show that

$$\rho \circ \text{PI} : L^2(\partial\Omega) \longrightarrow L^2(M).$$

(*Hint*: Look at Exercise 2 in §5 of Appendix A on functional analysis.)

5. Given $y \in \Omega$, let G_y be the "Green function," satisfying

$$\Delta G_y = \delta_y, \quad G_y = 0 \text{ on } \partial\Omega.$$

Show that, for $f \in C^\infty(\partial\Omega)$,

$$\text{PI } f(y) = \int_{\partial\Omega} f(x)\, \partial_\nu G_y(x)\, dS(x).$$

(*Hint*: Apply Green's formula to $(\text{PI } f, \Delta G_y) = (\text{PI } f, \Delta G_y) - (\Delta \text{ PI } f, G_y)$.)

6. Assume u is scalar, $\Delta u = f$, and w is a vector field on $\overline{\Omega}$. Show that

$$\begin{aligned}\int_{\partial\Omega} \langle \nu, w\rangle |\nabla u|^2\, dS &= 2\int_{\partial\Omega} (\nabla_w u)(\partial_\nu u)\, dS - 2\int_{\Omega} (\nabla_w u) f\, dV \\ &\quad + \int_{\Omega} (\text{div } w)|\nabla u|^2\, dV - 2\int_{\Omega} (\mathcal{L}_w g)(\nabla u, \nabla u)\, dV,\end{aligned} \tag{11.60}$$

where g is the metric tensor on $\overline{\Omega}$. This identity is a "Rellich formula."
(*Hint*: Compute $\operatorname{div}\big(\langle \nabla u, \nabla u\rangle w\big)$ and $2\operatorname{div}(\nabla_w u \cdot \nabla u)$, and apply the divergence theorem to the difference.)

7. In the setting of Exercise 6, assume w is a unit vector field and that $\langle \nu, w\rangle \ge a > 0$ on $\partial\Omega$. Deduce that

$$\begin{aligned}\frac{a}{2}\int\limits_{\partial\Omega} |\nabla u|^2\, dS \le \frac{2}{a}\int\limits_{\partial\Omega} |\partial_\nu u|^2\, dS + \int\limits_{\Omega} |f|^2\, dV \\ + \int\limits_{\Omega} \Big\{|\operatorname{div} w| + 2|\operatorname{Def} w| + 1\Big\} |\nabla u|^2\, dV.\end{aligned} \tag{11.61}$$

When $\Delta u = f = 0$, compare implications of (11.61) with implications of (11.36). See [Ver] for applications of Rellich's formula to analysis on domains with Lipschitz boundary.

8. What happens if, in Proposition 11.5, you allow $\mathcal{O} = \mathbb{R}^n \setminus \overline{\Omega}$ to have several connected components? Can you show that *one* of the operators in (11.46) is still an isomorphism?
9. Calculate $q(x_n, x', \xi')$ in (11.13) when $p(x, \xi) = \xi_j |\xi|^{-2}$. Relate this to the results (11.7) and (11.24) for $\mathcal{D}\ell\, f_\pm$ and $\partial_\nu\, \mathcal{S}\ell\, f_\pm$. (*Hint*. The calculation involves $\int (1 + \zeta^2)^{-1} e^{i\zeta\tau}\, d\zeta = \pi e^{-|\tau|}$.)
10. Let N and $N^\#$ be the operators given by (11.9) and (11.25). Show that $N^\# = N^*$, the L^2-adjoint of N.

12. Parametrix for regular elliptic boundary problems

Here we shall complement material on regular boundary problems for elliptic operators developed in §11 of Chap. 5, including in particular results promised after the statement of Proposition 11.16 in that chapter.

Suppose P is an elliptic differential operator of order m on a compact manifold $\overline{M}$ with boundary, with boundary operators B_j of order m_j, $1 \le j \le \ell$, satisfying the regularity conditions given in §11 of Chap. 5. In order to construct a parametrix for the solution to $Pu = f$, $B_j u|_{\partial M} = g_j$, we will use pseudodifferential operator calculus to manipulate P in ways that constant-coefficient operators $P(D)$ were manipulated in that section. To start, we choose a collar neighborhood $\mathcal{C}$ of ∂M, $\mathcal{C} \approx [0,1] \times \partial M$; use coordinates (y, x), $y \in [0,1], x \in \partial M$; and without loss of generality, consider

$$Pu = \frac{\partial^m u}{\partial y^m} + \sum_{j=0}^{m-1} A_j(y, x, D_x) \frac{\partial^j u}{\partial y^j}, \tag{12.1}$$

the order of $A_j(y, x, D_x)$ being $\le m - j$. We convert $Pu = f$ to a first-order system using $v = (v_1, \dots, v_m)^t$, with

$$v_1 = \Lambda^{m-1} u, \dots, v_j = \partial_y^{j-1} \Lambda^{m-j} u, \dots, v_m = \partial_y^{m-1} u, \tag{12.2}$$

as in (11.41) of Chap. 5. Here, Λ can be taken to be any elliptic, invertible operator in $OPS^1(\partial M)$, with principal symbol $|\xi|$ (with respect to some Riemannian metric put on ∂M). Then $Pu = f$ becomes, on $\mathcal{C}$, the system

$$\frac{\partial v}{\partial y} = K(y, x, D_x)v + F, \tag{12.3}$$

where $F = (0, \dots, 0, f)^t$ and

$$K = \begin{pmatrix} 0 & \Lambda & & & \\ & 0 & \Lambda & & \\ & & \ddots & \ddots & \\ & & & & \Lambda \\ C_0 & C_1 & C_2 & \dots & C_{m-1} \end{pmatrix}, \tag{12.4}$$

where

$$C_j(y, x, D_x) = -A_j(y, x, D_x)\Lambda^{1-(m-j)} \tag{12.5}$$

is a smooth family of operators in $OPS^1(\partial M)$, with y as a parameter. As in Lemma 11.3 of Chap. 5, we have that P is elliptic if and only if, for all $(x, \xi) \in T^*\partial M \setminus 0$, the principal symbol $K_1(y, x, \xi)$ has no purely imaginary eigenvalues.

We also rewrite the boundary conditions $B_j u = g_j$ at $y = 0$. If

$$B_j = \sum_{k \le m_j} b_{jk}(x, D_x) \frac{\partial^k}{\partial y^k} \tag{12.6}$$

at $y = 0$, then we have for v_j the boundary conditions

$$\sum_{k \le m_j} \tilde{b}_{jk}(x, D_x)\Lambda^{k-m_j} v_{k+1}(0) = \Lambda^{m-m_j-1} g_j = h_j, \quad 1 \le j \le \ell, \tag{12.7}$$

where $\tilde{b}_{jk}(x, D)$ has the same principal symbol as $b_{jk}(x, D)$. We write this as

$$B(x, D_x)v(0) = h, \quad B(x, D_x) \in OPS^0(\partial M). \tag{12.8}$$

We will construct a parametrix for the solution of (12.3), (12.8), with $F = 0$.

Generalizing (11.52) of Chap. 5, we construct $E_0(y, x, \xi)$ for $(x, \xi) \in T^*\partial M \setminus 0$, the projection onto the sum of the generalized eigenspaces of $K_1(y, x, \xi)$ corresponding to eigenvalues of positive real part, annihilating the other generalized eigenspaces, in the form

$$E_0(y, x, \xi) = \frac{1}{2\pi i} \int_\gamma \big(\zeta - K_1(y, x, \xi)\big)^{-1} d\zeta, \tag{12.9}$$

where $\gamma = \gamma(y, x, \xi)$ is a curve in the right half-plane of $\mathbb{C}$, encircling all the eigenvalues of $K_1(y, x, \xi)$ of positive real part. Then $E_0(y, x, \xi)$ is homogeneous of degree 0 in ξ, so it is the principal symbol of a family of operators in $OPS^0(\partial M)$.

Recall the statement of Proposition 11.9 of Chap. 5 on the regularity condition for $(P, B_j, 1 \le j \le \ell)$. One characterization is that, for $(x, \xi) \in T^*\partial M \setminus 0$,

$$B_0(x, \xi) : V(x, \xi) \longrightarrow \mathbb{C}^\lambda \text{ isomorphically,} \tag{12.10}$$

where $V(x, \xi) = \ker E_0(0, x, \xi)$, and $B_0(x, \xi) : \mathbb{C}^\nu \to \mathbb{C}^\lambda$ is the principal symbol of $B(x, D_x)$. Another, equivalent characterization is that, for any $\eta \in \mathbb{C}^\lambda$, $(x, \xi) \in T^*\partial M \setminus 0$, there exists a unique bounded solution on $y \in [0, \infty)$ to the ODE

$$\frac{\partial \varphi}{\partial y} - K_1(0, x, \xi)\varphi = 0, \quad B_0(x, \xi)\varphi(0) = \eta. \tag{12.11}$$

In that case, of course, $\varphi(0) = \varphi(0, x, \xi)$ belongs to $V(x, \xi)$, so $\varphi(y, x, \xi)$ is actually exponentially decreasing as $y \to +\infty$, for fixed (x, ξ), and it is exponentially decreasing as $|\xi| \to \infty$, for fixed $y > 0, x \in \partial M$.

On a conic neighborhood Γ of any $(x_0, \xi_0) \in T^*\partial M \setminus 0$, one can construct $U_0(y, x, \xi)$ smooth and homogeneous of degree 0 in ξ, so that

$$U_0 K_1 U_0^{-1} = \begin{pmatrix} E_1 & 0 \\ 0 & F_1 \end{pmatrix}, \tag{12.12}$$

where $E_1(y, x, \xi)$ has eigenvalues all in $\operatorname{Re} \zeta < 0$ and F_1 has all its eigenvalues in $\operatorname{Re} \zeta > 0$. If we set $w^{(0)} = U_0(y, x, D)v$, then the equation $\partial v/\partial y = K(y, x, D_x)v$ is transformed to

$$\frac{\partial w^{(0)}}{\partial y} = \begin{pmatrix} E & \\ & F \end{pmatrix} w^{(0)} + Aw^{(0)} = Gw^{(0)} + Aw^{(0)}, \tag{12.13}$$

where $E(y, x, D_x)$ and $F(y, x, D_x)$ have E_1 and F_1 as their principal symbols, respectively, and $A(y, x, D_x)$ is a smooth family of operators in the space $OPS^0(\partial M)$.

We want to decouple this equation more completely into two pieces. The next step is to decouple terms of order zero. Let $w^{(1)} = (I + V_1)w^{(0)}$, with $V_1 \in OPS^{-1}$ to be determined. We have

(12.14)
$$\begin{aligned}\frac{\partial w^{(1)}}{\partial y} &= (I + V_1)G(I + V_1)^{-1}w^{(1)} + (I + V_1)A(I + V_1)^{-1}w^{(1)} + \cdots \\ &= Gw^{(1)} + (V_1 G - G V_1 + A)w^{(1)} + \cdots ,\end{aligned}$$

where the remainder involves terms of order at most -1 operating on $w^{(1)}$. We would like to pick V_1 so that the off-diagonal terms of $V_1G - GV_1 + A$ vanish. We require V_1 to be of the form

$$V_1 = \begin{pmatrix} 0 & V_{12} \\ V_{21} & 0 \end{pmatrix}.$$

If A is put into 2×2 block form with entries A_{jk}, we are led to require that (on the symbol level)

$$\begin{aligned} V_{12}E_1 - F_1V_{12} &= -A_{12}, \\ V_{21}F_1 - E_1V_{21} &= -A_{21}. \end{aligned} \tag{12.15}$$

That we have unique solutions $V_{jk}(y, x, \xi)$ (homogeneous of degree -1 in ξ) is a consequence of the following lemma.

Lemma 12.1. *Let $F \in M_{\nu\times\nu}$, the set of $\nu \times \nu$ matrices, and $E \in M_{\mu\times\mu}$. Define $\psi : M_{\nu\times\mu} \to M_{\nu\times\mu}$ by*

$$\psi(T) = TF - ET.$$

Then ψ is bijective, provided E and F have disjoint spectra.

Proof. In fact, if $\{f_j\}$ are the eigenvalues of F and $\{e_k\}$ those of E, it is easily seen that the eigenvalues of ψ are $\{f_j - e_k\}$.

Thus we obtain solutions V_{12} and V_{21} to (12.15). With such a choice of the symbol of K_1, we have

$$\frac{\partial w^{(1)}}{\partial y} = Gw^{(1)} + \begin{pmatrix} A_1 & \\ & A_2 \end{pmatrix} w^{(1)} + Bw^{(1)}, \tag{12.16}$$

with $B \in OPS^{-1}$. To decouple the part of order -1, we try $w^{(2)} = (I + V_2)w^{(1)}$ with $V_2 \in OPS^{-2}$. We get

$$\frac{\partial w^{(2)}}{\partial y} = Gw^{(2)} + \begin{pmatrix} A_1 & \\ & A_2 \end{pmatrix} w^{(2)} + (V_2G - GV_2 + B)w^{(2)} + \cdots, \tag{12.17}$$

so we want to choose V_2 so that, on the symbol level, the off-diagonal terms of $V_2G - GV_2 + B$ vanish. This is the problem solved above, so we are in good shape.

From here we continue, defining $w^{(j)} = (I + V_j)w^{(j-1)}$ with $V_j \in OPS^{-j}$, decoupling further out along the line. Letting $w = (I + V)v$, with

$$I + V \sim \cdots (I + V_3)(I + V_2)(I + V_1), \quad V \in OPS^{-1}, \tag{12.18}$$

we have

$$\frac{\partial w}{\partial y} = \begin{pmatrix} E' & \\ & F' \end{pmatrix} w, \quad \text{mod } C^\infty, \tag{12.19}$$

with $E' = E$, $F' = F$ mod OPS^0. The system (12.3) is now completely decoupled.

We now concentrate on constructing a parametrix for an "elliptic evolution equation"

$$\frac{\partial u}{\partial y} = E(y, x, D_x)u, \quad u(0) = f, \tag{12.20}$$

where E is a $k \times k$ system of first-order pseudodifferential operators, whose principal symbol satisfies

$$\text{spec } E_1(y, x, \xi) \subset \{\zeta \in \mathbb{C} : \text{Re } \zeta \le -C_0|\xi| < 0\}, \quad \xi \neq 0, \tag{12.21}$$

for some $C_0 > 0$. We look for the parametrix in the form (in local coordinates on ∂M)

$$u(y) = \int A(y, x, \xi) e^{ix\cdot\xi} \hat{f}(\xi)\, d\xi, \tag{12.22}$$

with $A(y, x, \xi)$ in the form

$$A(y, x, \xi) \sim \sum_{j \ge 0} A_j(y, x, \xi), \tag{12.23}$$

and the $A_j(y, x, \xi)$ constructed inductively. We aim to obtain $A(y, x, \xi)$ bounded in $S^0_{1,0}$, for $y \in [0, 1]$, among other things. In such a case,

$$\Big(\frac{\partial}{\partial y} - E\Big)u = (2\pi)^{-n} \int \Big(\frac{\partial A}{\partial y} - L(y, x, \xi)\Big) e^{ix\cdot\xi} \hat{f}(\xi)\, d\xi, \tag{12.24}$$

where

$$L(y, x, \xi) \sim \sum_{\alpha \ge 0} \frac{1}{\alpha!} E^{(\alpha)}(y, x, \xi) A_{(\alpha)}(y, x, \xi). \tag{12.25}$$

We define $A_0(y, x, \xi)$ by the "transport equation"

$$\frac{\partial}{\partial y} A_0(y, x, \xi) = E(y, x, \xi) A_0(y, x, \xi), \quad A_0(0, x, \xi) = I. \tag{12.26}$$

If E is independent of y, the solution is

$$A_0(y,x,\xi) = e^{yE(x,\xi)}.$$

In general, $A_0(y,x,\xi)$ shares with this example the following important properties.

Lemma 12.2. *For $y \in [0,1]$, $k,\ell = 0,1,2,\dots$, we have*

$$y^k D_y^\ell A_0(y,x,\xi) \text{ bounded in } S^{-k+\ell}_{1,0}. \tag{12.27}$$

Proof. We can take $C_2 \in (0, C_0)$ and M large, so that $E(y,x,\xi)$ has spectrum in the half-space $\operatorname{Re}\zeta < -C_2|\xi|$, for $|\xi| \ge M$. Fixing $K \in (0, C_2)$, if $S(y,\sigma,x,\xi)$ is the solution operator to $\partial B/\partial y = E(y,x,\xi)B$, taking $B(\sigma,x,\xi)$ to $B(y,x,\xi)$, then, for $y > \sigma$,

$$|S(y,\sigma,x,\xi)B| \le C\, e^{-K(y-\sigma)|\xi|}|B|, \quad \text{for } |\xi| \ge M. \tag{12.28}$$

It follows that, for $y \in [0,1]$,

$$|A_0(y,x,\xi)| \le C\, e^{-Ky|\xi|}, \tag{12.29}$$

which implies

$$|y^k A_0(y,x,\xi)| \le C_k \langle\xi\rangle^{-k}\, e^{-Ky|\xi|/2}.$$

Now $A_{0j} = \partial A_0/\partial \xi_j$ satisfies

$$\frac{\partial}{\partial y} A_{0j} = E(y,x,\xi)A_{0j} + \frac{\partial E}{\partial \xi_j}(y,x,\xi)A_0, \quad A_{0j}(0,x,\xi) = 0,$$

so

$$A_{0j}(y,x,\xi) = \int_0^y S(y,\sigma,x,\xi)\frac{\partial E}{\partial \xi_j}(\sigma,x,\xi)A_0(\sigma,x,\xi)\, d\sigma, \tag{12.30}$$

which in concert with (12.28) and (12.29) yields

$$\left|\frac{\partial}{\partial \xi_j} A_0(y,x,\xi)\right| \le Cye^{-Ky|\xi|} \le C\langle\xi\rangle^{-1}e^{-Ky|\xi|/2}. \tag{12.31}$$

Inductively, one obtains estimates on $D_\xi^\alpha D_x^\beta A_0(y,x,\xi)$ leading to the $\ell = 0$ case of (12.27), and then use of (12.26) and induction on ℓ give (12.27) in general.

For $j \ge 1$, we define $A_j(y,x,\xi)$ inductively by

$$\frac{\partial A_j}{\partial y} = E(y,x,\xi)A_j(y,x,\xi) + R_j(y,x,\xi), \quad A_j(0,x,\xi) = 0, \tag{12.32}$$

where

$$R_j(y,x,\xi) = \sum_{\ell<j,\ell+|\alpha|=j} \frac{1}{\alpha!} E^{(\alpha)}(y,x,\xi) A_{\ell(\alpha)}(y,x,\xi). \tag{12.33}$$

Then, if, as above, $S(y,\sigma,x,\xi)$ is the solution operator to the equation $\partial B/\partial y = E(y,x,\xi)B$, we have

$$A_j(y,x,\xi) = \int_0^y S(y,\sigma,x,\xi) R_j(\sigma,x,\xi)\, d\sigma, \quad j \geq 1. \tag{12.34}$$

The arguments used to prove Lemma 12.2 also establish the following result.

Lemma 12.3. *For $y \in [0,1]$, $k, \ell = 0, 1, 2, \dots, j \geq 1$, we have*

$$y^k D_y^\ell A_j(y,x,\xi) \text{ bounded in } S_{1,0}^{-j-k+\ell}. \tag{12.35}$$

A symbol satisfying the condition (12.35) will be said to belong to $\mathcal{P}^{-j}$. In fact, it is convenient to use the following stronger property possessed by the symbols $A_j(y,x,\xi)$, for $j \geq 0$. Given the hypothesis (12.21) on spec $E_1(y,x,\xi)$, let $0 < C_1 < C_0$. Then

$$A_j(y,x,\xi) = B_j(y,x,\xi) e^{-C_1 y\langle\xi\rangle}, \quad \text{with } B_j(y,x,\xi) \in \mathcal{P}^{-j}. \tag{12.36}$$

We will say $A_j(y,x,\xi) \in \mathcal{P}_e^{-j}$ if this holds or, more generally, if it holds modulo a smooth family of symbols $S(y) \in S^{-\infty}$, $y \in [0,1]$. The associated families of operators will be denoted $OP\mathcal{P}^{-j}$ and $OP\mathcal{P}_e^{-j}$, respectively.

Operators formed from such symbols have the following mapping property, recapturing the Sobolev space regularity established for solutions to regular elliptic boundary problems in Chap. 5.

Proposition 12.4. *If $A = A(y,x,D_x)$ has symbol*

$$A(y,x,\xi) = B(y,x,\xi) e^{-C_1 y\langle\xi\rangle}, \quad B(y,x,\xi) \in \mathcal{P}^{-j},$$

then, for $s \geq -j - 1/2$,

$$A : H^s(\partial M) \longrightarrow H^{s+j+1/2}(I \times \partial M). \tag{12.37}$$

Proof. First consider the case $s = -1/2$, $j = 0$. As $B(y,x,D_x)$ is bounded in $\mathcal{L}(L^2(\partial M))$ for $y \in [0,1]$, we have, for $f \in H^{-1/2}(\partial M)$,

$$\begin{aligned}\int_0^1 \|A(y)f\|_{L^2(\partial M)}^2 dy &\leq C \int_0^1 \|e^{-C_1 y\Lambda} f\|_{L^2(\partial M)}^2 dx \\ &= C_2 \|\Lambda^{-1/2} f\|_{L^2(\partial M)}^2 - C_2 \|e^{-C_1\Lambda} \Lambda^{-1/2} f\|_{L^2(\partial M)}^2,\end{aligned}$$

with $C_2 = C/(2C_1)$, since

$$(e^{-C_1 y\Lambda} f, e^{-C_1 y\Lambda} f) = -\frac{1}{2C_1}\frac{d}{dy}(e^{-2C_1 y\Lambda} f, \Lambda^{-1} f).$$

This proves (12.37) in this case. The extension to $s = k - 1/2$ $(k = 1, 2, \dots)$, $j = 0$ is straightforward, and then the result for general $s \geq -1/2$, $j = 0$ follows by interpolation. The case of general j is reduced to that of $j = 0$ by forming $A(y, x, \xi)\langle\xi\rangle^{-j}$. One can take any $j \in \mathbb{R}$.

Having constructed operators with symbols in $\mathcal{P}_e^0$ as parametrices of (12.20), we now complete the construction of parametrices for the system (12.3), (12.8), when the regularity condition (12.10) holds. Using a partition of unity, write h as a sum $\sum h_j$, each term of which has wave front set in a conic set Γ_j on which the decoupling procedure (12.12) can be implemented. We drop the subscript j and just call the term h. Then, we construct a parametrix for $w = (I + V)U_0 v$, so that w solves (12.19), with $w(0) = (f, 0)^t$. Set $U = (I + V)U_0$, and let U^{-1} denote a parametrix of U. The solution $w(y)$ takes the form $w(y) = (w_1(y), 0)$, with

$$w_1(y) = A_1(y, x, D_x) f, \quad A_1(y, x, \xi) \in \mathcal{P}_e^0, \tag{12.38}$$

using the construction (12.22)–(12.34). Note that $v(0) = U^{-1}(f, 0)^t = U^{-1} J_1 f$, where here and below we set $J_1 f = (f, 0)^t$. Then

$$Bv(0) = BU^{-1} J_1 f, \tag{12.39}$$

so the boundary condition (12.8) is achieved (mod C^∞) provided f satisfies (mod C^∞)

$$BU^{-1} J_1 f = h. \tag{12.40}$$

The regularity condition (12.10) is precisely the condition that $BU^{-1} J_1$ is an elliptic $\lambda \times \lambda$ system, in $OPS^0(\partial M)$. Letting $Q \in OPS^0(\partial M)$ be a parametrix, we obtain

$$v(y) = U(y)^{-1} J_1 A_1(y) Q h = A^\#(y) h. \tag{12.41}$$

Recall that $Q \in OPS^0(\partial M)$, $U(y)^{-1}$ is a smooth family of operators in $OPS^0(\partial M)$, and $A_1(y) \in OP\mathcal{P}_e^0$. We can then say the following about the composition $A^\#(y) = A^\#(y, x, D_x)$.

Lemma 12.5. *Given $P_j(y)$, smooth families in $OPS^{m_j}(\partial M)$, and $A(y) \in OP\mathcal{P}_e^\mu$, we have*

$$P_1(y) A(y) P_2(y) = B(y) \in OP\mathcal{P}_e^{\mu + m_1 + m_2}. \tag{12.42}$$

The proof is a straightforward application of the results on products from §3.

Consequently, we have a solution mod C^∞ to (12.3), (12.8), constructed in the form $v(y) = A^\#(y)h$, with $A^\#(y) \in OP\mathcal{P}_e^0$. Finally, returning to the boundary problem for P, we have:

Theorem 12.6. *If $(P, B_j, 1 \le j \le \ell)$ is a regular elliptic boundary problem, then a parametrix (i.e., a solution mod C^∞) for*

$$Pu = 0 \text{ on } M, \quad B_j u = g_j \text{ on } \partial M \tag{12.43}$$

is constructed in the form

$$u = \sum_{j=1}^{\ell} Q_j g_j, \tag{12.44}$$

where $Q_j g_j$ is C^∞ on the interior of $\overline{M}$, and, on a collar neighborhood $\mathcal{C} = [0,1] \times \partial M$,

$$Q_j g_j = Q_j(y) g_j, \quad Q_j(y) \in OP\mathcal{P}_e^{-m_j}. \tag{12.45}$$

Recall that m_j is the order of B_j. Here, the meaning of solution mod C^∞ to (12.43) is that if $u^\#$ is given by (12.44), then

$$Pu^\# \in C^\infty(\overline{M}), \quad B_j u^\# - g_j \in C^\infty(\partial M). \tag{12.46}$$

Of course, the regularity results of Chap. 5 imply that if u is a genuine solution to (12.43), then $u - u^\# \in C^\infty(\overline{M})$.

The following is an easy route to localizing boundary regularity results.

Proposition 12.7. *Take $A(y, x, \xi) \in \mathcal{P}^{-j}$. Let $\varphi, \psi \in C^\infty(\partial M)$, and assume their supports are disjoint. Then*

$$f \in \mathcal{D}'(\partial M) \Longrightarrow \varphi A(y, x, D)\psi f \in C^\infty([0,1] \times \partial M). \tag{12.47}$$

Proof. Symbol calculus gives

$$\varphi A(y, x, D)\psi \in \mathcal{P}^{-k}, \quad \forall\, k \ge 0.$$

Hence this is a smooth family of elements of $OPS^{-\infty}(\partial M)$. This readily gives (12.47).

Proposition 12.7 immediately gives the following.

Corollary 12.8. *In the setting of Theorem 12.6, if $\mathcal{O} \subset \partial M$ is open and $g_j \in C^\infty(\mathcal{O})$ for each j, then $u \in C^\infty$ on a neighborhood in $\overline{M}$ of $\mathcal{O}$.*

Exercises

1. Suppose $A(y) \in OP\mathcal{P}^m$. Show that

$$\frac{\partial^j}{\partial y^j} A(y)\big|_{y=0} = Q_j f, \quad Q_j \in OPS^{m+j}_{1,0}(\partial M). \tag{12.48}$$

   If $A(y) \in OP\mathcal{P}^0_e$ is given by the construction (12.24)–(12.34), show that $Q_j \in OPS^j(\partial M)$.

2. Applying the construction of this section to the Dirichlet problem for Δ on $\overline{M}$, show that the Neumann operator $\mathcal{N}$, defined by (11.31)–(11.32), satisfies

$$\mathcal{N} \in OPS^1(\partial M), \tag{12.49}$$

   thus providing a proof different from that used in (11.36).

3. Show that $A(y,x,\xi)$ belongs to $\mathcal{P}^m_e$ if and only if, for some $\epsilon > 0$ and all $N < \infty$,

$$|D^\ell_y D^\beta_x D^\alpha_\xi A_0(y,x,\xi)| \le C_{\alpha\beta\ell}\, e^{-\epsilon y|\xi|}\, \langle\xi\rangle^{m+\ell-|\alpha|} + C_{N\alpha\beta\ell}\langle\xi\rangle^{-N}. \tag{12.50}$$

4. If $A(y,x,\xi) \in \mathcal{P}^{-j}_e$, show that, for some $\kappa > 0$, you can write

$$A(y,x,D) = e^{-\kappa y \Lambda} B(y,x,D), \quad B(y,x,\xi) \in \mathcal{P}^{-j},\ y \in [0,1], \tag{12.51}$$

   modulo a smooth family of smoothing operators.

5. If $u = \mathrm{PI}\, f$ is the solution to $\Delta u = 0$, $u\big|_{\partial\Omega} = f$, use Proposition 12.4 and Theorem 12.6 to show that

$$\mathrm{PI} : H^s(\partial\Omega) \longrightarrow H^{s+1/2}(\Omega), \quad \forall\, s \ge -\frac{1}{2}. \tag{12.52}$$

   Compare the regularity result of Propositions 11.14–11.15 in Chap. 5.

13. Parametrix for the heat equation

Let $L = L(x,D)$ be a second-order, elliptic differential operator, whose principal symbol $L_2(x,\xi)$ is a positive scalar function, though lower-order terms need not be scalar. We want to construct an approximate solution to the initial-value problem

$$\frac{\partial u}{\partial t} = -Lu, \quad u(0) = f, \tag{13.1}$$

in the form

$$u(t,x) = \int a(t,x,\xi) e^{ix\cdot\xi} \hat{f}(\xi)\, d\xi, \tag{13.2}$$

for f supported in a coordinate patch. The amplitude $a(t,x,\xi)$ will have an asymptotic expansion of the form

$$a(t,x,\xi) \sim \sum_{j\geq 0} a_j(t,x,\xi), \tag{13.3}$$

and the $a_j(t,x,\xi)$ will be defined recursively, as follows. By the Leibniz formula, write

$$\begin{aligned} L(a\, e^{ix\cdot\xi}) &= e^{ix\cdot\xi} \sum_{|\alpha|\leq 2} \frac{i^{|\alpha|}}{\alpha!} L^{(\alpha)}(x,\xi) D_x^\alpha a(t,x,\xi) \\ &= e^{ix\cdot\xi}\Big[L_2(x,\xi)a(t,x,\xi) + \sum_{\ell=1}^{2} B_{2-\ell}(x,\xi,D_x)a(t,x,\xi)\Big], \end{aligned} \tag{13.4}$$

where $B_{2-\ell}(x,\xi,D_x)$ is a differential operator (of order ℓ) whose coefficients are polynomials in ξ, homogeneous of degree $2-\ell$ in ξ.

Thus, we want the amplitude $a(t,x,\xi)$ in (13.2) to satisfy (formally)

$$\frac{\partial a}{\partial t} \sim -L_2 a - \sum_{\ell=1}^{2} B_{2-\ell}(x,\xi,D_x)a.$$

If a is taken to have the form (13.3), we obtain the following equations, called "transport equations," for a_j:

$$\frac{\partial a_0}{\partial t} = -L_2(x,\xi)a_0(t,x,\xi) \tag{13.5}$$

and, for $j \geq 1$,

$$\frac{\partial a_j}{\partial t} = -L_2(x,\xi)a_j(t,x,\xi) + \Omega_j(t,x,\xi), \tag{13.6}$$

where

$$\Omega_j(t,x,\xi) = -\sum_{\ell=1}^{2} B_{2-\ell}(x,\xi,D_x)a_{j-\ell}(t,x,\xi). \tag{13.7}$$

By convention we set $a_{-1} = 0$. So that (6.15) reduces to Fourier inversion at $t = 0$, we set

$$a_0(0,x,\xi) = 1, \quad a_j(0,x,\xi) = 0, \quad \text{for } j \geq 1. \tag{13.8}$$

Then we have

$$a_0(t,x,\xi) = e^{-tL_2(x,\xi)}, \tag{13.9}$$

and the solution to (13.6) is

$$a_j(t,x,\xi) = \int_0^t e^{(s-t)L_2(x,\xi)}\Omega_j(s,x,\xi)\,ds. \tag{13.10}$$

In view of (13.7), this defines $a_j(t,x,\xi)$ inductively in terms of $a_{j-1}(t,x,\xi)$ and $a_{j-2}(t,x,\xi)$.

We now make a closer analysis of these terms. Define $A_j(t,x,\xi)$ by

$$a_j(t,x,\xi) = A_j(t,x,\xi)e^{-tL_2(x,\xi)}. \tag{13.11}$$

The following result is useful; it applies to A_j for all $j \geq 1$.

Lemma 13.1. *If $\mu = 0,1,2,\dots$, $\nu \in \{1,2\}$, then $A_{2\mu+\nu}$ can be written in the form*

$$A_{2\mu+\nu}(t,x,\xi) = t^{\mu+1}A^{\#}_{2\mu+\nu}(x,\omega,\xi), \quad \textit{with } \omega = t^{1/2}\xi. \tag{13.12}$$

The factor $A^{\#}_{2\mu+\nu}(x,\omega,\xi)$ is a polynomial in both ω and ξ. It is homogeneous of degree $2-\nu$ in ξ (i.e., either linear or constant). Furthermore, as a polynomial in ω, each monomial has even order; equivalently, $A^{\#}_{2\mu+\nu}(x,-\omega,\xi) = A^{\#}_{2\mu+\nu}(x,\omega,\xi)$.

To prove the lemma, we begin by recasting (13.10). Let $\Gamma_j(t,x,\xi)$ be defined by

$$\Omega_j(t,x,\xi) = \Gamma_j(t,x,\xi)e^{-tL_2(x,\xi)}. \tag{13.13}$$

Then the recursion (13.7) yields

$$\Gamma_j e^{-tL_2} = -\sum_{\ell=1}^{2} B_{2-\ell}(x,\xi,D_x)\big(A_{j-\ell}e^{-tL_2}\big). \tag{13.14}$$

Applying the Leibniz formula gives

$$\Gamma_j = -\sum_{\ell=1}^{2}\sum_{|\gamma|\leq\ell} \Lambda_\ell(x,\omega)B^{[\gamma]}_{2-\ell}(x,\xi,D_x)A_{j-\ell}(t,x,\xi), \tag{13.15}$$

evaluated at $\omega = t^{1/2}\xi$, where

$$e^{tL_2(x,\xi)}D_x^\gamma e^{-tL_2(x,\xi)} = \Lambda_\gamma(x,t^{1/2}\xi). \tag{13.16}$$

Clearly, $\Lambda_\gamma(x, t^{1/2}\xi)$ is a polynomial in ξ and also a polynomial in t; hence $\Lambda_\gamma(x,\omega)$ is an even polynomial in ω. Note also that the differential operator $B^{[\gamma]}_{2-\ell}(x,\xi,D_x)$ is of order $\ell - |\gamma|$, and its coefficients are polynomials in ξ, homogeneous of degree $2-\ell$, as were those of $B_{2-\ell}(x,\xi,D_x)$. The factor A_j is given by

$$A_j(t,x,\xi) = \int_0^t \Gamma_j(s,x,\xi)\,ds. \tag{13.17}$$

The recursion (13.15)–(13.17) will provide an inductive proof of Lemma 13.1.

To carry this out, assume the lemma true for A_j, for all $j < 2\mu+\nu$. We then have

$$\begin{aligned}\Gamma_{2\mu+\nu}(t,x,\xi) = &\sum_{1\le\ell<\nu}\sum_{|\gamma|\le\ell} \Lambda_\ell(x,\omega)B^{[\gamma]}_{2-\ell}(x,\xi,D_x)A^{\#}_{2\mu+\nu-\ell}(x,\omega,\xi)t^{\mu+1}\\ &+\sum_{\nu\le\ell\le 2}\sum_{|\gamma|\le\ell} \Lambda_\ell(x,\omega)B^{[\gamma]}_{2-\ell}(x,\xi,D_x)A^{\#}_{2\mu+\nu-\ell}(x,\omega,\xi)t^{\mu}.\end{aligned} \tag{13.18}$$

The first sum is empty if $\nu = 1$. In the first sum, $A^{\#}_{2\mu+\nu-\ell}(x,\omega,\xi)$ is homogeneous of degree $2+\ell-\nu$ in ξ, so in the first sum

$$t^{\mu+1}\Lambda_\gamma(x,\omega)B^{[\gamma]}_{2-\ell}(x,\xi,D_x)A^{\#}_{2\mu+\nu-\ell}(x,\omega,\xi) = t^{\mu+1}H^{\#}_{\mu\nu\ell\gamma}(x,\omega,\xi), \tag{13.19}$$

where $H^{\#}_{\mu\nu\ell\gamma}(x,\omega,\xi)$ is a polynomial in ξ, homogeneous of degree $4-\nu$, and an even polynomial in ω. We can hence write

$$t^{\mu+1}H^{\#}_{\mu\nu\ell\gamma}(x,\omega,\xi) = t^{\mu}H_{\mu\nu\ell\gamma}(x,\omega,\xi), \tag{13.20}$$

where $H_{\mu\nu\ell\gamma}(x,\omega,\xi)$ is a polynomial in ξ, homogeneous of degree $2-\nu$, and an even polynomial in ω.

In the last sum in (13.18), $A^{\#}_{2\mu+\nu-\ell}$ is homogeneous in ξ of degree $\ell-\nu$, so in this sum

$$t^{\mu}\Lambda_\gamma(x,\omega)B^{[\gamma]}_{2-\ell}(x,\xi,D_x)A^{\#}_{2\mu+\nu-\ell}(x,\omega,\xi) = t^{\mu}H_{\mu\nu\ell\gamma}(x,\omega,\xi), \tag{13.21}$$

where, as in (13.20), $H_{\mu\nu\ell\gamma}(x,\omega,\xi)$ is a polynomial in ξ, homogeneous of degree $2-\nu$, and an even polynomial in ω. Thus

$$\Gamma_{2\mu+\nu}(t,x,\xi) = t^{\mu}\sum_{\ell,\gamma} H_{\mu\nu\ell\gamma}(x,\omega,\xi) = t^{\mu}K_{\mu\nu}(x,\omega,\xi), \tag{13.22}$$

where $K_{\mu\nu}$ is a polynomial in ξ, homogeneous of degree $2-\nu$, and an even polynomial in ω. It follows that

$$A_{2\mu+\nu}(t,x,\xi) = \int_0^t s^\mu K_{\mu\nu}(x, s^{1/2}\xi, \xi)\, ds \tag{13.23}$$

has the properties stated in Lemma 13.1, whose proof is complete.

The analysis of (13.12) yields estimates on $a_j(t,x,\xi)$, easily obtained by writing (for $j = 2\mu+\nu$, $\nu = 1$ or 2)

$$a_j(t,x,\xi) = t^{\mu+1} A_j^\#(x,\omega,\xi) e^{-L_2(x,\omega)/2} e^{-tL_2(x,\xi)/2}, \tag{13.24}$$

and using the simple estimates

$$|\omega|^L e^{-L_2(x,\omega)/2} \le C_L, \quad \bigl(t|\xi|^2\bigr)^\ell e^{-tL_2(x,\xi)/2} \le C_{2\ell}. \tag{13.25}$$

Note that $t^{\mu+1} = t^{j/2}$ if j is even; if j is odd, then $t^{\mu+1} = t^{j/2}\cdot t^{1/2}$, and the factor $t^{1/2}$ can be paired with the linear factor of ξ in $A_j^\#$. Thus we have estimates

$$|a_j(t,x,\xi)| \le C_j t^{j/2} \tag{13.26}$$

and

$$|a_j(t,x,\xi)| \le C_j \langle\xi\rangle^{-j}. \tag{13.27}$$

Derivatives are readily estimated by the same method, and we obtain:

Lemma 13.2. *For $0 \le t \le T$, $k \ge -j$, we have*

$$t^{k/2} D_t^\ell a_j(t,x,\xi) \quad \textit{bounded in } S^{2\ell-k-j}_{1,0}. \tag{13.28}$$

We can construct a function $a(t,x,\xi)$ such that each difference $a(t,x,\xi) - \sum_{\ell<j} a_\ell(t,x,\xi)$ has the properties (13.28), and then, for $u(t,x)$ given by (13.22), we have $u(0,x) = f(x)$ and

$$\Bigl(\frac{\partial}{\partial t} + L\Bigr) u(t,x) = r(t,x), \tag{13.29}$$

where $r(t,x)$ is smooth for $t \ge 0$ and rapidly decreasing as $t \searrow 0$. If the construction is made on a compact manifold M, energy estimates imply that the difference between $u(t,x)$ and $v(t,x) = e^{-tL} f(x)$ is smooth and rapidly decreasing as $t \searrow 0$, for all $f \in \mathcal{D}'(M)$. Consequently the "heat kernel" $H(t,x,y)$, given by

$$e^{-tL} f(x) = \int_M H(t,x,y)\, f(y)\, dV(y), \tag{13.30}$$

and the integral kernel $Q(t,x,y)$ of the operator constructed in the form (13.2) differ by a function $R(t,x,y)$, which is smooth on $[0,\infty)\times M\times M$ and rapidly decreasing as $t\searrow 0$.

Look at the integral kernel of the operator

$$Q_j(t,x,D)f=\int a_j(t,x,\xi)e^{ix\cdot\xi}\hat{f}(\xi)\,d\xi, \tag{13.31}$$

which is

$$Q_j(t,x,y)=(2\pi)^{-n}\int_{\mathbb{R}^n} a_j(t,x,\xi)e^{i(x-y)\cdot\xi}\,d\xi. \tag{13.32}$$

For $a_j(t,x,\xi)$ in the form (13.11)–(13.12), we obtain

$$Q_j(t,x,y)=t^{(j-n)/2}\,q_j\big(x,t^{-1/2}(x-y)\big), \tag{13.33}$$

where

$$q_0(x,z)=(2\pi)^{-n}\int_{\mathbb{R}^n} e^{-L_2(x,\xi)}e^{iz\cdot\xi}\,d\xi \tag{13.34}$$

and, for $j\ge 1$,

$$q_j(x,z)=(2\pi)^{-n}\int_{\mathbb{R}^n} A_j^{\#}(x,\xi,\xi)e^{-L_2(x,\xi)}e^{iz\cdot\xi}\,d\xi. \tag{13.35}$$

We can evaluate the Gaussian integral (13.34) via the method developed in Chap. 3. If, in the local coordinate system used in (13.2), $L_2(x,\xi)=\mathcal{L}(x)\xi\cdot\xi$, for a positive-definite matrix $\mathcal{L}(x)$, then

$$q_0(x,z)=\Big[\det\big(4\pi\mathcal{L}(x)\big)\Big]^{-1/2}\,e^{-\mathcal{G}(x)z\cdot z/4}, \tag{13.36}$$

where $\mathcal{G}(x)=\mathcal{L}(x)^{-1}$. Consequently,

$$Q_0(t,x,y)=(4\pi t)^{-n/2}\Big[\det\mathcal{L}(x)\Big]^{-1/2}e^{-\mathcal{G}(x)(x-y)\cdot(x-y)/4t}. \tag{13.37}$$

The integrals (13.35) can be computed in terms of

$$\begin{aligned}(2\pi)^{-n}\int \xi^\beta e^{-L_2(x,\xi)}e^{iz\cdot\xi}d\xi&=\Big[\det\big(4\pi\mathcal{L}(x)\big)\Big]^{-1/2}D_z^\beta e^{-\mathcal{G}(x)z\cdot z/4}\\&=p_\beta(x,z)\,e^{-\mathcal{G}(x)z\cdot z/4},\end{aligned} \tag{13.38}$$

where $p_\beta(x,z)$ is a polynomial of degree $|\beta|$ in z. Clearly, $p_\beta(x,z)$ is even or odd in z according to the parity of $|\beta|$. Note also that, in (13.35), $A_j^\#(x,\xi,\xi)$ is even or odd in ξ according to the parity of j. We hence obtain the following result.

Proposition 13.3. *If L is a second-order, elliptic differential operator with positive scalar principal symbol, then the integral kernel $H(t,x,y)$ of the operator e^{-tL} has the form*

$$H(t,x,y) \sim \sum_{j\ge 0} t^{(j-n)/2}\ p_j\big(x, t^{-1/2}(x-y)\big)\ e^{-\mathcal{G}(x)(x-y)\cdot(x-y)/4t}, \tag{13.39}$$

where $p_j(x,z)$ is a polynomial in z, which is even or odd in z according to the parity of j.

To be precise about the strong sense in which (13.39) holds, we note that, for any $\nu < \infty$, there is an $N < \infty$ such that the difference $R_N(t,x,y)$ between the left side of (13.39) and the sum over $j \le N$ of the right side belongs to $C^\nu([0,\infty)\times M \times M)$ and vanishes to order ν as $t \searrow 0$.

In particular, we have

$$H(t,x,x) \sim \sum_{j\ge 0} t^{-n/2+j}\ p_{2j}(x,0), \tag{13.40}$$

since $p_j(x,0) = 0$ for j odd. Consequently, the trace of the operator e^{-tL} has the asymptotic expansion

$$\operatorname{Tr} e^{-tL} \sim t^{-n/2}\big(a_0 + a_1 t + a_2 t^2 + \cdots\big), \tag{13.41}$$

with

$$a_j = \int_M p_{2j}(x,0)\ dV(x). \tag{13.42}$$

Further use will be made of this in Chaps. 8 and 10.

Note that the exponent in (13.39) agrees with $r(x,y)^2/4t$, up to $O(r^3/t)$, for x close to y, where $r(x,y)$ is the geodesic distance from x to y. In fact, when $L = -\Delta$, the integral operator with kernel

$$H_0(t,x,y) = (4\pi t)^{-n/2}\ e^{-r(x,y)^2/4t}, \quad t > 0, \tag{13.43}$$

is in some ways a better first approximation to e^{-tL} than is (13.2) with $a(t,x,\xi)$ replaced by $a_0(t,x,\xi) = e^{-tL_2(x,\xi)}$. (See Exercise 3 below.) It can be shown that

$$\Big(\frac{\partial}{\partial t} + L_x\Big) H_0(t,x,y) = Q(t,x,y), \quad t > 0, \tag{13.44}$$

is the integral kernel of an operator that is regularizing, and if one defines

$$H_0\#Q(t,x,y) = \int_0^t \int\limits_M H_0(t-s,x,z)Q(s,z,y)\,dV(z)\,ds, \tag{13.45}$$

then a parametrix that is as good as (13.39) can be obtained in the form

$$\sim H_0 - H_0\#Q + H_0\#Q\#Q - \cdots. \tag{13.46}$$

This approach, one of several alternatives to that used above, is taken in [MS].

One can also look at (13.43)–(13.46) from a pseudodifferential operator perspective, as done in [Gr]. The symbol of $\partial/\partial t + L$ is $i\tau + L(x,\xi)$, and

$$H_0(x,\tau,\xi) = \Big(i\tau + L_2(x,\xi)\Big)^{-1} \in S^{-1}_{1/2,0}(\mathbb{R}\times M). \tag{13.47}$$

The operator with integral kernel $H_0(t-s,x,y)$ given by (13.43) belongs to $OPS^{-1}_{1/2,0}(\mathbb{R}\times M)$ and has (13.47) as its principal symbol. This operator has two additional properties; it is causal, that is, if v vanishes for $t<T$, so does $H_0 v$, for any T, and it commutes with translations. Denote by $\mathcal{C}^m$ the class of operators in $OPS^m_{1/2,0}(\mathbb{R}\times M)$ with these two properties. One easily has $P_j \in \mathcal{C}^{m_j} \Rightarrow P_1P_2 \in \mathcal{C}^{m_1+m_2}$. The symbol computation gives

$$\Big(\frac{\partial}{\partial t} + L\Big)H_0 = I + Q, \quad Q \in \mathcal{C}^{-1}, \tag{13.48}$$

and from here one obtains a parametrix

$$H \in \mathcal{C}^{-1}, \quad H \sim H_0 - H_0Q + H_0Q^2 - \cdots. \tag{13.49}$$

The formulas (13.46) and (13.49) agree, via the correspondence of operators and their integral kernels.

One can proceed to construct a parametrix for the heat equation on a manifold with boundary. We sketch an approach, using a variant of the double-layer-potential method described for elliptic boundary problems in §11. Let Ω be an open domain, with smooth boundary, in M, a compact Riemannian manifold without boundary. We construct an approximate solution to

$$\frac{\partial u}{\partial t} = -Lu, \tag{13.50}$$

for $(t,x) \in \mathbb{R}^+ \times \Omega$, satisfying

$$u(0,x) = 0, \quad u(t,x) = h(t,x), \ \text{ for } x \in \partial\Omega, \tag{13.51}$$

in the form

$$u = \mathcal{D}\ell\, g(t,x) = \int_0^\infty \int_{\partial\Omega} g(s,y)\, \frac{\partial H}{\partial \nu_y}(t-s,x,y)\, dS(y)\, ds, \tag{13.52}$$

where $H(t,x,y)$ is the heat kernel on $\mathbb{R}^+ \times M$ studied above. For $x \in \partial\Omega$, denote by $\mathcal{D}\ell\, g_+(t,x)$ the limit of $\mathcal{D}\ell\, g$ from within $\mathbb{R}^+ \times \Omega$. As in (11.7), one can establish the identity

$$\mathcal{D}\ell\, g_+ = \frac{1}{2}(I+N)g, \tag{13.53}$$

where $(1/2)Ng$ is given by the double integral on the right side of (13.52), with y and x both in $\partial\Omega$. In analogy with (11.23), we have

$$N \in OPS^{-1/2}_{1/2,0}(\mathbb{R}^+ \times \partial\Omega).$$

For u to solve (13.50)–(13.51), we need

$$h = \frac{1}{2}(I+N)g. \tag{13.54}$$

Thus we have a parametrix for (13.50)–(13.51) in the form (13.52) with

$$g \sim 2(I - N + N^2 - \cdots)h. \tag{13.55}$$

We can use the analysis of (13.50)–(13.55) to construct a parametrix for the solution operator to

$$\frac{\partial u}{\partial t} = \Delta u, \text{ for } x \in \Omega, \quad u(0,x) = f(x), \quad u(t,x) = 0, \text{ for } x \in \partial\Omega. \tag{13.56}$$

To begin, let v solve

$$\frac{\partial v}{\partial t} = \Delta v \text{ on } \mathbb{R}^+ \times M, \quad v(0) = \widetilde{f}, \tag{13.57}$$

where

$$\begin{aligned} \widetilde{f}(x) &= f(x), \quad \text{for } x \in \Omega, \\ &\quad\ 0, \qquad \text{for } x \in M \setminus \Omega. \end{aligned} \tag{13.58}$$

One way to obtain u would be to subtract a solution to (13.50)–(13.51), with $-L = \Delta$, $h = v\big|_{\mathbb{R}^+\times\partial\Omega}$. This leads to a parametrix for the solution operator for (13.56) of the form

$$(13.59)\quad \begin{aligned} p(t,x,y) &= H(t,x,y) - \int_0^\infty \int_{\partial\Omega} h(s,z,y) \frac{\partial H}{\partial \nu_z}(t-s,x,z)\, dS(z)\, ds, \\ h(s,z,y) &\sim 2H(s,z,y) + \cdots, \end{aligned}$$

where, as above, $H(t,x,y)$ is the heat kernel on $\mathbb{R}^+ \times M$.

We mention an alternative treatment of (13.56) that has some advantages. We will apply a reflection to v. To do this, assume that $\overline{\Omega}$ is contained in a compact Riemannian manifold M, diffeomorphic to the double of $\overline{\Omega}$, and let $R : M \to M$ be a smooth involution of M, fixing $\partial\Omega$, which near $\partial\Omega$ is a reflection of each geodesic normal to $\partial\Omega$, about the point where the geodesic intersects $\partial\Omega$. Pulling back the metric tensor on M by R yields a metric tensor that agrees with the original on $\partial\Omega$. Now set

$$(13.60)\qquad u_1(t,x) = v(t,x) - v\big(t, R(x)\big), \quad x \in \Omega.$$

We see that u_1 satisfies

$$(13.61)\qquad \frac{\partial u_1}{\partial t} = \Delta u_1 + g, \quad u_1(0,x) = f, \quad u_1(t,x) = 0, \text{ for } x \in \partial\Omega,$$

where

$$(13.62)\qquad g = L^b \widetilde{v}\big|_{\mathbb{R}^+\times\Omega}, \quad \widetilde{v}(t,x) = v\big(t, R(x)\big),$$

and where L^b is a second-order differential operator, with smooth coefficients, whose principal symbol vanishes on $\partial\Omega$. Thus the difference $u - u_1 = w$ solves

$$(13.63)\qquad \frac{\partial w}{\partial t} = \Delta w - g, \quad w(0) = 0, \quad w(t,x) = 0, \text{ for } x \in \partial\Omega.$$

Next let v_2 solve

$$(13.64)\qquad \frac{\partial v_2}{\partial t} = \Delta v_2 - \widetilde{g} \text{ on } \mathbb{R}^+ \times M, \quad v_2(0) = 0,$$

where

$$(13.65)\qquad \begin{aligned} \widetilde{g}(t,x) = g(t,x), &\quad \text{for } x \in \Omega, \\ 0, &\quad \text{for } x \in M \setminus \Omega, \end{aligned}$$

and set

$$u_2 = v_2\big|_{\mathbb{R}^+\times\Omega}. \tag{13.66}$$

It follows that $w_2 = u - (u_1 + u_2)$ satisfies

$$\frac{\partial w_2}{\partial t} = \Delta w_2 \text{ on } \mathbb{R}^+ \times \Omega, \quad w_2(0) = 0, \quad w_2\big|_{\mathbb{R}^+\times\partial\Omega} = -v_2\big|_{\mathbb{R}^+\times\partial\Omega}. \tag{13.67}$$

Now we can obtain w_2 by the construction (13.52)–(13.55), with

$$h = -v_2\big|_{\mathbb{R}^+\times\partial\Omega}.$$

To illustrate the effect of this construction using reflection, suppose that, in (13.56),

$$f \in H_0^1(\Omega). \tag{13.68}$$

Then, in (13.57)–(13.58), $\widetilde{f} \in H^1(M)$, so $v \in C\big(\mathbb{R}^+, H^1(M)\big)$, and hence

$$u_1 \in C\big(\mathbb{R}^+, H_0^1(\Omega)\big). \tag{13.69}$$

Furthermore, given the nature of L^b and that of the heat kernel on $\mathbb{R}^+ \times M \times M$, one can show that, in (13.62),

$$g \in C\big(\mathbb{R}^+, L^2(\Omega)\big), \tag{13.70}$$

that is, L^b effectively acts like a first-order operator on $\widetilde{v}$, when one restricts to Ω. It follows that $\widetilde{g} \in C\big(\mathbb{R}^+, L^2(M)\big)$ and hence, via Duhamel's formula for the solution to (13.64), that $v_2 \in C\big(\mathbb{R}^+, H^{2-\epsilon}(M)\big)$, $\forall\, \epsilon > 0$. Therefore,

$$u_2 \in C\big(\mathbb{R}^+, H^{2-\epsilon}(\Omega)), \tag{13.71}$$

and, in (13.67), we have a PDE of the form (13.50)–(13.51), with $h \in C\big(\mathbb{R}^+, H^{3/2-\epsilon}(\partial\Omega)\big)$, for all $\epsilon > 0$. One can deduce from (13.52)–(13.55) that w_2 has as much regularity as that given for u_2 in (13.71).

It also follows directly from Duhamel's principle, applied to (13.63), that

$$w \in C\big(\mathbb{R}^+, H^{2-\epsilon}(\Omega)\big), \tag{13.72}$$

so we can see without analyzing (13.52)–(13.55) that w_2 has as much regularity as mentioned above. Either way, we see that when f satisfies (13.68), the principal singularities of the solution u to (13.56) are captured by u_1, defined by (13.60). Constructions of u_2 and, via (13.52)–(13.55), of w_2 yield smoother corrections, at least when smoothness is measured in the spaces used above.

The construction (13.56)–(13.67) can be compared with constructions in §7 of Chap. 13.

Exercises

1. Let L be a positive, self-adjoint, elliptic differential operator of order $2k > 0$ on a compact manifold M, with scalar principal symbol $L_{2k}(x,\xi)$. Show that a parametrix for $\partial u/\partial t = -Lu$ can be constructed in the form (13.2)–(13.3), with $a_j(t,x,\xi)$ of the following form, generalizing (13.11)–(13.12):

$$a_j(t,x,\xi) = A_j(t,x,\xi)e^{-tL_{2k}(x,\xi)},$$

where $A_0(t,x,\xi) = 1$ and if $\mu = 0,1,2,\dots$ and $\nu \in \{1,\dots,2k\}$, then

$$A_{2k\mu+\nu}(t,x,\xi) = t^{\mu+1}A^{\#}_{2k\mu+\nu}(x,\omega,\xi), \quad \omega = t^{1/2k}\xi,$$

where $A^{\#}_{2k\mu+\nu}(x,\omega,\xi)$ is a polynomial in ξ, homogeneous of degree $2k-\nu$, whose coefficients are polynomials in ω, each monomial of which has degree (in ω) that is an integral multiple of $2k$, so $A^{\#}_{2k\mu+\nu}(x,e^{\pi i/k}\omega,\xi) = A^{\#}_{2k\mu+\nu}(x,\omega,\xi)$.

2. In the setting of Exercise 1, show that

$$\operatorname{Tr} e^{-tL} \sim t^{-n/2k}\Big(a_0 + a_1t^{1/k} + a_2t^{2/k} + \cdots\Big),$$

generalizing (13.41).

3. Let $g_{jk}(y,x)$ denote the components of the metric tensor at x in a normal coordinate system centered at y. Suppose $-Lu(x) = \Delta u(x) = g^{jk}(y,x)\,\partial_j\partial_k u(x) + b^j(y,x)\,\partial_j u(x)$ in this coordinate system. With $H_0(t,x,y)$ given by (13.43), show that

$$\begin{aligned}\Big(\frac{\partial}{\partial t} + L_x\Big)&H_0(t,x,y)\\ &= H_0(t,x,y)\Big\{(2t)^{-2}\big[g^{jk}(x,x) - g^{jk}(y,x)\big](x_j-y_j)(x_k-y_k)\\ &\qquad - (2t)^{-1}\big[g^j{}_j(x,x) - g^j{}_j(y,x) - b^j(y,x)(x_j-y_j)\big]\Big\}\\ &= H_0(t,x,y)\left\{O\left(\frac{|x-y|^4}{t^2}\right) + O\left(\frac{|x-y|^2}{t}\right)\right\}.\end{aligned}$$

Compare formula (2.10) in Chap. 5. Note that $g_{jk}(y,y) = \delta_{jk}$, $\partial_\ell g_{jk}(y,y) = 0$, and $b^j(y,y) = 0$. Relate this calculation to the discussion involving (13.43)–(13.49).

4. Using the parametrix, especially (13.39), show that if M is a smooth, compact Riemannian manifold, without boundary, then

$$e^{t\Delta} : C^k(M) \longrightarrow C^k(M)$$

is a strongly continuous semigroup, for each $k \in \mathbb{Z}^+$.

14. The Weyl calculus

To define the Weyl calculus, we begin with a modification of the formula (1.10) for $a(x,D)$. Namely, we replace $e^{iq\cdot X}e^{ip\cdot D}$ by $e^{i(q\cdot X+p\cdot D)}$, and set

$$a(X,D)u = \int \hat{a}(q,p)e^{i(q\cdot X+p\cdot D)}\,u\,dq\,dp, \tag{14.1}$$

initially for $a(x,\xi) \in \mathcal{S}(\mathbb{R}^{2n})$. Note that $v(t,x) = e^{it(q\cdot X+p\cdot D)}u(x)$ solves the PDE

$$\frac{\partial v}{\partial t} = \sum_j p_j \frac{\partial v}{\partial x_j} + i(q\cdot x)v, \quad v(0,x) = u(x), \tag{14.2}$$

and the solution is readily obtained by integrating along the integral curves of $\partial/\partial t - \sum p_j\, \partial/\partial x_j$, which are straight lines. We get

$$e^{i(q\cdot X+p\cdot D)}u(x) = e^{iq\cdot x+iq\cdot p/2}\, u(x+p). \tag{14.3}$$

Note that this is equivalent to the identity

$$e^{i(q\cdot X+p\cdot D)} = e^{iq\cdot p/2}\, e^{iq\cdot X}\, e^{ip\cdot D}. \tag{14.4}$$

If we plug (14.3) into (14.1), a few manipulations using the Fourier inversion formula yield

$$a(X,D)u(x) = (2\pi)^{-n} \int a\Bigl(\frac{x+y}{2},\xi\Bigr) e^{i(x-y)\cdot\xi}\, u(y)\, dy\, d\xi, \tag{14.5}$$

which can be compared with the formula (1.3) for $a(x,D)$. Note that $a(X,D)$ is of the form (3.2) with $a(x,y,\xi) = a\bigl((x+y)/2,\xi\bigr)$, while $a(x,D)$ is of the form (3.2) with $a(x,y,\xi) = a(x,\xi)$. In particular, Proposition 3.1 is applicable; we have

$$a(X,D) = b(x,D), \tag{14.6}$$

where

$$b(x,\xi) = e^{iD_\xi\cdot D_y}\, a\Bigl(\frac{x+y}{2},\xi\Bigr)\Big|_{y=x} = e^{(i/2)D_\xi\cdot D_x}\, a(x,\xi). \tag{14.7}$$

If $a(x,\xi) \in S^m_{\rho,\delta}$, with $0 \le \delta < \rho \le 1$, then $b(x,\xi)$ also belongs to $S^m_{\rho,\delta}$ and, by (3.6),

$$b(x,\xi) \sim \sum_{\alpha\ge 0} \frac{i^{|\alpha|}}{\alpha!} 2^{-|\alpha|}\, D^\alpha_\xi D^\alpha_x a(x,\xi). \tag{14.8}$$

Of course this relation is invertible; we have $a(x,\xi) = e^{-(i/2)D_\xi\cdot D_x} b(x,\xi)$ and a corresponding asymptotic expansion. Thus, at least on a basic level, the two methods of assigning an operator, either $a(x,D)$ or $a(X,D)$, to a symbol $a(x,\xi)$ lead to equivalent operator calculi. However, they are not identical, and the differences sometimes lead to subtle advantages for the Weyl calculus.

One difference is that since the adjoint of $e^{i(q\cdot X+p\cdot D)}$ is $e^{-i(q\cdot X+p\cdot D)}$, we have the formula

$$a(X,D)^* = b(X,D), \quad b(x,\xi) = a(x,\xi)^*, \tag{14.9}$$

which is somewhat simpler than the formula (3.13)–(3.14) for $a(x,D)^*$.

Other differences can be traced to the fact that the Weyl calculus exhibits certain symmetries rather clearly. To explain this, we recall, from the exercises after §1, that the set of operators

$$e^{it}\, e^{iq\cdot X}\, e^{ip\cdot D} = \tilde{\pi}(t,q,p) \tag{14.10}$$

form a unitary group of operators on $L^2(\mathbb{R}^n)$, a representation of the group $\mathcal{H}^n$, with group law

$$(t,q,p)\circ(t',q',p') = (t+t'+p\cdot q', q+q', p+p'). \tag{14.11}$$

Now, using (14.4), one easily computes that

$$e^{i(t+q\cdot X+p\cdot D)}\, e^{i(t'+q'\cdot X+p'\cdot D)} = e^{i(s+u\cdot X+v\cdot D)}, \tag{14.12}$$

with $u = q+q'$, $v = p+p'$, and

$$s = t+t'+\frac{1}{2}(p\cdot q' - q\cdot p') = t+t'+\frac{1}{2}\sigma\big((p,q),(p',q')\big), \tag{14.13}$$

where σ is the natural symplectic form on $\mathbb{R}^n\times\mathbb{R}^n$. Thus

$$\pi(t,q,p) = e^{i(t+q\cdot X+p\cdot D)} \tag{14.14}$$

defines a unitary representation of a group we'll denote $\mathbf{H}^n$, which is $\mathbb{R}\times\mathbb{R}^{2n}$ with group law

$$(t,w)\cdot(t',w') = \Big(t+t'+\frac{1}{2}\sigma(w,w'), w+w'\Big), \tag{14.15}$$

where we have set $w = (q,p)$. Of course, the groups $\mathcal{H}^n$ and $\mathbf{H}^n$ are isomorphic; both are called the *Heisenberg group*. The advantage of using the group law (14.15) rather than (14.11) is that it makes transparent the existence of the action of the group $\mathrm{Sp}(n,\mathbb{R})$ of linear symplectic maps on $\mathbb{R}^{2n}$, as a group of automorphisms of $\mathbf{H}^n$. Namely, if $g:\mathbb{R}^{2n}\to\mathbb{R}^{2n}$ is a linear map preserving the symplectic form, so $\sigma(gw,gv) = \sigma(w,v)$ for $v,w\in\mathbb{R}^{2n}$, then

$$\alpha(g):\mathbf{H}^n\to\mathbf{H}^n, \quad \alpha(g)(t,w) = (t,gw) \tag{14.16}$$

defines an automorphism of $\mathbf{H}^n$, so

$$(t, w) \cdot (t', w') = (s, v) \Rightarrow (t, gw) \cdot (t', gw') = (s, gv) \tag{14.17}$$

and $\alpha(gg') = \alpha(g)\alpha(g')$. The associated action of $\mathrm{Sp}(n, \mathbb{R})$ on $\mathcal{H}^n$ has a formula that is less clean.

This leads to an action of $\mathrm{Sp}(n, \mathbb{R})$ on operators in the Weyl calculus. Setting

$$a_g(x, \xi) = a\big(g^{-1}(x, \xi)\big), \tag{14.18}$$

we have

$$a(X, D)b(X, D) = c(X, D) \Rightarrow a_g(X, D)b_g(X, D) = c_g(X, D), \tag{14.19}$$

for $g \in \mathrm{Sp}(n, \mathbb{R})$.

In fact, let us rewrite (14.1) as

$$a(X, D) = \int \hat{a}(w)\pi(0, w)\, dw.$$

Then

$$\begin{aligned} &a(X, D)b(X, D) \\ &= \iint \hat{a}(w)\hat{b}(w')\pi(0, w)\pi(0, w')\, dw\, dw' \\ &= \iint \hat{a}(w)\hat{b}(w')e^{\sigma(w,w')/2}\pi(0, w + w')\, dw\, dw', \end{aligned} \tag{14.20}$$

so $c(X, D)$ in (14.19) has symbol satisfying

$$\hat{c}(w) = (2\pi)^{-n} \int \hat{a}(w - w')\hat{b}(w')e^{i\sigma(w,w')/2}\, dw'. \tag{14.21}$$

The implication in (14.19) follows immediately from this formula. Let us write $c(x, \xi) = (a \circ b)(x, \xi)$ when this relation holds.

From (14.21), one easily obtains the product formula

$$(a \circ b)(x, \xi) = e^{(i/2)(D_y \cdot D_\xi - D_x \cdot D_\eta)} a(x, \xi)b(y, \eta)\Big|_{y=x,\eta=\xi}. \tag{14.22}$$

If $a \in S^m_{\rho,\delta}$, $b \in S^\mu_{\rho,\delta}$, $0 \le \delta < \rho \le 1$, we have the following asymptotic expansion:

$$(a \circ b)(x, \xi) \sim ab + \sum_{j \ge 1} \frac{1}{j!}\{a, b\}_j(x, \xi), \tag{14.23}$$

where

$$\{a,b\}_j(x,\xi) = \Bigl(-\frac{i}{2}\Bigr)^j \bigl(\partial_y\cdot\partial_\xi - \partial_x\cdot\partial_\eta\bigr)^j a(x,\xi)b(y,\eta)\Big|_{y=x,\eta=\xi}. \tag{14.24}$$

For comparison, recall the formula for

$$a(x,D)b(x,D) = (a\#b)(x,D) \tag{14.25}$$

given by (3.16)–(3.20):

$$\begin{aligned}(a\#b)(x,\xi) &= e^{iD_\eta\cdot D_y} a(x,\eta)b(y,\xi)\Big|_{y=x,\eta=\xi} \\ &\sim ab + \sum_{\alpha>0} \frac{(-i)^{|\alpha|}}{\alpha!} \partial_\xi^\alpha a(x,\xi)\partial_x^\alpha b(x,\xi).\end{aligned} \tag{14.26}$$

In the respective cases, $(a\circ b)(x,\xi)$ differs from the sum over $j < N$ by an element of $S^{m+\mu-N(\rho-\delta)}_{\rho,\delta}$ and $(a\#b)(x,\xi)$ differs from the sum over $|\alpha| < N$ by an element of the same symbol class.

In particular, for $\rho = 1$, $\delta = 0$, we have

$$(a\circ b)(x,\xi) = a(x,\xi)b(x,\xi) + \frac{i}{2}\{a,b\}(x,\xi) \bmod S^{m+\mu-2}_{1,0}, \tag{14.27}$$

where $\{a,b\}$ is the Poisson bracket, while

$$(a\#b)(x,\xi) = a(x,\xi)b(x,\xi) - i\sum \frac{\partial a}{\partial\xi_j}\frac{\partial b}{\partial x_j} \bmod S^{m+\mu-2}_{1,0}. \tag{14.28}$$

Consequently, in the scalar case,

$$\begin{aligned}[a(X,D),b(X,D)] &= [a(x,D),b(x,D)] \\ &= e(x,D) = e(X,D) \bmod OPS^{m+\mu-2}_{1,0},\end{aligned} \tag{14.29}$$

with

$$e(x,\xi) = i\{a,b\}(x,\xi). \tag{14.30}$$

Now we point out one of the most useful aspects of the difference between (14.27) and (14.28). Namely, one starts with an operator $A = a(X,D) = a_1(x,D)$, maybe a differential operator, and perhaps one wants to construct a parametrix for A, or perhaps a "heat semigroup" e^{-tA}, under appropriate hypotheses. In such a case, the leading term in the symbol of the operator $b(X,D) = b_1(x,D)$ used in (14.20) or (14.25) is a *function* of $a(x,\xi)$, for example, $a(x,\xi)^{-1}$, or $e^{-ta(x,\xi)}$. But then, at least when $a(x,\xi)$ is scalar, the last term in (14.27) vanishes! On the other hand, the last term in (14.28) generally does not vanish. From this it follows that, with a given amount of work, one can often construct a more accurate approximation to a parametrix using the Weyl calculus, instead of using the constructions of the previous sections.

In the remainder of this section, we illustrate this point by reconsidering the parametrix construction for the heat equation, made in §13. Thus, we look again at

$$\frac{\partial u}{\partial t} = -Lu, \quad u(0) = f. \tag{14.31}$$

This time, set

$$Lu = a(X, D)u + b(x)u, \tag{14.32}$$

where

$$\begin{aligned} a(x, \xi) &= \sum g^{jk}(x)\xi_j \xi_k + \sum \ell_j(x)\xi_j \\ &= g(x, \xi) + \ell(x, \xi). \end{aligned} \tag{14.33}$$

We assume $g(x, \xi)$ is scalar, while $\ell(x, \xi)$ and $b(x)$ can be $K \times K$ matrix-valued. As the notation indicates, we assume (g^{jk}) is positive-definite, defining an inner product on cotangent vectors, corresponding to a Riemannian metric (g_{jk}). We note that a symbol that is a polynomial in ξ also defines a differential operator in the Weyl calculus. For example,

$$\begin{aligned} \ell(x, D)u &= \sum \ell_j(x)\, \partial_j u \Longrightarrow \\ \ell(X, D)u &= \sum \ell_j(x)\, \partial_j u + \frac{1}{2} \sum (\partial_j \ell_j) u \end{aligned} \tag{14.34}$$

and

$$\begin{aligned} a(x, D) &= \sum a_{jk}(x) \partial_j \partial_k u \Longrightarrow \\ a(X, D)u &= \sum \Big[a_{jk}(x)\partial_j\partial_k u + (\partial_j a_{jk})\partial_k u + \frac{1}{4}(\partial_j\partial_k a_{jk})u \Big] \\ &= \sum \Big[\partial_j (a_{jk}\partial_k u) + \frac{1}{4}(\partial_j \partial_k a_{jk})u \Big]. \end{aligned} \tag{14.35}$$

We use the Weyl calculus to construct a parametrix for (14.31). We will begin by treating the case when all the terms in (14.33) are scalar, and then we will discuss the case when only $g(x, \xi)$ is assumed to be scalar.

We want to write an approximate solution to (14.31) as

$$u = E(t, X, D) f. \tag{14.36}$$

We write

$$E(t,x,\xi) \sim E_0(t,x,\xi) + E_1(t,x,\xi) + \cdots \tag{14.37}$$

and obtain the various terms recursively. The PDE (14.31) requires

$$\frac{\partial}{\partial t} E(t,X,D) = -LE(t,X,D) = -(L \circ E)(t,X,D), \tag{14.38}$$

where, by the Weyl calculus,

$$(L \circ E)(t,x,\xi) \sim L(x,\xi)E(t,x,\xi) + \sum_{j \geq 1} \frac{1}{j!}\{L, E\}_j(t,x,\xi). \tag{14.39}$$

It is natural to set

$$E_0(t,x,\xi) = e^{-ta(x,\xi)}, \tag{14.40}$$

as in (13.9). Note that the Weyl calculus applied to this term provides a better approximation than the previous calculus, because

$$\{a, e^{-ta}\}_1 = 0. \tag{14.41}$$

If we plug (14.37) into (14.39) and collect the highest order nonvanishing terms, we are led to define $E_1(t,x,\xi)$ as the solution to the "transport equation"

$$\frac{\partial E_1}{\partial t} = -aE_1 - \frac{1}{2}\{a, E_0\}_2 - b(x)E_0, \quad E_1(0,x,\xi) = 0. \tag{14.42}$$

Let us set

$$\Omega_1(t,x,\xi) = -\frac{1}{2}\{a, e^{-ta}\}_2 - b(x)e^{-ta(x,\xi)}. \tag{14.43}$$

Then the solution to (14.42) is

$$E_1(t,x,\xi) = \int_0^t e^{(s-t)a(x,\xi)}\Omega_1(s,x,\xi)\, ds. \tag{14.44}$$

Higher terms $E_j(t,x,\xi)$ are then obtained in a straightforward fashion. This construction is similar to (13.6)–(13.10), but there is the following important difference. Once you have $E_0(t,x,\xi)$ and $E_1(t,x,\xi)$ here, you have the first two terms in the expansion of the integral kernel of e^{-tL} on the diagonal:

$$K(t,x,x) \sim c_0(x)t^{-n/2} + c_1(x)t^{-n/2+1} + \cdots. \tag{14.45}$$

To get so far using the method of §13, it is necessary to go further and compute the solution $a_2(t,x,\xi)$ to the next transport equation. Since the formulas become rapidly more complicated, the advantage is with the method of this section. We proceed with an explicit determination of the first two terms in (14.45).

Thus we now evaluate the integral in (14.44). Clearly,

$$\int_0^t e^{(s-t)a(x,\xi)}b(x)e^{-sa(x,\xi)}\,ds = tb(x)e^{-ta(x,\xi)}. \tag{14.46}$$

Now, a straightforward calculation yields

$$\{a, e^{-sa}\}_2 = \frac{s}{2}Q(\nabla^2 a)e^{-sa} - \frac{s^2}{4}T(\nabla a, \nabla^2 a)e^{-sa}, \tag{14.47}$$

where

$$Q(\nabla^2 a) = \sum_{k,\ell}\Big\{(\partial_{\xi_k}\partial_{\xi_\ell}a)(\partial_{x_k}\partial_{x_\ell}a) - (\partial_{\xi_k}\partial_{x_\ell}a)(\partial_{x_k}\partial_{\xi_\ell}a)\Big\} \tag{14.48}$$

and

$$\begin{aligned} &T(\nabla a, \nabla^2 a) \\ &\quad = \sum_{k,\ell}\Big\{(\partial_{\xi_k}\partial_{\xi_\ell}a)(\partial_{x_k}a)(\partial_{x_\ell}a) \\ &\qquad + (\partial_{x_k}\partial_{x_\ell}a)(\partial_{\xi_k}a)(\partial_{\xi_\ell}a) - 2(\partial_{\xi_k}\partial_{x_\ell}a)(\partial_{x_k}a)(\partial_{\xi_\ell}a)\Big\}. \end{aligned} \tag{14.49}$$

Therefore,

$$\int_0^t e^{(s-t)a}\{a, e^{-sa}\}_2\,ds = \frac{t^2}{4}Q(\nabla^2 a)e^{-ta} - \frac{t^3}{12}T(\nabla a, \nabla^2 a)e^{-ta}. \tag{14.50}$$

We get $E_1(t,x,\xi)$ in (14.44) from (14.46) and (14.50).

Suppose for the moment that $\ell(x,\xi) = 0$ in (14.33), that is, $a(X,D) = g(X,D)$. Suppose also that, for some point x_0,

$$\nabla_x\, g^{jk}(x_0) = 0, \quad g^{jk}(x_0) = \delta_{jk}. \tag{14.51}$$

Then, at x_0,

$$\begin{aligned} Q(\nabla^2 a) &= \sum_{k,\ell}\big(\partial_{\xi_k}\partial_{\xi_\ell}a\big)\big(\partial_{x_k}\partial_{x_\ell}a\big) \\ &= 2\sum_{j,k,\ell}\frac{\partial^2 g^{jk}}{\partial x_\ell^2}(x_0)\xi_j\xi_k \end{aligned} \tag{14.52}$$

and

$$
\begin{aligned}
T(\nabla a, \nabla^2 a) &= \sum_{k,\ell} \big(\partial_{x_k}\partial_{x_\ell} a\big)\big(\partial_{\xi_k} a\big)\big(\partial_{\xi_\ell} a\big) \\
&= 4 \sum_{j,k,\ell,m} \frac{\partial^2 g^{jk}}{\partial x_\ell \partial x_m}(x_0)\xi_j\xi_k\xi_\ell\xi_m.
\end{aligned} \tag{14.53}
$$

Such a situation as (14.51) arises if $g^{jk}(x)$ comes from a metric tensor $g_{jk}(x)$, and one uses geodesic normal coordinates centered at x_0. Now the Laplace-Beltrami operator is given by

$$
\Delta u = g^{-1/2} \sum \partial_j g^{jk} g^{1/2} \partial_k\, u, \tag{14.54}
$$

where $g = \det(g_{jk})$. This is symmetric when one uses the Riemannian volume element $dV = \sqrt{g}\, dx_1 \cdots dx_n$. To use the Weyl calculus, we want an operator that is symmetric with respect to the Euclidean volume element $dx_1 \cdots dx_n$, so we conjugate Δ by multiplication by $g^{1/4}$:

$$
-Lu = g^{1/4}\, \Delta\big(g^{-1/4} u\big) = g^{-1/4} \sum \partial_j g^{jk} g^{1/2}\, \partial_k \big(g^{-1/4} u\big). \tag{14.55}
$$

Note that the integral kernel $k_L^t(x, y)$ of e^{tL} is $g^{1/4}(x) k_\Delta^t(x, y) g^{-1/4}(y)$; in particular, of course, the two kernels coincide on the diagonal $x = y$. To compare L with $g(X, D)$, note that

$$
-Lu = \sum \partial_j g^{jk}\, \partial_k u + \Phi(x) u, \tag{14.56}
$$

where

$$
\Phi(x) = \sum \partial_j \big(g^{jk} g^{1/2}\, \partial_k g^{-1/4}\big) - \sum g^{jk} g^{1/2} \big(\partial_j g^{-1/4}\big)\big(\partial_k g^{-1/4}\big). \tag{14.57}
$$

If $g^{jk}(x)$ satisfies (14.51), we see that

$$
\Phi(x_0) = \sum_j \partial_j^2 g^{-1/4}(x_0) = -\frac{1}{4} \sum_\ell \partial_\ell^2 g(x_0). \tag{14.58}
$$

Since $g(x_0 + he_\ell) = \det\big(\delta_{jk} + (1/2)h^2\, \partial_\ell^2 g_{jk}\big) + O(h^3)$, we have

$$
\Phi(x_0) = -\frac{1}{4} \sum_{j,\ell} \partial_\ell^2 g_{jj}(x_0). \tag{14.59}
$$

By comparison, note that, by (14.35),

$$
\begin{aligned}
g(X,D)u &= -\sum \partial_j g^{jk}\,\partial_k u + \Psi(x)u, \\
\Psi(x) &= -\frac{1}{4}\sum \partial_j\partial_k g^{jk}(x).
\end{aligned}
\tag{14.60}
$$

If x_0 is the center of a normal coordinate system, we can express these results in terms of curvature, using

$$
\partial_\ell\partial_m g^{jk}(x_0) = \frac{1}{3}R_{j\ell km}(x_0) + \frac{1}{3}R_{jmk\ell}(x_0),
\tag{14.61}
$$

in terms of the components of the Riemann curvature tensor, which follows from formula (3.51) of Appendix C. Thus we get

$$
\begin{aligned}
\Phi(x_0) &= -\frac{1}{4}\cdot\frac{2}{3}\sum_{j,\ell} R_{j\ell j\ell}(x_0) = -\frac{1}{6}S(x_0), \\
\Psi(x_0) &= -\frac{1}{4}\cdot\frac{1}{3}\sum_{j,k}\big[R_{jjkk}(x_0) + R_{jkkj}(x_0)\big] = \frac{1}{12}S(x_0).
\end{aligned}
\tag{14.62}
$$

Here S is the scalar curvature of the metric g_{jk}.

When $a(X,D) = g(X,D)$, we can express the quantities (14.52) and (14.53) in terms of curvature:

$$
Q(\nabla^2 g) = 2\cdot\frac{2}{3}\sum_{j,k,\ell} R_{j\ell k\ell}(x_0)\xi_j\xi_k = \frac{4}{3}\sum_{j,k}\mathrm{Ric}_{jk}(x_0)\xi_j\xi_k,
\tag{14.63}
$$

where Ric_{jk} denotes the components of the Ricci tensor, and

$$
T(\nabla g,\nabla^2 g) = 4\cdot\frac{2}{3}\sum_{j,k,\ell,m} R_{j\ell km}(x_0)\xi_j\xi_k\xi_\ell\xi_m = 0,
\tag{14.64}
$$

the cancelation here resulting from the antisymmetry of $R_{j\ell km}$ in (j,ℓ) and in (k,m).

Thus the heat kernel for (14.31) with

$$
Lu = g(X,D)u + b(x)u
\tag{14.65}
$$

is of the form (14.36)–(14.37), with $E_0 = e^{-tg(x,\xi)}$ and

$$
\begin{aligned}
E_1(t,x,\xi) &= \Big(-tb(x) - \frac{t^2}{8}Q(\nabla^2 g) + \frac{t^3}{24}T(\nabla g,\nabla^2 g)\Big)e^{-tg} \\
&= -\Big(tb(x) + \frac{t^2}{6}\,\mathrm{Ric}(\xi,\xi)\Big)e^{-tg(x,\xi)},
\end{aligned}
\tag{14.66}
$$

at $x = x_0$. Note that $g(x_0,\xi) = |\xi|^2$.

Now the integral kernel of $E_j(t, X, D)$ is

$$K_j(t,x,y) = (2\pi)^{-n} \int E_j\Big(t, \frac{x+y}{2}, \xi\Big) e^{i(x-y)\cdot\xi}\, d\xi. \tag{14.67}$$

In particular, on the diagonal we have

$$K_j(t,x,x) = (2\pi)^{-n} \int E_j(t,x,\xi)\, d\xi. \tag{14.68}$$

We want to compute these quantities, for $j = 0, 1$, and at $x = x_0$. First,

$$K_0(t,x_0,x_0) = (2\pi)^{-n} \int e^{-t|\xi|^2}\, d\xi = (4\pi t)^{-n/2}, \tag{14.69}$$

since, as we know, the Gaussian integral in (14.69) is equal to $(\pi/t)^{n/2}$. Next,

$$\begin{aligned} &(2\pi)^n K_1(t,x_0,x_0) \\ &\quad = -tb(x_0) \int e^{-t|\xi|^2} d\xi - \frac{t^2}{6} \sum \mathrm{Ric}_{jk}(x_0) \int \xi_j \xi_k e^{-t|\xi|^2} d\xi. \end{aligned} \tag{14.70}$$

We need to compute more Gaussian integrals. If $j \neq k$, the integrand is an odd function of ξ_j, so the integral vanishes. On the other hand,

$$\begin{aligned} \int \xi_j^2 e^{-t|\xi|^2} d\xi &= \frac{1}{n} \int |\xi|^2 e^{-t|\xi|^2} d\xi \\ &= -\frac{1}{n} \frac{d}{dt} \int e^{-t|\xi|^2} d\xi = \frac{1}{2} \pi^{n/2} t^{-n/2-1}. \end{aligned} \tag{14.71}$$

Thus

$$K_1(t,x_0,x_0) = -(4\pi t)^{-n/2} \Big(tb(x_0) + \frac{t}{12} S(x_0) \Big), \tag{14.72}$$

since $\sum \mathrm{Ric}_{jj}(x) = S(x)$.

As noted above, the Laplace operator Δ on scalar functions, when conjugated by $g^{1/4}$, has the form (14.65), with $b(x_0) = \Phi(x_0) - \Psi(x_0) = -S(x_0)/4$. Thus, for the keat kernel $e^{t\Delta}$ on scalars, we have

$$K_1(t,x_0,x_0) = (4\pi t)^{-n/2} \frac{t}{6} S(x_0). \tag{14.73}$$

We now generalize this, setting

$$a(x,\xi) = g(x,\xi) + \ell(x,\xi), \quad \ell(x,\xi) = \sum \ell_j(x)\xi_j. \tag{14.74}$$

Continue to assume that $a(x,\xi)$ is scalar and consider $L = a(X,D) + b(x)$. We have

$$E_0(t,x,\xi) = e^{-ta(x,\xi)} = e^{-t\ell(x,\xi)}\, e^{-tg(x,\xi)}, \tag{14.75}$$

and $E_1(t,x,\xi)$ is still given by (14.42)–(14.50). A point to keep in mind is that we can drop $\ell(x,\xi)$ from the computation involving $\{a, e^{-ta}\}_2$, altering $K_1(t,x,x)$ only by $o(t^{-n/2+1})$ as $t \searrow 0$. Thus, mod $o(t^{-n/2+1})$, $K_1(t,x_0,x_0)$ is still given by (14.73). To get $K_0(t,x_0,x_0)$, expand $e^{-t\ell(x,\xi)}$ in (14.75) in powers of t:

$$E_0(t,x,\xi) \sim \Bigl[1 - t\ell(x,\xi) + \frac{t^2}{2}\ell(x,\xi)^2 + \cdots\Bigr] e^{-tg(x,\xi)}. \tag{14.76}$$

When doing the ξ-integral, the term $t\ell(x,\xi)$ is obliterated, of course, while, by (14.71),

$$\frac{t^2}{2}\int \ell(x_0,\xi)^2 e^{-t|\xi|^2}\, d\xi = \frac{1}{4}\pi^{n/2} t^{-n/2+1} \sum \ell_j(x_0)^2. \tag{14.77}$$

Hence, in this situation,

$$\begin{aligned} &K_0(t,x_0,x_0) + K_1(t,x_0,x_0) \\ &\quad = (4\pi t)^{-n/2}\Bigl[1 + t\Bigl(\sum \ell_j(x_0)^2 - b(x_0) - \frac{1}{12}S(x_0)\Bigr) + O(t^2)\Bigr]. \end{aligned} \tag{14.78}$$

Finally, we drop the assumption that $\ell(x,\xi)$ in (14.74) be scalar. We still assume that $g(x,\xi)$ defines the metric tensor. There are several changes whose effects on (14.78) need to be investigated. In the first place, (14.41) is no longer quite true. We have

$$\{a, e^{-ta}\}_1 = \frac{i}{2}\sum \Bigl\{ \frac{\partial a}{\partial x_j}\frac{\partial}{\partial \xi_j} e^{-ta} - \frac{\partial a}{\partial \xi_j}\frac{\partial}{\partial x_j} e^{-ta} \Bigr\}. \tag{14.79}$$

In this case, with $a(x,\xi)$ matrix-valued, we have

$$\begin{aligned} \frac{\partial}{\partial x_j} e^{-ta} &= -te^{-ta}\ \Xi\bigl(\mathrm{ad}(-ta)\bigr)\Bigl(\frac{\partial a}{\partial x_j}\Bigr) \\ &= -te^{-ta}\ \Xi\bigl(\mathrm{ad}(-t\ell)\bigr)\Bigl(\frac{\partial a}{\partial x_j}\Bigr), \end{aligned} \tag{14.80}$$

where $\Xi(z) = (1 - e^{-z})/z$, so

$$\begin{aligned} \frac{\partial}{\partial x_j} e^{-ta} &= te^{-ta}\Bigl(\frac{\partial a}{\partial x_j} + \frac{t}{2}\Bigl[\ell, \frac{\partial \ell}{\partial x_j}\Bigr] + \cdots\Bigr) \\ &= -t\frac{\partial a}{\partial x_j} + O(t^2|\xi|)e^{-ta} + \cdots, \end{aligned} \tag{14.81}$$

and so forth. Hence

$$\{a, e^{-ta}\}_1 = -\frac{i}{2} t \sum \left[\frac{\partial \ell}{\partial x_j}, \frac{\partial \ell}{\partial \xi_j} \right] e^{-ta} + \cdots . \tag{14.82}$$

This is smaller than any of the terms in the transport equation (14.42) for E_1, so it could be put in a higher transport equation. It does not affect (14.78).

Another change comes from the following modification of (14.46):

$$\begin{aligned} &\int_0^t e^{(s-t)a(x,\xi)} b(x) e^{-sa(x,\xi)}\, ds \\ &\quad = \left[\int_0^t e^{(s-t)\ell(x,\xi)} b(x) e^{-s\ell(x,\xi)}\, ds \right] \cdot e^{-tg(x,\xi)}. \end{aligned} \tag{14.83}$$

This time, $b(x)$ and $\ell(x, \xi)$ may not commute. We can write the right side as

$$\begin{aligned} &\int_0^t e^{s \operatorname{ad} \ell(x,\xi)} \big[b(x) \big]\, ds\, e^{-t\ell(x,\xi)} e^{-tg(x,\xi)} \\ &\quad = t \Big\{ b(x) - \frac{t}{2} \big(\ell(x,\xi) b(x) + b(x) \ell(x,\xi) \big) + \cdots \Big\} e^{-tg(x,\xi)}. \end{aligned} \tag{14.84}$$

Due to the extra power of t with the anticommutator, this does not lead to a change in (14.78).

The other change in letting $\ell(x, \xi)$ be nonscalar is that the quantity $\ell(x, \xi)^2 = \sum \ell_j(x)\ell_k(x)\xi_j\xi_k$ generally has noncommuting factors, but this also does not affect (14.78). Consequently, allowing $\ell(x, \xi)$ to be nonscalar does not change (14.78). We state our conclusion:

Theorem 14.1. *If $Lu = a(X, D)u + b(x)u$, with*

$$a(x, \xi) = \sum g^{jk}(x)\xi_j\xi_k + \sum \ell_j(x)\xi_j, \tag{14.85}$$

where (g^{jk}) is the inverse of a metric tensor (g_{jk}), and $\ell_j(x)$ and $b(x)$ are matrix-valued, and if $g_{jk}(x_0) = \delta_{jk}$, $\nabla g_{jk}(x_0) = 0$, then the integral kernel $K(t, x, y)$ of e^{-tL} has the property

(14.86)

$$K(t, x_0, x_0) = (4\pi t)^{-n/2} \Big[1 + t \Big(\sum \ell_j(x_0)^2 - b(x_0) - \frac{1}{12} S(x_0) \Big) + O(t^2) \Big].$$

Exercises

1. If $a(x, \xi) = \sum a_\alpha(x)\xi^\alpha$ is a polynomial in ξ, so that $a(x, D)$ is a differential operator, show that $a(X, D)$ is also a differential operator, given by

$$a(X,D)u(x) = \sum_\alpha D_y^\alpha\Big[a_\alpha\Big(\frac{x+y}{2}\Big)u(y)\Big]\Big|_{y=x}$$
$$= \sum_\alpha \sum_{\beta+\gamma=\alpha} \binom{\alpha}{\beta} 2^{-|\gamma|}\, D^\gamma a_\alpha(x)\, D^\beta u(x).$$

Verify the formulas (14.34) and (14.35) as special cases.

2. If $p \in S^m_{1,0}$ and $q \in S^\mu_{1,0}$ are scalar symbols and $p \circ q$ is defined so that the product $p(X,D)q(X,D) = (p \circ q)(X,D)$, as in (14.22)–(14.23), show that

$$q \circ p \circ q = q^2 p \bmod S^{m+2\mu-2}_{1,0}$$

More generally, if $p_{jk} \in S^m_{1,0}$, $p_{jk} = p_{kj}$, and $q_j \in S^\mu_{1,0}$, show tha

$$\sum_{j,k} q_j \circ p_{jk} \circ q_k = \sum_{j,k} q_j p_{jk} q_k \bmod S^{m+2\mu-2}_{1,0}.$$

Relate this to the last identity in (14.35), comparing a second-order differential operator in the Weyl calculus and in divergence form.

15. Operators of harmonic oscillator type

In this section we study operators with symbols in $\mathcal{S}^m_1(\mathbb{R}^n)$, defined to consist of functions $p(x,\xi)$, smooth on $\mathbb{R}^{2n}$ and satisfying

$$|D_x^\beta D_\xi^\alpha p(x,\xi)| \le C_{\alpha\beta}(1+|x|+|\xi|)^{m-|\alpha|-|\beta|}. \tag{15.1}$$

This class has the property of treating x and ξ on the same footing. We define $OP\mathcal{S}^m_1(\mathbb{R}^n)$ to consist of operators $p(X,D)$ with $p(x,\xi) \in \mathcal{S}^m_1(\mathbb{R}^n)$. Here we use the Weyl calculus, (14.5). In this setting, the $Sp(n,\mathbb{R})$ action (14.18)–(14.19) can be well exploited. This action does not preserve $S^m_{1,0}(\mathbb{R}^n)$, but it does preserve $\mathcal{S}^m_1(\mathbb{R}^n)$. The class $OP\mathcal{S}^m_1(\mathbb{R}^n)$ has been studied in [GLS, Ho4], and [V], and played a role in microlocal analysis on the Heisenberg group in [T2].

Note that

$$\mathcal{S}^0_1(\mathbb{R}^n) \subset S^0_{1,0}(\mathbb{R}^n), \tag{15.2}$$

so it follows from Theorem 6.3, plus (14.6)–(14.8), that

$$P \in OP\mathcal{S}^0_1(\mathbb{R}^n) \Longrightarrow P : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n). \tag{15.3}$$

If $a \in \mathcal{S}^m_1$ and $b \in \mathcal{S}^\mu_1$, variants of methods of §3 and (14.22)–(14.24) give

$$a(X,D)b(X,D) = (a \circ b)(X,D) \in OP\mathcal{S}^{m+\mu}_1(\mathbb{R}^n), \tag{15.4}$$

with

$$(a \circ b)(x,\xi) \sim ab + \sum_{j\geq 1} \frac{1}{j!}\{a,b\}_j(x,\xi), \tag{15.5}$$

where $\{a,b\}_j$ is given by (14.24). Note that

$$\{a,b\}_j \in \mathcal{S}_1^{m+\mu-2j}(\mathbb{R}^n). \tag{15.6}$$

We mention that if either $a(x,\xi)$ or $b(x,\xi)$ is a polynomial in (x,ξ), then the sum in (15.5) is finite and provides an exact formula for $(a \circ b)(x,\xi)$.

The set of "classical" symbols in $\mathcal{S}_1^m(\mathbb{R}^n)$, denoted $\mathcal{S}^m(\mathbb{R}^n)$, is defined to consist of all $p(x,\xi) \in \mathcal{S}_1^m(\mathbb{R}^n)$ such that

$$p(x,\xi) \sim \sum_{j\geq 0} p_j(x,\xi), \tag{15.7}$$

with $p_j(x,\xi)$ smooth and, for $|x|^2 + |\xi|^2 \geq 1$, homogeneous of degree $m-2j$ in (x,ξ). The meaning of (15.7) is that for each N,

$$p(x,\xi) - \sum_{j=0}^{N-1} p_j(x,\xi) \in \mathcal{S}_1^{m-2N}(\mathbb{R}^n). \tag{15.8}$$

It follows from (15.4)–(15.6) that

$$\begin{aligned} a \in \mathcal{S}^m(\mathbb{R}^n),\ b \in \mathcal{S}^\mu(\mathbb{R}^n) \Longrightarrow a(X,D)b(X,D) \\ = (a \circ b)(X,D), a \circ b \in \mathcal{S}^{m+\mu}(\mathbb{R}^n). \end{aligned} \tag{15.9}$$

Sobolev spaces tailored to these operator classes are defined as follows, for $k \in \mathbb{Z}^+$.

$$\begin{aligned} \mathcal{H}^k(\mathbb{R}^n) &= \{u \in L^2(\mathbb{R}^n) : Pu \in L^2(\mathbb{R}^n),\ \forall\, P \in \mathcal{D}^k(\mathbb{R}^n)\}, \\ \mathcal{D}^k(\mathbb{R}^n) &= \text{ span of } x^\beta D_x^\alpha,\ \ |\alpha| + |\beta| \leq k. \end{aligned} \tag{15.10}$$

Note that $\mathcal{D}^k(\mathbb{R}^n) \subset OP\mathcal{S}^k(\mathbb{R}^n)$. The following Rellich type theorem is straightforward:

(15.11) The natural inclusion $\mathcal{H}^k(\mathbb{R}^n) \hookrightarrow L^2(\mathbb{R}^n)$ is compact, $\forall\, k \geq 1$.

The results (15.4) and (15.3) yield, for $k \in \mathbb{Z}^+$,

$$A \in OP\mathcal{S}_1^{-k}(\mathbb{R}^n) \Longrightarrow A : L^2(\mathbb{R}^n) \to \mathcal{H}^{-k}(\mathbb{R}^n). \tag{15.12}$$

We will obtain other Sobolev mapping properties below. These spaces will be seen to be natural settings for elliptic regularity results.

An operator $P = p(X,D) \in OP\mathcal{S}^m_1(\mathbb{R}^n)$ is said to be elliptic provided

$$|p(x,\xi)^{-1}| \le C(1+|x|+|\xi|)^{-m}, \tag{15.13}$$

for $|x|^2 + |\xi|^2$ sufficiently large. With the results (15.4)–(15.6) in hand, natural variants of the parametrix construction of §4 yield for such elliptic P,

$$\begin{aligned} &Q \in OP\mathcal{S}^{-m}_1(\mathbb{R}^n), \quad PQ = I + R_1, \quad QP = I + R_2, \\ &R_j \in OP\mathcal{S}^{-\infty}_1(\mathbb{R}^n) = \bigcap_{k\ge 1} OP\mathcal{S}^{-k}_1(\mathbb{R}^n). \end{aligned} \tag{15.14}$$

Clearly, for each $m \in \mathbb{R}$,

$$A_m(x,\xi) = (1+|x|^2+|\xi|^2)^{m/2} \tag{15.15}$$

is the symbol of an elliptic operator in $OP\mathcal{S}^m_1(\mathbb{R}^n)$. We have

$$A_m(X,D)A_{-m}(X,D) = I + R_m, \quad R_m \in OP\mathcal{S}^{-4}_1(\mathbb{R}^n). \tag{15.16}$$

In this situation, (15.5) applies, and $\{A_m, A_{-m}\}_1 = 0$.

We now introduce the central operator in this class, the harmonic oscillator,

$$H = -\Delta + |x|^2 = \sum_{j=1}^n \Big(-\frac{\partial^2}{\partial x_j^2} + x_j^2\Big). \tag{15.17}$$

This is an elliptic element of $OP\mathcal{S}^2(\mathbb{R}^n)$, with symbol $|x|^2 + |\xi|^2$. It defines a positive, self adjoint operator on $L^2(\mathbb{R}^n)$. Note that

$$\begin{aligned} L_j = \partial_j + x_j &\Longrightarrow L_j^* = -\partial_j + x_j \\ &\Longrightarrow L_j^* L_j = -\partial_j^2 + x_j^2 - 1 \\ &\Longrightarrow H = \sum L_j^* L_j + n, \end{aligned} \tag{15.18}$$

so H is positive definite, with H^{-1} bounded on $L^2(\mathbb{R}^n)$. The following result will be very useful.

Theorem 15.1. *For all $s \in (0,\infty)$, $H^{-s} \in OP\mathcal{S}^{-2s}(\mathbb{R}^n)$. With $A_m(x,\xi)$ as in (15.15),*

$$H^{-s} - A_{-2s}(X,D) \in OP\mathcal{S}^{-2s-2}(\mathbb{R}^n). \tag{15.19}$$

We postpone the proof of Theorem 15.1 and observe some of its consequences.

Proposition 15.2. *For $k \in \mathbb{Z}^+$,*

$$H^{-k/2} : L^2(\mathbb{R}^n) \longrightarrow \mathcal{H}^k(\mathbb{R}^n) \tag{15.20}$$

is an isomorphism.

Proof. The mapping property (15.20) follows from Theorem 15.1 and (15.12). If $k = 2\ell$ is even, the two sided inverse to (15.20) is

$$H^\ell : \mathcal{H}^{2\ell}(\mathbb{R}^n) \longrightarrow L^2(\mathbb{R}^n). \tag{15.21}$$

We need to show that if $k = 2\ell - 1$ is odd,

$$H^{k/2} = H^{\ell-1/2} : \mathcal{H}^{2\ell-1}(\mathbb{R}^n) \longrightarrow L^2(\mathbb{R}^n). \tag{15.22}$$

Indeed, take $u \in \mathcal{H}^{2\ell-1}(\mathbb{R}^n)$. Then

$$H^\ell u = \sum X_j u_j, \quad u_j \in L^2(\mathbb{R}^n),\ X_j \in \mathcal{D}^1(\mathbb{R}^n), \tag{15.23}$$

and hence

$$H^{\ell-1/2} u = \sum H^{-1/2} X_j u_j, \tag{15.24}$$

which belongs to $L^2(\mathbb{R}^n)$ since $H^{-1/2} X_j \in OP\mathcal{S}^0(\mathbb{R}^n)$.

Given Proposition 15.2, it is natural to set

$$\mathcal{H}^s(\mathbb{R}^n) = H^{-s/2} L^2(\mathbb{R}^n), \tag{15.25}$$

for $s \in \mathbb{R}$, and we have that this space agrees with (15.10) for $s = k \in \mathbb{Z}^+$. For $s > 0$, this says

$$\mathcal{H}^s(\mathbb{R}^n) = \mathcal{D}(H^{s/2}). \tag{15.26}$$

Thus, by Proposition 2.2 of Chap. 4, we can identify $\mathcal{H}^s(\mathbb{R}^n)$ with the complex interpolation space:

$$\mathcal{H}^s(\mathbb{R}^n) = [L^2(\mathbb{R}^n), \mathcal{H}^k(\mathbb{R}^n)]_\theta, \quad s = k\theta,\ \theta \in (0,1). \tag{15.27}$$

Also note that

$$\bigcap_{s<\infty} \mathcal{H}^s(\mathbb{R}^n) = \mathcal{S}(\mathbb{R}^n), \qquad \bigcup_{s>-\infty} \mathcal{H}^s(\mathbb{R}^n) = \mathcal{S}'(\mathbb{R}^n). \tag{15.28}$$

In fact, (15.10) gives $\cap_{k\in\mathbb{Z}^+} \mathcal{H}^k(\mathbb{R}^n) = \mathcal{S}(\mathbb{R}^n)$, and (15.25) gives $\mathcal{H}^s(\mathbb{R}^n)' = \mathcal{H}^{-s}(\mathbb{R}^n)$.

Given Theorem 15.1, it easily follows that

$$H^s \in OP\mathcal{S}^{2s}(\mathbb{R}^n), \quad \forall\, s \in \mathbb{R}. \tag{15.29}$$

In fact, given $s > 0$, take an integer $k > s$ and write $H^s = H^k H^{s-k}$. Also, given (15.25), we have, for all $m, s \in \mathbb{R}$,

$$P \in OP\mathcal{S}_1^m(\mathbb{R}^n) \Longrightarrow P : \mathcal{H}^s(\mathbb{R}^n) \to \mathcal{H}^{s-m}(\mathbb{R}^n). \tag{15.30}$$

Indeed, $P = H^{-(s-m)/2}(H^{(s-m)/2} P H^{-s/2}) H^{s/2}$, and $H^{(s-m)/2} P H^{-s/2} \in OP\mathcal{S}_1^0(\mathbb{R}^n)$ is bounded on $L^2(\mathbb{R}^n)$.

We will approach the proof of Theorem 15.1 via the identity

$$H^{-s} = \frac{1}{\Gamma(s)} \int_0^\infty e^{-tH} t^{s-1}\, dt, \quad s > 0. \tag{15.31}$$

Thus we have the task of writing

$$e^{-tH} = h_t(X, D) \tag{15.32}$$

and computing $h_t(x, \xi)$. We need to solve

$$\frac{\partial}{\partial t} h_t(X, D) = -H h_t(X, D), \quad h_0(x, \xi) = 1. \tag{15.33}$$

Taking

$$b_t(X, D) = H h_t(X, D), \quad H = Q(X, D), \quad Q(x, \xi) = |x|^2 + |\xi|^2, \tag{15.34}$$

since $Q(x, \xi)$ is a polynomial, the formula (15.5) for composition is a finite sum, and it is exact:

$$b_t(x, \xi) = Q(x, \xi) h_t(x, \xi) + \sum_{j=1}^{2} \frac{1}{j!} \{Q, h_t\}_j(x, \xi). \tag{15.35}$$

Now we make the "guess" that for each $t > 0$, $h_t(x, \xi)$ is a function of $|x|^2 + |\xi|^2 = Q$,

$$h_t(x, \xi) = g(t, Q). \tag{15.36}$$

In that case, $\{Q, h_t\}_1 = 0$, and (15.33)–(15.35) lead to the equation

$$\frac{\partial h_t}{\partial t}(x, \xi) = -(|x|^2 + |\xi|^2) h_t(x, \xi) + \frac{1}{4} \sum_k \Big(\frac{\partial^2}{\partial x_k^2} + \frac{\partial^2}{\partial \xi_k^2} \Big) h_t(x, \xi), \tag{15.37}$$

with initial condition $h_0(x,\xi) = 1$, or equivalently to solve

$$\frac{\partial g}{\partial t} = -Qg + Q\frac{\partial^2 g}{\partial Q^2} + n\frac{\partial g}{\partial Q}, \quad g(0,Q) = 1. \tag{15.38}$$

We now guess that (15.38) has a solution of the form

$$g(t,Q) = a(t)e^{b(t)Q}.$$

Then the left side of (15.38) is $(a'/a + b'Q)g$ and the right side is $(-Q + Qb^2 + nb)g$, so (15.38) is equivalent to

$$\frac{a'(t)}{a(t)} = nb(t), \quad b'(t) = b(t)^2 - 1. \tag{15.39}$$

We can solve the second equation for $b(t)$ by separation of variables. Since $g(0,Q) = 1$, we need $b(0) = 0$, and the unique solution is

$$b(t) = -\tanh t. \tag{15.40}$$

Then the equation $a'/a = -n \tanh t$ with $a(0) = 1$ gives

$$a(t) = (\cosh t)^{-n}. \tag{15.41}$$

We have our desired formula

$$h_t(x,\xi) = (\cosh t)^{-n} e^{-(\tanh t)(|x|^2+|\xi|^2)}. \tag{15.42}$$

We discuss briefly why the "guess" that $h_t(x,\xi)$ is a function of $|x|^2 + |\xi|^2$ was bound to succeed. It is related to the identity (14.19) for the composition of operators transformed by $a_g(x,\xi) = a(g^{-1}(x,\xi)),\ g \in Sp(n,\mathbb{R})$. If we identify $\mathbb{R}^{2n}$ with $\mathbb{C}^n$ and (x,ξ) with $x + i\xi$, then the unitary group $U(n)$ acts on $\mathbb{C}^n = \mathbb{R}^{2n}$, as a subgroup of $Sp(n,\mathbb{R})$, preserving $|x|^2 + |\xi|^2 = |x + i\xi|^2$. It follows from (14.9) that the set of operators whose symbols are invariant under this $U(n)$ action forms an algebra. From there, it is a short step to guess that e^{-tH} belongs to this algebra. For more details, see Chap. 1, §7 of [T3].

We return to the identity (15.31), which implies

$$H^{-s} = Q_{-s}(X,D) \tag{15.43}$$

with

$$Q_{-s}(x,\xi) = \frac{1}{\Gamma(s)} \int_0^\infty t^{s-1} (\cosh t)^{-n} e^{-(\tanh t)(|x|^2+|\xi|^2)}\, dt. \tag{15.44}$$

To complete the proof of Theorem 15.1, it remains to show that, whenever $s > 0$,

$$Q_{-s}(x,\xi) \in \mathcal{S}^{-2s}(\mathbb{R}^n), \quad \text{and} \quad Q_{-s}(x,\xi) - A_{-2s}(x,\xi) \in \mathcal{S}^{-2s-2}(\mathbb{R}^n). \tag{15.45}$$

To begin, it is clear by inspection that $Q_{-s} \in C^\infty(\mathbb{R}^{2n})$ whenever $s > 0$. Also, if we set

$$Q^b_{-s}(x,\xi) = \frac{1}{\Gamma(s)} \int_0^1 t^{s-1}(\cosh t)^{-n} e^{-(\tanh t)(|x|^2+|\xi|^2)}\, dt, \tag{15.46}$$

we easily get

$$Q_{-s}(x,\xi) - Q^b_{-s}(x,\xi) \in \mathcal{S}^{-\infty}_1(\mathbb{R}^n). \tag{15.47}$$

We can set $\tau = \tanh t$ and write

$$Q^b_{-s}(x,\xi) = \frac{1}{\Gamma(s)} \int_0^a \tau^{s-1}\varphi(\tau) e^{-\tau(|x|^2+|\xi|^2)}\, d\tau, \tag{15.48}$$

with $a = \tanh 1$ and $\varphi \in C^\infty([0,a])$, with power series

$$\varphi(\tau) \sim 1 + b_1\tau^2 + b_2\tau^4 + \cdots. \tag{15.49}$$

Thus, as $|x|^2 + |\xi|^2 \to \infty$, we have

$$Q_{-s}(x,\xi) \sim \sum_{j\ge 0} q_{-s,j}(x,\xi), \tag{15.50}$$

with

$$\begin{aligned} q_{-s,j}(x,\xi) &= \frac{b_j}{\Gamma(s)} \int_0^\infty e^{-\tau(|x|^2+|\xi|^2)} \tau^{s+2j-1}\, d\tau \\ &= b_j \frac{\Gamma(s+2j)}{\Gamma(s)} \bigl(|x|^2+|\xi|^2\bigr)^{-s-2j}, \end{aligned} \tag{15.51}$$

and $b_0 = 1$. This proves Theorem 15.1.

Remark: We can sharpen (15.45) as follows. Replace $A_{-2s}(x,\xi)$ by $\widetilde{A}_{-2s}(x,\xi)$, smooth on $\mathbb{R}^{2n}$ and equal to $(|x|^2+|\xi|^2)^{-2s}$ for $|x|^2+|\xi|^2 \ge 1$. Then

$$Q_{-s}(x,\xi) - \widetilde{A}_{-2s}(x,\xi) \in \mathcal{S}^{-2s-4}(\mathbb{R}^n). \tag{15.52}$$

We make a further specific study of the harmonic oscillator H in §6 of Chap. 8, including results on the eigenvalues and eigenfunctions of H, and an alternative approach to the analysis of the semigroup e^{-tH}.

We can extend the Rellich type result (15.11) as follows. By (15.11) and (15.20), we have H^{-1} compact on $L^2(\mathbb{R}^n)$, so H^{-1} has a discrete set of eigenvalues, tending to 0, and hence so does $H^{-\sigma}$ for all $\sigma > 0$. Thus $H^{-\sigma}$ is compact on $L^2(\mathbb{R}^n)$, and, by (15.26), also compact on $\mathcal{H}^s(\mathbb{R}^n)$, for all $s \in \mathbb{R}$. This gives the following.

Proposition 15.3. *Given $r < s \in \mathbb{R}$, the natural inclusion*

$$\mathcal{H}^s(\mathbb{R}^n) \hookrightarrow \mathcal{H}^r(\mathbb{R}^n) \tag{15.53}$$

is compact.

If $P \in OP\mathcal{S}_1^m(\mathbb{R}^n)$ is elliptic (say a $k \times k$ system), and $Q \in OP\mathcal{S}_1^{-m}(\mathbb{R}^n)$ is a parametrix, as in (15.14), we see that the operators R_j are compact on $\mathcal{H}^s(\mathbb{R}^n)$ for all s, so we have the following.

Proposition 15.4. *If $P \in OP\mathcal{S}_1^m(\mathbb{R}^n)$ is elliptic, then, for all $s \in \mathbb{R}$,*

$$P : \mathcal{H}^s(\mathbb{R}^n) \longrightarrow \mathcal{H}^{s-m}(\mathbb{R}^n) \ \text{ is Fredholm.} \tag{15.54}$$

Also

$$\text{Ker}\, P,\ \text{Ker}\, P^* \subset \mathcal{S}(\mathbb{R}^n), \tag{15.55}$$

and the index of P is independent of s.

Material on the index of elliptic operators in $OP\mathcal{S}^m(\mathbb{R}^n)$ will be covered in §11 of Chap. 10. See the exercises below for some preliminary results.

Exercises

1. In case $n = 1$, consider
$$D_1 = \partial_1 + x_1.$$
Show that $D_1 \in OP\mathcal{S}^1(\mathbb{R})$ is elliptic, and that
$$\text{Index}\, D_1 = 1.$$

2. In case $n = 2$, consider
$$D_2 = \begin{pmatrix} \partial_1 + x_1 & \partial_2 - x_2 \\ \partial_2 + x_2 & -\partial_1 + x_1 \end{pmatrix}.$$
Show that $D_2 \in OP\mathcal{S}^1(\mathbb{R}^2)$ is elliptic and that
$$\text{Index}\, D_2 = 1.$$

3. In the setting of Exercises 1–2, compute $D_j^* D_j$ and $D_j D_j^*$, and compare with H. This should help to compute the kernels of D_j and D_j^*.

References

[ADN] S. Agmon, A. Douglis, and L. Nirenberg, Estimates near the boundary for solutions of elliptic differential equations satisfying general boundary conditions, *CPAM* 12(1959), 623–727; II, *CPAM* 17(1964), 35–92

[Be] R. Beals, A general calculus of pseudo-differential operators, *Duke Math. J.* 42(1975), 1–42

[BF] R. Beals and C. Fefferman, Spatially inhomogeneous pseudodifferential operators, *Comm. Pure Appl. Math.* 27(1974), 1–24

[BGM] M. Berger, P. Gauduchon, and E. Mazet, *Le Spectre d'une Variété Riemannienne*, LNM #194, Springer, New York, 1971

[BJS] L. Bers, F. John, and M. Schechter, *Partial Differential Equations*, Wiley, New York, 1964

[Ca1] A. P. Calderon, Uniqueness in the Cauchy problem of partial differential equations, *Am. J. Math.* 80(1958), 16–36

[Ca2] A. P. Calderon, Singular integrals, *Bull. AMS* 72(1966), 427–465

[Ca3] A. P. Calderon, Cauchy integrals on Lipschitz curves and related operators, *Proc. NAS, USA* 74(1977), 1324–1327

[CV] A. P. Calderon and R. Vaillancourt, A class of bounded pseudodifferential operators, *Proc. NAS, USA* 69(1972), 1185–1187

[CZ] A. P. Calderon and A. Zygmund, Singular integral operators and differential equations, *Am. J. Math.* 79(1957), 901–921

[CMM] R. Coifman, A. McIntosh, and Y. Meyer, L'integrale de Cauchy definit un operateur borne sur L^2 pour les courbes Lipschitziennes, *Ann. Math.* 116(1982), 361–388

[Cor] H. O. Cordes, On compactness of commutators of multiplications and convolutions, and boundedness of pseudo-differential operators, *J. Funct. Anal.* 18(1975), 115–131

[Cor2] H. O. Cordes, *Elliptic Pseudodifferential Operators—An Abstract Theory*, LNM #756, Springer, New York, 1979

[Cor3] H. O. Cordes, *Spectral Theory of Linear Differential Operators and Comparison Algebras*, London Math. Soc. Lecture Notes #70, Cambridge University Press, London, 1987

[CH] H. O. Cordes and E. Herman, Gelfand theory of pseudo-differential operators, *Am. J. Math.* 90(1968), 681–717

[DK] B. Dahlberg and C. Kenig, Hardy spaces and the Neumann problem in L^p for Laplace's equation in Lipschitz domains, *Ann. Math.* 125(1987), 437–465

[D] G. David, Opérateurs d'intégrale singulières sur les surfaces régulières, *Ann. Scient. Ecole Norm. Sup.* 21(1988), 225–258

[DS] G. David and S. Semmes, *Analysis of and on Uniformly Rectifiable Sets*, American Math. Soc., Providence, RI, 1993

[Dui] J. J. Duistermaat, *Fourier Integral Operators*, Courant Institute Lecture Notes, New York, 1974

[Eg] Y. Egorov, On canonical transformations of pseudo-differential operators, *Uspehi Mat. Nauk.* 24(1969), 235–236

[FJR] E. Fabes, M. Jodeit, and N. Riviere, Potential techniques for boundary problems in C^1 domains, *Acta Math.* 141(1978), 165–186

[F] C. Fefferman, The Uncertainty Principle, *Bull. AMS* 9(1983), 129–266

[FP] C. Fefferman and D. Phong, On positivity of pseudo-differential operators, *Proc. NAS, USA* 75(1978), 4673–4674

[Fo2] G. Folland, *Harmonic Analysis on Phase Space,* Princeton University Press, Princeton, NJ, 1989

[Gå] L. Gårding, Dirichlet's problem for linear elliptic partial differential equations, *Math. Scand.* 1(1953), 55–72

[GT] D. Gilbarg and N. Trudinger, *Elliptic Partial Differential Equations of Second Order*, 2nd ed., Springer, New York, 1983

[Gr] P. Greiner, An asymptotic expansion for the heat equation, *Arch. Rat. Mech. Anal.* 41(1971), 163–218

[GLS] A. Grossman, G. Loupias, and E. Stein, An algebra of pseudodifferential operators and quantum mechanics in phase space, *Ann. Inst. Fourier* 18(1969), 343–368

[HMT] S. Hofmann, M. Mitrea, and M. Taylor, Singular integrals and elliptic boundary problems on regular Semmes-Kenig-Toro domains, *International Math. Research Notices* (2009), doi:10.1093/imrn/rnp214

[Ho1] L. Hörmander, Pseudo-differential operators, *Comm. Pure Appl. Math.* 18(1965), 501–517

[Ho2] L. Hörmander, Pseudodifferential operators and hypoelliptic equations, *Proc. Symp. Pure Math.* 10(1967), 138–183

[Ho3] L. Hörmander, Fourier integral operators I, *Acta Math.* 127(1971), 79–183

[Ho4] L. Hörmander, The Weyl calculus of pseudodifferential operators, *Comm. Pure Appl. Math.* 32(1979), 355–443

[Ho5] L. Hörmander, *The Analysis of Linear Partial Differential Operators*, Vols. 3 and 4, Springer, New York, 1985

[How] R. Howe, Quantum mechanics and partial differential equations, *J. Funct. Anal.* 38(1980), 188–254

[JK] D. Jerison and C. Kenig, The inhomogeneous Dirichlet problem in Lipschitz domains, *J. Funct. Anal.* 130(1995), 161–219

[K] T. Kato, Boundedness of some pseudo-differential operators, *Osaka J. Math.* 13(1976), 1–9

[Keg] O. Kellogg, *Foundations of Potential Theory*, Dover, New York, 1954

[KN] J. J. Kohn and L. Nirenberg, An algebra of pseudo-differential operators, *Comm. Pure Appl. Math.* 18(1965), 269–305

[Kg] H. Kumano-go, *Pseudodifferential Operators*, MIT, Cambridge, MA, 1981

[LM] J. Lions and E. Magenes, *Non-homogeneous Boundary Problems and Applications I, II*, Springer, New York, 1972

[MS] H. McKean and I. Singer, Curvature and the eigenvalues of the Laplacian, *J. Diff. Geom.* 1(1967), 43–69

[Mik] S. Mikhlin, *Multidimensional Singular Integral Equations,* Pergammon, New York, 1965

[MT1] M. Mitrea and M. Taylor, Boundary layer methods for Lipschitz domains in Riemannian manifolds, *J. Funct. Anal.* 163(1999), 181–251

[MT2] M. Mitrea and M. Taylor, Potential theory on Lipschitz domains in Riemannian manifolds: Sobolev-Besov space results and the Poisson problem, *J. Funct. Anal.* 176(2000), 1–79

[Miz] S. Mizohata, *The Theory of Partial Differential Equations,* Cambridge University Press, Cambridge, 1973

[Mus] N. Muskhelishvilli, *Singular Integral Equations*, P. Nordhoff, Groningen, 1953

[Ni] L. Nirenberg, *Lectures on Linear Partial Differential Equations*, Reg. Conf. Ser. in Math., no. 17, AMS, Providence, RI, 1972

[Pal] R. Palais, *Seminar on the Atiyah-Singer Index Theorem*, Princeton University Press, Princeton, NJ, 1965

[Po] J. Polking, Boundary value problems for parabolic systems of partial differential equations, *Proc. Symp. Pure Math*. 10(1967), 243–274

[RS] M. Reed and B. Simon, *Methods of Mathematical Physics*, Academic, New York, Vols. 1,2, 1975; Vols. 3,4, 1978

[Se1] R. Seeley, Refinement of the functional calculus of Calderon and Zygmund, *Proc. Acad. Wet. Ned. Ser. A* 68(1965), 521–531

[Se2] R. Seeley, Singular integrals and boundary value problems, *Am. J. Math*. 88(1966), 781–809

[Se3] R. Seeley, Complex powers of an elliptic operator, *Proc. Symp. Pure Math* 10(1967), 288–307

[So] S. Sobolev, *Partial Differential Equations of Mathematical Physics*, Dover, New York, 1964

[St] E. Stein, *Singular Integrals and the Differentiability of Functions*, Princeton University Press, Princeton, NJ, 1972

[St2] E. Stein, *Harmonic Analysis*, Princeton University Press, Princeton, NJ, 1993

[SW] E. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Space*, Princeton University Press, Princeton, NJ, 1971

[T1] M. Taylor, *Pseudodifferential Operators*, Princeton University Press, Princeton, NJ, 1981

[T2] M. Taylor, *Noncommutative Microlocal Analysis*, Memoirs AMS, no. 313, Providence, RI, 1984

[T3] M. Taylor, *Noncommutative Harmonic Analysis*, AMS, Providence, RI, 1986

[T4] M. Taylor, *Pseudodifferential Operators and Nonlinear PDE*, Birkhauser, Boston, 1991

[Tre] F. Treves, *Introduction to Pseudodifferential Operators and Fourier Integral Operators*, Plenum, New York, 1980

[Ver] G. Verchota, Layer potentials and regularity for the Dirichlet problem for Laplace's equation in Lipschitz domains, *J. Funct. Anal.* 39(1984), 572–611

[V] A. Voros, An algebra of pseudodifferential operators and asymptotics of quantum mechanics, *J. Funct. Anal.* 29(1978), 104–132

[Wey] H. Weyl, *The Theory of Groups and Quantum Mechanics*, Dover, New York, 1931

8

Spectral Theory

Introduction

This chapter is devoted to the spectral theory of self-adjoint, differential operators. We cover a number of different topics, beginning in §1 with a proof of the spectral theorem. It was an arbitrary choice to put that material here, rather than in Appendix A, on functional analysis. The main motivation for putting it here is to begin a line of reasoning that will be continued in subsequent sections, using the great power of studying unitary groups as a tool in spectral theory. After we show how easily this study leads to a proof of the spectral theorem in §1, in later sections we use it in various ways: as a tool to establish self-adjointness, as a tool for obtaining specific formulas, including basic identities among special functions, and in other capacities.

Sections 2 and 3 deal with some general questions in spectral theory, such as when does a differential operator define a self-adjoint operator, when does it have a compact resolvent, and what asymptotic properties does its spectrum have? We tackle the latter question, for the Laplace operator Δ, by examining the asymptotic behavior of the trace of the solution operator $e^{t\Delta}$ for the heat equation, showing that

$$\text{(0.1)} \qquad \operatorname{Tr} e^{t\Delta} = (4\pi t)^{-n/2} \operatorname{vol}\Omega + o(t^{-n/2}), \quad t \searrow 0,$$

when Ω is either a compact Riemannian manifold or a bounded domain in $\mathbb{R}^n$ (and has the Dirichlet boundary condition). Using techniques developed in §13 of Chap. 7, we could extend (0.1) to general compact Riemannian manifolds with smooth boundary and to other boundary conditions, such as the Neumann boundary condition. We use instead a different method here in §3, one that works without any regularity hypotheses on $\partial\Omega$. In such generality, (0.1) does not necessarily hold for the Neumann boundary problem.

The study of (0.1) and refinements got a big push from [Kac]. As pursued in [MS], it led to developments that we will discuss in Chap. 10. The problem of to what extent a Riemannian manifold is determined by the spectrum of its Laplace operator has led to much work, which we do not include here. Some is discussed

M.E. Taylor, *Partial Differential Equations II: Qualitative Studies of Linear Equations*, Applied Mathematical Sciences 116, DOI 10.1007/978-1-4419-7052-7_2,

in [Ber, Br, BGM], and [Cha]. We mention particularly some distinct regions in $\mathbb{R}^2$ whose Laplace operators have the same spectra, given in [GWW].

We have not included general results on the spectral behavior of Δ obtained via geometrical optics and its refinement, the theory of Fourier integral operators. Results of this nature can be found in Volume 3 of [Ho], in [Shu], and in Chap. 12 of [T1].

Sections 4–7 are devoted to specific examples. In §4 we study the Laplace operator on the unit spheres S^n. We specify precisely the spectrum of Δ and discuss explicit formulas for certain functions of Δ, particularly

$$A^{-1}\sin tA, \quad A = \Big(-\Delta + \frac{K}{4}(n-1)^2\Big)^{1/2}. \tag{0.2}$$

with $K = 1$, the sectional curvature of S^n. In §5 we obtain an explicit formula for (0.2), with $K = -1$, on hyperbolic space. In §6 we study the spectral theory of the harmonic oscillator

$$H = -\Delta + |x|^2. \tag{0.3}$$

We obtain an explicit formula for e^{-tH}, an analogue of which will be useful in Chap. 10. In §8 we study the operator

$$H = -\Delta - K|x|^{-1} \tag{0.4}$$

on $\mathbb{R}^3$, obtaining in particular all the eigenvalues of this operator. This operator arises in the simplest quantum mechanical model of the hydrogen atom. In §9 we study the Laplace operator on a cone. Studies done in these sections bring in a number of special functions, including Legendre functions, Bessel functions, and hypergeometric functions. We have included two auxiliary problem sets, one on confluent hypergeometric functions and one on hypergeometric functions.

1. The spectral theorem

Appendix A contains a proof of the spectral theorem for a compact, self-adjoint operator A on a Hilbert space H. In that case, H has an orthonormal basis $\{u_j\}$ such that $Au_j = \lambda_j u_j$, λ_j being real numbers having only 0 as an accumulation point. The vectors u_j are eigenvectors.

A general bounded, self-adjoint operator A may not have any eigenvectors, and the statement of the spectral theorem is somewhat more subtle. The following is a useful version.

Theorem 1.1. *If A is a bounded, self-adjoint operator on a separable Hilbert space H, then there is a σ-compact space Ω, a Borel measure μ, a unitary map*

$$W : L^2(\Omega, d\mu) \longrightarrow H, \tag{1.1}$$

and a real-valued function $a \in L^\infty(\Omega, d\mu)$ such that

$$W^{-1}AWf(x) = a(x)f(x), \quad f \in L^2(\Omega, d\mu). \tag{1.2}$$

Note that when A is compact, the eigenvector decomposition above yields (1.1) and (1.2) with (Ω, μ) a purely atomic measure space. Later in this section we will extend Theorem 1.1 to the case of unbounded, self-adjoint operators.

In order to prove Theorem 1.1, we will work with the operators

$$U(t) = e^{itA}, \tag{1.3}$$

defined by the power-series expansion

$$e^{itA} = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} A^n. \tag{1.4}$$

This is a special case of a construction made in §4 of Chap. 1. $U(t)$ is uniquely characterized as the solution to the differential equation

$$\frac{d}{dt}U(t) = iAU(t), \quad U(0) = I. \tag{1.5}$$

We have the group property

$$U(s+t) = U(s)U(t), \tag{1.6}$$

which follows since both sides satisfy the ODE $(d/ds)Z(s) = iAZ(s), Z(0) = U(t)$. If $A = A^*$, then applying the adjoint to (1.4) gives

$$U(t)^* = U(-t), \tag{1.7}$$

which is the inverse of $U(t)$ in view of (1.6). Thus $\{U(t) : t \in \mathbb{R}\}$ is a group of unitary operators.

For a given $v \in H$, let H_v be the closed linear span of $\{U(t)v : t \in \mathbb{R}\}$; we say H_v is the cyclic space generated by v. We say v is a cyclic vector for H if $H = H_v$. If H_v is not all of H, note that $H_v^\perp$ is invariant under $U(t)$, that is, $U(t)H_v^\perp \subset H_v^\perp$ for all t, since for a linear subspace V of H, generally

$$U(t)V \subset V \Longrightarrow U(t)^* V^\perp \subset V^\perp. \tag{1.8}$$

Using this observation, we can prove the next result.

Lemma 1.2. *If $U(t)$ is a unitary group on a separable Hilbert space H, then H is an orthogonal direct sum of cyclic subspaces.*

Proof. Let $\{w_j\}$ be a countable, dense subset of H. Take $v_1 = w_1$ and $H_1 = H_{v_1}$. If $H_1 \neq H$, let v_2 be the first nonzero element $P_1 w_j$, $j \geq 2$, where P_1 is the orthogonal projection of H onto $H_1^\perp$, and let $H_2 = H_{v_2}$. Continue.

In view of this, Theorem 1.1 is a consequence of the following:

Proposition 1.3. *If $U(t)$ is a strongly continuous, unitary group on H, having a cyclic vector v, then we can take $\Omega = \mathbb{R}$, and there exists a positive Borel measure μ on $\mathbb{R}$ and a unitary map $W : L^2(\mathbb{R}, d\mu) \to H$ such that*

$$W^{-1}U(t)W\, f(x) = e^{itx} f(x), \quad f \in L^2(\mathbb{R}, d\mu). \tag{1.9}$$

The measure μ on $\mathbb{R}$ will be the Fourier transform

$$\mu = \hat{\zeta}, \tag{1.10}$$

where

$$\zeta(t) = (2\pi)^{-1/2}\, (e^{itA}v, v). \tag{1.11}$$

It is not clear a priori that (1.10) defines a measure; since $\zeta \in L^\infty(\mathbb{R})$, we see that μ is a tempered distribution. We will show that μ is indeed a positive measure during the course of our argument. As for the map W, we first define

$$W : \mathcal{S}(\mathbb{R}) \longrightarrow H, \tag{1.12}$$

where $\mathcal{S}(\mathbb{R})$ is the Schwartz space of rapidly decreasing functions, by

$$W(f) = f(A)v, \tag{1.13}$$

where we define the operator $f(A)$ by the formula

$$f(A) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \hat{f}(t)e^{itA}\, dt. \tag{1.14}$$

The reason for this notation will become apparent shortly; see (1.20). Making use of (1.10), we have

$$\begin{aligned}
\big(f(A)v, g(A)v\big) &= (2\pi)^{-1}\Big(\int \hat{f}(s)e^{isA}v\, ds, \int \hat{g}(t)e^{itA}v\, dt\Big) \\
&= (2\pi)^{-1} \iint \hat{f}(s)\overline{\hat{g}(t)}\big(e^{i(s-t)A}v, v\big)\, ds\, dt \\
&= (2\pi)^{-1/2} \iint \hat{f}(s)\hat{\overline{g}}(\sigma - s)\zeta(\sigma)\, ds\, d\sigma \\
&= \langle \widehat{(f\overline{g})}, \zeta\rangle \\
&= \langle f\overline{g}, \mu\rangle.
\end{aligned} \tag{1.15}$$

Now, if $g = f$, the left side of (1.15) is $\|f(A)v\|^2$, which is ≥ 0. Hence

$$\langle |f|^2, \mu \rangle \geq 0, \quad \text{for all } f \in \mathcal{S}(\mathbb{R}). \tag{1.16}$$

With this, we can establish:

Lemma 1.4. *The tempered distribution μ, defined by (1.10)–(1.11), is a positive measure on $\mathbb{R}$.*

Proof. Apply (1.16) with $f = \sqrt{F_{s,\sigma}}$, where

$$F_{s,\sigma}(\tau) = (4\pi s)^{-1/2} e^{-(\tau-\sigma)^2/4s}, \quad s > 0, \ \sigma \in \mathbb{R}.$$

Note that this is a fundamental solution to the heat equation. For each $s > 0$, $F_{s,0} * \mu$ is a positive function. We saw in Chap. 3 that $F_{s,0} * \mu$ converges to μ in $\mathcal{S}'(\mathbb{R})$ as $s \to 0$, so this implies that μ is a positive measure.

Now we can finish the proof of Proposition 1.3. From (1.15) we see that W has a unique continuous extension

$$W : L^2(\mathbb{R}, d\mu) \longrightarrow H, \tag{1.17}$$

and W is an isometry. Since v is assumed to be cyclic, the range of W must be dense in H, so W must be unitary. From (1.14) it follows that if $f \in \mathcal{S}(\mathbb{R})$, then

$$e^{isA} f(A) = f_s(A), \ \text{ with } f_s(\tau) = e^{is\tau} f(\tau). \tag{1.18}$$

Hence, for $f \in \mathcal{S}(\mathbb{R})$,

$$W^{-1} e^{isA} W f = W^{-1} f_s(A) v = e^{is\tau} f(\tau). \tag{1.19}$$

Since $\mathcal{S}(\mathbb{R})$ is dense in $L^2(\mathbb{R}, d\mu)$, this gives (1.9). Thus the spectral theorem for bounded, self-adjoint operators is proved.

Given (1.9), we have from (1.14) that

$$W^{-1} f(A) W g(x) = f(x) g(x), \quad f \in \mathcal{S}(\mathbb{R}), \ g \in L^2(\mathbb{R}, d\mu), \tag{1.20}$$

which justifies the notation $f(A)$ in (1.14).

Note that (1.9) implies

$$W^{-1} A W f(x) = x f(x), \quad f \in L^2(\mathbb{R}, d\mu), \tag{1.21}$$

since $(d/dt)U(t) = iAU(t)$. The essential supremum of x on $(\mathbb{R}, \mu)$ is equal to $\|A\|$. Thus μ has compact support in $\mathbb{R}$ if A is bounded. If a self-adjoint operator A has the representation (1.21), one says A has *simple spectrum*. It follows from Proposition 1.3 that A has simple spectrum if and only if it has a cyclic vector.

One can generalize the results above to a k-tuple of commuting, bounded, self-adjoint operators $A = (A_1, \dots, A_k)$. In that case, for $t = (t_1, \dots, t_k) \in \mathbb{R}^k$, set

$$U(t) = e^{it\cdot A}, \quad t \cdot A = t_1 A_1 + \cdots + t_k A_k. \tag{1.22}$$

The hypothesis that the A_j all commute implies $U(t) = U_1(t_1) \cdots U_k(t_k)$, where $U_j(s) = e^{isA_j}$. $U(t)$ in (1.22) continues to satisfy the properties (1.6) and (1.7); we have a k-parameter unitary group. As above, for $v \in H$, we set H_v equal to the closed linear span of $\{U(t)v : t \in \mathbb{R}^k\}$, and we say v is a cyclic vector provided $H_v = H$. Lemma 1.2 goes through in this case. Furthermore, for $f \in \mathcal{S}(\mathbb{R}^k)$, we can define

$$f(A) = (2\pi)^{-k/2} \int \hat{f}(t) e^{it\cdot A}\, dt, \tag{1.23}$$

and if H has a cyclic vector v, the proof of Proposition 1.3 generalizes, giving a unitary map $W : L^2(\mathbb{R}^k, d\mu) \to H$ such that

$$W^{-1} U(t) W f(x) = e^{it\cdot x} f(x), \quad f \in L^2(\mathbb{R}^k, d\mu), \quad t \in \mathbb{R}^k. \tag{1.24}$$

Therefore, Theorem 1.1 has the following extension

Proposition 1.5. *If $A = (A_1, \dots, A_k)$ is a k-tuple of commuting, bounded, self-adjoint operators on H, there is a measure space (Ω, μ), a unitary map $W : L^2(\Omega, d\mu) \to H$, and real-valued $a_j \in L^\infty(\Omega, d\mu)$ such that*

$$W^{-1} A_j W f(x) = a_j(x) f(x), \quad f \in L^2(\Omega, d\mu),\ 1 \le j \le k. \tag{1.25}$$

A bounded operator $B \in \mathcal{L}(H)$ is said to be normal provided B and B^* commute. Equivalently, if we set

$$A_1 = \frac{1}{2}(B + B^*), \quad A_2 = \frac{1}{2i}(B - B^*), \tag{1.26}$$

then $B = A_1 + iA_2$, and (A_1, A_2) is a 2-tuple of commuting, self-adjoint operators. Applying Proposition 1.5 and setting $b(x) = a_1(x) + i a_2(x)$, we have:

Corollary 1.6. *If $B \in \mathcal{L}(H)$ is a normal operator, there is a unitary map $W : L^2(\Omega, d\mu) \to H$ and a (complex-valued) $b \in L^\infty(\Omega, d\mu)$ such that*

$$W^{-1} B W f(x) = b(x) f(x), \quad f \in L^2(\Omega, d\mu). \tag{1.27}$$

In particular, Corollary 1.6 holds when $B = U$ is unitary. We next extend the spectral theorem to an unbounded, self-adjoint operator A on a Hilbert space H, whose domain $\mathcal{D}(A)$ is a dense linear subspace of H. This extension, due to von Neumann, uses von Neumann's unitary trick, described in (8.18)–(8.19) of

Appendix A. We recall that, for such A, the following three properties hold:

$$\begin{aligned} A \pm i : \mathcal{D}(A) &\longrightarrow H \text{ bijectively}, \\ U = (A - i)(A + i)^{-1} &\text{ is unitary on } H, \\ A &= i(I + U)(I - U)^{-1}, \end{aligned} \tag{1.28}$$

where the range of $I - U = 2i(A + i)^{-1}$ is $\mathcal{D}(A)$. Applying Corollary 1.6 to $B = U$, we have the following theorem:

Theorem 1.7. *If A is an unbounded, self-adjoint operator on a separable Hilbert space H, there is a measure space (Ω, μ), a unitary map $W : L^2(\Omega, d\mu) \to H$, and a real-valued measurable function a on Ω such that*

$$W^{-1}AWf(x) = a(x)f(x), \quad Wf \in \mathcal{D}(A). \tag{1.29}$$

In this situation, given $f \in L^2(\Omega, d\mu)$, Wf belongs to $\mathcal{D}(A)$ if and only if the right side of (1.29) belongs to $L^2(\Omega, d\mu)$.

The formula (1.29) is called the "spectral representation" of a self-adjoint operator A. Using it, we can extend the functional calculus defined by (1.14) as follows. For a Borel function $f : \mathbb{R} \to \mathbb{C}$, define $f(A)$ by

$$W^{-1}f(A)Wg(x) = f(a(x))g(x). \tag{1.30}$$

If f is a bounded Borel function, this is defined for all $g \in L^2(\Omega, d\mu)$ and provides a bounded operator $f(A)$ on H. More generally,

$$\mathcal{D}\big(f(A)\big) = \big\{Wg \in H : g \in L^2(\Omega, d\mu) \text{ and } f(a(x))g \in L^2(\Omega, d\mu)\big\}. \tag{1.31}$$

In particular, we can define e^{itA}, for unbounded, self-adjoint A, by

$$W^{-1}e^{itA}Wg = e^{ita(x)}g(x)$$

Then e^{itA} is a strongly continuous unitary group, and we have the following result, known as Stone's theorem (stated as Proposition 9.5 in Appendix A):

Proposition 1.8. *If A is self-adjoint, then iA generates a strongly continuous, unitary group, $U(t) = e^{itA}$.*

Note that Lemma 1.2 and Proposition 1.3 are proved for a strongly continuous, unitary group $U(t) = e^{itA}$, without the hypothesis that A be bounded. This yields the following analogue of (1.2):

$$W^{-1}U(t)Wf(x) = e^{ita(x)}f(x), \quad f \in L^2(\Omega, d\mu), \tag{1.32}$$

for this more general class of unitary groups. Sometimes a direct construction, such as by PDE methods, of $U(t)$ is fairly easy. In such a case, the use of $U(t)$ can be a more convenient tool than the unitary trick involving (1.28).

We say a self-adjoint operator A is positive, $A \geq 0$, provided $(Au, u) \geq 0$, for all $u \in \mathcal{D}(A)$. In terms of the spectral representation, this says we have (1.29) with $a(x) \geq 0$ on Ω. In such a case, e^{-tA} is bounded for $t \geq 0$, even for complex t with Re $t \geq 0$, and also defines a strongly continuous semigroup. This proves Proposition 9.4 of Appendix A.

Given a self-adjoint operator A and a Borel set $S \subset \mathbb{R}$, define $P(S) = \chi_S(A)$, that is, using (1.29),

$$W^{-1}P(S)Wg = \chi_S(a(x))g(x), \quad g \in L^2(\Omega, d\mu), \tag{1.33}$$

where χ_S is the characteristic function of S. Then each $P(S)$ is an orthogonal projection. Also, if $S = \bigcup_{j\geq 1} S_j$ is a countable union of disjoint Borel sets S_j, then, for each $u \in H$,

$$\lim_{n\to\infty} \sum_{j=1}^{n} P(S_j)u = P(S)u, \tag{1.34}$$

with convergence in the H-norm. This is equivalent to the statement that

$$\sum_{j=1}^{n} \chi_{S_j}(a(x))g \to \chi_S(a(x))g \text{ in } L^2\text{-norm}, \quad \text{for each } g \in L^2(\Omega, d\mu),$$

which in turn follows from Lebesgue's dominated convergence theorem. By (1.34), $P(\cdot)$ is a strongly countably additive, projection-valued measure. Then (1.30) yields

$$f(A) = \int f(\lambda)\, P(d\lambda). \tag{1.35}$$

$P(\cdot)$ is called the *spectral measure* of A.

One useful formula for the spectral measure is given in terms of the jump of the resolvent $R_\lambda = (\lambda - A)^{-1}$, across the real axis. We have the following

Proposition 1.9. *For bounded, continuous $f : \mathbb{R} \to \mathbb{C}$,*

$$f(A)u = \lim_{\varepsilon\searrow 0} \frac{1}{2\pi i} \int_{-\infty}^{\infty} f(\lambda)\Big[(\lambda - i\varepsilon - A)^{-1} - (\lambda + i\varepsilon - A)^{-1}\Big]u\, d\lambda. \tag{1.36}$$

Proof. Since $W^{-1}f(A)W$ is multiplication by $f(a(x))$, (1.36) follows from the fact that

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\varepsilon f(\lambda)}{(\lambda-a(x))^2+\varepsilon^2}\,d\lambda \longrightarrow f(a(x)), \tag{1.37}$$

pointwise and boundedly, as $\varepsilon \searrow 0$.

An important class of operators $f(A)$ are the fractional powers $f(A) = A^\alpha,\ \alpha \in (0,\infty)$, defined by (1.30)–(1.31), with $f(\lambda) = \lambda^\alpha$, provided $A \geq 0$. Note that if $g \in C([0,\infty))$ satisfies $g(0) = 1,\ g(\lambda) = O(\lambda^{-\alpha})$ as $\lambda \to \infty$, then, for $u \in H$,

$$u \in \mathcal{D}(A^\alpha) \Longleftrightarrow \|A^\alpha g(\varepsilon A)u\|_H \text{ is bounded,}\quad \text{for } \varepsilon \in (0,1], \tag{1.38}$$

as follows easily from the characterization (1.31) and Fatou's lemma. We note that Proposition 2.2 of Chap. 4 applies to $\mathcal{D}(A^\alpha)$, describing it as an interpolation space.

We particularly want to identify $\mathcal{D}(A^{1/2})$, when A is a positive, self-adjoint operator on a Hilbert space H constructed by the Friedrichs method, as described in Proposition 8.7 of Appendix A. Recall that this arises as follows. One has a Hilbert space H_1, a continuous injection $J : H_1 \to H$ with dense range, and one defines A by

$$\big(A(Ju), Jv\big)_H = (u,v)_{H_1}, \tag{1.39}$$

with

$$\mathcal{D}(A) = \big\{Ju \in JH_1 \subset H : v \mapsto (u,v)_{H_1} \text{ is continuous in } Jv,\ \text{in the } H\text{-norm}\big\}. \tag{1.40}$$

We establish the following.

Proposition 1.10. *If A is obtained by the Friedrichs extension method (1.39)–(1.40), then*

$$\mathcal{D}(A^{1/2}) = J(H_1) \subset H. \tag{1.41}$$

Proof. $\mathcal{D}(A^{1/2})$ consists of elements of H that are limits of sequences in $\mathcal{D}(A)$, in the norm $\|A^{1/2}u\|_H + \|u\|_H$. As shown in the proof of Proposition 8.7 in Appendix A, $\mathcal{D}(A) = \mathcal{R}(JJ^*)$. Now

$$\|A^{1/2}JJ^*f\|_H^2 = (AJJ^*f, JJ^*f)_H = \|J^*f\|_{H_1}^2. \tag{1.42}$$

Thus a sequence (JJ^*f_n) converges in the $\mathcal{D}(A^{1/2})$-norm (to an element g) if and only if (J^*f_n) converges in the H_1-norm (to an element u), in which case $g = Ju$. Since $J^* : H \to H_1$ has dense range, precisely all $u \in H_1$ arise as limits of such (J^*f_n), so the proposition is proved.

Exercises

1. The definition (1.33) of the spectral measure $P(\cdot)$ of a self-adjoint operator A depends a priori on a choice of the spectral representation of A. Show that any two spectral representations of A yield the same spectral measure.
(*Hint*: For $f \in \mathcal{S}(\mathbb{R})$, $f(A)$ is well defined by (1.14), or alternatively by (1.36).)

2. Self-adjoint differential operators

In this section we present some examples of differential operators on a manifold Ω which, with appropriately specified domains, give unbounded, self-adjoint operators on $L^2(\Omega, dV)$, dV typically being the volume element determined by a Riemannian metric on Ω.

We begin with self-adjoint operators arising from the Laplacian, making use of material developed in Chap. 5. Let $\overline{\Omega}$ be a smooth, compact Riemannian manifold with boundary, or more generally the closure of an open subset Ω of a compact manifold M without boundary. Then, as shown in Chap. 5,

$$I - \Delta : H_0^1(\Omega) \longrightarrow H_0^1(\Omega)^* \tag{2.1}$$

is bijective, with inverse we denote T; if we restrict T to $L^2(\Omega)$,

$$T : L^2(\Omega) \longrightarrow L^2(\Omega) \text{ is compact and self-adjoint.} \tag{2.2}$$

Denote by $\mathcal{R}(T)$ the image of $L^2(\Omega)$ under T. We can apply Proposition 8.2 of Appendix A to deduce the following

Proposition 2.1. *If Ω is a region in a compact Riemannian manifold M, then Δ is self-adjoint on $L^2(\Omega)$, with domain $\mathcal{D}(\Delta) = \mathcal{R}(T) \subset H_0^1(\Omega)$ described above.*

For a further description of $\mathcal{D}(\Delta)$, note that

$$\mathcal{D}(\Delta) = \{u \in H_0^1(\Omega) : \Delta u \in L^2(\Omega)\}. \tag{2.3}$$

If $\partial\Omega$ is smooth, we can apply the regularity theory of Chap. 5 to obtain

$$\mathcal{D}(\Delta) = H_0^1(\Omega) \cap H^2(\Omega). \tag{2.4}$$

Instead of relying on Proposition 8.2, we could use the Friedrichs construction, given in Proposition 8.7 of Appendix A. This construction can be applied more generally. Let Ω be any Riemannian manifold, with Laplace operator Δ. We can define $H_0^1(\Omega)$ to be the closure of $C_0^\infty(\Omega)$ in the space $\{u \in L^2(\Omega) : du \in L^2(\Omega, \Lambda^1)\}$. The inner product on $H_0^1(\Omega)$ is

$$(u, v)_1 = (u, v)_{L^2} + (du, dv)_{L^2}. \tag{2.5}$$

We have a natural inclusion $H_0^1(\Omega) \hookrightarrow L^2(\Omega)$, and the Friedrichs method gives a self-adjoint operator A on $L^2(\Omega)$ such that

$$(Au, v)_{L^2} = (u, v)_1, \quad \text{for } u \in \mathcal{D}(A),\ v \in H_0^1(\Omega), \tag{2.6}$$

with

$$\mathcal{D}(A) = \{u \in H_0^1(\Omega) : v \mapsto (u, v)_1 \text{ extends from } H_0^1(\Omega) \to \mathbb{C} \text{ to a continuous linear functional } L^2(\Omega) \to \mathbb{C}\}, \tag{2.7}$$

that is,

$$\mathcal{D}(A) = \{u \in H_0^1(\Omega) : \exists f \in L^2(\Omega) \text{ such that } (u, v)_1 = (f, v)_{L^2}, \forall v \in H_0^1(\Omega)\}. \tag{2.8}$$

Integrating (2.5) by parts for $v \in C_0^\infty(\Omega)$, we see that $A = I - \Delta$ on $\mathcal{D}(A)$, so we have a self-adjoint extension of Δ in this general setting, with domain again described by (2.3).

The process above gives one self-adjoint extension of Δ, initially defined on $C_0^\infty(\Omega)$. It is not always the only self-adjoint extension. For example, suppose $\overline{\Omega}$ is compact with smooth boundary; consider $H^1(\Omega)$, with inner product (2.5), and apply the Friedrichs extension procedure. Again we have a self-adjoint operator A, extending $I - \Delta$, with (2.8) replaced by

$$\mathcal{D}(A) = \{u \in H^1(\Omega) : \exists f \in L^2(\Omega) \text{ such that } (u, v)_1 = (f, v)_{L^2}, \forall v \in H^1(\Omega)\}. \tag{2.9}$$

In this case, Proposition 7.2 of Chap. 5 yields the following

Proposition 2.2. *If $\overline{\Omega}$ is a smooth, compact manifold with boundary and Δ the self-adjoint extension just described, then*

$$\mathcal{D}(\Delta) = \{u \in H^2(\Omega) : \partial_\nu u = 0 \text{ on } \partial\Omega\}. \tag{2.10}$$

In case (2.10), we say $\mathcal{D}(\Delta)$ is given by the Neumann boundary condition, while in case (2.4) we say $\mathcal{D}(\Delta)$ is given by the Dirichlet boundary condition.

In both cases covered by Propositions 2.1 and 2.2, $(-\Delta)^{1/2}$ is defined as a self-adjoint operator. We can specify its domain using Proposition 1.10, obtaining the next result:

Proposition 2.3. *In case* (2.3), $\mathcal{D}((-\Delta)^{1/2}) = H_0^1(\Omega)$; *in case (2.10),* $\mathcal{D}((-\Delta)^{1/2}) = H^1(\Omega)$.

Though Δ on $C_0^\infty(\Omega)$ has several self-adjoint extensions when Ω has a boundary, it has only one when Ω is a complete Riemannian manifold. This is a classical result, due to Roelcke; we present an elegant proof due to Chernoff [Chn]. When an unbounded operator A_0 on a Hilbert space H, with domain $\mathcal{D}_0$, has exactly one self-adjoint extension, namely the closure of A_0, we say A_0 is *essentially self-adjoint* on $\mathcal{D}_0$.

Proposition 2.4. *If Ω is a complete Riemannian manifold, then Δ is essentially self-adjoint on $C_0^\infty(\Omega)$. Thus the self-adjoint extension with domain given by (2.3) is the closure of Δ on $C_0^\infty(\Omega)$.*

Proof. We will obtain this as a consequence of Proposition 9.6 of Appendix A, which states the following. Let $U(t) = e^{itA}$ be a unitary group on a Hilbert space H which leaves invariant a dense linear space $\mathcal{D}$; $U(t)\mathcal{D} \subset \mathcal{D}$. If A is an extension of A_0 and $A_0 : \mathcal{D} \to \mathcal{D}$, then A_0 and all its powers are essentially self-adjoint on $\mathcal{D}$.

In this case, $U(t)$ will be the solution operator for a wave equation, and we will exploit finite propagation speed. Set

$$iA_0 = \begin{pmatrix} 0 & I \\ \Delta - I & 0 \end{pmatrix}, \quad \mathcal{D}(A_0) = C_0^\infty(\Omega) \oplus C_0^\infty(\Omega). \tag{2.11}$$

The group $U(t)$ will be the solution operator for the wave equation

$$U(t)\begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} u(t) \\ u_t(t) \end{pmatrix}, \tag{2.12}$$

where $u(t,x)$ is determined by

$$\frac{\partial^2 u}{\partial t^2} - (\Delta - 1)u = 0; \quad u(0,x) = f,\ u_t(0,x) = g$$

It was shown in §2 of Chap. 6 that $U(t)$ is a unitary group on $H = H_0^1(\Omega) \oplus L^2(\Omega)$; its generator is an extension of (2.11), and finite propagation speed implies that $U(t)$ preserves $C_0^\infty(\Omega) \oplus C_0^\infty(\Omega)$ for all t, provided Ω is complete. Thus each A_0^k is essentially self-adjoint on this space. Since

$$-A_0^2 = \begin{pmatrix} \Delta - I & 0 \\ 0 & \Delta - I \end{pmatrix}, \tag{2.13}$$

we have the proof of Proposition 2.3. Considering A_0^{2k}, we deduce furthermore that each power Δ^k is essentially self-adjoint on $C_0^\infty(\Omega)$, when Ω is complete.

Though Δ is not essentially self-adjoint on $C_0^\infty(\Omega)$ when $\overline{\Omega}$ is compact, we do have such results as the following:

Proposition 2.5. *If $\overline{\Omega}$ is a smooth, compact manifold with boundary, then Δ is essentially self-adjoint on*

$$\{u \in C^\infty(\overline{\Omega}) : u = 0 \text{ on } \partial\Omega\}, \tag{2.14}$$

its closure having domain described by (2.3). *Also,* Δ *is essentially self-adjoint on*

$$\{u \in C^\infty(\overline{\Omega}) : \partial_\nu u = 0 \text{ on } \partial\Omega\}, \tag{2.15}$$

its closure having domain described by (2.10).

Proof. It suffices to note the simple facts that the closure of (2.14) in $H^2(\Omega)$ is (2.3) and the closure of (2.15) in $H^2(\Omega)$ is (2.10).

We note that when $\overline{\Omega}$ is a smooth, compact Riemannian manifold with boundary, and $\mathcal{D}(\Delta)$ is given by the Dirichlet boundary condition, then

$$\bigcap_{j=1}^{\infty} \mathcal{D}(\Delta^j) = \{u \in C^\infty(\overline{\Omega}) : \Delta^k u = 0 \text{ on } \partial\Omega,\ k = 0, 1, 2, \dots\}, \tag{2.16}$$

and when $\mathcal{D}(\Delta)$ is given by the Neumann boundary condition, then

$$\bigcap_{j=1}^{\infty} \mathcal{D}(\Delta^j) = \{u \in C^\infty(\overline{\Omega}) : \partial_\nu(\Delta^k u) = 0 \text{ on } \partial\Omega,\ k \geq 0\}. \tag{2.17}$$

We now derive a result that to some degree amalgamates Propositions 2.4 and 2.5. Let $\overline{\Omega}$ be a smooth Riemannian manifold with boundary, and set

$$C_c^\infty(\overline{\Omega}) = \{u \in C^\infty(\overline{\Omega}) : \text{supp } u \text{ is compact in } \overline{\Omega}\}; \tag{2.18}$$

we do not require elements of this space to vanish on $\partial\Omega$. We say that $\overline{\Omega}$ is complete if it is complete as a metric space.

Proposition 2.6. *If* $\overline{\Omega}$ *is a smooth Riemannian manifold with boundary which is complete, then* Δ *is essentially self-adjoint on*

$$\{u \in C_c^\infty(\overline{\Omega}) : u = 0 \text{ on } \partial\Omega\}. \tag{2.19}$$

In this case, the closure has domain given by (2.3).

Proof. Consider the following linear subspace of (2.19):

$$\mathcal{D}_0 = \{u \in C_c^\infty(\overline{\Omega}) : \Delta^j u = 0 \text{ on } \partial\Omega \text{ for } j = 0, 1, 2, \dots\}. \tag{2.20}$$

Let $U(t)$ be the unitary group on $H_0^1(\Omega) \oplus L^2(\Omega)$ defined as in (2.12), with u also satisfying the Dirichlet boundary condition, $u(t, x) = 0$ for $x \in \partial\Omega$. Then, by finite propagation speed, $U(t)$ preserves $\mathcal{D}_0 \oplus \mathcal{D}_0$, provided $\overline{\Omega}$ is complete, so as in the proof of Proposition 2.4, we deduce that Δ is essentially self-adjoint on $\mathcal{D}_0$; a fortiori it is essentially self-adjoint on the space (2.19).

By similar reasoning, we can show that if $\overline{\Omega}$ is complete, then Δ is essentially self-adjoint on

$$\{u \in C_c^\infty(\overline{\Omega}) : \partial_\nu u = 0 \text{ on } \partial\Omega\}. \tag{2.21}$$

The results of this section so far have involved only the Laplace operator Δ. It is also of interest to look at Schrödinger operators, of the form $-\Delta + V$, where the "potential" $V(x)$ is a real-valued function. In this section we will restrict attention to the case $V \in C^\infty(\Omega)$ and we will also suppose that V is bounded from below. By adding a constant to $-\Delta + V$, we may as well suppose

$$V(x) \geq 1 \quad \text{on } \Omega. \tag{2.22}$$

We can define a Hilbert space $H^1_{V0}(\Omega)$ to be the closure of $C_0^\infty(\Omega)$ in the space

$$H^1_V(\Omega) = \{u \in L^2(\Omega) : du \in L^2(\Omega, \Lambda^1),\ V^{1/2}u \in L^2(\Omega)\}, \tag{2.23}$$

with inner product

$$(u, v)_{1,V} = (du, dv)_{L^2} + (Vu, v)_{L^2}. \tag{2.24}$$

Then there is a natural injection $H^1_{V0}(\Omega) \hookrightarrow L^2(\Omega)$, and the Friedrichs extension method provides a self-adjoint operator A. Integration by parts in (2.24), with $v \in C_0^\infty(\Omega)$, shows that such A is an extension of $-\Delta + V$. For this self-adjoint extension, we have

$$\mathcal{D}(A^{1/2}) = H^1_{V0}(\Omega). \tag{2.25}$$

In case $\overline{\Omega}$ is a smooth, compact Riemannian manifold with boundary and $V \in C^\infty(\overline{\Omega})$, one clearly has $H^1_{V0}(\Omega) = H^1_0(\Omega)$. In such a case, we have an immediate extension of Proposition 2.1, including the characterization (2.4) of $\mathcal{D}(-\Delta + V)$. One can also easily extend Proposition 2.2 to $-\Delta + V$ in this case. It is of substantial interest that Proposition 2.4 also extends, as follows:

Proposition 2.7. *If Ω is a complete Riemannian manifold and the function $V \in C^\infty(\Omega)$ satisfies $V \geq 1$, then $-\Delta + V$ is essentially self-adjoint on $C_0^\infty(\Omega)$.*

Proof. We can modify the proof of Proposition 2.4; replace $\Delta - 1$ by $\Delta - V$ in (2.11) and (2.12). Then $U(t)$ gives a unitary group on $H^1_{V0}(\Omega) \oplus L^2(\Omega)$, and the finite propagation speed argument given there goes through. As before, all powers of $-\Delta + V$ are essentially self-adjoint on $C_0^\infty(\Omega)$.

Some important classes of potentials V have singularities and are not bounded below. In §7 we return to this, in a study of the quantum mechanical Coulomb problem.

We record here an important compactness property when $V \in C^\infty(\Omega)$ tends to $+\infty$ at infinity in Ω

Proposition 2.8. *If the Friedrichs extension method described above is used to construct the self-adjoint operator $-\Delta + V$ for smooth $V \geq 1$, as above, and if $V \to +\infty$ at infinity (i.e., for each $N < \infty$, $\Omega_N = \{x \in \Omega : V(x) \leq N\}$ is compact), then $-\Delta + V$ has compact resolvent.*

Proof. Given (2.25), it suffices to prove that the injection $H^1_{V0}(\Omega) \to L^2(\Omega)$ is compact, under the current hypotheses on V. Indeed, if $\{u_n\}$ is bounded in $H^1_{V0}(\Omega)$, with inner product (2.24), then $\{du_n\}$ and $\{V^{1/2}u_n\}$ are bounded in $L^2(\Omega)$. By Rellich's theorem and a diagonal argument, one has a subsequence $\{u_{n_k}\}$ whose restriction to each Ω_N converges in $L^2(\Omega_N)$-norm. The boundedness of $\{V^{1/2}u_n\}$ in $L^2(\Omega)$ then gives convergence of this subsequence in $L^2(\Omega)$-norm, proving the proposition.

The following result extends Proposition 2.4 of Chap. 5

Proposition 2.9. *Assume that Ω is connected and that either Ω is compact or $V \to +\infty$ at infinity. Denote by λ_0 the first eigenvalue of $-\Delta + V$. Then a λ_0-eigenfunction of $-\Delta + V$ is nowhere vanishing on Ω. Consequently, the λ_0-eigenspace is one-dimensional.*

Proof. Let u be a λ_0-eigenfunction of $-\Delta + V$. As in the proof of Proposition 2.4 of Chap. 5, we can write $u = u^+ + u^-$, where $u^+(x) = u(x)$ for $u(x) > 0$ and $u^-(x) = u(x)$ for $u(x) \leq 0$, and the variational characterization of the λ_0-eigenspace implies that $u^\pm$ are eigenfunctions (if nonzero). Hence it suffices to prove that if u is a λ_0-eigenfunction and $u(x) \geq 0$ on Ω, then $u(x) > 0$ on Ω. To this end, write

$$u(x) = e^{t(\Delta - V + \lambda_0)}u(x) = \int_\Omega p_t(x,y)u(y)\, dV(y)$$

We see that this forces $p_t(x,y) = 0$ for all $t > 0$, when

$$x \in \Sigma = \{x : u(x) = 0\}, \quad y \in \overline{\mathcal{O}}, \quad \mathcal{O} = \{x : u(x) > 0\},$$

since $p_t(x,y)$ is smooth and ≥ 0. The strong maximum principle (see Exercise 3 in §1 of Chap. 6 forces $\Sigma = \emptyset$.

Exercises

1. Let $H^1_V(\Omega)$ be the space (2.23). If $V \geq 1$ belongs to $C^\infty(\Omega)$, show that the Friedrichs extension also defines a self-adjoint operator A_1, equal to $-\Delta + V$ on $C_0^\infty(\Omega)$, such that $\mathcal{D}(A_1^{1/2}) = H^1_V(\Omega)$. If Ω is complete, show that this operator coincides with the extension A defined in (2.25). Conclude that, in this case, $H^1_V(\Omega) = H^1_{V0}(\Omega)$.

2. Let Ω be complete, $V \geq 1$ smooth. Show that if A is the self-adjoint extension of $-\Delta + V$ described in Proposition 2.7, then

$$\mathcal{D}(A) = \{u \in L^2(\Omega) : -\Delta u + Vu \in L^2(\Omega)\}, \tag{2.26}$$

   where a priori we regard $-\Delta u + Vu$ as an element of $\mathcal{D}'(\Omega)$.
3. Define $T : L^2(\Omega) \to L^2(\Omega, \Lambda^1) \oplus L^2(\Omega)$ by $\mathcal{D}(T) = H^1_{V0}(\Omega)$, $Tu = (du, V^{1/2}u)$. Show that

$$\mathcal{D}(T^*) = \{(v_1, v_2) \in L^2(\Omega, \Lambda^1) \oplus L^2(\Omega) : \delta v_1 \in L^2(\Omega),\ V^{1/2} v_2 \in L^2(\Omega)\}. \tag{2.27}$$

   Show that T^*T is equal to the self-adjoint extension A of $-\Delta + V$ defined by the Friedrichs extension, as in (2.25).
4. If Ω is complete, show that the self-adjoint extension A of $-\Delta + V$ in Proposition 2.7 satisfies

$$\mathcal{D}(A) = \{u \in L^2(\Omega) : \Delta u \in L^2(\Omega),\ Vu \in L^2(\Omega)\}. \tag{2.28}$$

   (*Hint*: Denote the right side by $\mathcal{W}$. Use Exercise 3 and $A = T^*T$ to show that $\mathcal{D}(A) \subset \mathcal{W}$. Use Exercise 2 to show that $\mathcal{W} \subset \mathcal{D}(A)$.)
5. Let $D = -i\, d/dx$ on $C^\infty(\mathbb{R})$, and let $B(x) \in C^\infty(\mathbb{R})$ be real-valued. Define the unbounded operator L on $L^2(\mathbb{R})$ by

$$\mathcal{D}(L) = \{u \in L^2(\mathbb{R}) : Du \in L^2(\mathbb{R}),\ Bu \in L^2(\mathbb{R})\}, \quad Lu = Du + iB(x)u. \tag{2.29}$$

   Show that $L^* = D - iB$, with

$$\mathcal{D}(L^*) = \{u \in L^2(\mathbb{R}) : Du - iBu \in L^2(\mathbb{R})\}$$

   Deduce that $A_0 = L^*L$ is given by $A_0 u = D^2 u + B^2 u + B'(x)u$ on

$$\mathcal{D}(A_0) = \{u \in L^2(\mathbb{R}) : Du \in L^2(\mathbb{R}),\ Bu \in L^2(\mathbb{R}),\ D^2u + B^2u + B'(x)u \in L^2(\mathbb{R})\}$$

6. Suppose that $|B'(x)| \leq \vartheta B(x)^2 + C$, for some $\vartheta < 1,\ C < \infty$. Show that

$$\mathcal{D}(A_0) = \{u \in L^2(\mathbb{R}) : D^2u + (B^2 + B')u \in L^2(\mathbb{R})\}$$

   (*Hint*: Apply Exercise 2 to $D^2 + (B^2 + B') = A$, and show that $\mathcal{D}(A^{1/2})$ is given by $\mathcal{D}(L)$, defined in (2.29).)
7. In the setting of Exercise 6, show that the operator L of Exercise 5 is closed. (*Hint*: $L^*\overline{L} = A$ is a self-adjoint extension of $D^2 + (B^2 + B')$. Show that $\mathcal{D}(A_1^{1/2}) = \mathcal{D}(L)$ and also $= \mathcal{D}(\overline{L})$.) Also show that $\mathcal{D}(L^*) = \mathcal{D}(L)$ in this case.

3. Heat asymptotics and eigenvalue asymptotics

In this section we will study the asymptotic behavior of the eigenvalues of the Laplace operator on a compact Riemannian manifold, with or without boundary.

We begin with the boundaryless case. Let M be a compact Riemannian manifold without boundary, of dimension n. In §13 of Chap. 7 we have constructed a parametrix for the solution operator $e^{t\Delta}$ of the heat equation

$$\left(\frac{\partial}{\partial t} - \Delta\right)u = 0 \text{ on } \mathbb{R}^+ \times M, \quad u(0,x) = f(x) \tag{3.1}$$

and deduced that

$$\operatorname{Tr} e^{t\Delta} \sim t^{-n/2}\left(a_0 + a_1 t + a_2 t^2 + \cdots\right), \quad t \searrow 0, \tag{3.2}$$

for certain constants a_j. In particular,

$$a_0 = (4\pi)^{-n/2} \operatorname{vol} M. \tag{3.3}$$

This is related to the behavior of the eigenvalues of Δ as follows. Let the eigenvalues of $-\Delta$ be $0 = \lambda_0 \le \lambda_1 \le \lambda_2 \le \cdots \nearrow \infty$. Then (3.2) is equivalent to

$$\sum_{j=0}^{\infty} e^{-t\lambda_j} \sim t^{-n/2}\left(a_0 + a_1 t + a_2 t^2 + \cdots\right), \quad t \searrow 0. \tag{3.4}$$

We will relate this to the counting function

$$N(\lambda) \sim \#\{\lambda_j : \lambda_j \le \lambda\}, \tag{3.5}$$

establishing the following:

Theorem 3.1. *The eigenvalues* $\{\lambda_j\}$ *of* $-\Delta$ *on the compact Riemannian manifold* M *have the behavior*

$$N(\lambda) \sim C(M)\lambda^{n/2}, \quad \lambda \to +\infty, \tag{3.6}$$

with

$$C(M) = \frac{a_0}{\Gamma(\frac{n}{2}+1)} = \frac{vol\ M}{\Gamma(\frac{n}{2}+1)(4\pi)^{n/2}}. \tag{3.7}$$

That (3.6) follows from (3.4) is a special case of a result known as *Karamata's Tauberian theorem*. The following neat proof follows one in [Si3]. Let μ be a positive (locally finite) Borel measure on $[0,\infty)$; in the example above, $\mu\big([0,\lambda]\big) = N(\lambda)$.

Proposition 3.2. *If* μ *is a positive measure on* $[0,\infty)$, $\alpha \in (0,\infty)$*, then*

$$\int_0^\infty e^{-t\lambda}\, d\mu(\lambda) \sim a t^{-\alpha}, \quad t \searrow 0, \tag{3.8}$$

implies

$$\int_0^x d\mu(\lambda) \sim bx^\alpha, \quad x \nearrow \infty, \tag{3.9}$$

with

$$b = \frac{a}{\Gamma(\alpha+1)}. \tag{3.10}$$

Proof. Let $d\mu_t$ be the measure given by $\mu_t(A) = t^\alpha \mu(t^{-1}A)$, and let $d\nu(\lambda) = \alpha\lambda^{\alpha-1}d\lambda$; then $\nu_t = \nu$. The hypothesis (3.8) becomes

$$\lim_{t\to 0} \int e^{-\lambda}\, d\mu_t(\lambda) = b \int e^{-\lambda}\, d\nu(\lambda), \tag{3.11}$$

with b given by (3.10), and the desired conclusion becomes

$$\lim_{t\to 0} \int \chi(\lambda)\, d\mu_t(\lambda) = b \int \chi(\lambda)\, d\nu(\lambda) \tag{3.12}$$

when χ is the characteristic function of $[0, 1]$. It would suffice to show that (3.12) holds for all continuous $\chi(\lambda)$ with compact support in $[0, \infty)$.

From (3.11) we deduce that the measures $e^{-\lambda}d\mu_t$ are uniformly bounded, for $t \in (0, 1]$. Thus (3.12) follows if we can establish

$$\lim_{t\to 0} \int g(\lambda)e^{-\lambda}\, d\mu_t(\lambda) = b \int g(\lambda)e^{-\lambda}\, d\nu(\lambda), \tag{3.13}$$

for g in a dense subspace of $C_0(\mathbb{R}^+)$, the space of continuous functions on $[0, \infty)$ that vanish at infinity. Indeed, the hypothesis implies that (3.13) holds for all g in $\mathfrak{A}$, the space of finite, linear combinations of functions of $\lambda \in [0, \infty)$ of the form $\varphi_s(\lambda) = e^{-s\lambda}$, $s \in (0, \infty)$, as can be seen by dilating the variables in (3.11). By the Stone-Weierstrass theorem, $\mathfrak{A}$ is dense in $C_o(\mathbb{R}^+)$, so the proof is complete.

We next want to establish similar results on $N(\lambda)$ for the Laplace operator Δ on a compact manifold $\overline{\Omega}$ with boundary, with Dirichlet boundary condition. At the end of §13 in Chap. 7 we sketched a construction of a parametrix for $e^{t\Delta}$ in this case which, when carried out, would yield an expansion

$$\operatorname{Tr} e^{t\Delta} \sim t^{-n/2}\big(a_0 + a_{1/2}t^{1/2} + a_1 t + \cdots\big), \quad t \searrow 0, \tag{3.14}$$

extending (3.2). However, we will be able to verify the hypothesis of Proposition 3.2 with less effort than it would take to carry out the details of this construction, and for a much larger class of domains.

For simplicity, we will restrict attention to bounded domains in $\mathbb{R}^n$ and to the flat Laplacian, though more general cases can be handled similarly. Now, let Ω be an arbitrary bounded, open subset of $\mathbb{R}^n$, with closure $\overline{\Omega}$. The Laplace operator on Ω, with Dirichlet boundary condition, was studied in §5 of Chap. 5

Lemma 3.3. *For any bounded, open $\Omega \subset \mathbb{R}^n$, Δ with Dirichlet boundary condition, $e^{t\Delta}$ is trace class for all $t > 0$.*

Proof. Let $\overline{\Omega} \subset B$, a large open ball. Then the variational characterization of eigenvalues shows that the eigenvalues $\lambda_j(\Omega)$ of $-\Delta$ on Ω and $\lambda_j(B)$ of $L = -\Delta$ on B, both arranged in increasing order, have the relation

$$\lambda_j(\Omega) \geq \lambda_j(B). \tag{3.15}$$

But we know that e^{-tL} has integral kernel in $C^\infty(\overline{B} \times \overline{B})$ for each $t > 0$, hence is trace class. Since $e^{-t\lambda_j(\Omega)} \leq e^{-t\lambda_j(B)}$, this implies that the positive self-adjoint operator $e^{t\Delta}$ is also trace class.

Limiting arguments, which we leave to the reader, allow one to show that, even in this generality, if $H(t,x,y) \in C^\infty(\Omega \times \Omega)$ is, for fixed $t > 0$, the integral kernel of $e^{t\Delta}$ on $L^2(\Omega)$, then

$$\operatorname{Tr} e^{t\Delta} = \int_\Omega H(t,x,x)\,dx. \tag{3.16}$$

See Exercises 1–5 at the end of this section.

Proposition 3.4. *If Ω is a bounded, open subset of $\mathbb{R}^n$ and Δ has the Dirichlet boundary condition, then*

$$Tr\, e^{t\Delta} \sim (4\pi t)^{-n/2}\, vol\,\Omega, \quad t \searrow 0. \tag{3.17}$$

Proof. We will compare $H(t,x,y)$ with $H_0(t,x,y) = (4\pi t)^{-n/2} e^{|x-y|^2/4t}$, the free-space heat kernel. Let $E(t,x,y) = H_0(t,x,y) - H(t,x,y)$. Then, for fixed $y \in \Omega$,

$$\frac{\partial E}{\partial t} - \Delta_x E = 0 \text{ on } \mathbb{R}^+ \times \Omega, \quad E(0,x,y) = 0, \tag{3.18}$$

and

$$E(t,x,y) = H_0(t,x,y), \text{ for } x \in \partial\Omega. \tag{3.19}$$

To make simple sense out of (3.19), one might assume that every point of $\partial\Omega$ is a regular boundary point, though a further limiting argument can be made to lift such a restriction. The maximum principle for solutions to the heat equation implies

$$0 \le E(t,x,y) \le \sup_{0\le s\le t, z\in\Omega} H_0(s,z,y) \le \sup_{0\le s\le t} (4\pi s)^{-n/2}\, e^{-\delta(y)^2/4s}, \tag{3.20}$$

where $\delta(y) = \operatorname{dist}(y, \partial\Omega)$. Now the function

$$\psi_\delta(s) = (4\pi s)^{-n/2} e^{-\delta^2/4s}$$

on $(0,\infty)$ vanishes at 0 and ∞ and has a unique maximum at $s = \delta^2/2n$; we have $\psi_\delta(\delta^2/2n) = C_n\delta^{-n}$. Thus

$$0 \le E(t,x,y) \le \max\Big((4\pi t)^{-n/2} e^{-\delta(y)^2/4t}, C_n\delta(y)^{-n}\Big). \tag{3.21}$$

Of course, $E(t,x,y) \le H_0(t,x,y)$ also.

Now, let $\overline{\mathcal{O}} \subset\subset \Omega$ be such that $\operatorname{vol}(\Omega \setminus \mathcal{O}) < \varepsilon$. For t small enough, namely for $t \le \delta_1^2/2n$ where $\delta_1 = \operatorname{dist}(\overline{\mathcal{O}}, \partial\Omega)$, we have

$$0 \le E(t,x,x) \le (4\pi t)^{-n/2} e^{-\delta(x)^2/4t}, \quad x \in \mathcal{O}, \tag{3.22}$$

while of course $0 \le E(t,x,x) \le (4\pi t)^{-n/2}$, for $x \in \Omega \setminus \mathcal{O}$. Therefore,

$$\limsup_{t\to 0}\, (4\pi t)^{n/2} \int_\Omega E(t,x,x)\, dx \le \varepsilon, \tag{3.23}$$

so

$$\begin{aligned} \operatorname{vol}\Omega - \varepsilon &\le \liminf_{t\to 0}\, (4\pi t)^{n/2} \int_\Omega H(t,x,x)\, dx \\ &\le \limsup_{t\to 0}\, (4\pi t)^{n/2} \int_\Omega H(t,x,x)\, dx \le \operatorname{vol}\Omega. \end{aligned} \tag{3.24}$$

As ε can be taken arbitrarily small, we have a proof of (3.17).

Corollary 3.5. *If Ω is a bounded, open subset of $\mathbb{R}^n$, $N(\lambda)$ the counting function of the eigenvalues of $-\Delta$, with Dirichlet boundary condition, then (3.6) holds.*

Note that if $\mathcal{O}_\varepsilon$ is the set of points in Ω of distance $\ge \varepsilon$ from $\partial\Omega$ and we define $v(\varepsilon) = \operatorname{vol}(\Omega \setminus \mathcal{O}_\varepsilon)$, then the estimate (3.24) can be given the more precise reformulation

$$0 \le \operatorname{vol}\Omega - (4\pi t)^{n/2} \operatorname{Tr} e^{t\Delta} \le \omega(\sqrt{2nt}), \tag{3.25}$$

where

$$\omega(\varepsilon) = v(\varepsilon) + \int_\varepsilon^\infty e^{-ns^2/2\varepsilon^2}\, dv(s). \tag{3.26}$$

The fact that such a crude argument works, and works so generally, is a special property of the Dirichlet problem. If one uses the Neumann boundary condition, then for bounded $\Omega \subset \mathbb{R}^n$ with nasty boundary, Δ need not even have compact resolvent. However, Theorem 3.1 does extend to the Neumann boundary condition provided $\partial\Omega$ is smooth. One can do this via the sort of parametrix for boundary problems sketched in §13 of Chap. 7.

We now look at the heat kernel $H(t, x, y)$ on the *complement* of a smooth, bounded region $K \subset \mathbb{R}^n$. We impose the Dirichlet boundary condition on ∂K. As before, $0 \le H(t, x, y) \le H_0(t, x, y)$, where $H_0(t, x, y)$ is the free-space heat kernel. We can extend $H(t, x, y)$ to be Lipschitz continuous on $(0, \infty) \times \mathbb{R}^n \times \mathbb{R}^n$ by setting $H(t, x, y) = 0$ when either $x \in K$ or $y \in K$. We now estimate $E(t, x, y) = H_0(t, x, y) - H(t, x, y)$. Suppose K is contained in the open ball of radius R centered at the origin.

Lemma 3.6. *For $|x - y| \le |y| - R$, we have*

$$E(t, x, y) \le C t^{-1/2} e^{-(|y|-R)^2/4t}. \tag{3.27}$$

Proof. With $y \in \Omega = \mathbb{R}^n \setminus K$, write

$$H(t, x, y) = (4\pi t)^{-1/2} \int_{-\infty}^{\infty} e^{-s^2/4t} \cos s\Lambda \, ds, \tag{3.28}$$

where $\Lambda = \sqrt{-\Delta}$ and Δ is the Laplace operator on Ω, with the Dirichlet boundary condition. We have a similar formula for $H_0(t, x, y)$, using instead $\Lambda_0 = \sqrt{-\Delta_0}$, with Δ_0 the free-space Laplacian. Now, by finite propagation speed,

$$\cos s\Lambda \; \delta_y(x) = \cos s\Lambda_0 \; \delta_y(x),$$

provided

$$|s| \le d = \text{dist}(y, \partial K), \quad \text{and } |x - y| \le d$$

Thus, as long as $|x - y| \le d$, we have

$$E(t, x, y) = (4\pi t)^{-1/2} \int_{|s| \ge d} e^{-s^2/4t} \big[\cos s\Lambda_0 \; \delta_y(x) - \cos s\Lambda \; \delta_y(x)\big] \, ds. \tag{3.29}$$

Then the estimate (3.27) follows easily, along the same lines as estimates on heat kernels discussed in Chap. 6, §2.

When we combine (3.27) with the obvious inequality

$$0 \le E(t, x, y) \le H_0(t, x, y) = (4\pi t)^{-n/2} e^{-|x-y|^2/4t}, \tag{3.30}$$

we see that, for each $t > 0$, $E(t, x, y)$ is rapidly decreasing as $|x| + |y| \to \infty$. Using this and appropriate estimates on derivatives, we can show that $E(t, x, y)$ is the integral kernel of a trace class operator on $L^2(\mathbb{R}^n)$. We can write

$$\text{Tr}\left(e^{t\Delta_0} - e^{t\Delta}P\right) = \int_{\mathbb{R}^n} E(t,x,x)\,dx, \tag{3.31}$$

where P is the projection of $L^2(\mathbb{R}^n)$ onto $L^2(\Omega)$ defined by restriction to Ω. Now, as $t \searrow 0$, $(4\pi t)^{n/2}E(t,x,x)$ approaches 1 on K and 0 on $\mathbb{R}^n \setminus K$. Together with the estimates (3.27) and (3.30), this implies

$$(4\pi t)^{n/2} \int_{\mathbb{R}^n} E(t,x,x)\,dx \longrightarrow \text{vol}\,K, \tag{3.32}$$

as $t \searrow 0$. This establishes the following:

Proposition 3.7. *If K is a closed, bounded set in $\mathbb{R}^n$, Δ is the Laplacian on $L^2(\mathbb{R}^n \setminus K)$, with Dirichlet boundary condition, and Δ_0 is the Laplacian on $L^2(\mathbb{R}^n)$, then $e^{t\Delta_0} - e^{t\Delta}P$ is trace class for each $t > 0$ and*

$$Tr\left(e^{t\Delta_0} - e^{t\Delta}P\right) \sim (4\pi t)^{-n/2}\,vol\,K, \tag{3.33}$$

as $t \searrow 0$.

This result will be of use in the study of scattering by an obstacle K, in Chap. 9. It is also valid for the Neumann boundary condition if ∂K is smooth.

Exercises

In Exercises 1–4, let $\Omega \subset \mathbb{R}^n$ be a bounded, open set and let $\mathcal{O}_j$ be open with smooth boundary such that

$$\mathcal{O}_1 \subset\subset \mathcal{O}_2 \subset\subset \cdots \subset\subset \mathcal{O}_j \subset\subset \cdots \nearrow \Omega.$$

Let L_j be $-\Delta$ on $\mathcal{L}_j$, with Dirichlet boundary condition; the corresponding operator on Ω is simply denoted $-\Delta$.

1. Using material developed in §5 of Chap. 5, show that, for any $t > 0$, $f \in L^2(\Omega)$,
$$e^{-tL_j}P_j f \longrightarrow e^{t\Delta} f \ \text{ strongly in } L^2(\Omega),$$
as $j \to \infty$, where P_j is multiplication by the characteristic function of $\mathcal{O}_j$.
Don't peek at Lemma 3.4 in Chap. 11!
2. If $\lambda_\nu(\mathcal{O}_j)$ are the eigenvalues of L_j, arranged in increasing order for each j, show that, for each ν,
$$\lambda_\nu(\mathcal{O}_j) \searrow \lambda_\nu(\Omega), \ \text{ as } j \to \infty.$$
3. Show that, for each $t > 0$,
$$\text{Tr}\,e^{-tL_j} \nearrow \text{Tr}\,e^{t\Delta}.$$
4. Let $H_j(t,x,y)$ be the heat kernel on $\mathbb{R}^+ \times \overline{\mathcal{O}}_j \times \overline{\mathcal{O}}_j$. Extend H_j to $\mathbb{R}^+ \times \Omega \times \Omega$ so as to vanish if x or y belongs to $\Omega \setminus \mathcal{O}_j$. Show that, for each $x \in \Omega$, $y \in \Omega$, $t > 0$,
$$H_j(t,x,y) \nearrow H(t,x,y), \ \text{ as } j \to \infty.$$

Deduce that, for each $t > 0$,

$$\int_{\mathcal{O}_j} H_j(t,x,x)\,dx \nearrow \int_{\Omega} H(t,x,x)\,dx$$

5. Using Exercises 1–4, give a detailed proof of (3.16) for general bounded $\Omega \subset \mathbb{R}^n$.
6. Give an example of a bounded, open, connected set $\Omega \subset \mathbb{R}^2$ (with rough boundary) such that Δ, with Neumann boundary condition, does not have compact resolvent.

4. The Laplace operator on S^n

A key tool in the analysis of the Laplace operator Δ_S on S^n is the formula for the Laplace operator on $\mathbb{R}^{n+1}$ in polar coordinates:

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{n}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta_S. \tag{4.1}$$

In fact, this formula is simultaneously the main source of interest in Δ_S and the best source of information about it.

To begin, we consider the Dirichlet problem for the unit ball in Euclidean space, $B = \{x \in \mathbb{R}^{n+1} : |x| < 1\}$:

$$\Delta u = 0 \text{ in } B, \quad u = f \text{ on } S^n = \partial B, \tag{4.2}$$

given $f \in \mathcal{D}'(S^n)$. In Chap. 5 we obtained the Poisson integral formula for the solution:

$$u(x) = \frac{1-|x|^2}{A_n} \int_{S^n} \frac{f(y)}{|x-y|^{n+1}}\,dS(y), \tag{4.3}$$

where A_n is the volume of S^n. Equivalently, if we set $x = r\omega$ with $r = |x|$, $\omega \in S^n$,

$$u(r\omega) = \frac{1-r^2}{A_n} \int_{S^n} \frac{f(\omega')}{(1 - 2r\omega\cdot\omega' + r^2)^{(n+1)/2}}\,dS(\omega'). \tag{4.4}$$

Now we can derive an alternative formula for the solution of (4.2) if we use (4.1) and regard $\Delta u = 0$ as an operator-valued ODE in r; it is an Euler equation, with solution

$$u(r\omega) = r^{A-(n-1)/2} f(\omega), \quad r \le 1, \tag{4.5}$$

where A is an operator on $\mathcal{D}'(S^n)$, defined by

$$A = \Big(-\Delta_S + \frac{(n-1)^2}{4}\Big)^{1/2}. \tag{4.6}$$

If we set $r = e^{-t}$ and compare (4.5) and (4.4), we obtain a formula for the semigroup e^{-tA} as follows. Let $\theta(\omega, \omega')$ denote the geodesic distance on S^n from ω to ω', so $\cos\theta(\omega, \omega') = \omega \cdot \omega'$. We can rewrite (4.4) as

$$\begin{aligned} u(r\omega) = \frac{2}{A_n} \sinh(\log r^{-1})\, r^{-(n-1)/2} \\ \cdot \int_{S^n} \frac{f(\omega')}{\big[2\cosh(\log r^{-1}) - 2\cos\theta(\omega, \omega')\big]^{-(n+1)/2}}\, dS(\omega'). \end{aligned} \tag{4.7}$$

In other words, by (4.5),

$$e^{-tA} f(\omega) = \frac{2}{A_n} \sinh t \int_{S^n} \frac{f(\omega')}{\big(2\cosh t - 2\cos\theta(\omega, \omega')\big)^{(n+1)/2}}\, dS(\omega'). \tag{4.8}$$

Identifying an operator on $\mathcal{D}'(S^n)$ with its Schwartz kernel in $\mathcal{D}'(S^n \times S^n)$, we write

$$e^{-tA} = \frac{2}{A_n} \frac{\sinh t}{(2\cosh t - 2\cos\theta)^{(n+1)/2}}, \quad t > 0. \tag{4.9}$$

Note that the integration of (4.9) from t to ∞ produces the formula

$$A^{-1} e^{-tA} = 2C_n (2\cosh t - 2\cos\theta)^{-(n-1)/2}, \quad t > 0, \tag{4.10}$$

provided $n \geq 2$, where

$$C_n = \frac{1}{(n-1)A_n} = \frac{1}{4} \pi^{-(n+1)/2} \Gamma\Big(\frac{n-1}{2}\Big).$$

With the exact formula (4.9) for the semigroup e^{-tA}, we can proceed to give formulas for fundamental solutions to various important PDE, particularly

$$\frac{\partial^2 u}{\partial t^2} - Lu = 0 \quad \text{(wave equation)} \tag{4.11}$$

and

$$\frac{\partial u}{\partial t} - Lu = 0 \quad \text{(heat equation)}, \tag{4.12}$$

where

$$L = \Delta_S - \frac{(n-1)^2}{4} = -A^2. \tag{4.13}$$

If we prescribe Cauchy data $u(0) = f,\ u_t(0) = g$ for (4.11), the solution is

$$u(t) = (\cos tA) f + A^{-1} (\sin tA) g. \tag{4.14}$$

Assume $n \geq 2$. We obtain formulas for these terms by analytic continuation of the formulas (4.9) and (4.10) to $\operatorname{Re} t > 0$ and then passing to the limit $t \in i\mathbb{R}$. This is parallel to the derivation of the fundamental solution to the wave equation on Euclidean space in §5 of Chap. 3. We have

$$\begin{aligned} A^{-1}e^{(it-\varepsilon)A} &= -2C_n\big[2\cosh(it-\varepsilon)-2\cos\theta\big]^{-(n-1)/2}, \\ e^{(it-\varepsilon)A} &= \frac{2}{A_n}\sinh(it-\varepsilon)\big[2\cosh(it-\varepsilon)-2\cos\theta\big]^{-(n+1)/2}. \end{aligned} \tag{4.15}$$

Letting $\varepsilon \searrow 0$, we have

$$\begin{aligned} A^{-1}\sin tA = \\ \lim_{\varepsilon\searrow 0} -2C_n \operatorname{Im}(2\cosh\varepsilon\cos t - 2i\sinh\varepsilon\sin t - 2\cos\theta)^{-(n-1)/2} \end{aligned} \tag{4.16}$$

and

$$\begin{aligned} \cos tA = \\ \lim_{\varepsilon\searrow 0}\frac{-2}{A_n}\operatorname{Im}(\sin t)(2\cosh\varepsilon\cos t - 2i\sinh\varepsilon\sin t - 2\cos\theta)^{-(n+1)/2}. \end{aligned} \tag{4.17}$$

For example, on S^2 we have, for $0 \leq t \leq \pi$,

$$\begin{aligned} A^{-1}\sin tA = -2C_2(2\cos\theta - 2\cos t)^{-1/2}, &\quad \theta < |t|, \\ 0, &\quad \theta > |t|, \end{aligned} \tag{4.18}$$

with an analogous expression for general t, determined by the identity

$$A^{-1}\sin(t+2\pi)A = -A^{-1}\sin tA \quad \text{on } \mathcal{D}'(S^{2k}), \tag{4.19}$$

plus the fact that $\sin tA$ is odd in t. The last line on the right in (4.18) reflects the well-known finite propagation speed for solutions to the hyperbolic equation (4.11).

To understand how the sign is determined in (4.19), note that, in (4.15), with $\varepsilon > 0$, for $t = 0$ we have a real kernel, produced by taking the $-(\mathrm{n}-1)/2 = -k + 1/2$ power of a positive quantity. As t runs from 0 to 2π, the quantity $2\cosh(it-\varepsilon) = 2\cosh\varepsilon\cos t - 2i\sinh\varepsilon\sin t$ moves once clockwise around a circle of radius $2(\cosh^2\varepsilon + \sinh^2\varepsilon)^{1/2}$, centered at 0, so $2\cosh\varepsilon\cos t - 2i\sinh\varepsilon\sin t - 2\cos\theta$ describes a curve winding once clockwise about the origin in $\mathbb{C}$. Thus taking a half-integral power of this gives one the negative sign in (4.14).

On the other hand, when n is odd, the exponents on the right side of (4.15)–(4.17) are integers. Thus

$$A^{-1}\sin(t+2\pi)A = A^{-1}\sin tA \quad \text{on } \mathcal{D}'(S^{2k+1}). \tag{4.20}$$

Also, in this case, the distributional kernel for $A^{-1}\sin tA$ must vanish for $|t| \neq \theta$. In other words, the kernel is supported on the shell $\theta = |t|$. This is the generalization to spheres of the strict Huygens principle.

In case $n = 2k+1$ is odd, we obtain from (4.16) and (4.17) that

$$A^{-1}\sin tA\ f(x) = \frac{1}{(2k-1)!!}\Big(\frac{1}{\sin s}\frac{\partial}{\partial s}\Big)^{k-1}\big(\sin^{2k-1} s\ \overline{f}(x,s)\big)_{s=t} \tag{4.21}$$

and

$$\cos tA\ f(x) = \frac{1}{(2k-1)!!}\ \sin s\ \Big(\frac{1}{\sin s}\frac{\partial}{\partial s}\Big)^{k}\big(\sin^{2k-1} s\ \overline{f}(x,s)\big)_{s=t}, \tag{4.22}$$

where, as in (5.66) of Chap. 3, $(2k-1)!! = 3\cdot 5\cdots(2k-1)$ and

$$\overline{f}(x,s) = \text{ mean value of } f \text{ on } \Sigma_s(x) = \{y \in S^n : \theta(x,y) = |s|\}. \tag{4.23}$$

We can examine general functions of the operator A by the functional calculus

$$g(A) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} \hat{g}(t)e^{itA}\,dt = (2\pi)^{-1/2}\int_{-\infty}^{\infty} \hat{g}(t)\cos tA\,dt, \tag{4.24}$$

where the last identity holds provided g is an even function. We can rewrite this, using the fact that $\cos tA$ has period 2π in t on $\mathcal{D}'(S^n)$ for n odd, period 4π for n even. In concert with (4.22), we have the following formula for the Schwartz kernel of $g(A)$ on $\mathcal{D}'(S^{2k+1})$, for g even:

$$g(A) = (2\pi)^{-1/2}\Big(-\frac{1}{2\pi}\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\Big)^k \sum_{k=-\infty}^{\infty} \hat{g}(\theta + 2k\pi). \tag{4.25}$$

As an example, we compute the heat kernel on odd-dimensional spheres. Take $g(\lambda) = e^{-t\lambda^2}$. Then $\hat{g}(s) = (2t)^{-1/2}e^{-s^2/4t}$ and

$$(2\pi)^{-1/2}\sum_k \hat{g}(s+2k\pi) = (4\pi t)^{-1/2}\sum_k e^{-(s+2k\pi)^2/4t} = \vartheta(s,t), \tag{4.26}$$

where $\vartheta(s,t)$ is a "theta function." Thus the kernel of e^{-tA^2} on S^{2k+1} is given by

$$e^{-tA^2} = \Big(-\frac{1}{2\pi}\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\Big)^k \vartheta(\theta,t). \tag{4.27}$$

A similar analysis on S^{2k} gives an integral, with the theta function appearing in the integrand.

The operator A has a compact resolvent on $L^2(S^n)$, and hence a discrete set of eigenvalues, corresponding to an orthonormal basis of eigenfunctions. Indeed, the spectrum of A has the following description

Proposition 4.1. *The spectrum of the self-adjoint operator A on $L^2(S^n)$ is*

$$\text{spec } A = \left\{\frac{1}{2}(n-1) + k : k = 0, 1, 2, \dots\right\}. \tag{4.28}$$

Proof. Since 0 is the smallest eigenvalue of $-\Delta_S$, the definition (4.6) shows that $(n-1)/2$ is the smallest eigenvalue of A. Also, (4.20) shows that all eigenvalues of A are integers if n is odd, while (4.19) implies that all eigenvalues of A are (nonintegral) half-integers if n is even. Thus spec A is certainly contained in the right side of (4.28).

Another way to see this containment is to note that since the function $u(x)$ given by (4.5) must be smooth at $x = 0$, the exponent of r in that formula can take only integer values.

Let V_k denote the eigenspace of A with eigenvalue $\nu_k = (n-1)/2 + k$. We want to show that $V_k \neq 0$ for $k = 0, 1, 2, \dots$. Moreover, we want to identify V_k. Now if $f \in V_k$, it follows that $u(x) = u(r\omega) = r^{A-(n-1)/2} f(\omega) = r^k f(\omega)$ is a harmonic function defined on all of $\mathbb{R}^{n+1}$, which, being homogeneous and smooth at $x = 0$, must be a harmonic polynomial, homogeneous of degree k in x. If $\mathcal{H}_k$ denotes the space of harmonic polynomials, homogeneous of degree k, restriction to $S^n \subset \mathbb{R}^{n+1}$ produces an isomorphism:

$$\rho : \mathcal{H}_k \xrightarrow{\approx} V_k. \tag{4.29}$$

To show that each $V_k \neq 0$, it suffices to show that each $\mathcal{H}_k \neq 0$.

Indeed, for $c = (c_1, \dots, c_{n+1}) \in \mathbb{C}^{n+1}$, consider

$$p_c(x) = (c_1 x_1 + \cdots + c_{n+1} x_{n+1})^k.$$

A computation gives

$$\begin{aligned} \Delta p_c(x) &= k(k-1)\langle c, c\rangle (c_1 x_1 + \cdots + c_k x_k)^{k-2}, \\ \langle c, c\rangle &= c_1^2 + \cdots + c_k^2. \end{aligned}$$

Hence $\Delta p_c = 0$ whenever $\langle c, c\rangle = 0$, so the proposition is proved.

We now want to specify the orthogonal projections E_k of $L^2(S^n)$ on V_k. We can attack this via (4.10), which implies

$$\sum_{k=0}^{\infty} \nu_k^{-1} e^{-t\nu_k} E_k(x, y) = 2C_n (2\cosh t - 2\cos\theta)^{-(n-1)/2}, \tag{4.30}$$

where $\theta = \theta(x, y)$ is the geodesic distance from x to y in S^n. If we set $r = e^{-t}$ and use $\nu_k = (n-1)/2 + k$, we get the generating function identity

$$\sum_{k=0}^{\infty} r^k \nu_k^{-1} E_k(x, y) = 2C_n(1 - 2r\cos\theta + r^2)^{-(n-1)/2} \tag{4.31}$$
$$= \sum_{k=0}^{\infty} r^k p_k(\cos\theta);$$

in particular,

$$E_k(x, y) = \nu_k \; p_k(\cos\theta). \tag{4.32}$$

These functions are polynomials in $\cos\theta$. To see this, set $t = \cos\theta$ and write

$$(1 - 2tr + r^2)^{-\alpha} = \sum_{k=0}^{\infty} C_k^\alpha(t)\, r^k, \tag{4.33}$$

thus defining coefficients $C_k^\alpha(t)$. To compute these, use

$$(1-z)^{-\alpha} = \sum_{j=0}^{\infty} \binom{j+\alpha-1}{j} z^j,$$

with $z = r(2t - r)$, to write the left side of (4.33) as

$$\begin{aligned}\sum_{j=0}^{\infty}\binom{\alpha}{j} r^j(2t-r)^j &= \sum_{j=0}^{\infty}\sum_{\ell=0}^{j}\binom{j+\alpha-1}{j}\binom{j}{\ell}(-1)^\ell r^{j+\ell}(2t)^{j-\ell} \\ &= \sum_{k=0}^{\infty}\sum_{\ell=0}^{[k/2]}(-1)^\ell\binom{k-\ell+\alpha-1}{k-\ell}\binom{k-\ell}{\ell}(2t)^{k-2\ell} r^k.\end{aligned}$$

Hence

$$C_k^\alpha(t) = \sum_{\ell=0}^{[k/2]}(-1)^\ell\binom{k-\ell+\alpha-1}{k-\ell}\binom{k-\ell}{\ell}(2t)^{k-2\ell}. \tag{4.34}$$

These are called *Gegenbauer polynomials*. Therefore, we have the following:

Proposition 4.2. *The orthogonal projection of $L^2(S^n)$ onto V_k has kernel*

$$E_k(x, y) = 2C_n\nu_k\; C_k^\alpha(\cos\theta), \quad \alpha = \frac{1}{2}(n-1), \tag{4.35}$$

with C_n as in (4.10).

In the special case $n = 2$, we have $C_2 = 1/4\pi$, and $\nu_k = k + 1/2$; hence

$$E_k(x, y) = \frac{2k+1}{4\pi} C_k^{1/2}(\cos\theta) = \frac{2k+1}{4\pi} P_k(\cos\theta), \tag{4.36}$$

where $C_k^{1/2}(t) = P_k(t)$ are the *Legendre polynomials*.

The trace of E_k is easily obtained by integrating (4.35) over the diagonal, to yield

$$\mathrm{Tr}\, E_k = 2C_n A_n \nu_k\, C_k^{(n-1)/2}(1) = \frac{2\nu_k}{n-1} C_k^{(n-1)/2}(1). \tag{4.37}$$

Setting $t = 1$ in (4.33), so $(1 - 2r + r^2)^{-\alpha} = (1 - r)^{-2\alpha}$, we obtain

$$C_k^\alpha(1) = \binom{k + 2\alpha - 1}{k}, \quad \text{e.g., } P_k(1) = 1. \tag{4.38}$$

Thus we have the dimensions of the eigenspaces V_k:

Corollary 4.3. *The eigenspace V_k of $-\Delta_S$ on S^n, with eigenvalue*

$$\lambda_k = \nu_k^2 - \frac{1}{4}(n-1)^2 = k^2 + (n-1)k,$$

satisfies

$$\dim V_k = \frac{2k+n-1}{n-1}\binom{k+n-2}{k} = \binom{k+n-2}{k-1} + \binom{k+n-1}{k}. \tag{4.39}$$

In particular, on S^2 we have $\dim V_k = 2k + 1$.

Another natural approach to E_k is via the wave equation. We have

$$\begin{aligned} E_k &= \frac{1}{2T}\int_{-T}^{T} e^{-i\nu_k t} e^{itA}\, dt \\ &= \frac{1}{2T}\int_{-T}^{T} \cos t(A - \nu_k)\, dt, \end{aligned} \tag{4.40}$$

where $T = \pi$ or 2π depending on whether n is odd or even. (In either case, one can take $T = 2\pi$.) In the special case of S^2, when (4.18) is used, comparison of (4.36) with the formula produced by this method produces the identity

$$P_k(\cos\theta) = \frac{1}{\pi}\int_{-\theta}^{\theta} \frac{\cos(k + \frac{1}{2})t}{(2\cos t - 2\cos\theta)^{1/2}}\, dt, \tag{4.41}$$

for the Legendre polynomials, known as the *Mehler-Dirichlet formula*.

Exercises

Exercises 1–5 deal with results that follow from symmetries of the sphere. The group $SO(n+1)$ acts as a group of isometries of $S^n \subset \mathbb{R}^{n+1}$, hence as a group of unitary operators on $L^2(S^n)$. Each eigenspace V_k of the Laplace operator is preserved by this action. Fix $p = (0, \dots, 0, 1) \in S^n$, regarded as the "north pole." The subgroup of $SO(n+1)$ fixing p is a copy of $SO(n)$.

1. Show that each eigenspace V_k has an element u such that $u(p) \neq 0$. Conclude by forming
$$\int_{SO(n)} u(gx)\, dg$$
that each eigenspace V_k of Δ_S has an element $z_k \neq 0$ such that $z_k(x) = z_k(gx)$, for all $g \in SO(n)$. Such a function is called a *spherical function*.
2. Suppose V_k has a proper subspace W invariant under $SO(n+1)$. (Hence $W^\perp \subset V_k$ is also invariant.) Show that W must contain a nonzero spherical function.
3. Suppose z_k and y_k are two nonzero spherical functions in V_k. Show that they must be multiples of each other. Hence the unique spherical functions (up to constant multiples) are given by (4.35), with $y = p$. (*Hint*: z_k and y_k are eigenfunctions of $-\Delta_S$, with eigenvalue $\lambda_k = k^2 + (n-1)k$. Pick a sequence of surfaces
$$\Sigma_j = \{x \in S^n : \theta(x, p) = \varepsilon_j\} \subset S^n,$$
with $\varepsilon_j \to 0$, on which $z_k = \alpha_j \neq 0$. With $\beta_j = y_k|_{\Sigma_j}$, it follows that $\beta_j z_k - \alpha_j y_k$ is an eigenfunction of $-\Delta_S$ that vanishes on Σ_j. Show that, for j large, this forces $\beta_j z_k - \alpha_j y_k$ to be identically zero.)
4. Using Exercises 2 and 3, show that the action of $SO(n+1)$ on each eigenspace V_k is irreducible, that is, V_k has no proper invariant subspaces.
5. Show that each V_k is equal to the linear span of the set of polynomials of the form $p_c(x) = (c_1 x_1 + \cdots + c_{n+1} x_{n+1})^k$, with $\langle c, c \rangle = 0$.
(*Hint*: Show that this linear span is invariant under $SO(n+1)$.)
6. Using (4.9), show that
$$\operatorname{Tr} e^{-tA} = \frac{2 \sinh t}{(2 \cosh t - 2)^{(n+1)/2}}. \tag{4.42}$$
Find the asymptotic behavior as $t \searrow 0$. Use Karamata's Tauberian theorem to determine the asymptotic behavior of the eigenvalues of A, hence of $-\Delta_S$. Compare this with the general results of §3 and also with the explicit results of Corollary 4.3.
7. Using (4.27), show that, for A on S^n with $n = 2k+1$,
$$\begin{aligned} \operatorname{Tr} e^{-tA^2} &= \frac{A_{2k+1}}{\sqrt{4\pi t}} \Big(-\frac{1}{2\pi} \frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \Big)^k e^{-\theta^2/4t} \Big|_{\theta=0} + O(t^\infty) \\ &= (4\pi t)^{-n/2} A_{2k+1} + O\big(t^{-n/2+1}\big), \end{aligned} \tag{4.43}$$
as $t \searrow 0$. Compare the general results of §3.
8. Show that
$$e^{-\pi i (A - (n-1)/2)} f(\omega) = f(-\omega), \quad f \in L^2(S^n). \tag{4.44}$$

(*Hint*: Check it for $f \in V_k$, the restriction to S^n of a homogeneous harmonic polynomial of degree k.)

Exercises 9–13 deal with analysis on S^n when $n = 2$. When doing them, look for generalizations to other values of n.

9. If $\Xi(A)$ has integral kernel $K_\Xi(x, y)$, show that when $n = 2$,

$$K_\Xi(x, y) = \frac{1}{4\pi} \sum_{\ell=0}^{\infty} (2\ell + 1)\Xi\Big(\ell + \frac{1}{2}\Big) P_\ell(\cos\theta), \tag{4.45}$$

where $\cos\theta = x \cdot y$ and $P_\ell(t)$ are the Legendre polynomials.

10. Demonstrate the Rodrigues formula for the Legendre polynomials:

$$P_k(t) = \frac{1}{2^k k!} \Big(\frac{d}{dt}\Big)^k (t^2 - 1)^k. \tag{4.46}$$

(*Hint*: Use Cauchy's formula to get

$$P_k(t) = \frac{1}{2\pi i} \int_\gamma (1 - 2zt + z^2)^{-1/2} z^{-k-1}\, dz$$

from (4.33); then use the change of variable $1 - uz = (1 - 2tz + z^2)^{1/2}$. Then appeal to Cauchy's formula again, to analyze the resulting integral.)

11. If $f \in L^2(S^2)$ has the form $f(x) = g(x \cdot y) = \sum \varphi_\ell P_\ell(x \cdot y)$, for some $y \in S^2$, show that

$$\varphi_\ell = \frac{2\ell + 1}{4\pi} \int_{S^2} f(z) P_\ell(y \cdot z)\, dS(z) = \Big(\ell + \frac{1}{2}\Big) \int_{-1}^{1} g(t) P_\ell(t)\, dt. \tag{4.47}$$

(*Hint*: Use $\int_{S^2} E_k(x, z) E_\ell(z, y)\, dS(z) = \delta_{k\ell}\, E_\ell(x, y)$.) Conclude that $g(x \cdot y)$ is the integral kernel of $\psi(A - 1/2)$, where

$$\psi(\ell) = \frac{4\pi}{2\ell + 1}\, \varphi_\ell = 2\pi \int_{-1}^{1} g(t) P_\ell(t)\, dt. \tag{4.48}$$

This result is known as the Funk-Hecke theorem.

12. Show that, for $x, y \in S^2$,

$$e^{ikx\cdot y} = \sum_{\ell=0}^{\infty} (2\ell + 1)\, i^\ell\, j_\ell(k)\, P_\ell(x \cdot y), \tag{4.49}$$

where

$$j_\ell(z) = \Big(\frac{\pi}{2z}\Big)^{1/2} J_{\ell+1/2}(z) = \frac{1}{2} \frac{1}{\ell!} \Big(\frac{z}{2}\Big)^\ell \int_{-1}^{1} (1 - t^2)^\ell\, e^{izt}\, dt. \tag{4.50}$$

(*Hint*: Take $g(t) = e^{ikt}$ in Exercise 11, apply the Rodrigues formula, and integrate by parts.) Thus $e^{ikx\cdot y}$ is the integral kernel of the operator

$$4\pi\, e^{(1/2)\pi i (A-1/2)}\, j_{A-1/2}(k)$$

For another approach, see Exercises 10 and 11 in §9 of Chap. 9.

13. Demonstrate the identities

$$\Big[(1-t^2)\frac{d}{dt}+\ell t\Big]P_\ell(t)=\ell P_{\ell-1}(t) \tag{4.51}$$

and

$$\frac{d}{dt}\Big[(1-t^2)\frac{d}{dt}\,P_\ell(t)\Big]+\ell(\ell+1)P_\ell(t)=0. \tag{4.52}$$

Relate (4.52) to the statement that, for fixed $y\in S^2$, $\varphi(x)=P_\ell(x\cdot y)$ belongs to the $\ell(\ell+1)$-eigenspace of $-\Delta_S$.

Exercises 14–19 deal with formulas for an orthogonal basis of V_k (for S^2). We will make use of the structure of irreducible representations of SO(3), obtained in §9 of Appendix B, Manifolds, Vector Bundles, and Lie Groups.

14. Show that the representation of SO(3) on V_k is equivalent to the representation D_k, for each $k=0,1,2,\dots$.
15. Show that if we use coordinates (θ,ψ) on S^2, where θ is the geodesic distance from $(1,0,0)$ and ψ is the angular coordinate about the x_1-axis in $\mathbb{R}^3$, then

$$L_1=\frac{\partial}{\partial\psi},\quad L_\pm=i\,e^{\pm i\psi}\Big[\pm\frac{\partial}{\partial\theta}+i\cot\theta\,\frac{\partial}{\partial\psi}\Big]. \tag{4.53}$$

16. Set

$$w_k(x)=(x_2+ix_3)^k=\sin^k\theta\,e^{ik\psi}. \tag{4.54}$$

Show that $w_k\in V_k$ and that it is the highest-weight vector for the representation, so

$$L_1w_k=ik\,w_k$$

17. Show that an orthogonal basis of V_k is given by

$$w_k,L_-w_k,\dots,L_-^{2k}w_k$$

18. Show that the functions $\zeta_{kj}=L_-^{k-j}w_k,\ j\in\{-k,-k+1,\dots,k-1,k\}$, listed in Exercise 17 coincide, up to nonzero constant factors, with z_{kj}, given by

$$z_{k0}=z_k,$$

the spherical function considered in Exercises 1–3, and, for $1\le j\le k$,

$$z_{k,-j}=L_-^jz_k,\quad z_{kj}=L_+^jz_k$$

19. Show that the functions z_{kj} coincide, up to nonzero constant factors, with

$$e^{ij\psi}\,P_k^j(\cos\theta),\quad -k\le j\le k, \tag{4.55}$$

where $P_k^j(t)$, called *associated Legendre functions*, are defined by

$$P_k^j(t)=(-1)^j(1-t^2)^{|j|/2}\Big(\frac{d}{dt}\Big)^{|j|}P_k(t). \tag{4.56}$$

5. The Laplace operator on hyperbolic space

The hyperbolic space $\mathcal{H}^n$ shares with the sphere S^n the property of having constant sectional curvature, but for $\mathcal{H}^n$ it is -1. One way to describe $\mathcal{H}^n$ is as a set of vectors with square length 1 in $\mathbb{R}^{n+1}$, not for a Euclidean metric, but rather for a Lorentz metric

$$\langle v, v\rangle = -v_1^2 - \cdots - v_n^2 + v_{n+1}^2, \tag{5.1}$$

namely,

$$\mathcal{H}^n = \{v \in \mathbb{R}^{n+1} : \langle v, v\rangle = 1,\ v_{n+1} > 0\}, \tag{5.2}$$

with metric tensor induced from (5.1). The connected component G of the identity of the group $\mathrm{O}(n, 1)$ of linear transformations preserving the quadratic form (5.1) acts transitively on $\mathcal{H}^n$, as a group of isometries. In fact, $\mathrm{SO}(n)$, acting on $\mathbb{R}^n \subset \mathbb{R}^{n+1}$, leaves invariant $p = (0, \dots, 0, 1) \in \mathcal{H}^n$ and acts transitively on the unit sphere in $T_p\mathcal{H}^n$. Also, if $A(u_1, \dots, u_n, u_{n+1})^t = (u_1, \dots, u_{n+1}, u_n)^t$, then e^{tA} is a one-parameter subgroup of $\mathrm{SO}(n, 1)$ taking p to the curve

$$\gamma = \{(0, \dots, 0, x_n, x_{n+1}) : x_{n+1}^2 - x_n^2 = 1, x_{n+1} > 0\}$$

Together these facts imply that $\mathcal{H}^n$ is a homogeneous space.

There is a map of $\mathcal{H}^n$ onto the unit ball in $\mathbb{R}^n$, defined in a fashion similar to the stereographic projection of S^n. The map

$$\mathfrak{s} : \mathcal{H}^n \longrightarrow B^n = \{x \in \mathbb{R}^n : |x| < 1\} \tag{5.3}$$

is defined by

$$\mathfrak{s}(x, x_{n+1}) = (1 + x_{n+1})^{-1}x. \tag{5.4}$$

The metric on $\mathcal{H}^n$ defined above then yields the following metric tensor on B^n:

$$ds^2 = 4\big(1 - |x|^2\big)^{-2} \sum_{j=1}^{n} dx_j^2. \tag{5.5}$$

Another useful representation of hyperbolic space is as the upper half space $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_n > 0\}$, with a metric we will specify shortly. In fact, with $e_n = (0, \dots, 0, 1)$,

$$\tau(x) = |x + e_n|^{-2}(x + e_n) - \frac{1}{2}e_n \tag{5.6}$$

defines a map of the unit ball B^n onto $\mathbb{R}^n_+$, taking the metric (5.5) to

$$ds^2 = x_n^{-2} \sum_{j=1}^{n} dx_j^2. \tag{5.7}$$

The Laplace operator for the metric (5.7) has the form

$$\Delta u = \sum_{j=1}^{n} x_n^n \, \partial_j \big(x_n^{2-n} \, \partial_j u\big) \tag{5.8}$$
$$= x_n^2 \sum_{j=1}^{n} \partial_j^2 u + (2-n) x_n \, \partial_n u.$$

which is convenient for a number of computations, such as (5.9) in the following:

Proposition 5.1. *If Δ is the Laplace operator on $\mathcal{H}^n$, then Δ is essentially self-adjoint on $C_0^\infty(\mathcal{H}^n)$, and its natural self-adjoint extension has the property*

$$spec(-\Delta) \subset \Big[\frac{1}{4}(n-1)^2, \infty\Big). \tag{5.9}$$

Proof. Since $\mathcal{H}^n$ is a complete Riemannian manifold, the essential self-adjointness on $C_0^\infty(\mathcal{H}^n)$ follows from Proposition 2.4. To establish (5.9), it suffices to show that

$$(-\Delta u, u)_{L^2(\mathcal{H}^n)} \geq \frac{(n-1)^2}{4} \|u\|^2_{L^2(\mathcal{H}^n)},$$

for all $u \in C_0^\infty(\mathcal{H}^n)$. Now the volume element on $\mathcal{H}^n$, identified with the upper half-space with the metric (5.7), is $x_n^{-n} \, dx_1 \cdots dx_n$, so for such u we have

$$\Big(\Big(-\Delta - \frac{1}{4}(n-1)^2\Big)u, u\Big)_{L^2} \tag{5.10}$$
$$= \int \Big[(\partial_n u)^2 - \Big(\frac{(n-1)u}{2x_n}\Big)^2\Big] x_n^{2-n} \, dx_1 \cdots dx_n$$
$$+ \sum_{j=1}^{n-1} \int (\partial_j u)^2 x_n^{2-n} \, dx_1 \cdots dx_n.$$

Now, by an integration by parts, the first integral on the right is equal to

$$\int_{\mathbb{R}^n_+} \Big[\partial_n \big(x_n^{-(n-1)/2} u\big)\Big]^2 x_n \, dx_1 \cdots dx_n. \tag{5.11}$$

Thus the expression (5.10) is ≥ 0, and (5.9) is proved.

We next describe how to obtain the fundamental solution to the wave equation on $\mathcal{H}^n$. This will be obtained from the formula for S^n, via an analytic continuation in the metric tensor. Let p be a fixed point (e.g., the north pole) in S^n, taken to be the origin in geodesic normal coordinates. Consider the one-parameter family of metrics given by dilating the sphere, which has constant curvature $K = 1$. Spheres dilated to have radius > 1 have constant curvature $K \in (0, 1)$. On such a space, the fundamental kernel $A^{-1} \sin tA \, \delta_p(x)$, with

$$A = \Big(-\Delta + \frac{K}{4}(n-1)^2\Big)^{1/2}, \tag{5.12}$$

can be obtained explicitly from that on the unit sphere by a change of scale. The explicit representation so obtained continues analytically to all real values of K and at $K = -1$ gives a formula for the wave kernel,

$$A^{-1}\ \sin tA\ \delta_p(x) = R(t,p,x), \quad A = \Big(-\Delta - \frac{1}{4}(n-1)^2\Big)^{1/2}. \tag{5.13}$$

We have

$$R(t,p,x) = \lim_{\varepsilon \searrow 0} -2C_n \ \mathrm{Im}\ \big[2\cos(it-\varepsilon) - 2\cosh r\big]^{-(n-1)/2}, \tag{5.14}$$

where $r = r(p,x)$ is the geodesic distance from p to x. Here, as in (4.10), $C_n = 1/(n-1)A_n$. This exhibits several properties similar to those in the case of S^n discussed in §4. Of course, for $r > |t|$, the limit vanishes, exhibiting the finite propagation speed phenomenon. Also, if n is odd, the exponent $(n-1)/2$ is an integer, which implies that (5.14) is supported on the shell $r = |t|$.

In analogy with (4.25), we have the following formula for $g(A)\delta_p(x)$, for $g \in \mathcal{S}(\mathbb{R})$, when acting on $L^2(\mathcal{H}^n)$, with $n = 2k+1$:

$$g(A) = (2\pi)^{-1/2}\Big(-\frac{1}{2\pi}\frac{1}{\sinh r}\frac{\partial}{\partial r}\Big)^k\ \hat{g}(r). \tag{5.15}$$

If $n = 2k$, we have

$$g(A) = \frac{1}{\pi^{1/2}}\int_r^\infty \Big(-\frac{1}{2\pi}\frac{1}{\sinh s}\frac{\partial}{\partial s}\Big)^k\ \hat{g}(s)\big(\cosh s - \cosh r\big)^{-1/2}\sinh s\ ds. \tag{5.16}$$

Exercises

1. If $n = 2k+1$, show that the Schwartz kernel of $\big(-\Delta - (n-1)^2/4 - z^2\big)^{-1}$ on $\mathcal{H}^n$, for $z \in \mathbb{C} \setminus [0,\infty)$, is

$$G_z(x,y) = -\frac{1}{2iz}\Big(-\frac{1}{2\pi}\frac{1}{\sinh r}\frac{\partial}{\partial r}\Big)^k e^{izr},$$

where $r = r(x,y)$ is geodesic distance, and the integral kernel of $e^{t(\Delta+(n-1)^2/4)}$, for $t > 0$, is

$$H_t(x,y) = \frac{1}{\sqrt{4\pi t}}\Big(-\frac{1}{2\pi}\frac{1}{\sinh r}\frac{\partial}{\partial r}\Big)^k e^{-r^2/4t}$$

6. The harmonic oscillator

We consider the differential operator $H = -\Delta + |x|^2$ on $L^2(\mathbb{R}^n)$. By Proposition 2.7, H is essentially self-adjoint on $C_0^\infty(\mathbb{R}^n)$. Furthermore, as a special case of Proposition 2.8, we know that H has compact resolvent, so $L^2(\mathbb{R}^n)$ has an orthonormal basis of eigenfunctions of H. To work out the spectrum, it suffices to work with the case $n = 1$, so we consider $H = D^2 + x^2$, where $D = -i\, d/dx$.

The spectral analysis follows by some simple algebraic relations, involving the operators

$$\begin{aligned} a &= D - ix = \frac{1}{i}\Big(\frac{d}{dx} + x\Big), \\ a^+ &= D + ix = \frac{1}{i}\Big(\frac{d}{dx} - x\Big). \end{aligned} \tag{6.1}$$

Note that on $\mathcal{D}'(\mathbb{R})$,

$$H = aa^+ - I = a^+a + I, \tag{6.2}$$

and

$$[H, a] = -2a, \quad [H, a^+] = 2a^+. \tag{6.3}$$

Suppose that $u_j \in C^\infty(\mathbb{R})$ is an eigenfunction of H, that is,

$$u_j \in \mathcal{D}(H), \quad Hu_j = \lambda_j u_j. \tag{6.4}$$

Now, by material developed in §2,

$$\begin{aligned} \mathcal{D}(H^{1/2}) &= \{u \in L^2(\mathbb{R}) : Du \in L^2(\mathbb{R}),\ xu \in L^2(\mathbb{R})\}, \\ \mathcal{D}(H) &= \{u \in L^2(\mathbb{R}) : D^2u + x^2u \in L^2(\mathbb{R})\}. \end{aligned} \tag{6.5}$$

Since certainly each u_j belongs to $\mathcal{D}(H^{1/2})$, it follows that au_j and a^+u_j belong to $L^2(\mathbb{R})$. By (6.3), we have

$$H(au_j) = (\lambda_j - 2)au_j, \quad H(a^+u_j) = (\lambda_j + 2)a^+u_j. \tag{6.6}$$

It follows that au_j and a^+u_j belong to $\mathcal{D}(H)$ and are eigenfunctions. Hence, if

$$\text{Eigen}(\lambda, H) = \{u \in \mathcal{D}(H) : Hu = \lambda u\}, \tag{6.7}$$

we have, for all $\lambda \in \mathbb{R}$,

$$
\begin{aligned}
a^+ &: \mathrm{Eigen}(\lambda, H) \to \mathrm{Eigen}(\lambda + 2, H), \\
a &: \mathrm{Eigen}(\lambda + 2, H) \to \mathrm{Eigen}(\lambda, H).
\end{aligned} \tag{6.8}
$$

From (6.2) it follows that $(Hu, u) \geq \|u\|_{L^2}^2$, for all $u \in C_0^\infty(\mathbb{R})$; hence, in view of essential self-adjointness,

$$
\mathrm{spec}\, H \subset [1, \infty), \quad \text{for } n = 1. \tag{6.9}
$$

Now each space $\mathrm{Eigen}(\lambda, H)$ is a finite-dimensional subspace of $C^\infty(\mathbb{R})$, and, by (6.2), we conclude that, in (6.8), a^+ is an isomorphism of $\mathrm{Eigen}(\lambda_j, H)$ onto $\mathrm{Eigen}(\lambda_j + 2, H)$, for each $\lambda_j \in \mathrm{spec}\, H$. Also, a is an isomorphism of $\mathrm{Eigen}(\lambda_j, H)$ onto $\mathrm{Eigen}(\lambda_j - 2, H)$, for all $\lambda_j > 1$. On the other hand, a must annihilate $\mathrm{Eigen}(\lambda_0, H)$ when λ_0 is the smallest element of spec H, so

$$
\begin{aligned}
u_0 \in \mathrm{Eigen}(\lambda_0, H) &\Longrightarrow u_0'(x) = -x u_0(x) \\
&\Longrightarrow u_0(x) = K\, e^{-x^2/2}.
\end{aligned} \tag{6.10}
$$

Thus

$$
\lambda_0 = 1, \quad \mathrm{Eigen}(1, H) = \mathrm{span}\big(e^{-x^2/2}\big). \tag{6.11}
$$

Since $e^{-x^2/2}$ spans the null space of a, acting on $C^\infty(\mathbb{R})$, and since each nonzero space $\mathrm{Eigen}(\lambda_j, H)$ is mapped by some power of a to this null space, it follows that, for $n = 1$,

$$
\mathrm{spec}\, H = \{2k + 1 : k = 0, 1, 2, \dots\} \tag{6.12}
$$

and

$$
\mathrm{Eigen}(2k + 1, H) = \mathrm{span}\left(\left(\frac{\partial}{\partial x} - x\right)^k e^{-x^2/2}\right). \tag{6.13}
$$

One also writes

$$
\left(\frac{\partial}{\partial x} - x\right)^k e^{-x^2/2} = H_k(x)\, e^{-x^2/2}, \tag{6.14}
$$

where $H_k(x)$ are the *Hermite polynomials*, given by

$$
\begin{aligned}
H_k(x) &= (-1)^k e^{x^2} \left(\frac{d}{dx}\right)^k e^{-x^2} \\
&= \sum_{j=0}^{[k/2]} (-1)^j \frac{k!}{j!(k - 2j)!} (2x)^{k-2j}.
\end{aligned} \tag{6.15}
$$

We define eigenfunctions of H:

$$h_k(x) = c_k\Big(\frac{\partial}{\partial x} - x\Big)^k e^{-x^2/2} = c_k H_k(x) e^{-x^2/2}, \tag{6.16}$$

where c_k is the unique positive number such that $\|h_k\|_{L^2(\mathbb{R})} = 1$. To evaluate c_k, note that

$$\|a^+ h_k\|_{L^2}^2 = (aa^+ h_k, h_k)_{L^2} = 2(k+1)\|h_k\|_{L^2}^2. \tag{6.17}$$

Thus, if $\|h_k\|_{L^2} = 1$, in order for $h_{k+1} = \gamma_k a^+ h_k$ to have unit norm, we need $\gamma_k = (2k+2)^{-1/2}$. Hence

$$c_k = \big[\pi^{1/2} 2^k (k!)\big]^{-1/2}. \tag{6.18}$$

Of course, given the analysis above of H on $L^2(\mathbb{R})$, then for $H = -\Delta + |x|^2$ on $L^2(\mathbb{R}^n)$, we have

$$\text{spec}\, H = \{2k + n : k = 0, 1, 2, \dots\}. \tag{6.19}$$

In this case, an orthonormal basis of $\text{Eigen}(2k+n, H)$ is given by

$$c_{k_1} \cdots c_{k_n}\, H_{k_1}(x_1) \cdots H_{k_n}(x_n) e^{-|x|^2/2}, \quad k_1 + \cdots + k_n = k, \tag{6.20}$$

where $k_\nu \in \{0, \dots, k\}$, the $H_{k_\nu}(x_\nu)$ are the Hermite polynomials, and the c_{k_ν} are given by (6.18). The dimension of this eigenspace is the same as the dimension of the space of homogeneous polynomials of degree k in n variables.

We now want to derive a formula for the semigroup e^{-tH}, $t > 0$, called the *Hermite semigroup*. Again it suffices to treat the case $n = 1$. To some degree paralleling the analysis of the eigenfunctions above, we can produce this formula via some commutator identities, involving the operators

$$X = D^2 = -\partial_x^2, \quad Y = x^2, \quad Z = x\partial_x + \partial_x x = 2x\,\partial_x + 1. \tag{6.21}$$

Note that $H = X + Y$. The commutator identities are

$$[X, Y] = -2Z, \quad [X, Z] = 4X, \quad [Y, Z] = -4Y. \tag{6.22}$$

Thus, X, Y, and Z span a three-dimensional, real Lie algebra. This is isomorphic to $\text{sl}(2, \mathbb{R})$, the Lie algebra consisting of 2×2 real matrices of trace zero, spanned by

$$n_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad n_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \alpha = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{6.23}$$

We have

$$[n_+, n_-] = \alpha, \quad [n_+, \alpha] = -2n_+, \quad [n_-, \alpha] = 2n_-. \tag{6.24}$$

The isomorphism is implemented by

$$X \leftrightarrow 2n_+, \quad Y \leftrightarrow 2n_-, \quad Z \leftrightarrow -2\alpha. \tag{6.25}$$

Now we will be able to write

$$e^{-t(2n_+ + 2n_-)} = e^{-2\sigma_1(t)n_+}\, e^{-2\sigma_3(t)\alpha}\, e^{-2\sigma_2(t)n_-}, \tag{6.26}$$

as we will see shortly, and, once this is accomplished, we will be motivated to suspect that also

$$e^{-tH} = e^{-\sigma_1(t)X}\, e^{\sigma_3(t)Z}\, e^{-\sigma_2(t)Y}. \tag{6.27}$$

To achieve (6.26), write

$$\begin{aligned} e^{-2\sigma_1 n_+} &= \begin{pmatrix} 1 & -2\sigma_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \\ e^{-2\sigma_3 \alpha} &= \begin{pmatrix} e^{-2\sigma_3} & 0 \\ 0 & e^{2\sigma_3} \end{pmatrix} = \begin{pmatrix} y & 0 \\ 0 & 1/y \end{pmatrix}, \\ e^{-2\sigma_2 n_-} &= \begin{pmatrix} 1 & 0 \\ -2\sigma_2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix}, \end{aligned} \tag{6.28}$$

and

$$e^{-2t(n_+ + n_-)} = \begin{pmatrix} \cosh 2t & -\sinh 2t \\ -\sinh 2t & \cosh 2t \end{pmatrix} = \begin{pmatrix} u & v \\ v & u \end{pmatrix}. \tag{6.29}$$

Then (6.26) holds if and only if

$$y = \frac{1}{u} = \frac{1}{\cosh 2t}, \quad x = z = \frac{v}{u} = -\tanh 2t, \tag{6.30}$$

so the quantities $\sigma_j(t)$ are given by

$$\sigma_1(t) = \sigma_2(t) = \frac{1}{2}\tanh 2t, \quad e^{2\sigma_3(t)} = \cosh 2t. \tag{6.31}$$

Now we can compute the right side of (6.27). Note that

$$
\begin{aligned}
e^{-\sigma_1 X}u(x) &= (4\pi\sigma_1)^{-1/2}\int e^{-(x-y)^2/4\sigma_1}u(y)\,dy,\\
e^{-\sigma_2 Y}u(x) &= e^{-\sigma_2 x^2}u(x),\\
e^{\sigma_3 Z}u(x) &= e^{\sigma_3}\,u(e^{2\sigma_3}x).
\end{aligned}
\tag{6.32}
$$

Upon composing these operators we find that, for $n = 1$,

$$
e^{-tH}u(x) = \int K_t(x,y)u(y)\,dy,
\tag{6.33}
$$

with

$$
K_t(x,y) = \frac{\exp\Big\{\big[-\frac{1}{2}(\cosh 2t)(x^2+y^2)+xy\big]\big/\sinh 2t\Big\}}{\big(2\pi\,\sinh 2t\big)^{1/2}}.
\tag{6.34}
$$

This is known as Mehler's formula for the Hermite semigroup. Clearly, for general n, we have

$$
e^{-tH}u(x) = \int K_n(t,x,y)u(y)\,dy,
\tag{6.35}
$$

with

$$
K_n(t,x,y) = K_t(x_1,y_1)\cdots K_t(x_n,y_n).
\tag{6.36}
$$

The idea behind passing from (6.26) to (6.27) is that the Lie algebra homomorphism defined by (6.25) should give rise to a Lie group homomorphism from (perhaps a covering group G of) $\mathrm{SL}(2,\mathbb{R})$ into a group of operators. Since this involves an infinite-dimensional representation of G (not necessarily by bounded operators here, since e^{-tH} is bounded only for $t \geq 0$), there are analytical problems that must be overcome to justify this reasoning. Rather than take the space to develop such analysis here, we will instead just give a direct justification of (6.33)–(6.34).

Indeed, let $v(t,x)$ denote the right side of (6.33), with $u \in L^2(\mathbb{R})$ given. The rapid decrease of $K_t(x,y)$ as $|x|+|y| \to \infty$, for $t > 0$, makes it easy to show that

$$
u \in L^2(\mathbb{R}) \Longrightarrow v \in C^\infty\big((0,\infty),\mathcal{S}(\mathbb{R})\big).
\tag{6.37}
$$

Also, it is routine to verify that

$$
\frac{\partial v}{\partial t} = -Hv.
\tag{6.38}
$$

Simple estimates yielding uniqueness then imply that, for each $s > 0$,

$$v(t+s,\cdot) = e^{-tH} v(s,\cdot). \tag{6.39}$$

Indeed, if $w(t,\cdot)$ denotes the difference between the two sides of (6.39), then we have $w(0) = 0$, $w \in C(\mathbb{R}^+, \mathcal{D}(H))$, $\partial w/\partial t \in C(\mathbb{R}^+, L^2(\mathbb{R}))$, and

$$\frac{d}{dt} \|w(t)\|_{L^2}^2 = -2(Hw, w) \le 0,$$

so $w(t) = 0$, for all $t \ge 0$.

Finally, as $t \searrow 0$, we see from (6.31) that each $\sigma_j(t) \searrow 0$. Since $v(t,x)$ is also given by the right side of (6.27), we conclude that

$$v(t,\cdot) \to u \text{ in } L^2(\mathbb{R}), \quad \text{as } t \searrow 0. \tag{6.40}$$

Thus we can let $s \searrow 0$ in (6.39), obtaining a complete proof that $e^{-tH}u$ is given by (6.33) when $n = 1$.

It is useful to write down the formula for e^{-tH} using the Weyl calculus, introduced in §14 of Chap. 7. We recall that it associates to $a(x,\xi)$ the operator

$$\begin{aligned} a(X,D)u &= (2\pi)^{-n} \int \hat{a}(q,p) e^{i(q\cdot X + p\cdot D)} u(x)\, dq\, dp \\ &= (2\pi)^{-n} \int a\Big(\frac{x+y}{2}, \xi\Big) e^{i(x-y)\cdot\xi} u(y)\, dy\, d\xi. \end{aligned} \tag{6.41}$$

In other words, the operator $a(X,D)$ has integral kernel $K_a(x,y)$, for which

$$a(X,D)u(x) = \int K_a(x,y) u(y)\, dy,$$

given by

$$K_a(x,y) = (2\pi)^{-n} \int a\Big(\frac{x+y}{2}, \xi\Big) e^{i(x-y)\cdot\xi}\, d\xi$$

Recovery of $a(x,\xi)$ from $K_a(x,y)$ is an exercise in Fourier analysis. When it is applied to the formulas (6.33)–(6.36), this exercise involves computing a Gaussian integral, and we obtain the formula

$$e^{-tH} = h_t(X,D) \tag{6.42}$$

on $L^2(\mathbb{R}^n)$, with

$$h_t(x,\xi) = (\cosh t)^{-n}\, e^{-(\tanh t)(|x|^2+|\xi|^2)}. \tag{6.43}$$

It is interesting that this formula, while equivalent to (6.33)–(6.36), has a simpler and more symmetrical appearance.

In fact, the formula (6.43) was derived in §15 of Chap. 7, by a different method, which we briefly recall here. For reasons of symmetry, involving the identity (14.19), one can write

$$h_t(x,\xi) = g(t,Q), \quad Q(x,\xi) = |x|^2 + |\xi|^2. \tag{6.44}$$

Note that (6.42) gives $\partial_t\ h_t(X,D) = -Hh_t(X,D)$. Now the composition formula for the Weyl calculus implies that $h_t(x,\xi)$ satisfies the following evolution equation:

$$\begin{aligned}\frac{\partial}{\partial t}h_t(x,\xi) &= -(Q \circ h_t)(x,\xi) \\ &= -Q(x,\xi)h_t(x,\xi) - \frac{1}{2}\{Q,h_t\}_2(x,\xi) \\ &= -(|x|^2+|\xi|^2)h_t(x,\xi) + \frac{1}{4}\sum_k \big(\partial_{x_k}^2 + \partial_{\xi_k}^2\big)h_t(x,\xi).\end{aligned} \tag{6.45}$$

Given (6.44), we have for $g(t,Q)$ the equation

$$\frac{\partial g}{\partial t} = -Qg + Q\frac{\partial^2 g}{\partial Q^2} + n\frac{\partial g}{\partial Q}. \tag{6.46}$$

It is easy to verify that (6.43) solves this evolution equation, with $h_0(x,\xi) = 1$.

We can obtain a formula for

$$e^{-tQ(X,D)} = h_t^Q(X,D), \tag{6.47}$$

for a general positive-definite quadratic form $Q(x,\xi)$. First, in the case

$$Q(x,\xi) = \sum_{j=1}^n \mu_j(x_j^2+\xi_j^2), \quad \mu_j > 0, \tag{6.48}$$

it follows easily from (6.43) and multiplicativity, as in (6.36), that

$$h_t^Q(x,\xi) = \prod_{j=1}^n \big(\cosh t\mu_j\big)^{-1} \cdot \exp\left\{-\sum_{j=1}^n (\tanh t\mu_j)\Big(x_j^2+\xi_j^2\Big)\right\}. \tag{6.49}$$

Now any positive quadratic form $Q(x,\xi)$ can be put in the form (6.48) via a linear symplectic transformation, so to get the general formula we need only rewrite (6.49) in a symplectically invariant fashion. This is accomplished using the "Hamilton map" F_Q, a skew-symmetric transformation on $\mathbb{R}^{2n}$ defined by

$$Q(u,v) = \sigma(u, F_Q v), \quad u,v \in \mathbb{R}^{2n}, \tag{6.50}$$

where $Q(u, v)$ is the bilinear form polarizing Q, and σ is the symplectic form on $\mathbb{R}^{2n}$; $\sigma(u, v) = x \cdot \xi' - x' \cdot \xi$ if $u = (x, \xi)$, $v = (x', \xi')$. When Q has the form (6.48), F_Q is a sum of 2×2 blocks $\begin{pmatrix} 0 & \mu_j \\ -\mu_j & 0 \end{pmatrix}$, and we have

$$\prod_{j=1}^{n} \bigl(\cosh t\mu_j\bigr)^{-1} = \bigl(\det \cosh itF_Q\bigr)^{-1/2}. \tag{6.51}$$

Passing from F_Q to

$$A_Q = \bigl(-F_Q^2\bigr)^{1/2}, \tag{6.52}$$

the unique positive-definite square root, means passing to blocks

$$\begin{pmatrix} \mu_j & 0 \\ 0 & \mu_j \end{pmatrix},$$

and when Q has the form (6.48), then

$$\sum_{j=1}^{n} (\tanh t\mu_j)(x_j^2 + \xi_j^2) = tQ\bigl(\vartheta(tA_Q)\zeta, \zeta\bigr), \tag{6.53}$$

where $\zeta = (x, \xi)$ and

$$\vartheta(t) = \frac{\tanh t}{t}. \tag{6.54}$$

Thus the general formula for (6.47) is

$$h_t^Q(x, \xi) = \bigl(\cosh tA_Q\bigr)^{-1/2} e^{-tQ(\vartheta(tA_Q)\zeta, \zeta)}. \tag{6.55}$$

Exercises

1. Define an unbounded operator A on $L^2(\mathbb{R})$ by

$$\mathcal{D}(A) = \{u \in L^2(\mathbb{R}) : Du \in L^2(\mathbb{R}),\ xu \in L^2(\mathbb{R})\}, \quad Au = Du - ixu.$$

   Show that A is closed and that the self-adjoint operator H satisfies

$$H = A^*A + I = AA^* - I$$

   (*Hint*: Note Exercises 5–7 of §2.)
2. If $H_k(x)$ are the Hermite polynomials, show that there is the generating function identity

$$\sum_{k=0}^{\infty} \frac{1}{k!} H_k(x) s^k = e^{2xs-s^2}$$

(*Hint*: Use the first identity in (6.15).)

3. Show that Mehler's formula (6.34) is equivalent to the identity

$$\sum_{j=0}^{\infty} h_j(x) h_j(y) s^j =$$
$$\pi^{-1/2}(1-s^2)^{-1/2} \exp\Big\{(1-s^2)^{-1}\big[2xys-(x^2+y^2)s^2\big]\Big\} \cdot e^{-(x^2+y^2)/2},$$

for $0 \le s < 1$. Deduce that

$$\sum_{j=0}^{\infty} H_j(x)^2 \frac{s^j}{2^j j!} = (1-s^2)^{-1/2} e^{2sx^2/(1+s)}, \quad |s|<1.$$

4. Using

$$H^{-s} = \frac{1}{\Gamma(s)} \int_0^\infty e^{-tH}\, t^{s-1}\, dt, \quad \operatorname{Re} s > 0,$$

find the integral kernel $A_s(x,y)$ such that

$$H^{-s}u(x) = \int A_s(x,y) u(y)\, dy.$$

Writing $\operatorname{Tr} H^{-s} = \int A_s(x,x)\, dx$, Re $s > 1,\ n = 1$, show that

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{y^{s-1}}{e^y - 1}\, dy$$

See [Ing], pp. 41–44, for a derivation of the functional equation for the Riemann zeta function, using this formula.

5. Let $H_\omega = -d^2/dx^2 + \omega^2 x^2$. Show that e^{-tH_ω} has integral kernel

$$K_t^\omega(x,y) = (4\pi t)^{-1/2}\, \gamma(2\omega t)^{1/2}\, e^{-\gamma(2\omega t)[(\cosh 2\omega t)(x^2+y^2)-2xy]/4t},$$

where

$$\gamma(z) = \frac{z}{\sinh z}.$$

6. Consider the operator

$$\begin{aligned} Q(X,D) &= -\Big(\frac{\partial}{\partial x_1} - i\omega x_2\Big)^2 - \Big(\frac{\partial}{\partial x_2} + i\omega x_1\Big)^2 \\ &= -\Delta + \omega^2|x|^2 + 2i\omega\Big(x_2\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_2}\Big). \end{aligned}$$

Note that $Q(x,\xi)$ is nonnegative, but not definite. Study the integral kernel $K_t^Q(x,y)$ of $e^{-tQ(X,D)}$. Show that

$$K_t^Q(x,0) = (4\pi t)^{-1}\, \gamma(2\omega t)\, e^{-\tau(2\omega t)|x|^2/4t},$$

where

$$\tau(z) = z \coth z.$$

7. Let (ω_{jk}) be an invertible, $n \times n$, skew-symmetric matrix of real numbers (so n must be even). Suppose

$$L = -\sum_{j=1}^{n} \left(\frac{\partial}{\partial x_j} - i \sum_k \omega_{jk} x_k \right)^2 .$$

Evaluate the integral kernel $K_t^L(x, y)$, particularly at $y = 0$.

8. In terms of the operators a, a^+ given by (6.1) and the basis of $L^2(\mathbb{R})$ given by (6.16)–(6.18), show that

$$a^+ h_k = \sqrt{2k+2}\, h_{k+1}, \quad a h_k = \sqrt{2k}\, h_{k-1}.$$

7. The quantum Coulomb problem

In this section we examine the operator

$$Hu = -\Delta u - K|x|^{-1}u, \tag{7.1}$$

acting on functions on $\mathbb{R}^3$. Here, K is a positive constant.

This provides a quantum mechanical description of the Coulomb force between two charged particles. It is the first step toward a quantum mechanical description of the hydrogen atom, and it provides a decent approximation to the observed behavior of such an atom, though it leaves out a number of features. The most important omitted feature is the spin of the electron (and of the nucleus). Giving rise to further small corrections are the nonzero size of the proton, and relativistic effects, which confront one with great subtleties since relativity forces one to treat the electromagnetic field quantum mechanically. We refer to texts on quantum physics, such as [Mes], [Ser], [BLP], and [IZ], for work on these more sophisticated models of the hydrogen atom.

We want to define a self-adjoint operator via the Friedrichs method. Thus we want to work with a Hilbert space

$$\mathcal{H} = \left\{ u \in L^2(\mathbb{R}^3) : \nabla u \in L^2(\mathbb{R}^3), \int |x|^{-1}|u(x)|^2\, dx < \infty \right\}, \tag{7.2}$$

with inner product

$$(u, v)_{\mathcal{H}} = (\nabla u, \nabla v)_{L^2} + A(u, v)_{L^2} - K \int |x|^{-1} u(x)\overline{v(x)}\, dx, \tag{7.3}$$

where A is a sufficiently large, positive constant. We must first show that A can be picked to make this inner product positive-definite. In fact, we have the following:

Lemma 7.1. *For all $\varepsilon \in (0,1]$, there exists $C(\varepsilon) < \infty$ such that*

$$\int |x|^{-1}|u(x)|^2\, dx \le \varepsilon \|\nabla u\|_{L^2}^2 + C(\varepsilon)\|u\|_{L^2}^2, \tag{7.4}$$

for all $u \in H^1(\mathbb{R}^3)$.

Proof. Here and below we will use the inclusion

$$H^s(\mathbb{R}^n) \subset L^p(\mathbb{R}^n), \quad \forall\, p \in \Big[2, \frac{2n}{n-2s}\Big), \quad 0 \le s < \frac{n}{2}, \tag{7.5}$$

from (2.42) of Chap. 4. In Chap. 13 we will establish the sharper result that $H^s(\mathbb{R}^n) \subset L^{2n/(n-2s)}(\mathbb{R}^n)$; for example, $H^1(\mathbb{R}^3) \subset L^6(\mathbb{R}^3)$. We will also cite this stronger result in some arguments below, though that could be avoided.

We also use the fact that (if $B = \{|x| < 1\}$ and $\chi_B(x)$ is its characteristic function),

$$\chi_B V \in L^q(\mathbb{R}^3), \quad \text{for all } q < 3$$

Here and below we will use $V(x) = |x|^{-1}$. Thus the left side of (7.4) is bounded by

$$\|\chi_B V\|_{L^q} \cdot \|u\|_{L^{2q'}}^2 + \|u\|_{L^2}^2 \le C\|u\|_{H^\sigma(\mathbb{R}^3)}^2 + \|u\|_{L^2(\mathbb{R}^3)}^2, \tag{7.6}$$

where we can take any $q' > 3/2$; take $q' \in (3/2, 3)$. Then (7.6) holds for some $\sigma < 1$, for which $L^{2q'}(\mathbb{R}^3) \supset H^\sigma(\mathbb{R}^3)$. From this, (7.4) follows immediately.

Thus the Hilbert space $\mathcal{H}$ in (7.2) is simply $H^1(\mathbb{R}^3)$, and we see that indeed, for some $A > 0$, (7.3) defines an inner product equivalent to the standard one on $H^1(\mathbb{R}^3)$. The Friedrichs method then defines a positive, self-adjoint operator $H + AI$, for which

$$\mathcal{D}\big((H + AI)^{1/2}\big) = H^1(\mathbb{R}^3). \tag{7.7}$$

Then

$$\mathcal{D}(H) = \{u \in H^1(\mathbb{R}^3) : -\Delta u - K|x|^{-1}u \in L^2(\mathbb{R}^3)\}, \tag{7.8}$$

where $-\Delta u - K|x|^{-1}u$ is a priori regarded as an element of $H^{-1}(\mathbb{R}^3)$ if $u \in H^1(\mathbb{R}^3)$. Since $H^2(\mathbb{R}^3) \subset L^\infty(\mathbb{R}^3)$, we have

$$u \in H^2(\mathbb{R}^3) \Longrightarrow |x|^{-1}u \in L^2(\mathbb{R}^3), \tag{7.9}$$

so

$$\mathcal{D}(H) \supset H^2(\mathbb{R}^3). \tag{7.10}$$

Indeed, we have:

Proposition 7.2. *For the self-adjoint extension H of $-\Delta - K|x|^{-1}$ defined above,*

$$\mathcal{D}(H) = H^2(\mathbb{R}^3). \tag{7.11}$$

Proof. Pick λ in the resolvent set of H; for instance, $\lambda \in \mathbb{C} \setminus \mathbb{R}$. If $u \in \mathcal{D}(H)$ and $(H - \lambda)u = f \in L^2(\mathbb{R}^3)$, we have

$$u - KR_\lambda V u = R_\lambda f = g_\lambda, \tag{7.12}$$

where $V(x) = |x|^{-1}$ and $R_\lambda = (-\Delta - \lambda)^{-1}$. Now the operator of multiplication by $V(x) = |x|^{-1}$ has the property

$$M_V : H^1(\mathbb{R}^3) \longrightarrow L^{2-\varepsilon}(\mathbb{R}^3), \tag{7.13}$$

for all $\varepsilon > 0$, since $H^1(\mathbb{R}^3) \subset L^6(\mathbb{R}^3) \cap L^2(\mathbb{R}^3)$ and $V \in L^{3-\varepsilon}$ on $|x| < 1$. Hence

$$M_V : H^1(\mathbb{R}^3) \longrightarrow H^{-\varepsilon}(\mathbb{R}^3),$$

for all $\varepsilon > 0$. Let us apply this to (7.12). We know that $u \in \mathcal{D}(H) \subset \mathcal{D}(H^{1/2}) = H^1(\mathbb{R}^3)$, so $KR_\lambda V u \in H^{2-\varepsilon}(\mathbb{R}^3)$. Thus $u \in H^{2-\varepsilon}(\mathbb{R}^3)$, for all $\varepsilon > 0$. But, for $\varepsilon > 0$ small enough,

$$M_V : H^{2-\varepsilon}(\mathbb{R}^3) \longrightarrow L^2(\mathbb{R}^3), \tag{7.14}$$

so then $u = KR_\lambda(Vu) + R_\lambda f \in H^2(\mathbb{R}^3)$. This proves that $\mathcal{D}(H) \subset H^2(\mathbb{R}^3)$ and gives (7.11).

Since H is self-adjoint, its spectrum is a subset of the real axis, $(-\infty, \infty)$. We next show that there is only point spectrum in $(-\infty, 0)$

Proposition 7.3. *The part of* spec H *lying in* $\mathbb{C} \setminus [0, \infty)$ *is a bounded, discrete subset of* $(-\infty, 0)$*, consisting of eigenvalues of finite multiplicity and having at most* $\{0\}$ *as an accumulation point.*

Proof. Consider the equation $(H - \lambda)u = f \in L^2(\mathbb{R}^3)$, that is,

$$(-\Delta - \lambda)u - KVu = f, \tag{7.15}$$

with $V(x) = |x|^{-1}$ as before. Applying $R_\lambda = (-\Delta - \lambda)^{-1}$ to both sides, we again obtain (7.12):

$$(I - KR_\lambda M_V)u = g_\lambda = R_\lambda f. \tag{7.16}$$

Note that R_λ is a holomorphic function of $\lambda \in \mathbb{C} \setminus [0, \infty)$, with values in $\mathcal{L}(L^2(\mathbb{R}^3), H^2(\mathbb{R}^3))$. A key result in the analysis of (7.16) is the following:

Lemma 7.4. *For* $\lambda \in \mathbb{C} \setminus [0, \infty)$,

$$R_\lambda M_V \in \mathcal{K}(L^2(\mathbb{R}^3)), \tag{7.17}$$

where $\mathcal{K}$ *is the space of compact operators.*

We will establish this via the following basic tool. For $\lambda \in \mathbb{C} \setminus [0,\infty)$, $\varphi \in C_0(\mathbb{R}^3)$, the space of continuous functions vanishing at infinity, we have

$$M_\varphi R_\lambda \in \mathcal{K}(L^2) \text{ and } R_\lambda M_\varphi \in \mathcal{K}(L^2). \tag{7.18}$$

To see this, note that, for $\varphi \in C_0^\infty(\mathbb{R}^3)$, the first inclusion in (7.18) follows from Rellich's theorem. Then this inclusion holds for uniform limits of such φ, hence for $\varphi \in C_0(\mathbb{R}^3)$. Taking adjoints yields the rest of (7.18).

Now, to establish (7.17), write

$$V = V_1 + V_2, \tag{7.19}$$

where $V_1 = \psi V$, $\psi \in C_0^\infty(\mathbb{R}^3)$, $\psi(x) = 1$ for $|x| \le 1$. Then $V_2 \in C_0(\mathbb{R}^3)$, so $R_\lambda M_{V_2} \in \mathcal{K}$. We have $V_1 \in L^q(\mathbb{R}^3)$, for all $q \in [1,3)$, so, taking $q = 2$, we have

$$M_{V_1} : L^2(\mathbb{R}^3) \longrightarrow L^1(\mathbb{R}^3) \subset H^{-3/2-\varepsilon}(\mathbb{R}^3), \tag{7.20}$$

for all $\varepsilon > 0$, hence

$$R_\lambda M_{V_1} : L^2(\mathbb{R}^3) \longrightarrow H^{1/2-\varepsilon}(\mathbb{R}^3) \subset L^2(\mathbb{R}^3). \tag{7.21}$$

Given V_1 supported on a ball B_R, the operator norm in (7.21) is bounded by a constant times $\|V_1\|_{L^2}$. You can approximate V_1 in L^2-norm by a sequence $w_j \in C_0^\infty(\mathbb{R}^3)$. It follows that $R_\lambda M_{V_1}$ is a norm limit of a sequence of compact operators on $L^2(\mathbb{R}^3)$, so it is also compact, and (7.17) is established.

The proof of Proposition 7.4 is finished by the following result, which can be found as Proposition 7.4 in Chap. 9

Proposition 7.5. *Let $\mathcal{O}$ be a connected, open set in $\mathbb{C}$. Suppose $C(\lambda)$ is a compact-operator-valued holomorphic function of $\lambda \in \mathcal{O}$. If $I - C(\lambda)$ is invertible at one point $p \in \mathcal{O}$, then it is invertible except at most on a discrete set in $\mathcal{O}$, and $(I - C(\lambda))^{-1}$ is meromorphic on $\mathcal{O}$.*

This applies to our situation, with $C(\lambda) = KR_\lambda M_V$; we know that $I - C(\lambda)$ is invertible for all $\lambda \in \mathbb{C} \setminus \mathbb{R}$ in this case.

One approach to analyzing the negative eigenvalues of H is to use polar coordinates. If $-K|x|^{-1}$ is replaced by any radial potential $\mathcal{V}(|x|)$, the eigenvalue equation $Hu = -Eu$ becomes

$$\frac{\partial^2 u}{\partial r} + \frac{2}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\Delta_S u - \mathcal{V}(r)u = Eu. \tag{7.22}$$

We can use separation of variables, writing $u(r\theta) = v(r)\varphi(\theta)$, where φ is an eigenfunction of Δ_S, the Laplace operator on S^2,

$$\Delta_S \varphi = -\lambda\varphi, \quad \lambda = \left(k + \frac{1}{2}\right)^2 - \frac{1}{4} = k^2 + k. \tag{7.23}$$

Then we obtain for $v(r)$ the ODE

$$v''(r) + \frac{2}{r}v'(r) + f(r)v(r) = 0, \quad f(r) = -E - \frac{\lambda}{r^2} - \mathcal{V}(r). \tag{7.24}$$

One can eliminate the term involving v' by setting

$$w(r) = rv(r). \tag{7.25}$$

Then

$$w''(r) + f(r)w(r) = 0. \tag{7.26}$$

For the Coulomb problem, this becomes

$$w''(r) + \Big[-E + \frac{K}{r} - \frac{\lambda}{r^2}\Big]w(r) = 0. \tag{7.27}$$

If we set $W(r) = w(\beta r),\ \beta = 1/2\sqrt{E}$, we get a form of Whittaker's ODE:

$$W''(z) + \Big[-\frac{1}{4} + \frac{\varkappa}{z} + \frac{\frac{1}{4} - \mu^2}{z^2}\Big]W(z) = 0, \tag{7.28}$$

with

$$\varkappa = \frac{K}{2\sqrt{E}}, \quad \mu^2 = \lambda + \frac{1}{4} = \Big(k + \frac{1}{2}\Big)^2. \tag{7.29}$$

This in turn can be converted to the confluent hypergeometric equation

$$z\psi''(z) + (b - z)\psi'(z) - a\psi(z) = 0 \tag{7.30}$$

upon setting

$$W(z) = z^{\mu+1/2}\, e^{-z/2}\, \psi(z), \tag{7.31}$$

with

$$\begin{aligned} a &= \mu - \varkappa + \frac{1}{2} = k + 1 - \frac{K}{2\sqrt{E}}, \\ b &= 2\mu + 1 = 2k + 2. \end{aligned} \tag{7.32}$$

Note that ψ and v are related by

$$v(r) = (2\sqrt{E})^{k+1}\, r^k e^{-2\sqrt{E}r}\psi(2\sqrt{E}r). \tag{7.33}$$

Looking at (7.28), we see that there are two independent solutions, one behaving roughly like $e^{-z/2}$ and the other like $e^{z/2}$, as $z \to +\infty$. Equivalently, (7.30) has two linearly independent solutions, a "good" one growing more slowly than exponentially and a "bad" one growing like e^z, as $z \to +\infty$. Of course, for a solution to give rise to an eigenfunction, we need $v \in L^2(\mathbb{R}^+, r^2\,dr)$, that is, $w \in L^2(\mathbb{R}^+, dr)$. We need to have simultaneously $w(z) \sim ce^{-z/2}$ (roughly) as $z \to +\infty$ and w square integrable near $z = 0$. In view of (7.8), we also need $v' \in L^2(\mathbb{R}^+, r^2\,dr)$.

To examine the behavior near $z = 0$, note that the Euler equation associated with (7.28) is

$$z^2 W''(z) + \Big(\frac{1}{4} - \mu^2\Big) W(z) = 0, \tag{7.34}$$

with solutions $z^{1/2+\mu}$ and $z^{1/2-\mu}$, i.e., z^{k+1} and z^{-k}, $k = 0, 1, 2, \dots$. If $k = 0$, both are square integrable near 0, but for $k \ge 1$ only one is. Going to the confluent hypergeometric equation (7.30), we see that two linearly independent solutions behave respectively like z^0 and $z^{-2\mu} = z^{-2k-1}$ as $z \to 0$.

As a further comment on the case $k = 0$, note that a solution W behaving like z^0 at $z = 0$ gives rise to $v(r) \sim C/r$ as $r \to 0$, with $c \neq 0$, hence $v'(r) \sim -C/r^2$. This is not square integrable near $r = 0$, with respect to $r^2\,dr$, so also this case does not produce an eigenfunction of H.

If $b \notin \{0, -1, -2, \dots\}$, which certainly holds here, the solution to (7.30) that is "good" near $z = 0$ is given by the confluent hypergeometric function

$${}_1F_1(a; b; z) = \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n} \frac{z^n}{n!}, \tag{7.35}$$

an entire function of z. Here, $(a)_n = a(a+1)\cdots(a+n-1)$; $(a)_0 = 1$. If also $a \notin \{0, -1, -2, \dots\}$, it can be shown that

$${}_1F_1(a; b; z) \sim \frac{\Gamma(b)}{\Gamma(a)}\, e^z\, z^{-(b-a)}, \quad z \to +\infty. \tag{7.36}$$

See the exercises below for a proof of this. Thus the "good" solution near $z = 0$ is "bad" as $z \to +\infty$, *unless* a is a nonpositive integer, say $a = -j$. In that case, as is clear from (7.35), ${}_1F_1(-j; b; z)$ is a polynomial in z, thus "good" as $z \to +\infty$. Thus the negative eigenvalues of H are given by $-E$, with

$$\frac{K}{2\sqrt{E}} = j + k + 1 = n, \tag{7.37}$$

that is, by

$$E = \frac{K^2}{4n^2}, \quad n = 1, 2, 3 \dots. \tag{7.38}$$

Note that, for each value of n, one can write $n = j + k + 1$ using n choices of $k \in \{0, 1, 2, \dots, n-1\}$. For each such k, the $(k^2 + k)$-eigenspace of Δ_S has dimension $2k + 1$, as established in Corollary 4.3. Thus the eigenvalue $-E = -K^2/4n^2$ of H has multiplicity

$$\sum_{k=0}^{n-1} (2k+1) = n^2. \tag{7.39}$$

Let us denote by V_n the n^2-dimensional eigenspace of H, associated to the eigenvalue $\lambda_n = -K^2/4n^2$.

The rotation group SO(3) acts on each V_n, via

$$\rho(g) f(x) = f(g^{-1}x), \quad g \in SO(3),\ x \in \mathbb{R}^3$$

By the analysis leading to (7.39), this action on V_n is not irreducible, but rather has n irreducible components. This suggests that there is an extra symmetry, and indeed, as W. Pauli discovered early in the history of quantum mechanics, there is one, arising via the Lenz vector (briefly introduced in §16 of Chap. 1), which we proceed to define.

The angular momentum vector $\mathbf{L} = \mathbf{x} \times \mathbf{p}$, with $\mathbf{p}$ replaced by the vector operator $(\partial/\partial x_1, \partial/\partial x_2, \partial/\partial x_3)$, commutes with H as a consequence of the rotational invariance of H. The components of $\mathbf{L}$ are

$$L_\ell = x_j \frac{\partial}{\partial x_k} - x_k \frac{\partial}{\partial x_j}, \tag{7.40}$$

where (j, k, ℓ) is a cyclic permutation of $(1, 2, 3)$. Then the Lenz vector is defined by

$$\mathbf{B} = \frac{1}{K}\Big(\mathbf{L} \times \mathbf{p} - \mathbf{p} \times \mathbf{L}\Big) - \frac{\mathbf{x}}{r}, \tag{7.41}$$

with components B_j, $1 \le j \le 3$, each of which is a second-order differential operator, given explicitly by

$$B_j = \frac{1}{K}(L_k \partial_\ell + \partial_\ell L_k - L_\ell \partial_k - \partial_k L_\ell) - \frac{x_j}{r}, \tag{7.42}$$

where (j, k, ℓ) is a cyclic permutation of $(1, 2, 3)$. A calculation gives

$$[H, B_j] = 0, \tag{7.43}$$

in the sense that these operators commute on $C^\infty(\mathbb{R}^3 \setminus 0)$.

It follows that if $u \in V_n$, then $B_j u$ is annihilated by $H - \lambda_n$, on $\mathbb{R}^3 \setminus 0$. Now, we have just gone through an argument designed to glean from all functions that are so annihilated, those that are actually eigenfunctions of H. In view of that, it is important to establish the next lemma

Lemma 7.6. *We have*

$$B_j : V_n \longrightarrow V_n. \tag{7.44}$$

Proof. Let $u \in V_n$. We know that $u \in \mathcal{D}(H) = H^2(\mathbb{R}^3)$. Also, from the analysis of the ODE (7.28), we know that $u(x)$ decays as $|x| \to \infty$, roughly like $e^{-|\lambda_n|^{1/2}|x|}$. It follows from (7.42) that $B_j u \in L^2(\mathbb{R}^3)$. It will be useful to obtain a bit more regularity, using $V_n \subset \mathcal{D}(H^2)$ together with the following.

Proposition 7.7. *If $u \in \mathcal{D}(H^2)$, then, for all $\varepsilon > 0$,*

$$u \in H^{5/2-\varepsilon}(\mathbb{R}^3). \tag{7.45}$$

Furthermore,

$$g \in \mathcal{S}(\mathbb{R}^3),\ g(0) = 0 \Longrightarrow gu \in H^{7/2-\varepsilon}(\mathbb{R}^3). \tag{7.46}$$

Proof. We proceed along the lines of the proof of Proposition 7.2, using (7.12), i.e.,

$$u = KR_\lambda Vu + R_\lambda f, \tag{7.47}$$

where $f = (H - \lambda)u$, with λ chosen in $\mathbb{C} \setminus \mathbb{R}$. We know that $f = (H - \lambda)u$ belongs to $\mathcal{D}(H)$, so $R_\lambda f \in H^4(\mathbb{R}^3)$. We know that $u \in H^2(\mathbb{R}^3)$. Parallel to (7.13), we can show that, for all $\varepsilon > 0$,

$$M_V : H^2(\mathbb{R}^3) \longrightarrow H^{1/2-\varepsilon}(\mathbb{R}^3), \tag{7.48}$$

so $KR_\lambda Vu \in H^{5/2-\varepsilon}(\mathbb{R}^3)$. This gives (7.45).

Now, multiply (7.47) by g and write

$$gu = KR_\lambda gVu + K[M_g, R_\lambda]Vu + gR_\lambda f. \tag{7.49}$$

This time we have

$$M_{gV} : H^2(\mathbb{R}^3) \longrightarrow H^{3/2-\varepsilon}(\mathbb{R}^3),$$

so $R_\lambda gVu \in H^{7/2-\varepsilon}(\mathbb{R}^3)$. Furthermore,

$$[M_g, R_\lambda] = R_\lambda\, [\Delta, M_g]\, R_\lambda : H^s(\mathbb{R}^3) \longrightarrow H^{s+3}(\mathbb{R}^3), \tag{7.50}$$

so $[M_g, R_\lambda]Vu \in H^{7/2-\varepsilon}(\mathbb{R}^3)$. This establishes (7.46).

We can now finish the proof of Lemma 7.6. Note that the second-order derivatives in B_j have a coefficient vanishing at 0. Keep in mind the known exponential decay of $u \in V_n$. Also note that $M_{x_j/r} : H^2(\mathbb{R}^3) \to H^{3/2-\varepsilon}(\mathbb{R}^3)$. Therefore,

$$u \in V_n \Longrightarrow B_j u \in H^{3/2-\varepsilon}(\mathbb{R}^3). \tag{7.51}$$

Consequently,

$$\Delta(B_j u) \in H^{-1/2-\varepsilon}(\mathbb{R}^3), \text{ and } V(B_j u) \in L^1(\mathbb{R}^3) + L^2(\mathbb{R}^3). \tag{7.52}$$

Thus $(H-\lambda_n)(B_j u)$, which we know vanishes on $\mathbb{R}^3 \setminus 0$, must vanish completely, since (7.52) does not allow for a nonzero quantity supported on $\{0\}$. Using (7.8), we conclude that $B_j u \in \mathcal{D}(H)$, and the lemma is proved.

With Lemma 7.6 established, we can proceed to study the action of B_j and L_j on V_n. When (j,k,ℓ) is a cyclic permutation of $(1,2,3)$, we have

$$[L_j, L_k] = L_\ell, \tag{7.53}$$

and, after a computation,

$$[L_j, B_k] = B_\ell, \quad [B_j, B_k] = -\frac{4}{K} H L_\ell. \tag{7.54}$$

Of course, (7.52) is the statement that L_j span the Lie algebra so(3) of SO(3). The identities (7.54), when L_j and B_j act on V_n, can be rewritten as

$$[L_j, A_k] = A_\ell, \quad [A_j, A_k] = A_\ell, \quad A_j = \frac{K}{2\sqrt{-\lambda_n}} B_j. \tag{7.55}$$

If we set

$$\mathbf{M} = \frac{1}{2}(\mathbf{L} + \mathbf{A}), \quad \mathbf{N} = \frac{1}{2}(\mathbf{L} - \mathbf{A}), \tag{7.56}$$

we get, for cyclic permutations (j,k,ℓ) of $(1,2,3)$,

$$[M_j, M_k] = M_\ell, \quad [N_j, N_k] = N_\ell, \quad [M_j, N_{j'}] = 0, \tag{7.57}$$

which is clearly the set of commutation relations for the Lie algebra so(3) $\oplus$ so(3). We next aim to show that this produces an irreducible representation of SO(4) on V_n, and to identify this representation. A priori, of course, one certainly has a representation of SU(2) $\times$ SU(2) on V_n.

We now examine the behavior on V_n of the Casimir operators $M^2 = M_1^2 + M_2^2 + M_3^2$ and N^2. A calculation using the definitions gives $\mathbf{B} \cdot \mathbf{L} = 0$, hence $\mathbf{A} \cdot \mathbf{L} = 0$, so, on V_n,

$$M^2 = N^2 = \frac{1}{4}(A^2 + L^2) = \frac{1}{4}\Big(L^2 - \frac{K^2}{4\lambda_n}B^2\Big). \tag{7.58}$$

We also have the following key identity:

$$K^2(B^2 - I) = 4H(L^2 + I), \tag{7.59}$$

which follows from the definitions by a straightforward computation. If we compare (7.58) and (7.59) on V_n, where $H = \lambda_n$, we get

$$4M^2 = 4N^2 = -\Big(1 + \frac{K^2}{4\lambda_n}\Big)I \quad \text{on } V_n. \tag{7.60}$$

Now the representation σ_n we get of SU(2) × SU(2) on V_n is a direct sum (possibly with only one summand) of representations $D_{j/2} \otimes D_{j/2}$, where $D_{j/2}$ is the standard irreducible representation of SU(2) on $\mathbb{C}^{j+1}$, defined in §9 of Appendix B. The computation (7.60) implies that all the copies in this sum are isomorphic, that is, for some $j = j(n)$,

$$\sigma_n = \bigoplus_{\ell=1}^{\mu} D_{j(n)/2} \otimes D_{j(n)/2}. \tag{7.61}$$

A dimension count gives $\mu\big(j(n)+1\big)^2 = n^2$. Note that on $D_{j/2} \otimes D_{j/2}$, we have $M^2 = N^2 = (j/2)(j/2+1)$. Thus (7.60) implies $j(j+2) = -1 + K^2/4\lambda_n$, or

$$\lambda_n = -\frac{K^2}{4(j+1)^2}, \quad j = j(n). \tag{7.62}$$

Comparing (7.38), we have $(j+1)^2 = n^2$, that is,

$$j(n) = n - 1. \tag{7.63}$$

Since we know that $\dim V_n = n^2$, this implies that there is just one summand in (7.61), so

$$\sigma_n = D_{(n-1)/2} \otimes D_{(n-1)/2}. \tag{7.64}$$

This is an irreducible representation of SU(2) × SU(2), which is a double cover of SO(4),

$$\kappa : \mathrm{SU}(2) \times \mathrm{SU}(2) \longrightarrow \mathrm{SO}(4).$$

It is clear that σ_n is the identity operator on both elements in $\ker \kappa$, and so σ_n actually produces an irreducible representation of SO(4).

Let ρ_n denote the restriction to V_n of the representation ρ of SO(3) on $L^3(\mathbb{R}^3)$, described above. If we regard this as a representation of SU(2), it is clear that ρ_n is the composition of σ_n with the diagonal map $\mathrm{SU}(2) \to \mathrm{SU}(2) \times \mathrm{SU}(2)$. Results established in §9 of Appendix B imply that such a tensor-product representation of SU(2) has the decomposition into irreducible representations:

$$\rho_n \approx \bigoplus_{k=0}^{n-1} D_k. \tag{7.65}$$

This is also precisely the description of ρ_n given by the analysis leading to (7.39).

There are a number of other group-theoretic perspectives on the quantum Coulomb problem, which can be found in [Eng] and [GS2]. See also [Ad] and [Cor], Vol. 2.

Exercises

1. For $H = -\Delta - K|x|^{-1}$ with domain given by (7.8), show that

$$\mathcal{D}(H) = \{u \in L^2(\mathbb{R}^3) : -\Delta u - K|x|^{-1}u \in L^2(\mathbb{R}^3)\}, \tag{7.66}$$

where a priori, if $u \in L^2(\mathbb{R}^3)$, then $\Delta u \in H^{-2}(\mathbb{R}^3)$ and $|x|^{-1}u \in L^1(\mathbb{R}^3) + L^2(\mathbb{R}^3) \subset H^{-2}(\mathbb{R}^3)$.
(*Hint*: Parallel the proof of Proposition 7.2. If u belongs to the right side of (7.66), and if you pick $\lambda \in \mathbb{C} \setminus \mathbb{R}$, then, as in (7.12),

$$u - KR_\lambda Vu = R_\lambda f \in H^2(\mathbb{R}^3).) \tag{7.67}$$

Complement (7.13) with

$$\begin{aligned} M_V &: L^2(\mathbb{R}^3) \longrightarrow \bigcap_{\varepsilon>0} H^{-3/2-\varepsilon}(\mathbb{R}^3), \\ M_V &: \bigcap_{\varepsilon>0} H^{1/2-\varepsilon}(\mathbb{R}^3) \longrightarrow \bigcap_{\delta>0} H^{-3/4-\delta}(\mathbb{R}^3). \end{aligned} \tag{7.68}$$

(Indeed, sharper results can be obtained.) Then deduce from (7.67) first that $u \in H^{1/2-\varepsilon}(\mathbb{R}^3)$ and then that $u \in H^{5/4-\delta}(\mathbb{R}^3) \subset H^1(\mathbb{R}^3)$.)
2. As a variant of (7.4), show that, for $u \in H^1(\mathbb{R}^3)$,

$$\int |x|^{-2}|u(x)|^2\,dx \le 4\int |\nabla u(x)|^2\,dx. \tag{7.69}$$

Show that 4 is the best possible constant on the right. (*Hint*: Use the Mellin transform to show that the spectrum of $r\,d/dr - 1/2$ on $L^2(\mathbb{R}^+, r^{-1}dr)$ (which coincides with the spectrum of $r\,d/dr$ on $L^2(\mathbb{R}^+, dr)$) is $\{is - 1/2 : s \in \mathbb{R}\}$, hence

$$\int_0^\infty |u(r)|^2 dr \le 4\int_0^\infty |u'(r)|^2 r^2\,dr. \tag{7.70}$$

This is sometimes called an "uncertainty principle" estimate. Why might that be? (Cf. [RS], Vol. 2, p. 169.)

3. Show that $H = -\Delta - K/|x|$ has no non-negative eigenvalues, i.e., only continuous spectrum in $[0, \infty)$. (*Hint*: Study the behavior as $r \to +\infty$ of solutions to the ODE (7.28), when $-E$ is replaced by $+E \in [0, \infty)$. Consult [Olv] for techniques. See also [RS], Vol. 4, for general results.)
4. Generalize the propositions of this section, with modifications as needed, to other classes of potentials $V(x)$, such as

$$V \in L^2 + \varepsilon L^\infty,$$

the set of functions V such that, for each $\varepsilon > 0$, one can write $V = V_1 + V_2,\ V_1 \in L^2,\ \|V_2\|_{L^\infty} \le \varepsilon$. Consult [RS], Vols. 2–4, for further generalizations.

Exercises on the confluent hypergeometric function

1. Taking (7.35) as the definition of ${}_1F_1(a; b; z)$, show that

$$(7.71)\qquad {}_1F_1(a;b;z) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \int_0^1 e^{zt} t^{a-1} (1-t)^{b-a-1}\, dt, \quad \operatorname{Re} b > \operatorname{Re} a > 0.$$

(*Hint*: Use the beta function identity, (A.23)–(A.24) of Chap. 3.) Show that (7.71) implies the asymptotic behavior (7.36), provided $\operatorname{Re} b > \operatorname{Re} a > 0$, but that this is insufficient for making the deduction (7.37).

Exercises 2–5 deal with the analytic continuation of (7.71) in a and b, and a complete justification of (7.36). To begin, write

$$(7.72)\qquad {}_1F_1(a;b;z) = \frac{\Gamma(b)}{\Gamma(b-a)} A_\psi(a, -z) + \frac{\Gamma(b)}{\Gamma(a)} A_\varphi(b-a, z) e^z,$$

where, for $\operatorname{Re} c > 0,\ \psi \in C^\infty\big([0, 1/2]\big)$, we set

$$(7.73)\qquad A_\psi(c, z) = \frac{1}{\Gamma(c)} \int_0^{1/2} e^{-zt} \psi(t) t^{c-1}\, dt,$$

and, in (7.72),

$$\psi(t) = (1-t)^{b-a-1}, \quad \varphi(t) = (1-t)^{a-1}.$$

2. Given $\operatorname{Re} c > 0$, show that

$$(7.74)\qquad A_\psi(c, z) \sim \psi(0) z^{-c}, \quad z \to +\infty,$$

and

$$(7.75)\qquad A_\psi(c, -z) \sim \frac{\psi(\frac{1}{2})}{\Gamma(c)} z^{-1} e^{z/2}, \quad z \to +\infty.$$

3. For $j = 0, 1, 2, \dots,$ set

$$(7.76)\qquad A_j(c, t) = \frac{1}{\Gamma(c)} \int_0^{1/2} e^{-zt} t^j t^{c-1}\, dt,$$

so $A_j(c,z) = A_\psi(c,z)$, with $\psi(t) = t^j$. Show that

$$A_j(c,z) = \frac{\Gamma(c+j)}{\Gamma(c)} z^{-c-j} - \frac{1}{\Gamma(c)} \int_{1/2}^{\infty} e^{-zt} t^{c+j-1}\, dt,$$

for $\operatorname{Re} z > 0$. Deduce that $A_j(c,t)$ is an entire function of c, for $\operatorname{Re} z > 0$, and that

$$A_j(c,z) \sim \frac{\Gamma(c+j)}{\Gamma(c)} z^{-c-j}, \quad z \to +\infty,$$

if $c \notin \{0, -1, -2, \dots\}$.

4. Given $k = 1, 2, 3, \dots$, write

$$\psi(t) = a_0 + a_1 t + \cdots + a_{k-1} t^{k-1} + \psi_k(t) t^k, \quad \psi_k \in C^\infty\Big(\Big[0, \frac{1}{2}\Big]\Big)$$

Thus

$$A_\psi(c,z) = \sum_{j=0}^{k-1} a_j A_j(c,z) + \frac{1}{\Gamma(c)} \int_0^{1/2} e^{-zt} \psi_k(t) t^{k+c-1}\, dt. \tag{7.77}$$

Deduce that $A_\psi(c,z)$ can be analytically continued to $\operatorname{Re} c > -k$ when $\operatorname{Re} z > 0$ and that (7.74) continues to hold if $c \notin \{0, -1, -2, \dots\}$, $a_0 \neq 0$.

5. Using $t^{c-1} = c^{-1}(d/dt)t^c$ and integrating by parts, show that

$$A_0(c,z) = z A_0(c+1,z) - \frac{1}{2^c \Gamma(c+1)} e^{-z/2}, \tag{7.78}$$

for $\operatorname{Re} c > 0$, all $z \in \mathbb{C}$. Show that this provides an entire analytic continuation of $A_0(c,z)$ and that (7.74)–(7.75) hold, for $\psi(t) = 1$. Using

$$A_j(c,z) = \frac{\Gamma(c+j)}{\Gamma(c)} A_0(c+j, z)$$

and (7.77), verify (7.75) for all $\psi \in C^\infty([0, 1/2])$. (Also again verify (7.74)). Hence, verify the asymptotic expansion (7.36).

The approach given above to (7.36) is one the author learned from conversations with A. N. Varchenko. In Exercises 6–15 below, we introduce another solution to the confluent hypergeometric equation and follow a path to the expansion (7.36) similar to one described in [Leb] and in [Olv].

6. Show that a solution to the ODE (7.30) is also given by

$$z^{1-b}\ {}_1F_1(1+a-b; 2-b; z),$$

in addition to ${}_1F_1(a;b;z)$, defined by (7.35). Assume $b \neq 0, -1, -2, \dots$. Set

$$\begin{aligned}\Psi(a;b;z) = {} & \frac{\Gamma(1-b)}{\Gamma(1+a-b)}\ {}_1F_1(a;b;z) \\ & + \frac{\Gamma(b-1)}{\Gamma(a)}\ z^{1-b}\ {}_1F_1(1+a-b; 2-b; z).\end{aligned} \tag{7.79}$$

Show that the Wronskian is given by

$$W\big({}_1F_1(a;b;z), \Psi(a;b;z)\big) = -\frac{\Gamma(b)}{\Gamma(a)} z^{-b} e^z.$$

7. Show that

$$(7.80)\qquad {}_1F_1(a;b;z) = e^z\, {}_1F_1(b-a;b;-z), \quad b \notin \{0,-1,-2,\dots\}$$

(*Hint*: Use the integral in Exercise 1, and set $s = 1-t$, for the case $\operatorname{Re} b > \operatorname{Re} a > 0$.)

8. Show that

$$(7.81)\qquad \Psi(a;b;z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zt} t^{a-1} (1+t)^{b-a-1}\, dt, \quad \operatorname{Re} a > 0,\ \operatorname{Re} z > 0.$$

(*Hint*: First show that the right side solves (7.30). Then check the behavior as $z \to 0$.)

9. Show that

$$(7.82)\qquad \Psi(a;b;z) = z\Psi(a+1;b+1;z) + (1-a-b)\Psi(a+1;b;z).$$

(*Hint*: To get this when $\operatorname{Re} a > 0$, use the integral expression (7.81) for $\Psi(a+1;b+1;z)$, write $ze^{-zt} = -(d/dt)e^{-zt}$, and integrate by parts.)

10. Show that

$$(7.83)\qquad \begin{aligned} {}_1F_1(a;b;z) = {} & \frac{\Gamma(b)}{\Gamma(b-a)} e^{\pm\pi a i}\, \Psi(a;b;z) \\ & + \frac{\Gamma(b)}{\Gamma(a)} e^{\pm\pi(a-b)i}\, e^z\, \Psi(b-a;b;-z), \end{aligned}$$

where $-z = e^{\mp\pi i} z$, $b \neq 0, -1, -2, \dots$. (*Hint*: Make use of (7.80) as well as (7.79).)

11. Using the integral representation (7.81), show that under the hypotheses $\delta > 0$, $b \notin \{0,-1,-2,\dots\}$, and $\operatorname{Re} a > 0$, we have

$$(7.84)\qquad \Psi(a;b;z) \sim z^{-\alpha}, \quad |z| \to \infty,$$

in the sector

$$(7.85)\qquad |\operatorname{Arg} z| \le \frac{\pi}{2} - \delta.$$

12. Extend (7.84) to the sector $|\operatorname{Arg} z| \le \pi - \delta$. (*Hint*: Replace (7.81) by an integral along the ray $\gamma = \{e^{i\alpha}s : 0 \le s < \infty\}$, given $|\alpha| < \pi/2$.)
13. Further extend (7.84) to the case where no restriction is placed on $\operatorname{Re} a$. (*Hint*: Use (7.82).)
14. Extend (7.84) still further, to be valid for

$$(7.86)\qquad |\operatorname{Arg} z| \le \frac{3\pi}{2} - \delta.$$

(*Hint*: See Theorem 2.2 on p. 235 of [Olv], and its application to this problem on p. 256 of [Olv].)

15. Use (7.83)–(7.86) to prove (7.36), that is,

$$(7.87)\qquad {}_1F_1(a;b;z) \sim \frac{\Gamma(b)}{\Gamma(a)}\, e^z\, z^{-(b-a)}, \quad z \to +\infty,$$

provided $a, b \notin \{0,-1,-2,\dots\}$.

Remarks: For the analysis of $\Psi(b-a; b; -z)$ as $z \to +\infty$, the result of Exercise 14 suffices, but the result of Exercise 13 does not. This point appears to have been neglected in the discussion of (7.87) on p. 271 of [Leb].

8. The Laplace operator on cones

Generally, if N is any compact Riemannian manifold of dimension m, possibly with boundary, the cone over N, $C(N)$, is the space $\mathbb{R}^+ \times N$ together with the Riemannian metric

$$dr^2 + r^2 g, \tag{8.1}$$

where g is the metric tensor on N. In particular, a cone with vertex at the origin in $\mathbb{R}^{m+1}$ can be described as the cone over a subdomain Ω of the unit sphere S^m in $\mathbb{R}^{m+1}$. Our purpose is to understand the behavior of the Laplace operator Δ, a negative, self-adjoint operator, on $C(N)$. If $\partial N \neq \emptyset$, we impose Dirichlet boundary conditions on $\partial C(N)$, though many other boundary conditions could be equally easily treated. The analysis here follows [CT].

The initial step is to use the method of separation of variables, writing Δ on $C(N)$ in the form

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{m}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta_N, \tag{8.2}$$

where Δ_N is the Laplace operator on the base N. Let μ_j, $\varphi_j(x)$ denote the eigenvalues and eigenfunctions of $-\Delta_N$ (with Dirichlet boundary condition on ∂N if $\partial N \neq \emptyset$), and set

$$\nu_j = (\mu_j + \alpha^2)^{1/2}, \quad \alpha = -\frac{m-1}{2}. \tag{8.3}$$

If

$$g(r, x) = \sum_j g_j(r)\varphi_j(x),$$

with $g_j(r)$ well behaved, and if we define the second-order operator L_μ by

$$L_\mu g(r) = \left(\frac{\partial^2}{\partial r^2} + \frac{m}{r}\frac{\partial}{\partial r} - \frac{\mu}{r^2}\right) g(r), \tag{8.4}$$

then we have

$$\Delta g(r, x) = \sum_j L_{\mu_j} g_j(r)\varphi_j(x). \tag{8.5}$$

In particular,

$$\Delta(g_j\varphi_j) = -\lambda^2 g_j\varphi_j \tag{8.6}$$

provided

$$g_j(r) = r^{-(m-1)/2} J_{\nu_j}(\lambda r). \tag{8.7}$$

Here $J_\nu(z)$ is the Bessel function, introduced in §6 of Chap. 3; there in (6.6) it is defined to be

$$J_\nu(z) = \frac{(z/2)^\nu}{\Gamma(\frac{1}{2})\Gamma(\nu+\frac{1}{2})} \int_{-1}^{1} (1-t^2)^{\nu-1/2} e^{izt}\, dt, \tag{8.8}$$

for Re $\nu > -1/2$; in (6.11) we establish Bessel's equation

$$\left[\frac{d^2}{dz^2} + \frac{1}{z}\frac{d}{dz} + \left(1 - \frac{\nu^2}{z^2}\right)\right] J_\nu(z) = 0, \tag{8.9}$$

which justifies (8.6); and in (6.19) we produced the formula

$$J_\nu(z) = \left(\frac{z}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\Gamma(k+\nu+1)} \left(\frac{z}{2}\right)^{2k}. \tag{8.10}$$

We also recall, from (6.56) of Chap. 3, the asymptotic behavior

$$J_\nu(r) \sim \left(\frac{2}{\pi r}\right)^{1/2} \cos\left(r - \frac{\pi\nu}{2} - \frac{\pi}{4}\right) + O(r^{-3/2}), \quad r \to +\infty. \tag{8.11}$$

This suggests making use of the Hankel transform, defined for $\nu \in \mathbb{R}^+$ by

$$H_\nu(g)(\lambda) = \int_0^\infty g(r) J_\nu(\lambda r) r\, dr. \tag{8.12}$$

Clearly, $H_\nu : C_0^\infty\big((0,\infty)\big) \to L^\infty(\mathbb{R}^+)$. We will establish the following:

Proposition 8.1. *For $\nu \geq 0$, H_ν extends uniquely from $C_0^\infty\big((0,\infty)\big)$ to*

$$H_\nu : L^2(\mathbb{R}^+, r\, dr) \longrightarrow L^2(\mathbb{R}^+, \lambda\, d\lambda), \quad \textit{unitary}. \tag{8.13}$$

Furthermore, for each $g \in L^2(\mathbb{R}^+, r\, dr)$,

$$H_\nu \circ H_\nu g = g. \tag{8.14}$$

To prove this, it is convenient to consider first

$$\widetilde{H}_\nu f(\lambda) = \int_0^\infty f(r) \frac{J_\nu(\lambda r)}{(\lambda r)^\nu} r^{2\nu+1}\, dr, \tag{8.15}$$

since, by (8.10), $(\lambda r)^{-\nu} J_\nu(\lambda r)$ is a smooth function of λr. Set

$$\mathcal{S}(\mathbb{R}^+) = \{ f|_{\mathbb{R}^+} : f \in \mathcal{S}(\mathbb{R}) \text{ is even} \}. \tag{8.16}$$

Lemma 8.2. *If* $\nu \ge -1/2$, *then*

$$\widetilde{H}_\nu : \mathcal{S}(\mathbb{R}^+) \longrightarrow \mathcal{S}(\mathbb{R}^+). \tag{8.17}$$

Proof. By (8.10), $J_\nu(\lambda r)/(\lambda r)^\nu$ is a smooth function of λr. The formula (8.8) yields

$$\Bigl|\frac{J_\nu(\lambda r)}{(\lambda r)^\nu}\Bigr| \le C_\nu < \infty, \tag{8.18}$$

for $\lambda r \in [0,\infty),\ \nu > -1/2$, a result that, by the identity

$$J_{-1/2}(z) = \Bigl(\frac{2}{\pi z}\Bigr)^{1/2} \cos z, \tag{8.19}$$

established in (6.35) of Chap. 3, also holds for $\nu = -1/2$. This readily yields

$$\widetilde{H}_\nu : \mathcal{S}(\mathbb{R}^+) \longrightarrow L^\infty(\mathbb{R}^+), \tag{8.20}$$

whenever $\nu \ge -1/2$. Now consider the differential operator $\widetilde{L}_\nu$, given by

$$\begin{aligned}\widetilde{L}_\nu f(r) &= -r^{-2\nu-1}\frac{\partial}{\partial r}\Bigl(r^{2\nu+1}\frac{\partial f}{\partial r}\Bigr)\\ &= -\frac{\partial^2 f}{\partial r^2} - \frac{2\nu+1}{r}\frac{\partial f}{\partial r}.\end{aligned} \tag{8.21}$$

Using Bessel's equation (8.9), we have

$$\widetilde{L}_\nu\Bigl(\frac{J_\nu(\lambda r)}{(\lambda r)^\nu}\Bigr) = \lambda^2 \frac{J_\nu(\lambda r)}{(\lambda r)^\nu}, \tag{8.22}$$

and, for $f \in \mathcal{S}(\mathbb{R}^+)$,

$$\begin{aligned}\widetilde{H}_\nu(\widetilde{L}_\nu f)(\lambda) &= \lambda^2 \widetilde{H}_\nu f(\lambda),\\ \widetilde{H}_\nu(r^2 f)(\lambda) &= \widetilde{L}_\nu \widetilde{H}_\nu f(\lambda).\end{aligned} \tag{8.23}$$

Since $f \in L^\infty(\mathbb{R}^+)$ belongs to $\mathcal{S}(\mathbb{R}^+)$ if and only if arbitrary iterated applications of $\widetilde{L}_\nu$ and multiplication by r^2 to f yield elements of $L^\infty(\mathbb{R}^+)$, the result (8.17) follows. We also have that this map is continuous with respect to the natural Frechet space structure on $\mathcal{S}(\mathbb{R}^+)$.

Lemma 8.3. *Consider the elements* $E_b \in \mathcal{S}(\mathbb{R}^+)$, *given for* $b > 0$ *by*

$$E_b(r) = e^{-br^2}. \tag{8.24}$$

We have

$$\widetilde{H}_\nu E_{1/2}(\lambda) = E_{1/2}(\lambda), \tag{8.25}$$

and more generally

$$\widetilde{H}_\nu E_b(\lambda) = (2b)^{-\nu-1} E_{1/4b}(\lambda). \tag{8.26}$$

Proof. To establish (8.25), plug the power series (8.10) for $J_\nu(z)$ into (8.15) and integrate term by term, to get

$$\widetilde{H}_\nu E_{1/2}(\lambda) = \sum_{k=0}^\infty \frac{(-1)^k 2^{-\nu-2k}}{k!\Gamma(k+\nu+1)} \lambda^{2k} \int_0^\infty r^{2k+2\nu+1} e^{-r^2/2}\, dr. \tag{8.27}$$

This last integral is seen to equal $2^{k+\nu}\Gamma(k+\nu+1)$, so we have

$$\widetilde{H}_\nu E_{1/2}(\lambda) = \sum_{k=0}^\infty \frac{1}{k!}\Big(-\frac{\lambda^2}{2}\Big)^k = e^{-\lambda^2/2} = E_{1/2}(\lambda). \tag{8.28}$$

Having (8.25), we get (8.26) by an easy change of variable argument.

In more detail, set $r^2/2 = bs^2$, or $s = r/\sqrt{2b}$. Then set $\mu = \sqrt{2b}\lambda$, so $\lambda r = \mu s$. Then (8.28), which we can write as

$$\int_0^\infty e^{-r^2/2} J_\nu(\lambda r) r^{\nu+1}\, dr = \lambda^\nu e^{-\lambda^2/2}, \tag{8.29}$$

translates to

$$\int_0^\infty e^{-bs^2} J_\nu(\mu s)(2b)^{(\nu+1)/2} s^{\nu+1} (2b)^{1/2}\, ds = (2b)^{-\nu/2} \mu^\nu e^{-\mu^2/4b}, \tag{8.30}$$

or, changing notation back,

$$\int_0^\infty e^{-bs^2} J_\nu(\lambda s) s^{\nu+1}\, ds = (2b)^{-\nu-1} \lambda^\nu e^{-\lambda^2/4b}, \tag{8.31}$$

which gives (8.26).

From (8.26) we have, for each $b > 0$,

$$\widetilde{H}_\nu \widetilde{H}_\nu E_b = (2b)^{-\nu-1} \widetilde{H}_\nu E_{1/4b} = E_b, \tag{8.32}$$

which verifies our stated Hankel inversion formula for $f = E_b,\ b > 0$. To get the inversion formula for general $f \in \mathcal{S}(\mathbb{R}^+)$, it suffices to establish the following.

Lemma 8.4. *The space*

$$\mathcal{V} = Span\{E_b : b > 0\} \tag{8.33}$$

is dense in $\mathcal{S}(\mathbb{R}^+)$.

Proof. Let $\overline{\mathcal{V}}$ denote the closure of $\mathcal{V}$ in $\mathcal{S}(\mathbb{R}^+)$. From

$$\frac{1}{\varepsilon}\big(e^{-br^2} - e^{-(b+\varepsilon)r^2}\big) \to r^2 e^{-br^2}, \tag{8.34}$$

we deduce that $r^2 e^{-br^2} \in \overline{\mathcal{V}}$, and inductively, we get

$$r^{2j} e^{-br^2} \in \overline{\mathcal{V}}, \quad \forall\, j \in \mathbb{Z}^+. \tag{8.35}$$

From here, one has

$$(\cos \xi r)e^{-r^2} \in \overline{\mathcal{V}}, \quad \forall\, \xi \in \mathbb{R}. \tag{8.36}$$

Now each even $\omega \in \mathcal{S}'(\mathbb{R})$ annihilating (8.36) for all $\xi \in \mathbb{R}$ has the property that $e^{-r^2}\omega$ has Fourier transform zero, which implies $\omega = 0$. The assertion (8.33) then follows by the Hahn-Banach theorem.

Putting the results of Lemmas 8.2–8.4 together, we have

Proposition 8.5. *Given* $\nu \ge -1/2$*, we have*

$$\widetilde{H}_\nu \widetilde{H}_\nu f = f, \tag{8.37}$$

for all $f \in \mathcal{S}(\mathbb{R}^+)$.

We promote this to

Proposition 8.6. *If* $\nu \ge -1/2$*, we have a unique extension of* $\widetilde{H}_\nu$ *from* $\mathcal{S}(\mathbb{R}^+)$ *to*

$$\widetilde{H}_\nu : L^2(\mathbb{R}^+, r^{2\nu+1}\, dr) \longrightarrow L^2(\mathbb{R}^+, \lambda^{2\nu+1}\, d\lambda), \tag{8.38}$$

as a unitary operator, and (8.37) *holds for all* $f \in L^2(\mathbb{R}^+, r^{2\nu+1}\, dr)$.

Proof. Take $f, g \in \mathcal{S}(\mathbb{R}^+)$, and use the inner product

$$(f, g) = \int_0^\infty f(r)\overline{g(r)}r^{2\nu+1}\, dr. \tag{8.39}$$

Using Fubini's theorem and the fact that $J_\nu(\lambda r)/(\lambda r)^\nu$ is real valued and symmetric in (λ, r), we get the first identity in

$$(\widetilde{H}_\nu f, \widetilde{H}_\nu g) = (\widetilde{H}_\nu \widetilde{H}_\nu f, g) = (f, g), \tag{8.40}$$

the second identity following by Proposition 8.5. From here, given that the linear space $\mathcal{S}(\mathbb{R}^+) \subset L^2(\mathbb{R}^+, r^{2\nu+1}\,dr)$ is dense, the assertions of Proposition 8.6 are apparent.

We return to the Hankel transform (8.12). Note that

$$H_\nu(r^\nu f)(\lambda) = \lambda^\nu \widetilde{H}_\nu f(\lambda), \tag{8.41}$$

and that $M_\nu f(r) = r^\nu f(r)$ has the property that

$$M_\nu : L^2(\mathbb{R}^+, r^{2\nu+1}\,dr) \longrightarrow L^2(\mathbb{R}^+, r\,dr) \ \text{ is unitary.} \tag{8.42}$$

Thus Proposition 8.6 yields Proposition 8.1.

Another proof is sketched in the exercises. An elaboration of Hankel's original proof is given on pp. 456–464 of [Wat].

In view of (8.23) and (8.41), we have

$$\begin{aligned} H_\nu(r^{-\alpha} L_\mu g) &= \int_0^\infty L_\mu(r^\alpha J_\nu(\lambda r)) g\, r^m\, dr \\ &= -\lambda^2 \int_0^\infty g r^\alpha J_\nu(\lambda r) r^m\, dr \\ &= -\lambda^2 H_\nu(r^{-\alpha} g). \end{aligned} \tag{8.43}$$

Now from (8.5)–(8.13), it follows that the map $\mathcal{H}$ given by

$$\mathcal{H}g = \Big(H_{\nu_0}(r^{-\alpha} g_0), H_{\nu_1}(r^{-\alpha} g_1), \dots \Big) \tag{8.44}$$

provides an isometry of $L^2(C(N))$ onto $L^2(\mathbb{R}^+, \lambda\, d\lambda, \ell^2)$, such that Δ is carried into multiplication by $-\lambda^2$. Thus (8.44) provides a spectral representation of Δ. Consequently, for well-behaved functions f, we have

$$\begin{aligned} &f(-\Delta) g(r, x) \\ &= r^\alpha \sum_j \int_0^\infty f(\lambda^2) J_{\nu_j}(\lambda r) \lambda \int_0^\infty s^{1-\alpha} J_{\nu_j}(\lambda s) g_j(s)\, ds\, d\lambda\, \varphi_j(x). \end{aligned} \tag{8.45}$$

Now we can interpret (8.45) in the following fashion. Define the *operator* ν on N by

$$\nu = \big(-\Delta_N + \alpha^2\big)^{1/2}. \tag{8.46}$$

Thus $\nu\varphi_j = \nu_j \varphi_j$. Identifying operators with their distributional kernels, we can describe the kernel of $f(-\Delta)$ as a function on $\mathbb{R}^+ \times \mathbb{R}^+$ taking values in operators on N, by the formula

(8.47)
$$f(-\Delta) = (r_1 r_2)^{\alpha} \int_0^\infty f(\lambda^2) J_\nu(\lambda r_1) J_\nu(\lambda r_2) \lambda \, d\lambda \\ = K(r_1, r_2, \nu),$$

since the volume element on $C(N)$ is $r^m \, dr \, dS(x)$ if the m-dimensional area element of N is $dS(x)$.

At this point it is convenient to have in hand some calculations of Hankel transforms, including some examples of the form (8.47). We establish some here; many more can be found in [Wat]. Generalizing (8.31), we can compute $\int_0^\infty e^{-br^2} J_\nu(\lambda r) r^{\mu+1} \, dr$ in a similar fashion, replacing the integral in (8.27) by

(8.48)
$$\int_0^\infty r^{2k+\mu+\nu+1} e^{-br^2} \, dr = \frac{1}{2} b^{-k-\mu/2-\nu/2-1} \Gamma\Big(\frac{\mu}{2} + \frac{\nu}{2} + k + 1\Big).$$

We get

(8.49)
$$\int_0^\infty e^{-br^2} J_\nu(\lambda r) r^{\mu+1} \, dr \\ = \lambda^\nu 2^{-\nu-1} b^{-\mu/2-\nu/2-1} \sum_{k=0}^\infty \frac{\Gamma(\frac{\mu}{2} + \frac{\nu}{2} + k + 1)}{k!\Gamma(k+\nu+1)} \Big(-\frac{\lambda^2}{4b}\Big)^k.$$

We can express the infinite series in terms of the confluent hypergeometric function, introduced in §7. A formula equivalent to (7.35) is

(8.50)
$${}_1F_1(a; b; z) = \frac{\Gamma(b)}{\Gamma(a)} \sum_{k=0}^\infty \frac{\Gamma(a+k)}{\Gamma(b+k)} \frac{z^k}{k!},$$

since $(a)_k = a(a+1)\cdots(a+k-1) = \Gamma(a+k)/\Gamma(a)$. We obtain, for $\operatorname{Re} b > 0$, $\operatorname{Re}(\mu + \nu) > -2$,

(8.51)
$$\int_0^\infty e^{-br^2} J_\nu(\lambda r) r^{\mu+1} \, dr \\ = \lambda^\nu 2^{-\nu-1} b^{-\mu/2-\nu/2-1} \frac{\Gamma(\frac{\mu}{2} + \frac{\nu}{2} + 1)}{\Gamma(\nu+1)} {}_1F_1\Big(\frac{\mu}{2} + \frac{\nu}{2} + 1; \nu + 1; -\frac{\lambda^2}{4b}\Big).$$

We can apply a similar attack when e^{-br^2} is replaced by e^{-br}, obtaining

(8.52)
$$\int_0^\infty e^{-br} J_\nu(\lambda r) r^{\mu-1} \, dr \\ = \Big(\frac{\lambda}{2}\Big)^\nu b^{-\mu-\nu} \sum_{k=0}^\infty \frac{\Gamma(\mu+\nu+2k)}{k!\Gamma(\nu+k+1)} \Big(-\frac{\lambda^2}{2b^2}\Big)^k,$$

at least provided $\operatorname{Re} b > |\lambda|, \ \nu \ge 0$, and $\mu + \nu > 0$; here we use

$$\int_0^\infty e^{-br} r^{2k+\mu+\nu-1}\, dr = b^{-2k-\mu-\nu}\Gamma(\mu+\nu+2k). \tag{8.53}$$

The duplication formula for the gamma function (see (A.22) of Chap. 3) implies

$$\Gamma(2k+\mu+\nu) = \pi^{-1/2} 2^{2k+\mu+\nu-1}\Gamma\Big(\frac{\mu}{2}+\frac{\nu}{2}+k\Big)\Gamma\Big(\frac{\mu}{2}+\frac{\nu}{2}+k+\frac{1}{2}\Big), \tag{8.54}$$

so the right side of (8.52) can be rewritten as

$$\pi^{-1/2}\lambda^\nu 2^{\mu-1} b^{-\mu-\nu} \sum_{k=0}^\infty \frac{\Gamma(\frac{\mu}{2}+\frac{\nu}{2}+k)\Gamma(\frac{\mu}{2}+\frac{\nu}{2}+\frac{1}{2}+k)}{k!\Gamma(\nu+1+k)}\Big(-\frac{\lambda^2}{b^2}\Big)^k. \tag{8.55}$$

This infinite series can be expressed in terms of the hypergeometric function, defined by

$$\begin{aligned} {}_2F_1(a_1,a_2;b;z) &= \sum_{k=0}^\infty \frac{(a_1)_k (a_2)_k}{(b)_k}\frac{z^k}{k!} \\ &= \frac{\Gamma(b)}{\Gamma(a_1)\Gamma(a_2)}\sum_{k=0}^\infty \frac{\Gamma(a_1+k)\Gamma(a_2+k)}{\Gamma(b+k)}\frac{z^k}{k!}, \end{aligned} \tag{8.56}$$

for $a_1, a_2 \notin \{0,-1,-2,\dots\},\ |z|<1$. If we put the sum in (8.55) into this form, and use the duplication formula, to write

$$\Gamma(a_1)\Gamma(a_2) = \Gamma\Big(\frac{\mu}{2}+\frac{\nu}{2}\Big)\Gamma\Big(\frac{\mu}{2}+\frac{\nu}{2}+\frac{1}{2}\Big) = \pi^{1/2}2^{-\mu-\nu+1}\Gamma(\mu+\nu),$$

we obtain

$$\begin{aligned} &\int_0^\infty e^{-br} J_\nu(\lambda r) r^{\mu-1}\, dr \\ &= \Big(\frac{\lambda}{2}\Big)^\nu b^{-\mu-\nu}\frac{\Gamma(\mu+\nu)}{\Gamma(\nu+1)}\cdot {}_2F_1\Big(\frac{\mu}{2}+\frac{\nu}{2}, \frac{\mu}{2}+\frac{\nu}{2}+\frac{1}{2}; \nu+1; -\frac{\lambda^2}{b^2}\Big). \end{aligned} \tag{8.57}$$

This identity, established so far for $|\lambda| < \operatorname{Re} b$ (and $\nu \ge 0,\ \mu+\nu>0$), continues analytically to λ in a complex neighborhood of $(0,\infty)$.

To evaluate the integral (8.47) with $f(\lambda^2) = e^{-t\lambda^2}$, we can use the power series (8.10) for $J_\nu(\lambda r_1)$ and for $J_\nu(\lambda r_2)$ and integrate the resulting double series term by term using (8.48). We get

(8.58)

$$\begin{aligned} &\int_0^\infty e^{-t\lambda^2} J_\nu(r_1\lambda) J_\nu(r_2\lambda)\lambda\, d\lambda \\ &= \frac{1}{2t}\Big(\frac{r_1 r_2}{4t}\Big)^\nu \times \sum_{j,k\ge 0}\frac{\Gamma(\nu+j+k+1)}{\Gamma(\nu+j+1)\Gamma(\nu+k+1)}\frac{1}{j!k!}\Big(-\frac{r_1^2}{4t}\Big)^j\Big(-\frac{r_2^2}{4t}\Big)^k, \end{aligned}$$

for any $t, r_1, r_2 > 0,\ \nu \geq 0$. This can be written in terms of the modified Bessel function $I_\nu(z)$, given by

$$I_\nu(z) = \Big(\frac{z}{2}\Big)^\nu \sum_{k=0}^\infty \frac{1}{k!\Gamma(\nu+k+1)} \Big(\frac{z}{2}\Big)^{2k}. \tag{8.59}$$

One obtains the following, known as the Weber identity.

Proposition 8.7. *For* $t, r_1, r_2 > 0$,

$$\int_0^\infty e^{-t\lambda^2} J_\nu(r_1\lambda) J_\nu(r_2\lambda)\lambda\, d\lambda = \frac{1}{2t}\, e^{-(r_1^2+r_2^2)/4t}\, I_\nu\Big(\frac{r_1 r_2}{2t}\Big). \tag{8.60}$$

Proof. The left side of (8.60) is given by (8.58). Meanwhile, by (8.59), the right side of (8.60) is equal to $(1/2t)(r_1 r_2/4t)^\nu$ times

$$\sum_{\ell,m\geq 0} \frac{1}{\ell!m!}\Big(-\frac{r_1^2}{4t}\Big)^\ell \Big(-\frac{r_2^2}{4t}\Big)^m \sum_{n=0}^\infty \frac{1}{n!\Gamma(\nu+n+1)}\Big(\frac{r_1 r_2}{4t}\Big)^{2n}. \tag{8.61}$$

If we set $y_j = -r_j^2/4t$, we see that the asserted identity (8.60) is equivalent to the identity

$$\begin{aligned} &\sum_{j,k\geq 0} \frac{\Gamma(\nu+j+k+1)}{\Gamma(\nu+j+1)\Gamma(\nu+k+1)} \frac{1}{j!k!}\, y_1^j y_2^k \\ &\qquad = \sum_{\ell,m,n\geq 0} \frac{1}{\ell!m!} \frac{1}{n!\Gamma(\nu+n+1)}\, y_1^{\ell+n} y_2^{m+n}. \end{aligned} \tag{8.62}$$

We compare coefficients of $y_1^j y_2^k$ in (8.62). Since both sides of (8.62) are symmetric in (y_1, y_2), it suffices to treat the case

$$j \leq k, \tag{8.63}$$

which we assume henceforth. Then we take $\ell + n = j,\ m + n = k$ and sum over $n \in \{0, \dots, j\}$, to see that (8.62) is equivalent to the validity of

$$\sum_{n=0}^{j} \frac{1}{(j-n)!(k-n)!n!\Gamma(\nu+n+1)} = \frac{\Gamma(\nu+j+k+1)}{\Gamma(\nu+j+1)\Gamma(\nu+k+1)} \frac{1}{j!k!}, \tag{8.64}$$

whenever $0 \leq j \leq k$. Using the identity

$$\Gamma(\nu+j+1) = (\nu+j)\cdots(\nu+n+1)\Gamma(\nu+n+1)$$

and its analogues for the other Γ-factors in (8.64), we see that (8.64) is equivalent to the validity of

$$\sum_{n=0}^{j} \frac{j!k!}{(j-n)!(k-n)!n!}(\nu+j)\cdots(\nu+n+1) = (\nu+j+k)\cdots(\nu+k+1), \tag{8.65}$$

for $0 \le j \le k$. Note that the right side of (8.65) is a polynomial of degree j in ν, and the general term on the left side of (8.65) is a polynomial of degree $j-n$ in ν.

In order to establish (8.65), it is convenient to set

$$\mu = \nu + j \tag{8.66}$$

and consider the associated polynomial identity in μ. With

$$\begin{gathered} p_0(\mu) = 1, \quad p_1(\mu) = \mu, \quad p_2(\mu) = \mu(\mu-1), \dots \\ p_j(\mu) = \mu(\mu-1)\cdots(\mu-j+1), \end{gathered} \tag{8.67}$$

we see that $\{p_0, p_1, \dots, p_j\}$ is a basis of the space $\mathcal{P}_j$ of polynomials of degree j in μ, and our task is to write

$$p_j(\mu+k) = (\mu+k)(\mu+k-1)\cdots(\mu+k-j+1) \tag{8.68}$$

as a linear combination of $p_0, \dots, p_j$. To this end, define

$$T : \mathcal{P}_j \longrightarrow \mathcal{P}_j, \quad Tp(\mu) = p(\mu+1). \tag{8.69}$$

By explicit calculation,

$$\begin{aligned} p_1(\mu+1) &= p_1(\mu) + p_0(\mu), \\ p_2(\mu+1) &= (\mu+1)\mu = \mu(\mu-1) + 2\mu = p_2(\mu) + 2p_1(\mu), \end{aligned} \tag{8.70}$$

and an inductive argument gives

$$Tp_i = p_i + i p_{i-1}. \tag{8.71}$$

By convention we set $p_i = 0$ for $i < 0$. Our goal is to compute $T^k p_j$. Note that

$$T = I + N, \quad Np_i = i p_{i-1}, \tag{8.72}$$

and

$$T^k = \sum_{n=0}^{j} \binom{k}{n} N^n, \tag{8.73}$$

if $j \le k$. By (8.72),

$$N^n p_i = i(i-1)\cdots(i-n+1)p_{i-n}, \tag{8.74}$$

so we have

$$\begin{aligned} T^k p_j &= \sum_{n=0}^{j} \binom{k}{n} j(j-1)\cdots(j-n+1)p_{j-n} \\ &= \sum_{n=0}^{j} \frac{k!}{(k-n)!n!} \frac{j!}{(j-n)!} p_{j-n}. \end{aligned} \tag{8.75}$$

This verifies (8.65) and completes the proof of (8.60).

Similarly we can evaluate (8.47) with $f(\lambda^2) = e^{-t\lambda}/\lambda$, as an infinite series, using (8.53) to integrate each term of the double series. We get

(8.76)

$$\begin{aligned} &\int_0^\infty e^{-t\lambda} J_\nu(r_1\lambda) J_\nu(r_2\lambda)\, d\lambda \\ &= \frac{1}{t}\Big(\frac{r_1 r_2}{t^2}\Big)^\nu \sum_{j,k\ge 0} \frac{\Gamma(2\nu+2j+2k+1)}{\Gamma(\nu+j+1)\Gamma(\nu+k+1)} \frac{1}{j!k!} \Big(-\frac{r_1^2}{4t^2}\Big)^j \Big(-\frac{r_2^2}{4t^2}\Big)^k, \end{aligned}$$

provided $t > r_j > 0$. It is possible to express this integral in terms of the Legendre function $Q_{\nu-1/2}(z)$.

Proposition 8.8. *One has, for all* $y, r_1, r_2 > 0,\ \nu \ge 0$,

(8.77)

$$\int_0^\infty e^{-y\lambda} J_\nu(r_1\lambda) J_\nu(r_2\lambda)\, d\lambda = \frac{1}{\pi}(r_1 r_2)^{-1/2} Q_{\nu-1/2}\left(\frac{r_1^2 + r_2^2 + y^2}{2r_1 r_2}\right).$$

The Legendre functions $P_{\nu-1/2}(z)$ and $Q_{\nu-1/2}(z)$ are solutions to

$$\frac{d}{dz}\Big[(1-z^2)\frac{d}{dz}u(z)\Big] + \Big(\nu^2 - \frac{1}{4}\Big)u(z) = 0; \tag{8.78}$$

Compare with (4.52). Extending (4.41), we can set

$$P_{\nu-1/2}(\cos\theta) = \frac{2}{\pi}\int_0^\theta \big(2\cos s - 2\cos\theta\big)^{-1/2} \cos \nu s\, ds, \tag{8.79}$$

and $Q_{\nu-1/2}(z)$ can be defined by the integral formula

$$Q_{\nu-1/2}(\cosh\eta) = \int_\eta^\infty \big(2\cosh s - 2\cosh\eta\big)^{-1/2} e^{-s\nu}\, ds. \tag{8.80}$$

The identity (8.77) is known as the *Lipschitz-Hankel integral formula*.

Proof of Proposition 8.8. We derive (8.77) from the Weber identity (8.60). Recall

$$I_\nu(y) = e^{-\pi i\nu/2}\, J_\nu(iy), \quad y > 0. \tag{8.81}$$

To work with (8.60), we use the subordination identity

$$e^{-y\lambda} = \frac{\lambda}{\sqrt{\pi}} \int_0^\infty e^{-y^2/4t} e^{-t\lambda^2} t^{-1/2}\, dt; \tag{8.82}$$

cf. Chap. 3, (5.31) for a proof. Plugging this into the left side of (8.77), and using (8.60), we see that the left side of (8.77) is equal to

$$\frac{1}{2\sqrt{\pi}} \int_0^\infty e^{-(r_1^2+r_2^2+y^2)/4t}\, I_\nu\Big(\frac{r_1 r_2}{2t}\Big) t^{-3/2}\, dt. \tag{8.83}$$

The change of variable $s = r_1 r_2/2t$ gives

$$\sqrt{\frac{1}{2\pi}}\,(r_1 r_2)^{-1/2} \int_0^\infty e^{-s(r_1^2+r_2^2+y^2)/2r_1 r_2}\, I_\nu(s) s^{-1/2}\, ds. \tag{8.84}$$

Thus the asserted identity (8.77) follows from the identity

$$\int_0^\infty e^{-sz} I_\nu(s) s^{-1/2}\, ds = \sqrt{\frac{2}{\pi}}\, Q_{\nu-1/2}(z), \quad z > 0. \tag{8.85}$$

As for the validity of (8.85), we mention two identities. Recall from (8.57) that

(8.86)
$$\int_0^\infty e^{-sz} J_\nu(\lambda s) s^{\mu-1}\, ds = \Big(\frac{\lambda}{2}\Big)^\nu z^{-\mu-\nu} \frac{\Gamma(\mu+\nu)}{\Gamma(\nu+1)} \cdot {}_2F_1\Big(\frac{\mu}{2}+\frac{\nu}{2}+\frac{1}{2}, \frac{\mu}{2}+\frac{\nu}{2}; \nu+1; -\frac{\lambda^2}{z^2}\Big).$$

Next, there is the classical representation of the Legendre function $Q_{\nu-1/2}(z)$ as a hypergeometric function:

(8.87)
$$Q_{\nu-1/2}(z) = \frac{\Gamma\big(\frac{1}{2}\big)\Gamma\big(\nu+\frac{1}{2}\big)}{\Gamma(\nu+1)} (2z)^{-\nu-1/2}\, {}_2F_1\Big(\frac{\nu}{2}+\frac{3}{4}, \frac{\nu}{2}+\frac{1}{4}; \nu+1; \frac{1}{z^2}\Big);$$

cf. [Leb], (7.3.7) If we apply (8.86) with $\lambda = i,\ \mu = 1/2$ (keeping (8.81) in mind), then (8.85) follows.

Remark: Formulas (8.77) and (8.60) are proven in the opposite order in [W].

By analytic continuation, we can treat $f(\lambda^2) = e^{-\varepsilon\lambda}\lambda^{-1}\sin\lambda t$ for any $\varepsilon > 0$. We apply this to (8.47). Letting $\varepsilon \searrow 0$, we get for the fundamental solution to the wave equation:

$$(8.88) \qquad \begin{aligned}&(-\Delta)^{-1/2}\sin t(-\Delta)^{1/2}\\ &= -\lim_{\varepsilon\searrow 0}(r_1r_2)^{\alpha}\,\mathrm{Im}\int_0^{\infty} e^{-(\varepsilon+it)\lambda}\,J_\nu(\lambda r_1)J_\nu(\lambda r_2)\,d\lambda\\ &= -\frac{1}{\pi}(r_1r_2)^{\alpha-1/2}\lim_{\varepsilon\searrow 0}\,\mathrm{Im}\,Q_{\nu-1/2}\Big(\frac{r_1^2+r_2^2+(\varepsilon+it)^2}{2r_1r_2}\Big).\end{aligned}$$

Using the integral formula (8.80), where the path of integration is a suitable path from η to $+\infty$ in the complex plane, one obtains the following alternative integral representation of $(-\Delta)^{-1/2}\sin t(-\Delta)^{1/2}$. The Schwartz kernel is equal to

$$(8.89) \qquad 0, \quad \text{if } t < |r_1 - r_2|,$$

$$(8.90) \qquad \frac{1}{\pi}(r_1r_2)^{\alpha}\int_0^{\beta_1}\big[t^2 - (r_1^2 + r_2^2 - 2r_1r_2\cos s)\big]^{-1/2}\,\cos\nu s\,ds,$$

if $|r_1 - r_2| < t < r_1 + r_2$, and

$$(8.91) \qquad \frac{1}{\pi}(r_1r_2)^{\alpha}\cos\pi\nu\int_{\beta_2}^{\infty}\big[r_1^2 + r_2^2 + 2r_1r_2\cosh s - t^2\big]^{-1/2}e^{-s\nu}\,ds,$$

if $t > r_1 + r_2$, where

$$(8.92) \qquad \beta_1 = \cos^{-1}\Big(\frac{r_1^2+r_2^2-t^2}{2r_1r_2}\Big), \quad \beta_2 = \cosh^{-1}\Big(\frac{t^2-r_1^2-r_2^2}{2r_1r_2}\Big).$$

Recall that $\alpha = -(m-1)/2$, where $m = \dim N$.

We next show how formulas (8.89)–(8.91) lead to an analysis of the classical problem of diffraction of waves by a slit along the positive x-axis in the plane $\mathbb{R}^2$. In fact, if waves propagate in $\mathbb{R}^2$ with this ray removed, on which Dirichlet boundary conditions are placed, we can regard the space as the cone over an interval of length 2π, with Dirichlet boundary conditions at the endpoints. By the method of images, it suffices to analyze the case of the cone over a circle of circumference 4π (twice the circumference of the standard unit circle). Thus $C(N)$ is a double cover of $\mathbb{R}^2 \setminus 0$ in this case. We divide up the spacetime into regions I, II, and III, respectively, as described by (8.89), (8.90), and (8.91). Region I contains only points on $C(N)$ too far away from the source point to be influenced by time t; that the fundamental solution is 0 here is consistent with finite propagation speed.

Since the circle has dimension $m = 1$, we see that

$$(8.93) \qquad \nu = (-\Delta_N)^{1/2} = \Big(-\frac{d^2}{d\theta^2}\Big)^{1/2}$$

in this case if $\theta \in \mathbb{R}/(4\pi\mathbb{Z})$ is the parameter on the circle of circumference 4π. On the line, we have

$$\cos s\nu\ \delta_{\theta_1}(\theta_2) = \frac{1}{2}\big[\delta(\theta_1 - \theta_2 + s) + \delta(\theta_1 - \theta_2 - s)\big]. \tag{8.94}$$

To get $\cos s\nu$ on $\mathbb{R}/(4\pi\mathbb{Z})$, we simply make (8.94) periodic by the method of images. Consequently, from (8.90), the wave kernel $(-\Delta)^{-1/2}\sin t(-\Delta)^{1/2}$ is equal to

$$\begin{matrix} (2\pi)^{-1}\big[t^2 - r_1^2 - r_2^2 + 2r_1r_2\cos(\theta_1 - \theta_2)\big]^{-1/2} & \text{if } |\theta_1 - \theta_2| \le \pi, \\ 0 & \text{if } |\theta_1 - \theta_2| > \pi, \end{matrix} \tag{8.95}$$

in region II. Of course, for $|\theta_1 - \theta_2| < \pi$ this coincides with the free space fundamental solution, so (8.95) also follows by finite propagation speed.

We turn now to an analysis of region III. In order to make this analysis, it is convenient to make simultaneous use both of (8.91) and of another formula for the wave kernel in this region, obtained by choosing another path from η to ∞ in the integral representation (8.80). The formula (8.91) is obtained by taking a horizontal line segment; see Fig. 8.1.

If instead we take the path indicated in Fig. 8.2, we obtain the following formula for $(-\Delta)^{-1/2}\sin t(-\Delta)^{1/2}$ in region III:

$$\begin{aligned} \pi^{-1}(r_1r_2)^{-(m-1)/2}\bigg\{ &\int_0^\pi \big(t^2 - r_1^2 - r_2^2 + 2r_1r_2\cos s\big)^{-1/2}\cos s\nu\, ds \\ &- \sin\pi\nu \int_0^{\beta_2} \big(t^2 - r_1^2 - r_2^2 - 2r_1r_2\cosh s\big)^{-1/2} e^{-s\nu}\, ds\bigg\}. \end{aligned} \tag{8.96}$$

The operator ν on $\mathbb{R}/(4\pi\mathbb{Z})$ given by (8.93) has spectrum consisting of

$$\operatorname{Spec}\nu = \Big\{0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots\Big\}, \tag{8.97}$$

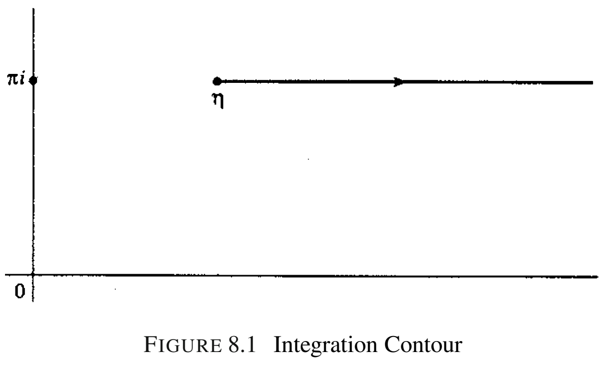

FIGURE 8.1 Integration Contour

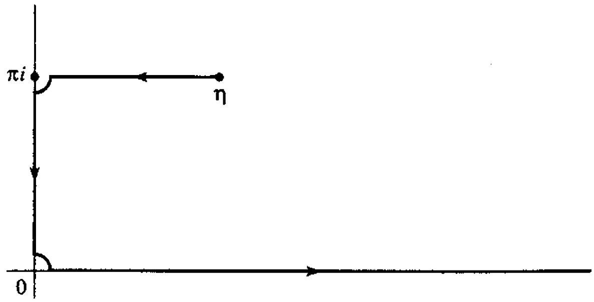

FIGURE 8.2 Alternative Contour

all the eigenvalues except for 0 occurring with multiplicity 2. The formula (8.91) shows the contribution coming from the half-integers in Spec ν vanishes, since $\cos\frac{1}{2}\pi n = 0$ if n is an odd integer. Thus we can use formula (8.96) and compose with the projection onto the sum of the eigenspaces of ν with integer spectrum. This projection is given by

$$P = \cos^2 \pi\nu \tag{8.98}$$

on $\mathbb{R}/(4\pi\mathbb{Z})$. Since $\sin\pi n = 0$, in the case $N = \mathbb{R}/(4\pi\mathbb{Z})$ we can rewrite (8.96) as

$$\pi^{-1}(r_1r_2)^{-(m-1)/2}\int_0^{\pi}\big(t^2 - r_1^2 - r_2^2 + 2r_1r_2\cos s\big)^{-1/2}\, P\cos s\nu\, ds. \tag{8.99}$$

In view of the formulas (8.94) and (8.96), we have

$$\begin{aligned} &P\ \cos s\nu\ \delta_{\theta_1}(\theta_2) \\ &\quad= \frac{1}{4}\big[\delta(\theta_1-\theta_2+s)+\delta(\theta_1-\theta_2-s) \\ &\qquad+\delta(\theta_1-\theta_2+2\pi+s)+\delta(\theta_1-\theta_2+2\pi-s)\big]\quad \text{mod } 4\pi. \end{aligned} \tag{8.100}$$

Thus, in region III, we have for the wave kernel $(-\Delta)^{-1/2}\sin t(-\Delta)^{1/2}$ the formula

$$(4\pi)^{-1}\big(t^2 - r_1^2 - r_2^2 + 2r_1r_2\cos(\theta_1-\theta_2)\big)^{-1/2}. \tag{8.101}$$

Thus, in region III, the value of the wave kernel at points $(r_1,\theta_1), (r_2,\theta_2)$ of the double cover of $\mathbb{R}^2 \setminus 0$ is given by half the value of the wave kernel on $\mathbb{R}^2$ at the image points. The jump in behavior from (8.95) to (8.101) gives rise to a diffracted wave.

We depict the singularities of the fundamental solution to the wave equation for $\mathbb{R}^2$ minus a slit in Figs. 8.3 and 8.4. In Fig. 8.3 we have the situation $|t| < r_1$,

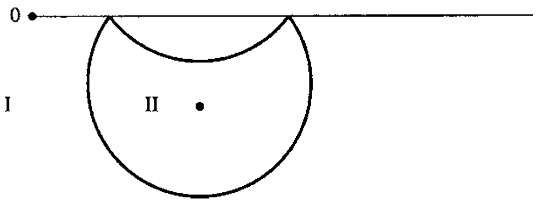

FIGURE 8.3 Reflected Wave Front

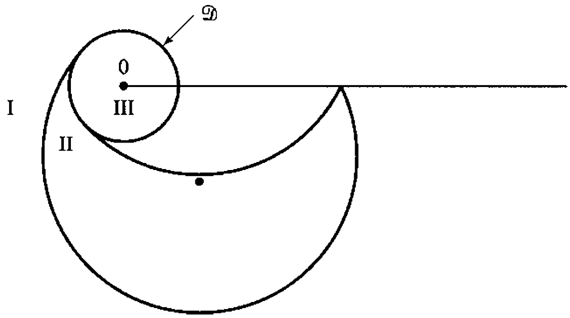

FIGURE 8.4 Reflected and Diffracted Wave Fronts

where no diffraction has occurred, and region III is empty. In Fig. 8.4 we have a typical situation for $|t| > r_1$, with the diffracted wave labeled by a "$\mathcal{D}$."

This diffraction problem was first treated by Sommerfeld [Som] and was the first diffraction problem to be rigorously analyzed. For other approaches to the diffraction problem presented above, see [BSU] and[Stk].

Generally, the solution (8.89)–(8.91) contains a diffracted wave on the boundary between regions II and III. In Fig. 8.5 we illustrate the diffraction of a wave by a wedge; here N is an interval of length $\ell < 2\pi$. We now want to provide, for general N, a description of the behavior of the distribution $v = (-\Delta)^{-1/2} \sin t(-\Delta)^{1/2}\ \delta_{(r_2,x_2)}$ near this diffracted wave, that is, a study of the limiting behavior as $r_1 \searrow t - r_2$ and as $r_1 \nearrow t - r_2$.

We begin with region II. From (8.90), we have v equal to

$$\frac{1}{2}(r_1 r_2)^{\alpha-1/2} P_{\nu-1/2}(\cos\beta_1)\ \delta_{x_2} \quad \text{in region II,} \tag{8.102}$$

where $P_{\nu-1/2}$ is the Legendre function defined by (8.79) and β_1 is given by (8.92). Note that as $r_1 \searrow t - r_2,\ \beta_1 \nearrow \pi$.

To analyze (8.102), replace s by $\pi - s$ in (8.79), and, with $\delta_1 = \pi - \beta_1$, write

$$\begin{aligned} \frac{\pi}{2} P_{\nu-1/2}(\cos\beta_1) = {} & \cos\pi\nu \int_{\delta_1}^{\pi} \bigl(2\cos\delta_1 - 2\cos s\bigr)^{-1/2} \cos s\nu\, ds \\ & + \sin\pi\nu \int_{\delta_1}^{\pi} \bigl(2\cos\delta_1 - 2\cos s\bigr)^{-1/2} \sin s\nu\, ds. \end{aligned} \tag{8.103}$$

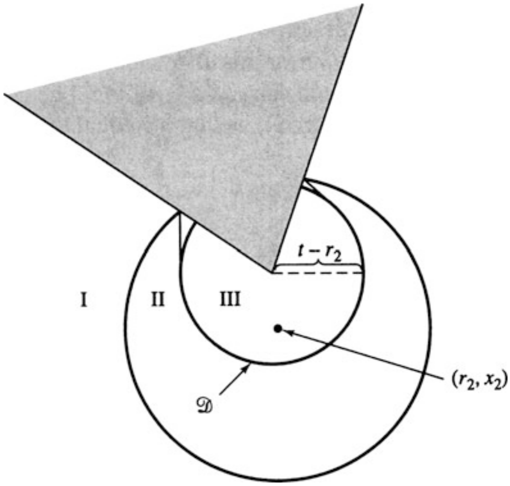

FIGURE 8.5 Diffraction by a Wedge

As $\delta_1 \searrow 0$, the second term on the right tends in the limit to

$$\sin \pi\nu \int_0^\pi \frac{\sin s\nu}{\sin \frac{1}{2}s}\, ds. \tag{8.104}$$

Write the first term on the right side of (8.103) as

$$\begin{aligned} &\cos \pi\nu \int_{\delta_1}^\pi (2\cos\delta_1 - 2\cos s)^{-1/2}(\cos s\nu - 1)\, ds \\ &+ \cos\pi\nu \int_{\delta_1}^\pi (2\cos\delta_1 - 2\cos s)^{-1/2}\, ds. \end{aligned} \tag{8.105}$$

As $\delta_1 \searrow 0$, the first term here tends in the limit to

$$\cos \pi\nu \int_0^\pi \frac{\cos s\nu - 1}{\sin \frac{1}{2}s}\, ds. \tag{8.106}$$

The second integral in (8.105) is a scalar, independent of ν, and it is easily seen to have a logarithmic singularity. More precisely,

$$\begin{aligned} &\int_{\delta_1}^\pi (2\cos\delta_1 - 2\cos s)^{-\frac{1}{2}}\, ds \\ &\sim \Big(\log\frac{2}{\delta_1}\Big) \sum_{j=0}^\infty A_j \delta_1^j + \sum_{j=1}^\infty B_j \delta_1^j, \quad A_0 = 1. \end{aligned} \tag{8.107}$$

Consequently, one derives the following.

Proposition 8.9. *Fix (r_2, x_2) and t. Then, as $r_1 \searrow t - r_2$,*

$$(8.108)\quad \begin{aligned}(-\Delta)^{-1/2} &\sin t(-\Delta)^{1/2}\, \delta_{(r_2,x_2)} \\ &= \frac{1}{\pi}(r_1 r_2)^{\alpha-1/2} \Big\{ \log \frac{2}{\delta_1}\ \cos \pi\nu\, \delta_{x_2} \\ &\quad + \int_0^\pi \frac{\cos s\nu - \cos \pi\nu}{2\cos \frac{1}{2}s}\, ds\, \delta_{x_2} + R_1\, \delta_{x_2} \Big\},\end{aligned}$$

where, for $s > (m+1)/2$,

$$(8.109)\quad \|R_1\, \delta_{x_2}\|_{\mathcal{D}^{-s-1}} \le C\delta_1\ \log \frac{1}{\delta_1}, \quad \text{as } \delta_1 \searrow 0.$$

The following result analyzes the second term on the right in (8.108).

Proposition 8.10. *We have*

$$(8.110)\quad \begin{aligned}\int_0^\pi &\Big(2\cos \frac{1}{2}s\Big)^{-1} (\cos s\nu - \cos \pi\nu)\, ds \\ &= \cos \pi\nu \Big\{ -\log \nu + \sum_{j=0}^K a_j \nu^{-2j} \Big\} + \frac{\pi}{2} \sin \pi\nu + S_K(\nu),\end{aligned}$$

where $S_K(\nu) : \mathcal{D}^s \to \mathcal{D}^{s+2K}$, for all s.

The spaces $\mathcal{D}^s$ are spaces of generalized functions on N, introduced in Chap. 5, Appendix A.

We turn to the analysis of v in region III. Using (8.91), we can write v as

$$(8.111)\quad \frac{1}{\pi}(r_1 r_2)^{\alpha-1/2} \cos \pi\nu\ Q_{\nu-1/2}(\cosh \beta_2)\, \delta_{x_2}, \qquad \text{in region III},$$

where $Q_{\nu-1/2}$ is the Legendre function given by (8.80) and β_2 is given by (8.92). It is more convenient to use (8.96) instead; this yields for v the formula

$$(8.112)\quad \begin{aligned}\frac{1}{\pi}(r_1 r_2)^{\alpha-1/2} \Big\{ &\int_0^\pi (2\cosh \beta_2 + 2\cos s)^{-1/2} \cos s\nu\, ds \\ &- \sin \pi\nu \int_0^{\beta_2} (2\cosh \beta_2 - 2\cosh s)^{-1/2} e^{-s\nu}\, ds \Big\}.\end{aligned}$$

Note that as $r_1 \nearrow t - r_2,\ \beta_2 \searrow 0$.

The first integral in (8.112) has an analysis similar to that arising in (8.103); first replace s by $\pi - s$ to rewrite the integral as

$$\cos \pi\nu \int_0^\pi (2\cosh\beta_2 - 2\cos s)^{-1/2} \cos s\nu \, ds \\ + \sin \pi\nu \int_0^\pi (2\cosh\beta_2 - 2\cos s)^{-1/2} \sin s\nu \, ds. \tag{8.113}$$

As $\beta_2 \searrow 0$, the second term in (8.113) tends to the limit (8.104), and the first term also has an analysis similar to (8.105)–(8.107), with (8.107) replaced by

$$\int_0^\pi (2\cosh\beta_2 - 2\cos s)^{-1/2} \, ds \\ \sim \Bigl(\log \frac{2}{\beta_2}\Bigr) \sum_{j\ge 0} A'_j \beta_2^j + \sum_{j\ge 1} B'_j \beta_2^j, \quad A'_0 = 1. \tag{8.114}$$

It is the second term in (8.112) that leads to the jump across $r_1 = t - r_2$, hence to the diffracted wave. We have

$$\int_0^{\beta_2} (2\cosh\beta_2 - 2\cosh s)^{-1/2} e^{-s\nu} \, ds \sim \int_0^{\beta_2} \frac{ds}{\sqrt{\beta_2^2 - s^2}} = \frac{\pi}{2}. \tag{8.115}$$

Thus we have the following:

Proposition 8.11. *For $r_1 \nearrow t - r_2$,*

$$(-\Delta)^{-1/2} \sin t(-\Delta)^{1/2} \, \delta_{(r_2,x_2)} \\ = \frac{1}{\pi}(r_1 r_2)^{\alpha-1/2} \Bigl\{ \log \frac{2}{\beta_2} \ \cos\pi\nu \, \delta_{x_2} \\ + \int_0^\pi \frac{\cos s\nu - \cos\pi\nu}{2\cos\frac{1}{2}s} \, ds \, \delta_{x_2} - \frac{\pi}{2} \sin\pi\nu \, \delta_{x_2} + \widetilde{R}_1 \delta_{x_2} \Bigr\}, \tag{8.116}$$

where, for $s > (m+1)/2$,

$$\| R_1 \delta_{x_2} \|_{\mathcal{D}^{-s-1}} \le C\beta_2 \log \frac{1}{\beta_2}, \quad \text{as } \beta_2 \searrow 0. \tag{8.117}$$

Note that (8.116) differs from (8.108) by the term $\pi^{-1}(r_1 r_2)^{\alpha-1/2}$ times

$$-\frac{\pi}{2} \sin\pi\nu \, \delta_{x_2}. \tag{8.118}$$

This contribution represents a jump in the fundamental solution across the diffracted wave $\mathcal{D}$. There is also the logarithmic singularity, $(r_1 r_2)^{\alpha-1/2}$ times

$$\frac{1}{\pi} \log \frac{2}{\delta} \ \cos\pi\nu \, \delta_{x_2}, \tag{8.119}$$

where $\delta = \delta_1$ in (8.108) and $\delta = \beta_2$ in (8.116). In the special case where N is an interval $[0, L]$, so $\dim C(N) = 2$, $\cos \pi\nu\, \delta_{x_2}$ is a sum of two delta functions. Thus its manifestation in such a case is subtle.

We also remark that if N is a subdomain of the unit sphere S^{2k} (of *even* dimension), then $\cos \pi\nu\, \delta_{x_2}$ vanishes on the set $N \setminus N_0$, where

$$N_0 = \{x_1 \in N : \text{for some } y \in \partial N,\ \text{dist}(x_2, y) + \text{dist}(y, x_1) \le \pi\}. \tag{8.120}$$

Thus the log blow-up disappears on $N \setminus N_0$. This follows from the fact that $\cos \pi\nu_0 = 0$, where ν_0 is the operator (8.46) on S^{2k}, together with a finite propagation speed argument.

While Propositions 8.9–8.11 contain substantial information about the nature of the diffracted wave, this information can be sharpened in a number of respects. A much more detailed analysis is given in [CT].

Exercises

1. Using (7.36) and (7.80), work out the asymptotic behavior of ${}_1F_1(a; b; -z)$ as $z \to +\infty$, given $b, b-a \notin \{0, -1, -2, \dots\}$. Deduce from (8.51) that whenever $\nu \ge 0,\ s \in \mathbb{R}$,

$$\lim_{b \searrow 0} \int_0^\infty e^{-br^2} J_\nu(r) r^{-is}\, dr = 2^{-is} \frac{\Gamma\big(\frac{1}{2}(\nu + 1 - is)\big)}{\Gamma\big(\frac{1}{2}(\nu + 1 + is)\big)}. \tag{8.121}$$

2. Define operators

$$M_r f(r) = r f(r), \quad \mathcal{J} f(r) = f(r^{-1}). \tag{8.122}$$

Show that

$$\begin{aligned} M_r : L^2(\mathbb{R}^+, r\, dr) &\longrightarrow L^2(\mathbb{R}^+, r^{-1}\, dr), \quad \mathcal{J} : L^2(\mathbb{R}^+, r^{-1}\, dr) \\ &\longrightarrow L^2(\mathbb{R}^+, r^{-1}\, dr) \end{aligned} \tag{8.123}$$

are unitary. Show that

$$H_\nu^\# = \mathcal{J} M_r H_\nu M_r^{-1} \tag{8.124}$$

is given by

$$H_\nu^\# f(\lambda) = (f \star \ell_\nu)(\lambda), \tag{8.125}$$

where $\star$ denotes the natural convolution on $\mathbb{R}^+$, with Haar measure $r^{-1} dr$:

$$(f \star g)(\lambda) = \int_0^\infty f(r) g(r^{-1}\lambda) r^{-1}\, dr, \tag{8.126}$$

and

$$\ell_\nu(r) = r^{-1} J_\nu(r^{-1}). \tag{8.127}$$

3. Consider the Mellin transform:

$$\mathcal{M}^{\#} f(s) = \int_0^\infty f(r) r^{is-1}\, dr. \tag{8.128}$$

As shown in (A.17)–(A.20) of Chap. 3, we have

$$(2\pi)^{-1/2}\mathcal{M}^{\#} : L^2(\mathbb{R}^+, r^{-1}\, dr) \longrightarrow L^2(\mathbb{R}, ds), \quad \text{unitary.} \tag{8.129}$$

Show that

$$\mathcal{M}^{\#}(f \star g)(s) = \mathcal{M}^{\#} f(s) \cdot \mathcal{M}^{\#} g(s), \tag{8.130}$$

and deduce that

$$\mathcal{M}^{\#} H_\nu^{\#} f(s) = \Psi(s) \mathcal{M}^{\#} f(s), \tag{8.131}$$

where

(8.132)
$$\Psi(s) = \int_0^\infty J_\nu(r^{-1}) r^{is-2}\, dr = \int_0^\infty J_\nu(r) r^{-is}\, dr = 2^{-is} \frac{\Gamma\big(\frac{1}{2}(\nu+1-is)\big)}{\Gamma\big(\frac{1}{2}(\nu+1+is)\big)}.$$

4. From (8.126)–(8.132), give another proof of the unitarity (8.13) of H_ν. Using symmetry, deduce that spec $H_\nu = \{-1, 1\}$, and hence deduce again the inversion formula (8.14).
5. Verify the asymptotic expansion (8.107). (*Hint*: Write $2\cos\delta - 2\cos s = (s^2 - \delta^2)$ $F(s,\delta)$ with F smooth and positive, $F(0,0) = 1$. Then, with $G(s,\delta) = F(s,\delta)^{-1/2}$,

$$\int_\delta^\pi (2\cos\delta - 2\cos s)^{-1/2}\, ds = \int_\delta^\pi G(s,\delta)\, \frac{ds}{\sqrt{s^2-\delta^2}}. \tag{8.133}$$

Write $G(s,\delta) = g(s) + \delta H(s,\delta),\ g(0) = 1$, and verify that (8.133) is equal to $A_1 + A_2$, where

$$A_1 = \int_\delta^\pi G(s,\delta)\, \frac{ds}{s} = g(0) \log\frac{1}{\delta} + O\Big(\delta \log\frac{1}{\delta}\Big),$$

$$A_2 = \int_\delta^\pi g(s)\Big[\frac{1}{\sqrt{s^2-\delta^2}} - \frac{1}{s}\Big]\, ds + O(\delta) = B_2 + O(\delta).$$

Show that

$$B_2 = g(0) \int_1^{\pi/\delta} \Big[\frac{1}{\sqrt{t^2-1}} - \frac{1}{t}\Big]\, dt + O(\delta) = C_2 + O(\delta),$$

with

$$C_2 = \int_1^\infty \Big[\frac{1}{\sqrt{t^2-1}} - \frac{1}{t}\Big]\, dt$$

Use the substitution $t = \cosh u$ to do this integral and get $C_2 = \log 2$.)
Next, verify the expansion (8.114).

Exercises on the hypergeometric function

1. Show that ${}_2F_1(a_1, a_2; b; z)$, defined by (8.56), satisfies

$$ {}_2F_1(a_1, a_2; b; z) = \frac{\Gamma(b)}{\Gamma(a_2)\Gamma(b-a_2)} \int_0^1 t^{a_2-1}(1-t)^{b-a_2-1}(1-tz)^{-a_1}\, dt, \tag{8.134} $$

for $\operatorname{Re} b > \operatorname{Re} a_2 > 0$, $|z| < 1$. (*Hint*: Use the beta function identity, (A.23)–(A.24) of Chap. 3, to write

$$ \frac{(a_2)_k}{(b)_k} = \frac{\Gamma(b)}{\Gamma(a_2)\Gamma(b-a_2)} \int_0^1 t^{a_2-1+k}(1-t)^{b-a_2-1}\, dt, \quad k = 0, 1, 2, \dots, $$

and substitute this into (8.39). Then use

$$ \sum_{k=0}^{\infty} \frac{(a_1)_k}{k!}(zt)^k = (1-tz)^{-a_1}, \quad 0 \le t \le 1, \quad |z| < 1.) $$

2. Show that, given $\operatorname{Re} b > \operatorname{Re} a_2 > 0$, (8.134) analytically continues in z to $z \in \mathbb{C} \setminus [1, \infty)$.
3. Show that the function (8.134) satisfies the ODE

$$ z(1-z)\frac{d^2u}{dz^2} + \{b - (a_1 + a_2 + 1)z\}\frac{du}{dz} - a_1 a_2 u = 0 $$

Note that $u(0) = 1$, $u'(0) = a_1 a_2 / b$, but zero is a singular point for this ODE. Show that another solution is

$$ u(z) = z^{1-b}\, {}_2F_1(a_1 - b + 1, a_2 - b + 1; 2 - b; z). $$

4. Show that

$$ {}_2F_1(a_1, a_2; b; z) = (1-z)^{-a_1}\, {}_2F_1\big(a_1, b - a_2; b; (z-1)^{-1}z\big). $$

(*Hint*: Make a change of variable $s = 1 - t$ in (8.134).)
For many other important transformation formulas, see [Leb] or [WW].
5. Show that

$$ {}_1F_1(a; b; z) = \lim_{c \nearrow \infty} {}_2F_1(a, c; b; c^{-1}z). $$

We mention the generalized hypergeometric function, defined by

$$ {}_pF_q(a; b; z) = \sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k} \frac{z^k}{k!}, $$

where $p \le q + 1$, $a = (a_1, \dots, a_p)$, $b = (b_1, \dots, b_q)$, $b_j \in \mathbb{C} \setminus \{0, -1, -2, \dots\}$, $|z| < 1$, and

$$ (a)_k = (a_1)_k \cdots (a_p)_k, \quad (b)_k = (b_1)_k \cdots (b_q)_k, $$

and where, as before, for $c \in \mathbb{C}$, $(c)_k = c(c+1)\cdots(c+k-1)$. For more on this class of functions, see [Bai].

6. The Legendre function $Q_{\nu-1/2}(z)$ satisfies the identity (8.87), for $\nu \geq 0$, $|z| > 1$, and $|\mathrm{Arg}\, z| < \pi$; cf. (7.3.7) of [Leb]. Take $z = (r_1^2 + r_2^2 + t^2)/2r_1 r_2$, and compare the resulting power series for the right side of (8.77) with the power series in (8.76).

References

[Ad] B. Adams, *Algebraic Approach to Simple Quantum Systems*, Springer, New York, 1994

[Ag] S. Agmon, *Lectures on Elliptic Boundary Problems*, Van Nostrand, New York, 1964

[Bai] W. Bailey, *Generalized Hypergeometric Series*, Hafner, New York, 1972

[Ber] P. Berard, Variétés Riemanniennes isospectrales non isométriques, *Sem. Bourbaki*, #705, 1988–89

[BLP] V. Berestetskii, E. Lifsitz, and L. Pitaevskii, *Relativistic Quantum Theory*, Pergammon, Oxford, 1971

[BGM] M. Berger, P. Gauduchon, and E. Mazet, *Le Spectre d'une Variété Riemannienne*, LNM #194, Springer, New York, 1971

[BSU] J. Bowman, T. Senior, and P. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes*, North-Holland, Amsterdam, 1969

[Br] R. Brooks, Constructing isospectral manifolds, *Am. Math. Monthly* 95(1988), 823–839

[BrC] J. Brossard and R. Carmona, Can one hear the dimension of a fractal? *Commun. Math. Phys*. 104(1986), 103–122

[Cha] I. Chavel, *Eigenvalues in Riemannian Geometry*, Academic, New York, 1984

[Che] J. Cheeger, Spectral geometry of spaces with cone-like singularities, *Proc. NAS, USA* 76(1979), 2103–2106

[CT] J. Cheeger and M. Taylor, Diffraction of waves by conical singularities, *Comm. Pure Appl. Math*. 35(1982), 275–331, 487–529

[Chn] P. Chernoff, Essential self-adjointness of powers of generators of hyperbolic equations, *J. Funct. Anal*. 12(1973), 401–414

[CCTV] L. Colzani, A. Crespi, G. Travaglini, and M. Vignati, Equiconvergence theorems for Fourier-Bessel expansions with applications to the harmonic analysis of radial functions in Euclidean and non-Euclidean spaces, *Trans. AMS* 338(1993), 43–55

[Cor] J. Cornwell, *Group Theory in Physics*, Academic, London, Vols. 1, 2, 1984; Vol. 3, 1989

[CyS] H. Cycon, R. Froese, W. Kirsch, and B. Simon, *Schrödinger Operators*, Springer, New York, 1987

[DST] E. B. Davies, B. Simon, and M. Taylor, L^p spectral theory of Kleinian groups, *J. Funct. Anal*. 78(1988), 116–136

[DS] N. Dunford and J. Schwartz, *Linear Operators, I, II*, Wiley-Interscience, New York, 1958, 1963

[Eng] M. Englefield, *Group Theory and the Coulomb Problem*, Wiley-Interscience, New York, 1972

[Fo] G. Folland, *Harmonic Analysis on Phase Space*, Princeton University Press, Princeton, NJ, 1989

[GWW] C. Gordon, D. Webb, and S. Wolpert, Isospectral plane domains and surfaces via Riemannian orbifolds, *Invent. Math*. 110(1992), 1–22

[Gr] P. Greiner, An asymptotic expansion for the heat equation, *Arch. Rat. Mech. Anal.* 41(1971), 163–218

[GS1] V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics*, Cambridge University Press, Cambridge, 1984

[GS2] V. Guillemin and S. Sternberg, *Variations on a Theme by Kepler*, AMS Colloq. Publ., Vol. 42, Providence, RI, 1990

[Hej] D. Hejhal, *The Selberg trace formula for* $PSL(2,\mathbb{R})$*, I*, LNM #548, *II*, LNM #1001, Springer, New York, 1976, 1983

[Helg] S. Helgason, *Analysis on Lie Groups and Symmetric Spaces*, CBMS Reg. Conf. Ser. in Math. #14, AMS, Providence, RI, 1972

[Ho] L. Hörmander, *The Analysis of Linear Partial Differential Operators*, Vols. 3 and 4, Springer, New York, 1985

[Ing] A. Ingham, *The Distribution of Prime Numbers*, Cambridge University Press, Cambridge, 1932

[IZ] C. Itzykson and J. Zuber, *Quantum Field Theory*, McGraw-Hill, New York, 1980

[Kac] M. Kac, Can one hear the shape of a drum?, *Am. Math. Monthly* 73(1966), 1–23

[Lap] M. Lapidus, Fractal drum, inverse spectral problems for elliptic operators, and a partial resolution of the Weyl-Berry conjecture, *Trans. AMS* 325(1991), 465–529

[LP] P. Lax and R. Phillips, *Scattering Theory for Automorphic Functions*, Princeton University Press, Princeton, NJ, 1976

[Leb] N. Lebedev, *Special Functions and Their Applications*, Dover, New York, 1972

[MS] H. McKean and I. Singer, Curvature and the eigenvalues of the Laplacian, *J. Diff. Geom.* 1(1967), 43–69

[Mes] A. Messiah, *Quantum Mechanics*, North Holland, Amsterdam, 1961

[Mil] W. Miller, *Symmetry and Separation of Variables*, Addison-Wesley, Reading, MA, 1977

[Miz] S. Mizohata, *The Theory of Partial Differential Equations*, Cambridge University Press, Cambridge, 1973

[Olv] F. Olver, *Asymptotics and Special Functions*, Academic, New York, 1974

[RS] M. Reed and B. Simon, *Methods of Mathematical Physics*, Academic, New York, Vols. 1, 2, 1975; Vols. 3, 4, 1978

[Ser] G. W. Series, *Spectrum of Atomic Hydrogen*, Oxford University Press, Oxford, 1957

[Shu] M. Shubin, *Pseudodifferential Operators and Spectral Theory*, Springer, New York, 1987

[Si1] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms*, Princeton University Press, Princeton, NJ, 1971

[Si2] B. Simon, *Trace Ideals and Their Applications*, Cambridge University Press, Cambridge, 1979

[Si3] B. Simon, *Functional Integration and Quantum Physics*, Academic, New York, 1979

[Sob] S. Sobolev, *Partial Differential Equations of Mathematical Physics*, Dover, New York, 1964

[Som] A. Sommerfeld, Mathematische theorie der diffraktion, *Math. Ann.* 47(1896), 317–374

[Stk] I. Stakgold, *Boundary Value Problems of Mathematical Physics*, Macmillan, New York, 1968

[Str] R. Strichartz, Analysis of the Laplacian on a complete Riemannian manifold, *J. Func. Anal.* 52(1983), 48–79

[Sun] T. Sunada, Riemannian coverings and isospectral manifolds, *Ann. Math.* 121(1985), 169–186

[T1] M. Taylor, *Pseudodifferential Operators*, Princeton University Press, Princeton, NJ, 1981

[T2] M. Taylor, *Noncommutative Harmonic Analysis*, AMS, Providence, RI, 1986

[Ter] A. Terras, *Harmonic Analysis on Symmetric Spaces and Applications*, Springer, New York, 1985

[Wat] G. N. Watson, *A Treatise on the Theory of Bessel Functions,* Cambridge University Press, Cambridge, 1945

[WW] E. Whittaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge University Press, Cambridge, 1927

9

Scattering by Obstacles

Introduction

In this chapter we study the phenomenon of scattering by a compact obstacle in Euclidean space $\mathbb{R}^3$. We restrict attention to the three-dimensional case, though a similar analysis can be given for obstacles in $\mathbb{R}^n$ whenever n is odd. The Huygens principle plays an important role in part of the analysis, and for that part the situation for n even is a little more complicated, though a theory exists there also.

The basic scattering problem is to solve the boundary problem

$$(\Delta + k^2)v = 0 \text{ in } \Omega, \quad v = f \text{ on } \partial K, \tag{0.1}$$

where $\Omega = \mathbb{R}^3 \setminus K$ is the complement of a compact set K. (We also assume Ω is connected.) We place on v the "radiation condition"

$$r\left(\frac{\partial v}{\partial r} - ikv\right) \longrightarrow 0, \text{ as } r \longrightarrow \infty, \tag{0.2}$$

in case k is real. We establish the existence and uniqueness of solutions to (0.1) and (0.2) in §1. Motivation for the condition (0.2) is also given there.

Special choices of the boundary value f give rise to the construction of the Green function $G(x, y, k)$ and of "eigenfunctions" $u_\pm(x, k\omega)$. In §2 we study analogues of the Fourier transform, arising from such eigenfunctions, providing $\Phi_\pm$, unitary operators from $L^2(\Omega)$ to $L^2(\mathbb{R}^3)$ which intertwine the Laplace operator on Ω, with the Dirichlet boundary condition, and multiplication by $|\xi|^2$ on $L^2(\mathbb{R}^3)$.

For any smooth f on ∂K, the solution to (0.1) and (0.2) has the following asymptotic behavior:

$$v(r\theta) = r^{-1}e^{ikr}\alpha(f, \theta, k) + o(r^{-1}), \quad r \longrightarrow \infty, \tag{0.3}$$

known as the "far field expansion." In case $f(x) = -e^{ik\omega\cdot x}$ on ∂K, the coefficient is denoted by $a(\omega, \theta, k)$ and called the "scattering amplitude." This is one of the

M.E. Taylor, *Partial Differential Equations II: Qualitative Studies of Linear Equations*, Applied Mathematical Sciences 116, DOI 10.1007/978-1-4419-7052-7_3,

fundamental objects of scattering theory; in §3 it is related to the unitary operator $S = \Phi_+ \Phi_-^{-1}$ on $L^2(\mathbb{R}^3)$, the "scattering operator."

The term "scattering" refers to the scattering of waves. Connection with the wave equation is made in §§4 and 5, where the scattering operator is related to the long-time behavior of the solution operator for the wave equation, in counterpoint to the long-distance characterization of the scattering amplitude given in §1. In the study of the wave-equation approach to scattering theory, a useful tool is a semigroup $Z(t)$, introduced by Lax and Phillips, which is described in §6.

Section 7 considers the meromorphic continuation in k of the solution operator to (0.1) and (0.2). This operator has poles in the lower half-plane, called *scattering poles*. The analytical method used to effect this construction involves the classical use of integral equations. We also relate the scattering poles to the spectrum of the Lax–Phillips semigroup $Z(t)$. In §8 we derive "trace formulas," further relating the poles and $Z(t)$. In §9 we illustrate material of earlier sections by explicit calculations for scattering by the unit sphere in $\mathbb{R}^3$.

In §§10 and 11, we discuss the "inverse" problem of determining an obstacle K, given scattering data. Section 10 focuses on uniqueness results, asserting that exact measurements of certain scattering data will uniquely determine K. In §11 we discuss some methods that have been used to determine K approximately, given approximate measurements of scattering data. This leads us to a discussion of "ill-posed" problems and how to regularize them.

In §12 we present some material on scattering by a rough obstacle, pointing out similarities and differences with the smooth cases considered in the earlier sections. Appendix A at the end of this chapter is devoted to the proof of a trace identity used in §8.

We have confined attention to the Dirichlet boundary condition. The scattering problem with the Neumann boundary condition, and for electromagnetic waves, with such boundary conditions as discussed in Chap. 5, are of equal interest. There are also studies of scattering for the equations of linear elasticity, with boundary conditions of the sort considered in Chap. 5. Many of the results in such cases can be obtained with only minor modifications of the techniques used here, while other results require further work. For further material on the theory of scattering by obstacles, consult [LP1], [Rm], [CK], and [Wil].

Another important setting for scattering theory is the Schrödinger operator $-\Delta + V$; see [RS], [New], and [Ho] for material on this. We include some exercises on some of the simplest problems in this quantum scattering theory. These exercises indicate that very similar techniques to those for scattering by a compact obstacle apply to scattering by a compactly supported potential. It would not take a much greater modification to handle potentials $V(x)$ that decay very rapidly as $|x| \to \infty$. Such potentials, with exponential fall-off, are used in crude models of two-body interactions involving nuclear forces. It takes more substantial modifications to treat long-range potentials, such as those that arise from the Coulomb force. The most interesting quantum scattering problems involve multiparticle interactions, and the analysis of these requires a much more elaborate set-up.

1. The scattering problem

In this section we establish the existence and uniqueness for the following boundary problem. Let $K \subset \mathbb{R}^3$ be a compact set with smooth boundary and connected complement Ω. Let $f \in H^s(\partial K)$ be given, and let $k > 0$. We want to solve

$$(\Delta + k^2)v = 0 \quad \text{on } \Omega, \tag{1.1}$$

$$v = f \quad \text{on } \partial K. \tag{1.2}$$

In addition, we impose a "radiation condition," of the following form:

$$|rv(x)| \le C, \quad r\left(\frac{\partial v}{\partial r} - ikv\right) \longrightarrow 0, \ \text{ as } r \longrightarrow \infty, \tag{1.3}$$

where $r = |x|$. This condition will hold provided v satisfies the integral identity

$$v(x) = \int_{\partial K} \left[f(y)\frac{\partial g}{\partial \nu_y}(x,y,k) - g(x,y,k)\frac{\partial v}{\partial \nu}(y) \right] dS(y), \tag{1.4}$$

for $x \in \Omega$, where

$$g(x,y,k) = \big(4\pi|x-y|\big)^{-1} e^{ik|x-y|}. \tag{1.5}$$

Our existence proof will utilize the following fact. If $k > 0$ is replaced by $k + i\varepsilon$, $\varepsilon > 0$, then $-(k+i\varepsilon)^2$ belongs to the resolvent set of the Laplace operator Δ on Ω, with Dirichlet boundary conditions on ∂K. Hence, for $s \ge 3/2$, (1.1)–(1.2) (with k replaced by $k + i\varepsilon$) has a unique solution $v_\varepsilon \in L^2(\Omega)$. To obtain this, extend f to $f^\# \in H^2(\Omega)$, and set $\varphi = \big(\Delta + (k+i\varepsilon)^2\big) f^\# \in L^2(\Omega)$. Then

$$v_\varepsilon = f^\# - \big(\Delta + (k+i\varepsilon)^2\big)^{-1}\varphi.$$

Furthermore, in this case, the integral formula (1.4) does hold, as a consequence of Green's theorem, with $g(x,y,k)$ replaced by

$$g(x,y,k+i\varepsilon) = \big(4\pi|x-y|\big)^{-1} e^{(ik-\varepsilon)|x-y|}, \tag{1.6}$$

which, as we saw in Chap. 3, is (the negative of) the resolvent kernel for $(\Delta + (k + i\varepsilon)^2)^{-1}$ on free space $\mathbb{R}^3$, a kernel that converges to (1.5) as $\varepsilon \searrow 0$. The strategy will be to show that, as $\varepsilon \searrow 0$, v_ε converges to the solution to (1.1)–(1.3).

Before tackling the existence proof, we first establish the uniqueness of solutions to (1.1)–(1.3), as this uniqueness result will play an important role in the existence proof.

Proposition 1.1. *Given $k > 0$, if v satisfies (1.1)–(1.3) with $f = 0$, then $v = 0$.*

Proof. Let S_R denote the sphere $\{|x| = R\}$ in $\mathbb{R}^3$; for R large, $S_R \subset \Omega$, and, with $v_r = \partial v/\partial r$, we have

$$\int_{S_R} |v_r - ikv|^2 \, dS = \int_{S_R} \big(|v_r|^2 + k^2|v|^2\big) \, dS - ik \int_{S_R} \big(v\overline{v}_r - \overline{v}v_r\big) \, dS. \tag{1.7}$$

Now Green's theorem applied to v and $\overline{v}$ implies

$$\int_{S_R} \big(v\overline{v}_r - \overline{v}v_r\big) \, dS = \int_{\partial K} \Big(v\frac{\partial \overline{v}}{\partial \nu} - \overline{v}\frac{\partial v}{\partial \nu}\Big) \, dS = 0, \tag{1.8}$$

provided $v|_{\partial K} = 0$. Since the hypothesis (1.3) implies

$$\int_{S_R} \big|v_r - ikv\big|^2 \, dS \longrightarrow 0, \quad \text{as } R \longrightarrow \infty, \tag{1.9}$$

we deduce from (1.7) that

$$\int_{S_R} |v|^2 \, dS \longrightarrow 0, \quad \text{as } R \longrightarrow \infty. \tag{1.10}$$

The proof of Proposition 1.1 is completed by the following result.

Lemma 1.2. *If v satisfies $(\Delta + k^2)v = 0$ for $|x| \geq R_0$ and (1.10) holds, then $v(x) = 0$ for $|x| \geq R_0$.*

Proof. It suffices to prove that, for $r \geq R_0$,

$$V(r) = \int_{S^2} v(r\omega)\varphi(\omega) \, dS(\omega) \tag{1.11}$$

is identically zero, for each eigenfunction φ of the Laplace operator Δ_S on the unit sphere S^2:

$$(\Delta_S + \mu^2)\varphi = 0 \quad (\mu \geq 0). \tag{1.12}$$

In view of the formula for Δ on $\mathbb{R}^3$ in polar coordinates,

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta_S, \tag{1.13}$$

it follows that $V(r)$ satisfies the ODE

$$V''(r) + \frac{2}{r}V'(r) + \Big(k^2 - \frac{\mu^2}{r^2}\Big)V(r) = 0, \quad r \geq R_0. \tag{1.14}$$

This ODE has two linearly independent solutions of the form $r^{-1/2}H_\nu^{(j)}(kr)$, $j=1,2$, where $H_\nu^{(1)}(z)$ and $H_\nu^{(2)}(z)$ are the Hankel functions discussed in Chap. 3, and $\nu^2 = \mu^2 + 1/4$. In view of the integral formulas given there, it follows that the asymptotic behavior of these two solutions is of the form

$$V_\pm(r) = C_\pm\, r^{-1}e^{\pm ikr} + o(r^{-1}), \quad r \longrightarrow \infty. \tag{1.15}$$

Clearly, no nontrivial linear combination of these two is $o(r^{-1})$ as $r \to \infty$. Since the hypothesis implies that $V(r) = o(r^{-1})$, we deduce that $V = 0$.

Applying Lemma 1.2, we see that under the hypotheses of Proposition 1.1, $v = 0$ for $|x| \geq R_0$, given that $K \subset \{x : |x| \leq R_0\}$. Since v satisfies the unique continuation property in Ω, this implies $v = 0$ in Ω, so Proposition 1.1 is proved.

Remark: The uniqueness proof above really used (1.9), which is formally weaker than the radiation condition (1.3). Consequently, (1.9) is sometimes called the radiation condition. On the other hand, the existence theorem, to which we turn next, shows that the formally stronger condition (1.3) holds.

The following result, which establishes the existence of solutions to (1.1)–(1.3), is known as the *limiting absorption principle*.

Theorem 1.3. *Let $s \geq 3/2$, and suppose that as $\varepsilon \searrow 0$,*

$$f_\varepsilon \longrightarrow f \quad \text{in } H^s(\partial K). \tag{1.16}$$

Let v_ε be the unique element of $L^2(\Omega)$ satisfying

$$\big(\Delta + (k+i\varepsilon)^2\big)v_\varepsilon = 0 \quad \text{in } \Omega, \tag{1.17}$$

$$v_\varepsilon = f_\varepsilon \quad \text{on } \partial K. \tag{1.18}$$

Then, as $\varepsilon \searrow 0$, we have a unique limit

$$v_\varepsilon \longrightarrow v = \mathcal{B}(k)f, \tag{1.19}$$

satisfying (1.1)–(1.3)*. Convergence occurs in the norm topology of the space $L^2(\Omega, \langle x\rangle^{-1-\delta}dx)$ for any $\delta > 0$, as well as in $H^{s+1/2}_{loc}(\Omega)$, and the limit v satisfies the identity* (1.4)*.*

It is convenient to divide the proof into two parts. Fix R such that $K \subset \{|x| < R\}$ and let $\mathcal{O}_R = \Omega \cap \{|x| < R\}$.

Lemma 1.4. *Assume $v_\varepsilon\big|_{\mathcal{O}_R}$ is bounded in $L^2(\mathcal{O}_R)$ as $\varepsilon \searrow 0$. Then the conclusions of Theorem 1.3 hold.*

Proof. Fix $S < R$ with $K \subset \{|x| < S\}$. The elliptic estimates of Chap. 5 imply that if $\|v_\varepsilon\|_{L^2(\mathcal{O}_R)}$ is bounded, then

$$\|v_\varepsilon\|_{H^{s+1/2}(\mathcal{O}_S)} \le C_k + C_k \|f_\varepsilon\|_{H^s(\partial K)}. \tag{1.20}$$

Passing to a subsequence, which we continue to denote by v_ε, we have

$$v_\varepsilon \longrightarrow v \quad \text{weakly in } H^{s+1/2}(\mathcal{O}_S), \tag{1.21}$$

for some $v \in H^{s+1/2}(\mathcal{O}_S)$. The trace theorem implies weak convergence

$$v_\varepsilon\big|_{\partial K} \longrightarrow v\big|_{\partial K} \quad \text{in } H^s(\partial K), \tag{1.22}$$

and

$$\frac{\partial v_\varepsilon}{\partial \nu} \longrightarrow \frac{\partial v}{\partial \nu} \quad \text{in } H^{s-1}(\partial K). \tag{1.23}$$

Since each v_ε satisfies

$$v_\varepsilon(x) = \int_{\partial K} \Big(f_\varepsilon(y) \frac{\partial g_\varepsilon}{\partial \nu} - g_\varepsilon \frac{\partial v_\varepsilon(y)}{\partial \nu} \Big)\, dS(y), \quad x \in \Omega, \tag{1.24}$$

with $g_\varepsilon = g(x, y, k + i\varepsilon)$ given by (1.6), we deduce from (1.22) and (1.23) that the right side of (1.24) converges locally uniformly in $x \in \overline{\Omega}$, as $\varepsilon \searrow 0$, to a limit, call it v, that coincides with the limit (1.21) on $\mathcal{O}_S$. Furthermore, in view of the formula (1.6), we have the estimate

$$|v_\varepsilon(x)| \le C_k \langle x \rangle^{-1}, \quad x \in \Omega, \tag{1.25}$$

with C_k independent of ε. Thus the limit v satisfies this estimate, and we have $v_\varepsilon \to v$ in $L^2(\Omega, \langle x \rangle^{-1-\delta}\, dx)$ for any $\delta > 0$. Furthermore, the limit v satisfies the identity (1.4), so the radiation condition (1.3) holds.

So far we have convergence for subsequences, but in view of the uniqueness result of Proposition 1.1, this limit v is unique, so Lemma 1.4 is proved.

The proof of Theorem 1.3 is completed by the following argument.

Lemma 1.5. *The hypotheses* (1.16)–(1.18) *of Theorem 1.3 imply that* $\{v_\varepsilon\}$ *is bounded in* $L^2(\Omega, \langle x \rangle^{-1-\delta}\, dx)$, *for any* $\delta > 0$.

Proof. Fix such a δ. Suppose $N_\varepsilon = \|v_\varepsilon\|_{L^2(\Omega, \langle x \rangle^{-1-\delta} dx)} \to \infty$ for a subsequence $\varepsilon_n \searrow 0$. Set $w_\varepsilon = N_\varepsilon^{-1} v_\varepsilon$. Then Lemma 1.4 applies to w_ε, with $w_\varepsilon\big|_{\partial K} = f_\varepsilon^\# = N_\varepsilon^{-1} f_\varepsilon \to 0$ in $H^s(\partial K)$. Thus the conclusion of Lemma 1.4 gives

$$w_\varepsilon \longrightarrow w \ \text{ strongly in } L^2(\Omega, \langle x \rangle^{-1-\delta}\, dx).$$

The limit w satisfies the scattering problem (1.1)–(1.3) with $f = 0$, so our uniqueness result implies $w = 0$. This contradicts the fact that each w_ε has norm 1 in $L^2(\Omega, \langle x \rangle^{-1-\delta}\, dx)$, so the proof is complete.

Remark: Considering the dense subspace $H^{s+1}(\partial K)$ of $H^s(\partial K)$, we can improve weak convergence of $v_\varepsilon \to v$ in $H^{s+1/2}_{\text{loc}}(\Omega)$ to strong convergence in this space.

We draw a couple of conclusions from Theorem 1.3. The first concerns the limiting behavior as $\varepsilon \searrow 0$ of the Green function $G(x, y, k + i\varepsilon)$, the kernel for the resolvent $\big(\Delta + (k + i\varepsilon)^2\big)^{-1}$ on Ω, which is of the form

$$G(x, y, k + i\varepsilon) = g(x, y, k + i\varepsilon) + h(x, y, k + i\varepsilon), \tag{1.26}$$

where $g(x, y, k + i\varepsilon)$ is the free-space Green kernel (1.6) and $h(x, y, k + i\varepsilon)$ is, for each $y \in \Omega$, the element of $L^2(\Omega)$ satisfying

$$\begin{aligned} &\big(\Delta_x + (k + i\varepsilon)^2\big)h = 0, \\ &\quad h(x, y, k + i\varepsilon) = -g(x, y, k + i\varepsilon), \quad \text{for } x \in \partial K. \end{aligned} \tag{1.27}$$

Clearly, as $\varepsilon \searrow 0$, $g(x, y, k + i\varepsilon) \to g(x, y, k)$, given by (1.5). On the other hand, for any $y \in \Omega$, Theorem 1.3 applies to $f_\varepsilon(x) = -g(x, y, k + i\varepsilon)$, and we have

$$h(x, y, k + i\varepsilon) \to h(x, y, k), \tag{1.28}$$

where $h(x, y, k)$ solves the scattering problem (1.1)–(1.3), with $h(x, y, k) = -g(x, y, k)$ for $x \in \partial K$. Consequently, as $\varepsilon \searrow 0$,

$$G(x, y, k + i\varepsilon) \longrightarrow G(x, y, k), \tag{1.29}$$

where

$$G(x, y, k) = g(x, y, k) + h(x, y, k). \tag{1.30}$$

Another important family of functions defined by a scattering problem is the following. Note that we have

$$(\Delta + |\xi|^2)e^{-ix\cdot\xi} = 0 \quad \text{on } \mathbb{R}^3, \tag{1.31}$$

for any $\xi \in \mathbb{R}^3$. We define the functions $u(x, \xi)$ on $\Omega \times \mathbb{R}^3$ by

$$u(x, \xi) = e^{-ix\cdot\xi} + v(x, \xi), \tag{1.32}$$

where $v(x, \xi)$ satisfies the scattering problem (1.1)–(1.3), with $k^2 = |\xi|^2$ and

$$v(x, \xi) = -e^{-ix\cdot\xi} \quad \text{on } \partial K. \tag{1.33}$$

As we will see in the next section, $u(x, \xi)$ plays a role on Ω of generalized eigenfunction of the Laplace operator on Ω, with Dirichlet boundary conditions, analogous to the role played by $u_0(x, \xi) = e^{-ix\cdot\xi}$ on $\mathbb{R}^3$.

There is an interesting relation between the Green function $G(x, y, k)$ and the "eigenfunctions" $u(x, \xi)$, which we give here, which will play an important role in the analysis in the next section. It involves the behavior of $G(x, y, k)$ as $|y| \to \infty$.

Proposition 1.6. *For $y = r\omega,\ \omega \in S^2,\ r \to \infty$, and any fixed $k > 0$,*

$$G(x, r\omega, k) = (4\pi r)^{-1} e^{ikr} u(x, k\omega) + O(r^{-2}). \tag{1.34}$$

This is uniformly valid for (x, ω, k) in any bounded subset of $\Omega \times S^2 \times \mathbb{R}^+$.

Proof. Write $G(x, r\omega, k) = g(x, r\omega, k) + h(x, r\omega, k)$, as in (1.30). Thus $h_r(x) = h(x, r\omega, k)$ satisfies

$$(\Delta + k^2) h_r(x) = 0, \quad h_r\big|_{\partial K} = -g(x, r\omega, k), \tag{1.35}$$

together with the radiation condition as $|x| \to \infty$. Now, in view of (1.5), as $r \to \infty$, we have, for $x \in \partial K$, or indeed for x in any bounded subset of $\mathbb{R}^3$,

$$g(x, r\omega, k) = (4\pi r)^{-1} e^{ikr} e^{-ik\omega\cdot x} + O(r^{-2}), \tag{1.36}$$

where the remainder is $O(r^{-2})$ in $C^\ell(\partial K)$ for any ℓ. Thus, in view of the estimates established in the proof of Theorem 1.3, we have

$$h_r = (4\pi r)^{-1} e^{ikr} v(x, k\omega) + O(r^{-2}), \quad r \longrightarrow \infty, \tag{1.37}$$

with $v(x, \xi)$ defined above. This gives the desired result (1.34).

We remark that a similar argument gives

$$\frac{\partial}{\partial r} G(x, r\omega, k) = (4\pi r)^{-1} ik\, e^{ikr} u(x, k\omega) + O(r^{-2}), \tag{1.38}$$

as $r \to \infty$.

Note that, for any $f \in C^\infty(\partial K)$, by (1.4) we have an asymptotic behavior of the form

$$v(r\theta) = r^{-1} e^{ikr} \alpha(f, \theta, k) + o(r^{-1}), \quad r \to \infty, \tag{1.39}$$

with $\theta \in S^2$, for the solution to the scattering problem (1.1)–(1.3), with a smooth coefficient $\alpha(f, \cdot, \cdot)$. Also,

$$\frac{\partial}{\partial r} v(r\theta) = \frac{ik}{r} e^{ikr} \alpha(f, \theta, k) + o(r^{-1}). \tag{1.40}$$

In particular, the function $v(x, \xi)$ given by (1.33) has the asymptotic behavior

$$v(r\theta, k\omega) \sim r^{-1} e^{ikr} a(-\omega, \theta, k), \quad r \to \infty, \tag{1.41}$$

for fixed $\theta, \omega \in S^2$, $k \in \mathbb{R}^+$, and its r-derivative has an analogous behavior. The coefficient $a(\omega, \theta, k)$ is called the *scattering amplitude* and is one of the fundamental objects of scattering theory. We will relate this to the scattering operator in §3.

The radiation condition (1.3) is more specifically called the "outgoing radiation condition." It has a counterpart, the "incoming radiation condition":

$$|rv(x)| \le C, \quad r\Big(\frac{\partial v}{\partial r} + ikv\Big) \longrightarrow 0, \ \text{ as } r \longrightarrow \infty. \tag{1.42}$$

Clearly there is a parallel treatment of the scattering problem (1.1), (1.2), (1.42). Indeed, if $v(x)$ satisfies (1.1)–(1.3), then $\overline{v(x)}$ satisfies the incoming scattering problem, with f replaced by $\overline{f}$, and conversely. In particular, we can define

$$u_-(x, \xi) = e^{-ix\cdot\xi} + v_-(x, \xi), \tag{1.43}$$

where $v_-(x, \xi)$ satisfies the scattering problem (1.1), (1.2), (1.42), with

$$v_-(x, \xi) = -e^{-ix\cdot\xi} \quad \text{on } \partial K, \tag{1.44}$$

and we clearly have

$$v_-(x, \xi) = \overline{v(x, -\xi)}, \quad u_-(x, \xi) = \overline{u(x, -\xi)}. \tag{1.45}$$

In analogy with (1.41), we have the asymptotic behavior

$$v_-(r\theta, k\omega) \sim r^{-1} e^{-ikr} a_-(\omega, \theta, k), \quad r \to \infty, \tag{1.46}$$

with

$$a_-(\omega, \theta, k) = \overline{a(\omega, \theta, k)}. \tag{1.47}$$

Sometimes, to emphasize the relation between these functions, we use the notation $u_+(x, \xi)$, $v_+(x, \xi)$ and $a_+(\omega, \theta, k)$ for the functions defined by (1.32) and (1.33) and by (1.41).

We note that while the discussion above has dealt with $k > 0$, the case $k = 0$ can also be included. In this case, the proof of Proposition 1.1 does not apply; for example, (1.7) no longer implies (1.10). However, the existence and uniqueness of a solution to

$$\Delta v = 0 \text{ on } \Omega, \quad v = f \text{ on } \partial K, \tag{1.48}$$

satisfying

$$|rv(x)| \le C, \quad |r^2\, \partial_r v| \le C, \ \text{ as } r \to \infty, \tag{1.49}$$

is easily established, as follows. We can assume that the origin $0 \in \mathbb{R}^3$ is in the interior of K. Then the inversion $\psi(x) = x/|x|^2$ interchanges 0 and the point at infinity, and the transformation

$$v(x) = |x|^{-1} w(|x|^{-2} x) \tag{1.50}$$

preserves harmonicity. We let w be the unique harmonic function on the bounded domain $\psi(\Omega)$, with boundary value $w(x) = |x|^{-1} f(\psi(x))$ on $\partial\psi(\Omega) = \psi(\partial K)$. It is easily verified that $v(x)$ satisfies (1.49) in this case. Conversely, if $v(x)$ satisfies (1.48) and w is defined by (1.50), then w is harmonic on $\psi(\Omega) \setminus 0$ and equal to $f \circ \psi$ on $\psi(\partial K)$. If v also satisfies (1.49), then w is bounded near 0, and so is $r\, \partial w/\partial r$.

Now the boundedness of w near 0 implies that 0 is a removable singularity of w, since $\Delta w \in \mathcal{D}'(\psi(\Omega))$ is a distribution supported at 0, hence a finite linear combination of derivatives of $\delta(x)$, which implies that w is the sum of a function harmonic on $\psi(\Omega)$ and a finite sum of derivatives of $|x|^{-1}$, and the latter cannot be bounded unless it is identically zero. Similarly, $r\, \partial w/\partial r$ is harmonic on $\psi(\Omega) \setminus 0$, and if it is bounded near 0 then it extends to be harmonic on $\psi(\Omega)$, and this in turn implies that w extends to be harmonic on $\psi(\Omega)$. Therefore, either one of the two conditions in (1.49) gives uniqueness. Of course, if $f \in C(\partial K)$ the uniqueness of solutions to (1.48), satisfying the first condition in (1.49), follows from the maximum principle.

With this result established, the limiting absorption principle, Theorem 1.3, also holds for $k = 0$. We also note that the proof of Theorem 1.3 continues to work if instead of using $k + i\varepsilon$ ($\varepsilon \searrow 0$) in (1.17), one replaces $k + i\varepsilon$ by any $\lambda(\varepsilon)$ approaching $k \in [0, \infty)$ from the upper half-plane. Furthermore, the limit v depends continuously on k. In particular, the functions $u_\pm(x, \xi)$ defined above are continuous in $\xi \in \mathbb{R}^3$, and $a_\pm(\omega, \theta, k)$ is continuous on $S^2 \times S^2 \times [0, \infty)$.

There is a natural fashion in which $u_+(x, \xi)$ and $u_-(x, \xi)$ fit together, which we describe. This will be useful in §4. Namely, for $k \in \mathbb{R},\ \omega \in S^2$, set

$$U_\pm(x, k, \omega) = e^{-ikx\cdot\omega} + V_\pm(x, k, \omega), \tag{1.51}$$

where V_+ satisfies (1.1)–(1.3) and V_- satisfies (1.1), (1.2), (1.42), with the boundary condition $V_\pm = -e^{-ikx\cdot\omega}$ for $x \in \partial K$. In each case, k is not restricted to be positive; we take any $k \in \mathbb{R}$ (using (1.49) for $k = 0$). It is easy to see that, for any $k > 0$, $V_\pm(x, k, \omega) = v_\pm(x, k\omega)$, while, for $k < 0$, $V_\pm(x, k, \omega) = v_\mp(x, -|k|\omega) = v_\mp(x, k\omega)$. Consequently,

$$\begin{aligned} k > 0 &\Longrightarrow U_\pm(x, k, \omega) = u_\pm(x, k\omega), \\ k < 0 &\Longrightarrow U_\pm(x, k, \omega) = u_\mp(x, k\omega). \end{aligned} \tag{1.52}$$

Similarly, we can define $A_\pm(\omega, \theta, k)$ for $k \in \mathbb{R}$. Note that as $r \to +\infty$,

$$k > 0 \Longrightarrow V_\pm(r\theta, k, \omega) \sim \frac{e^{\pm ikr}}{r} a_\pm(\mp\omega, \theta, k),$$
$$k < 0 \Longrightarrow V_\pm(r\theta, k, \omega) \sim \frac{e^{\mp ikr}}{r} a_\mp(\pm\omega, \theta, k). \tag{1.53}$$

Exercises

1. Let v solve (1.1)–(1.3), with $f \in H^1(\partial K)$, with $k > 0$. Show that
$$\Phi = \pi \operatorname{Im} \int_{\partial K} \frac{\partial v}{\partial \nu} \overline{v}\, dS$$
satisfies
$$\Phi = \pi \operatorname{Im} \int_{|x|=R} \frac{\partial v}{\partial \nu} \overline{v}\, dS,$$
for all R such that $K \subset B_R(0)$, and that
$$\Phi = \lim_{R\to\infty} \frac{\pi}{2k} \int_{|x|=R} \left(\left|\frac{\partial v}{\partial \nu}\right|^2 + k^2|v|^2 \right) dS = \pi k \int_{S^2} \left|\alpha(f, \theta, k)\right|^2 d\theta.$$
The quantity Φ is called the *flux* of the solution v. Show that $\Phi = 0$ implies $v = 0$. (*Hint*: Refer to the proof of Proposition 1.1.)
2. Investigate solutions of (1.1)–(1.3) for $f \in H^s(\partial K)$ with $s < 3/2$. (*Hint*: When extending f to $f^\# \in H^{s+1/2}(\Omega)$, use a parametrix construction for the Dirichlet problem for $\Delta + k^2$.)
3. If $(\Delta + k^2)v(x) = 0$ for $x \in \mathcal{O}$, open in $\mathbb{R}^n$, note that $w(x, y) = v(x)e^{ky}$ is harmonic on $\mathcal{O} \times \mathbb{R} \subset \mathbb{R}^{n+1}$. Deduce that v must be real analytic on $\mathcal{O}$, as asserted in the unique continuation argument used to prove Proposition 1.1.
4. With $a(\omega, \theta, k)$ defined for $k \in \mathbb{R}$ so that (1.53) holds, show that
$$k > 0 \Longrightarrow a(\omega, \theta, -k) = \overline{a(\omega, \theta, k)}. \tag{1.54}$$
Relate this to (1.47).
5. If the obstacle K_2 is obtained from K_1 by translation, $K_2 = K_1 + \eta$, show that the scattering amplitudes are related by
$$a_{K_2}(\omega, \theta, k) = e^{ik(\omega-\theta)\cdot\eta}\, a_{K_1}(\omega, \theta, k).$$

The following exercises deal with the operator $H = -\Delta + V$ on $\mathbb{R}^3$, assuming $V(x)$ is a real-valued function in $C_0^\infty(\mathbb{R}^3)$. We consider the following variant of (1.1)–(1.3), given $f \in L^2_{\text{comp}}(\mathbb{R}^3)$:
$$(\Delta - V + k^2)v = f \quad \text{on } \mathbb{R}^3, \tag{1.55}$$

$$|rv(x)| \le C, \quad r\left(\frac{\partial v}{\partial r} - ikv\right) \to 0, \ \text{ as } r \to \infty. \tag{1.56}$$

6. Show that if $k > 0$ and v satisfies (1.55)–(1.56) and $f = 0$, then $v = 0$. (*Hint*: Modify the proof of Proposition 1.1, to get $v(x) = 0$ on $\mathbb{R}^3 \setminus B_R$, given V supported on B_R. Then use the following unique continuation result:

   Theorem UCP. *If L is a second-order, real, scalar, elliptic operator on a connected region Ω, $Lv = 0$ on Ω, and $v = 0$ on a nonempty open set $\mathcal{O} \subset \Omega$, then $v = 0$ on Ω.*

   A proof of this theorem can be found in [Ho], or in Chap. 14 of [T3].)
7. Show that H has no positive eigenvalues. (*Hint*: Use similar reasoning, with an appropriate variant of Lemma 1.2.) Obtain an analogue of Proposition 7.3 of Chap. 8, regarding negative eigenvalues.
8. Modify the proof of Theorem 1.3 to obtain a (unique) solution of (1.55)–(1.56), as a limit of $\bigl(-H + (k + i\varepsilon)^2\bigr)^{-1} f$, as $\varepsilon \searrow 0$, given $k > 0$. Show that (parallel to (1.4)) the solution v satisfies

$$v(x) = -\int \bigl(V(y)v(y) + f(y)\bigr) g(x,y,k)\, dy = \mathcal{R}(k)(Vv + f). \tag{1.57}$$

   This is called the Lippman–Schwinger equation.
9. Let $u(x,\xi) = e^{-ix\cdot\xi} + v(x,\xi)$, where v satisfies

$$(\Delta - V + k^2)v = V(x)e^{-ix\cdot\xi}, \quad k^2 = |\xi|^2,$$

   and (1.56). Establish an analogue of Proposition 1.6 and an analogue of (1.41), yielding $a(-\omega,\theta,k)$. Note the following case of (1.57):

$$v_+(x,\xi) = -\int V(y)u_+(x,\xi)g(x,y,k)\, dy. \tag{1.58}$$

10. Note that the argument involving (1.48)–(1.50) has no analogue for the $k = 0$ case of (1.55)–(1.56). Reconsider this fact after looking at Exercise 9 of §9.

2. Eigenfunction expansions

The Laplace operator on Ω with the Dirichlet boundary condition, that is, with domain

$$\mathcal{D}(\Delta) = H_0^1(\Omega) \cap H^2(\Omega),$$

is self-adjoint and negative, so by the spectral theorem there is a projection-valued measure $dE(\lambda)$ such that

$$\varphi(-\Delta)v = \int_0^\infty \varphi(\lambda)\, dE(\lambda)v, \tag{2.1}$$

for any bounded continuous function φ. Furthermore, this spectral measure is given in terms of the jump of the resolvent across the real axis:

$$\varphi(-\Delta)v = \lim_{\varepsilon\to 0} \frac{1}{2\pi i} \int \varphi(\lambda)\bigl[(\Delta + \lambda - i\varepsilon)^{-1} - (\Delta + \lambda + i\varepsilon)^{-1}\bigr] v\, d\lambda. \tag{2.2}$$

Using the kernel $G(x, y, k + i\varepsilon)$ for $\big(\Delta + (k + i\varepsilon)^2\big)^{-1}$, we can write this as

$$\varphi(-\Delta)v(x) = \lim_{\varepsilon \searrow 0} \frac{2}{\pi} \int_0^\infty \int_\Omega \varphi(k^2) \operatorname{Im} G(x, y, k + i\varepsilon)v(y)\, dy\, k\, dk. \tag{2.3}$$

From the limiting behavior

$$G(x, y, k + i\varepsilon) \longrightarrow G(x, y, k)$$

established in §1, we can draw the following conclusion.

Proposition 2.1. *The operator Δ on Ω has only absolutely continuous spectrum. For any continuous φ with compact support, we have*

$$\varphi(\sqrt{-\Delta})v(x) = \frac{2}{\pi} \int_0^\infty \int_\Omega \mathit{Im}\, G(x, y, k)v(y)\, dy\, \varphi(k)\, k\, dk. \tag{2.4}$$

The meaning of the first statement of the proposition is that the spectral measure is absolutely continuous with respect to Lebesgue measure.

The primary goal of this section is to give the spectral decomposition of the Laplace operator on Ω in terms of the "eigenfunctions" $u(x, \xi)$ defined by (1.32)–(1.33). We use a modified version of an approach taken in [Rm]. In view of (2.4), the following result plays a key role in achieving the spectral decomposition.

Proposition 2.2. *We have the identity*

$$\mathit{Im}\, G(x, y, k) = \frac{k}{16\pi^2} \int_{S^2} u(x, k\omega)\, \overline{u(y, k\omega)}\, d\omega. \tag{2.5}$$

Proof. We obtain this identity from the asymptotic result of Proposition 1.6, as follows. Applying Green's theorem to $G(x, y, k)$ and $\overline{G(x, y, k)}$, and using the fact that they both vanish for $x \in \partial K$, we have

$$\begin{aligned} &\operatorname{Im} G(x, y, k) \\ &= \frac{1}{2i} \int_{S_R} \Big[G(x, y, k) \frac{\partial}{\partial |z|} \overline{G(z, y, k)} - \overline{G(x, y, k)} \frac{\partial}{\partial |z|} G(z, y, k) \Big]\, dS(z), \end{aligned} \tag{2.6}$$

for R large, where $S_R = \{z \in \mathbb{R}^3 : |z| = R\}$. Letting $R \to \infty$, and using (1.34) and (1.38), gives (2.5) in the limit.

In view of (2.5), we can write the identity (2.4) as

$$\varphi(\sqrt{-\Delta})v(x) = (2\pi)^{-3} \int_{\mathbb{R}^3} \int_\Omega u(x, \xi)\overline{u(y, \xi)}v(y)\varphi(|\xi|)\, dy\, d\xi. \tag{2.7}$$

Therefore, we are motivated to define the following analogues of the Fourier transform:

$$(\Phi v)(\xi) = (2\pi)^{-3/2} \int_{\Omega} v(y)\overline{u(y,\xi)}\, dy \tag{2.8}$$

and

$$(\Phi^* w)(x) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} u(x,\xi) w(\xi)\, d\xi. \tag{2.9}$$

We aim to prove that Φ defines a unitary transformation from $L^2(\Omega)$ onto $L^2(\mathbb{R}^3)$, with inverse Φ^*. Note that §1 gives the estimate

$$|u(x,\xi)| \le 1 + C(\xi)\langle x\rangle^{-1}, \tag{2.10}$$

with $C(\xi)$ locally bounded, but we have obtained no bound on $C(\xi)$ as $|\xi| \to \infty$, so our analysis of Φ and Φ^* will require some care. The following results on Φ and Φ^* are elementary.

Lemma 2.3. *We have*

$$\Phi : C_0^\infty(\Omega) \longrightarrow C(\mathbb{R}^3), \tag{2.11}$$

$$\Phi^* : L^\infty_{comp}(\mathbb{R}^3) \longrightarrow L^\infty(\Omega) \cap C^\infty(\overline{\Omega}), \tag{2.12}$$

and

$$(\Phi^* w, v) = (w, \Phi v), \ \ \textit{for } v \in C_0^\infty(\Omega),\ w \in L^\infty_{comp}(\mathbb{R}^3). \tag{2.13}$$

We also note that (2.7) gives

$$\varphi(\sqrt{-\Delta})v = \Phi^*\big(\varphi(|\xi|)\Phi v\big), \ \ \text{for } v \in C_0^\infty(\Omega),\ \varphi \in C_0^\infty(\mathbb{R}). \tag{2.14}$$

Using these results, we will be able to establish the following.

Proposition 2.4. *If $v \in C_0^\infty(\Omega)$, then $\Phi v \in L^2(\mathbb{R}^3)$ and*

$$\|\Phi v\|_{L^2(\mathbb{R}^3)} = \|v\|_{L^2(\Omega)}. \tag{2.15}$$

Consequently, Φ has a unique extension to an isometric map

$$\Phi : L^2(\Omega) \longrightarrow L^2(\mathbb{R}^3), \tag{2.16}$$

and Φ^* has a unique continuous extension to a continuous map

$$\Phi^* : L^2(\mathbb{R}^3) \longrightarrow L^2(\Omega), \tag{2.17}$$

the adjoint of (2.16).

Proof. Given $\varphi \in C_0^\infty(\mathbb{R}),\ v \in C_0^\infty(\Omega)$, we have

$$\begin{aligned}\big(\varphi(|\xi|)\Phi v, \Phi v\big) &= \big(\Phi^* \varphi(|\xi|)\Phi v, v\big) \qquad &\text{(by (2.13))}\\ &= (\varphi(\sqrt{-\Delta})v, v). \qquad &\text{(by (2.14))}\end{aligned} \tag{2.18}$$

In other words,

$$\int_{\mathbb{R}^3} \varphi(|\xi|)\big|\Phi v(\xi)\big|^2\, d\xi = \big(\varphi(\sqrt{-\Delta})v, v\big). \tag{2.19}$$

Now let $\varphi \nearrow 1$. The monotone convergence theorem applies, so

$$\int_{\mathbb{R}^3} \big|\Phi v(\xi)\big|^2\, d\xi = (v, v). \tag{2.20}$$

This proves the proposition.

In order to prove that (2.16) is surjective—hence unitary—we will need to know that Φ^* in (2.17) is injective. Before proving this, it will be useful to establish the following.

Proposition 2.5. *For any even $\varphi \in C_o(\mathbb{R})$ (i.e., φ continuous and $\varphi(t) \to 0$ as $|t| \to \infty$), and for any $w \in L^2(\mathbb{R}^3)$,*

$$\Phi^*\big(\varphi(|\xi|)w\big) = \varphi(\sqrt{-\Delta})\Phi^* w. \tag{2.21}$$

Proof. It suffices to establish this identity for $w \in C_0^\infty(\mathbb{R}^3)$. For such w, we have

$$(1-\Delta)\Phi^* w(x) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} u(x,\xi)\langle\xi\rangle^2 w(\xi)\, d\xi = \Phi^*\big(\langle\xi\rangle^2 w\big),$$

the left side a priori a distribution on Ω. By (2.17), we know that $\Phi^*(\langle\xi\rangle^2 w) \in L^2(\Omega)$. The integral above clearly belongs to $C^\infty(\overline{\Omega})$ and vanishes on $\partial\Omega$. Thus $\Phi^* w(x)$ belongs to the domain of Δ_c^*, where

$$\mathcal{D}(\Delta_c) = \{u \in C^\infty(\overline{\Omega}) : u = 0 \text{ on } \partial\Omega,\ \operatorname{supp} u \text{ bounded}\}.$$

It follows from Proposition 2.6 of Chap. 8 that Δ is essentially self-adjoint on $\mathcal{D}(\Delta_c)$, so we conclude that $\Phi^* w(x)$ belongs to the domain of Δ, namely, to $H_0^1(\Omega) \cap H^2(\Omega)$.

An inductive argument shows that $\Phi^* w$ belongs to the domain of each self-adjoint operator $(1-\Delta)^K$ and

$$(1-\Delta)^K \Phi^* w(x) = \Phi^*\big(\langle\xi\rangle^{2K} w\big).$$

Replacing w by $\langle\xi\rangle^{-2K}w$, we deduce

$$\Phi^*\big(\langle\xi\rangle^{-2K}w\big) = (1-\Delta)^{-K}\Phi^* w,$$

for all $w \in C_0^\infty(\mathbb{R}^3)$. From this we get

$$\Phi^*\big(|\xi|^{2j}\langle\xi\rangle^{-2K}w\big) = \Delta^j(1-\Delta)^{-K}\Phi^* w.$$

Consequently, the identity (2.21) is valid for any $\varphi(t) = t^{2j}\langle t\rangle^{-2K}$, $j < K$. Now the space of finite linear combinations of such φ is dense in the space of even elements of $C_o(\mathbb{R})$, with the sup norm, by the Stone-Weierstrass theorem, so (2.21) holds in general.

We also have the following dual result.

Proposition 2.6. *For $v \in L^2(\Omega)$, $\varphi \in C_o(\mathbb{R})$ even, we have*

$$\Phi\big(\varphi(\sqrt{-\Delta})v\big)(\xi) = \varphi(|\xi|)(\Phi v)(\xi). \tag{2.22}$$

Proof. Since, by Proposition 2.4, Φ and Φ^* are L^2-continuous and adjoints of each other, this follows directly from (2.21).

We now prove the asserted unitarity of Φ and Φ^*.

Proposition 2.7. *The map Φ^* is injective on $L^2(\mathbb{R}^3)$. Hence the maps* (2.16) *and* (2.17) *are unitary and are inverses of each other.*

Proof. By Proposition 2.5, if $w \in \ker\Phi^*$, then $\varphi(|\xi|)w \in \ker\Phi^*$, for any $\varphi \in C_0^\infty(\mathbb{R})$. Hence if $\ker\Phi^*$ is nonzero, it contains an element with compact support. Let w denote such an element. Then

$$0 = \int\limits_{\mathbb{R}^3} u(y,\xi)\varphi(|\xi|)w(\xi)\,d\xi, \quad \text{for all } y \in \Omega, \tag{2.23}$$

for any continuous φ, the integral being absolutely convergent. This being the case, we can take

$$\varphi(|\xi|) = g(x,y,|\xi|). \tag{2.24}$$

Also, we can use $\varphi(|\xi|) = \partial g(x,y,|\xi|)/\partial|y|$, and we can also replace $u(y,\xi)$ by $\partial u(y,\xi)/\partial|y|$. Consequently, for all $r > R_0$, such that $K \subset \{x \in \mathbb{R}^3 : |x| \le R_0\}$, we have

$$\int\int\limits_{|y|=r} w(\xi)\Big[u(y,\xi)\frac{\partial g}{\partial|y|}(x,y,|\xi|) - g(x,y,|\xi|)\frac{\partial u}{\partial|y|}(y,\xi)\Big]\,dS(y)\,d\xi = 0, \tag{2.25}$$

for all $x \in \mathbb{R}^3$. In the limit $r \to \infty$, this gives

$$0 = \int w(\xi) e^{-ix\cdot\xi}\, d\xi, \quad \text{for all } x \in \mathbb{R}^3. \tag{2.26}$$

In other words, the Fourier transform of w vanishes identically. This implies $w = 0$ and completes the proof.

If we replace $u(x, \xi) = u_+(x, \xi)$ by $u_-(x, \xi)$, given by (1.43)–(1.45), we can define the operator Φ_- by

$$(\Phi_- v)(\xi) = (2\pi)^{-3/2} \int_\Omega v(y)\overline{u_-(y, \xi)}\, dy. \tag{2.27}$$

The arguments as above show that Φ_- provides a unitary operator from $L^2(\Omega)$ onto $L^2(\mathbb{R}^3)$, and the intertwining property (2.22) also holds for Φ_-. The relation between Φ_- and Φ is important in scattering theory; often we denote Φ by Φ_+ to emphasize this.

Exercises

1. If $\varphi \in C_0^\infty(\mathbb{R})$ is even, show that the Schwartz kernel of $\varphi(\sqrt{-\Delta})$ is given by

$$K_\varphi(x, y) = (2\pi)^{-3} \int_{\mathbb{R}^3} u(x, \xi)\overline{u(y, \xi)}\varphi(|\xi|)\, d\xi. \tag{2.28}$$

In particular,

$$K_\varphi(x, x) = (2\pi)^{-3} \int_{\mathbb{R}^3} |u(x, \xi)|^2 \varphi(|\xi|)\, d\xi. \tag{2.29}$$

2. Show that (2.29) is also valid for $\varphi(\lambda) = \varphi_t(\lambda) = e^{-t\lambda^2}$, given $t > 0$. (*Hint*: Let $\varphi_j \in C_0^\infty(\mathbb{R})$, $\varphi_j \nearrow \varphi_t$.)
3. Show that the heat kernel $H_t(x, y)$ on $\Omega \times \Omega$ of $e^{t\Delta}$, with Dirichlet boundary condition, has the pointwise bound

$$H_t(x, y) \le (4\pi t)^{-3/2}\, e^{-|x-y|^2/4t}.$$

Deduce that, for each $x \in \Omega$,

$$(2\pi)^{-3} \int_{\mathbb{R}^3} |u(x, \xi)|^2\, e^{-t|\xi|^2}\, d\xi \le (4\pi t)^{-3/2},$$

and hence

$$\int_{|\xi| \le R} |u(x, \xi)|^2\, d\xi \le C\, R^3. \tag{2.30}$$

4. Deduce that (2.28) and (2.29) remain valid for even $\varphi \in \mathcal{S}(\mathbb{R})$, indeed, for continuous even φ satisfying $|\varphi(\lambda)| \le C\langle\lambda\rangle^{-4-\varepsilon}$, $\varepsilon > 0$.
5. Verify that (2.26) follows from (2.25). (*Hint*: If $e^{-iy\cdot\xi}$ is substituted for $u(y,\xi)$ in (2.25), Green's formula applies. If $v(y,\xi)$ is substituted, use the asymptotic behavior to show that the inner integral tends to 0 as $r \to \infty$.)
6. Produce results parallel to those of this section for $H = -\Delta + V$, given $V \in C_0^\infty(\mathbb{R}^3)$, real, $u(x,\xi)$ as in Exercise 9 of §1. Show that $\Phi : \mathcal{H}_c \to L^2(\mathbb{R}^3)$ is unitary, where $\mathcal{H}_c$ is the orthogonal complement of the set of eigenfunctions of H (with negative eigenvalue, if any). To what extent does $k = 0$ cause a problem?

3. The scattering operator

In §2 we produced the two unitary operators

$$\Phi_\pm : L^2(\Omega) \longrightarrow L^2(\mathbb{R}^3), \tag{3.1}$$

defined for $f \in C_0^\infty(\Omega)$ by

$$(\Phi_\pm f)(\xi) = (2\pi)^{-3/2} \int_\Omega \overline{u_\pm(y,\xi)}\, f(y)\, dy. \tag{3.2}$$

From these one constructs the unitary operator

$$S = \Phi_+ \Phi_-^* : L^2(\mathbb{R}^3) \longrightarrow L^2(\mathbb{R}^3), \tag{3.3}$$

called the scattering operator. Recall that Φ_+ and Φ_- intertwine $\varphi(\sqrt{-\Delta})$ on $L^2(\Omega)$ with multiplication by $\varphi(|\xi|)$ on $L^2(\mathbb{R}^3)$, for $\varphi \in C_o(\mathbb{R})$. It follows that S commutes with such $\varphi(|\xi|)$:

$$S\varphi(|\xi|) = \varphi(|\xi|) S. \tag{3.4}$$

From the definition (3.3) we see that S is uniquely characterized by the property

$$S\big(\varphi(|\xi|)\overline{u_-(y,\cdot)}\big) = \varphi(|\xi|)\overline{u_+(y,\cdot)}, \quad \text{for all } y \in \Omega, \tag{3.5}$$

for all $\varphi \in C_0^\infty(\mathbb{R})$. We will relate the operator S to "wave operators" in §5.

We aim to establish the following formula for S, in terms of the scattering amplitude $a(\omega,\theta,k)$ defined in §1.

Proposition 3.1. *For $g \in C_0^\infty(\mathbb{R}^3)$, we can write*

$$(Sg)(\xi) = S(k)\, g(k\omega), \quad \xi = k\omega, \quad \omega \in S^2, \tag{3.6}$$

where, for each $k \in \mathbb{R}^+$, $S(k)$ *is a unitary operator on* $L^2(S^2)$ *given by*

$$S(k)f(\omega) = f(\omega) + \frac{k}{2\pi i} \int_{S^2} a(\omega, \theta, k) f(\theta)\, d\theta. \tag{3.7}$$

Proof. Let

$$w(y, k\omega) = u_+(y, k\omega) - u_-(y, k\omega) = v_+(y, k\omega) - v_-(y, k\omega). \tag{3.8}$$

The assertion above is equivalent to the integral identity

$$w(y, k\omega) = -\frac{k}{2\pi i} \int_{S^2} \overline{a(\omega, \theta, k)} u_+(y, k\theta)\, d\theta. \tag{3.9}$$

In order to prove this, note that, since $w(x, k\omega) = 0$ for $x \in \partial K$, Green's theorem gives, for $R > |y|$,

$$w(y, k\omega) = \int_{S_R} \Big[w(x, k\omega) \frac{\partial G}{\partial |x|}(y, x, k) - G(y, x, k) \frac{\partial w}{\partial |x|}(x, k\omega) \Big]\, dS(x), \tag{3.10}$$

where $S_R = \{x : |x| = R\}$. Now let $R \to \infty$. Using the asymptotic behavior (1.34) and (1.38) for $G(y, x, k)$ and its radial derivative (with x and y interchanged and ω replaced by θ) and the asymptotic behavior, for $|x| = R \to \infty$, $x = R\theta$,

$$\begin{aligned} w(x, k\omega) &\sim \frac{e^{ikR}}{R}\, a(-\omega, \theta, k) - \frac{e^{-ikR}}{R}\, a_-(\omega, \theta, k), \\ \frac{\partial w}{\partial |x|}(x, k\omega) &\sim ik \frac{e^{ikR}}{R}\, a(-\omega, \theta, k) + ik \frac{e^{-ikR}}{R}\, a_-(\omega, \theta, k), \end{aligned} \tag{3.11}$$

with

$$a_-(\omega, \theta, k) = \overline{a(\omega, \theta, k)}, \tag{3.12}$$

as in (1.46)–(1.47), we see that the integrand in (3.10) is asymptotic to

$$\frac{2ik}{4\pi R^2}\, a_-(\omega, \theta, k) u_+(y, k\omega) + o(R^{-2}), \tag{3.13}$$

(the terms involving e^{2ikR} canceling out), so passing to the limit $R \to \infty$ gives (3.9) and proves the proposition.

We can rewrite the formula (3.7) as

$$S(k) = I + \frac{k}{2\pi i} A(k), \tag{3.14}$$

with

$$A(k) f(\omega) = \int_{S^2} a(\omega, \theta, k) f(\theta)\, d\theta. \tag{3.15}$$

Note that unitarity of $S(k)$ on $L^2(S^2)$ is equivalent to the identity

$$\frac{1}{2i}\big[A(k)^* - A(k)\big] = \frac{k}{4\pi} A(k)^* A(k), \tag{3.16}$$

that is, to the integral identity

$$\frac{1}{2i}\big[\overline{a(\theta, \omega, k)} - a(\omega, \theta, k)\big] = \frac{k}{4\pi} \int_{S^2} \overline{a(\eta, \omega, k)} a(\eta, \theta, k)\, d\eta. \tag{3.17}$$

The special case of this where $\omega = \theta$ is known as the *optical theorem*:

$$\operatorname{Im} a(\omega, \omega, k) = -\frac{k}{4\pi} \int_{S^2} |a(\eta, \omega, k)|^2\, d\eta. \tag{3.18}$$

It is useful to know integral identities for the scattering amplitude. We note one that follows from the characterization

$$v(r\theta, k\omega) \sim r^{-1} e^{ikr} a(-\omega, \theta, k), \quad r \to \infty \tag{3.19}$$

and the integral identity (a consequence of Green's identity)

$$v(x, k\omega) = \int_{\partial K} \Big[v(y, k\omega) \frac{\partial g}{\partial \nu_y}(x, y, k) - g(x, y, k) \frac{\partial v}{\partial \nu}(y, k\omega) \Big]\, dS(y). \tag{3.20}$$

We evaluate the integrand on the right as $x = r\theta,\ r \to \infty$. Using (1.36), that is,

$$g(x, y, k) \sim -(4\pi r)^{-1} e^{ikr} e^{-ik\theta\cdot y}, \quad x = r\theta,\ r \to \infty, \tag{3.21}$$

we find from (3.19) and (3.20) that

$$\begin{aligned} a(\omega, \theta, k) = &-\frac{1}{4\pi} \int_{\partial K} e^{ik\omega\cdot y} \frac{\partial}{\partial \nu} e^{-ik\theta\cdot y}\, dS(y) \\ &+ \frac{1}{4\pi} \int_{\partial K} e^{-ik\theta\cdot y} \frac{\partial}{\partial \nu} v(y, -k\omega)\, dS(y). \end{aligned} \tag{3.22}$$

The first term on the right side of (3.22) can be written as

$$\frac{ik}{4\pi}\int_{\partial K}\big(\nu(y)\cdot\theta\big)\,e^{ik(\omega-\theta)\cdot y}\,dS(y)=\frac{ik}{4\pi}\theta\cdot\hat{A}_K\big(k(\omega-\theta)\big),\tag{3.23}$$

where, for $\xi\in\mathbb{R}^3$,

$$\hat{A}_K(\xi)=\int_{\partial K}\nu(y)\,e^{i\xi\cdot y}\,dS(y).\tag{3.24}$$

The function $\hat{A}_K(\xi)$ clearly extends to an entire analytic function of $\xi\in\mathbb{C}^3$. For $\xi\in\mathbb{R}^3$ tending to infinity, one can (typically) find the asymptotic behavior of $\hat{A}_K(\xi)$ via the stationary phase method. Note that

$$\hat{A}_K(0)=0.\tag{3.25}$$

One way of writing the last term in (3.22) is the following. For any real k, or more generally for $\operatorname{Im} k\geq 0$, define the Neumann operator $\mathcal{N}(k)$ on $f\in H^1(\partial K)$ to be the value of $\partial v/\partial\nu$ in $L^2(\partial K)$, where v is the unique solution to the scattering problem (1.1)–(1.3). Define the functions e_ξ on ∂K by

$$e_\xi(y)=e^{iy\cdot\xi},\quad y\in\partial K.\tag{3.26}$$

Then the last term in (3.22) is

$$\frac{1}{4\pi}\big(\mathcal{N}(k)e_{k\omega},e_{k\theta}\big)_{L^2(\partial K)}.\tag{3.27}$$

Consequently, the formula for the scattering amplitude can be written as

$$a(\omega,\theta,k)=\frac{ik}{4\pi}\theta\cdot\hat{A}_K\big(k(\omega-\theta)\big)+\frac{1}{4\pi}\big(\mathcal{N}(k)e_{k\omega},e_{k\theta}\big)_{L^2(\partial K)}.\tag{3.28}$$

We will investigate the Neumann operator further in §7.

We can produce a variant of the formula (3.22) by using $G(x,y,k)$ instead of $g(x,y,k)$ in (3.20). We then get

$$v(x,k\omega)=-\int_{\partial K}e^{-ik\omega\cdot y}\,\frac{\partial G}{\partial\nu_y}(x,y,k)\,dS(y).\tag{3.29}$$

Using the limiting behavior for $G(x,y,k)$ as $|x|\to\infty$, which follows from (1.34), we have

$$a(\omega,\theta,k)=-\frac{1}{4\pi}\int_{\partial K}e^{ik\omega\cdot y}\,\frac{\partial u}{\partial\nu}(y,k\theta)\,dS(y).\tag{3.30}$$

If we write $u(y, k\theta) = e^{-ik\theta\cdot y} + v(y, k\theta)$, this becomes a sum of two terms. The first is identical to the first term in (3.22), while the second differs from the second term in (3.22) precisely by the replacement of (ω, θ) by $(-\theta, -\omega)$. From this observation, we can derive the following identity, called the *reciprocity relation*:

$$a(\omega, \theta, k) = a(-\theta, -\omega, k). \tag{3.31}$$

To see this, it suffices to show that $k(\omega + \theta) \cdot \hat{A}_K\big(k(\omega - \theta)\big) = 0$. Since $\omega + \theta$ and $\omega - \theta$ are orthogonal for unit ω and θ, this is equivalent to the observation that

$$\hat{A}_K(\xi) \text{ is parallel to } \xi, \quad \text{for } \xi \in \mathbb{R}^3, \tag{3.32}$$

and this follows easily from Green's theorem.

Exercises

1. Show that (3.5) follows from

$$\int u_-(x, \xi)\varphi(|\xi|)\overline{u_-(y, \xi)}\, d\xi = \int u_+(x, \xi)\varphi(|\xi|)\overline{u_+(y, \xi)}\, d\xi,$$

which in turn follows from (2.28).

2. Fill in the details on the identities (3.25) and (3.32) for $\hat{A}_K(\xi)$, and then on the reciprocity relation (3.31). What is the intuitive content of (3.31)?
3. If you set $S_- = \Phi_-\Phi_+^*$, obtain an analogue of (3.7), with $a(\omega, \theta, k)$ replaced by $a_-(\omega, \theta, k)$.
4. In case $f = -e^{-ikx\cdot\omega}\big|_{\partial K}$, with corresponding scattered wave v, show that the flux Φ studied in Exercise 1 of §1 is given by

$$\begin{aligned} \sigma(\omega, k) &= \lim_{r\to\infty} \pi k \int\limits_{|x|=r} |v(x, k\omega)|^2\, dS(x) \\ &= \pi k \int\limits_{S^2} |a(-\omega, \theta, k)|^2\, d\theta. \end{aligned}$$

We call $\sigma(\omega, k)$ the scattering cross section. Using the optical theorem and the reciprocity relation (3.31), show that

$$\sigma(\omega, k) = -4\pi^2 \operatorname{Im} a(\omega, \omega, k).$$

5. Generalizing (3.22), show that, for $f \in H^s(\partial K)$,

$$\mathcal{B}_K(k) f(r\theta) \sim r^{-1} e^{ikr}\, \mathcal{A}_K(k) f(\theta) + o(r^{-1}), \tag{3.33}$$

as $r \to \infty$, where

$$\mathcal{A}_K(k) f(\theta) = \frac{1}{4\pi} \int\limits_{\partial K} e^{-ik\theta\cdot y}\big[ik\big(\nu(y)\cdot\theta\big) f(y) + \mathcal{N}(k) f(y)\big]\, dS(y). \tag{3.34}$$

6. Make a parallel study of the scattering operator for $H = -\Delta + V,\ V \in C_0^\infty(\mathbb{R}^3)$, real-valued, using results from the exercises in §§1 and 2. To begin, use the unitary operators $\Phi_\pm : \mathcal{H}_c \to L^2(\mathbb{R}^3)$ to construct $S = \Phi_+ \Phi_-^*$. Show that, parallel to (3.22),

$$a(-\omega, \theta, k) = -\frac{1}{4\pi} \int V(y) u(y, k\omega) e^{-ik\theta\cdot y}\, dy,$$

or equivalently,

$$a(\omega, \theta, k) = -\Big(\frac{\pi}{2}\Big)^{1/2} \widehat{V}\big(k(\theta - \omega)\big) - \frac{1}{4\pi} \int V(y) v(y, -k\omega) e^{-ik\theta\cdot y}\, dy. \tag{3.35}$$

4. Connections with the wave equation

The initial-value problem for the wave equation on $\mathbb{R} \times \Omega$, with Dirichlet boundary conditions on $\mathbb{R} \times \partial K$, is of the following form:

$$\frac{\partial^2 u}{\partial t^2} - \Delta u = 0, \tag{4.1}$$

$$u(0, x) = f(x), \quad u_t(0, x) = g(x), \tag{4.2}$$

for $t \in \mathbb{R},\ x \in \Omega$, with

$$u(t, x) = 0, \quad \text{for } x \in \partial K. \tag{4.3}$$

As we know, given $f \in H_0^1(\Omega),\ g \in L^2(\Omega)$, there is a unique solution u belonging to $C(\mathbb{R}, H_0^1(\Omega)) \cap C^1(\mathbb{R}, L^2(\Omega))$ to this problem, given in terms of functions of the self-adjoint operator Δ on $L^2(\Omega)$, with domain $H_0^1(\Omega) \cap H^2(\Omega)$, as

$$u(t, x) = (\cos t\Lambda) f(x) + (\Lambda^{-1} \sin t\Lambda) g(x), \tag{4.4}$$

where

$$\Lambda = (-\Delta)^{1/2} \tag{4.5}$$

is the unique nonnegative, self-adjoint square root of $-\Delta$. Recall that the domain of Λ is precisely $\mathcal{D}(\Lambda) = H_0^1(\Omega)$. Alternatively, we can write

$$\begin{pmatrix} u \\ u_t \end{pmatrix} = U(t) \begin{pmatrix} f \\ g \end{pmatrix}, \tag{4.6}$$

where $U(t)$ is the one-parameter group of operators on $H_0^1(\Omega) \oplus L^2(\Omega)$ given by

$$U(t) = \begin{pmatrix} \cos t\Lambda & \Lambda^{-1} \sin t\Lambda \\ -\Lambda \sin t\Lambda & \cos t\Lambda \end{pmatrix}. \tag{4.7}$$

Using either of the unitary operators

$$\Phi_\pm : L^2(\Omega) \longrightarrow L^2(\mathbb{R}^3), \tag{4.8}$$

we can write

$$\begin{aligned}(\cos t\Lambda)f &= \Phi_\pm^{-1} \cos t|\xi|\Phi_\pm f,\\ (\Lambda^{-1}\sin t\Lambda)g &= \Phi_\pm^{-1}|\xi|^{-1}\sin t|\xi|\Phi_\pm g.\end{aligned} \tag{4.9}$$

Note that $\Phi_\pm$ also provide isomorphisms

$$\Phi_\pm : H_0^1(\Omega) \longrightarrow L^2(\mathbb{R}^3, \langle\xi\rangle^2 d\xi). \tag{4.10}$$

The group $U(t)$ is not a uniformly bounded group of operators on the Hilbert space $H_0^1(\Omega) \oplus L^2(\Omega)$. Indeed, with $f = 0$, we see from (4.4) that the best uniform estimate on $\|u(t,\cdot)\|_{L^2(\Omega)}$ is

$$\|u(t,\cdot)\|_{L^2(\Omega)} \le |t|\ \|g\|_{L^2(\Omega)}. \tag{4.11}$$

There is another Hilbert space on which $U(t)$ naturally acts as a group of unitary operators, namely the space

$$\mathcal{E} = \mathcal{H} \oplus L^2(\Omega), \tag{4.12}$$

where $\mathcal{H}$ is the completion of $H_0^1(\Omega)$ with respect to the norm given by

$$\|f\|_{\mathcal{H}}^2 = \|\Lambda f\|_{L^2(\Omega)}^2 = \int_\Omega |\nabla f(x)|^2\, dx. \tag{4.13}$$

(Recall that $\|f\|^2_{H_0^1(\Omega)} = \|f\|^2_{L^2(\Omega)} + \|\Lambda f\|^2_{L^2(\Omega)}$.) If we equip $\mathcal{H}$ with this norm, then $\Phi_\pm$ extend to unitary operators

$$\Phi_\pm : \mathcal{H} \longrightarrow L^2(\mathbb{R}^3, |\xi|^2 d\xi). \tag{4.14}$$

Since unitary operators are special, it is natural to use the Hilbert space (4.12) rather than $H_0^1(\Omega) \oplus L^2(\Omega)$. We will denote an element of $\mathcal{E}$ by $\langle f, g\rangle$; $f \in \mathcal{H}$, $g \in L^2(\Omega)$. When $U(t)$ is applied, this is treated as a column vector, as in (4.6); we will also use the column vector notation for elements of $\mathcal{E}$ when convenient.

Elements of $\mathcal{H}$ need not belong to $L^2(\Omega)$, though they do belong to $L^2_{\text{loc}}(\overline{\Omega})$. In fact, if B is a bounded subset of $\overline{\Omega}$, the estimate

$$\|u\|_{L^2(B)} \le C_B \|u\|_{\mathcal{H}} \tag{4.15}$$

can be established by the argument used to prove Proposition 5.2 in Chap. 4, provided K has nonempty interior. Since clearly $\int_B |\nabla u|^2\, dx \le \int_\Omega |\nabla u|^2\, dx$, we hence have

$$\|u\|_{H^1(B)} \le C'_B \|u\|_{\mathcal{H}}. \tag{4.16}$$

Further estimates are given in the exercises.

The unitarity of $U(t)$ on $\mathcal{E}$ reflects the conservation of total energy, given by

$$E(u(t)) = \|\langle u, u_t\rangle\|_{\mathcal{E}}^2 = \int_\Omega \left(|\nabla_x u(t,x)|^2 + |u_t(t,x)|^2\right) dx. \tag{4.17}$$

There is also the notion of local energy, given as follows. For a bounded subset B of $\overline{\Omega}$, set

$$E_B(u(t)) = \int_B \left(|\nabla_x u(t,x)|^2 + |u_t(t,x)|^2\right) dx. \tag{4.18}$$

Using the absolute continuity of the spectrum of Δ on $L^2(\Omega)$ established in §2, or more precisely, the absolute continuity of the spectrum of a related operator specified below, we will establish the following result on local energy decay.

Proposition 4.1. *Given $\langle f, g\rangle \in \mathcal{E}$, $\langle u, u_t\rangle = U(t)\langle f, g\rangle$, we have*

$$E_B(u(t)) \longrightarrow 0, \quad \text{as } |t| \longrightarrow \infty, \tag{4.19}$$

for any bounded $B \subset \overline{\Omega}$.

Before starting the proof of this proposition, we will make some further comments on the infinitesimal generator of the unitary group $U(t)$ on $\mathcal{E}$. This is a skew-adjoint operator, and it has the form

$$B = \begin{pmatrix} 0 & I \\ -A & 0 \end{pmatrix}, \tag{4.20}$$

where, for $f \in \mathcal{D}(A) \subset \mathcal{H}$,

$$Af = -\Delta f \tag{4.21}$$

in the distributional sense. Then B^2 is a self-adjoint operator of the form

$$-B^2 = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \tag{4.22}$$

where A_1 is self-adjoint on $\mathcal{H}$, A_2 is self-adjoint on $L^2(\Omega)$, and they both satisfy (4.21), on their respective domains. Note that the unitary operators

$$\Phi_\pm : \mathcal{H} \oplus L^2(\Omega) \longrightarrow L^2(\mathbb{R}^3, |\xi|^2\, d\xi) \oplus L^2(\mathbb{R}^3)$$

intertwine (4.20) with multiplication (on each factor) by $|\xi|^2$ and

$$\mathcal{D}(\Phi_\pm B^2 \Phi_\pm^{-1}) = L^2(\mathbb{R}^3, |\xi|^2 \langle\xi\rangle^4 \, d\xi) \oplus L^2(\mathbb{R}^3, \langle\xi\rangle^4 \, d\xi).$$

In particular, the operators A_1 and A_2 have only absolutely continuous spectrum. Let

$$L_j = A_j^{1/2} \tag{4.23}$$

be their unique nonnegative, self-adjoint square roots. Both L_1 and L_2 are intertwined via $\Phi_\pm$ with multiplication by $|\xi|$, so we can identify them, denoting them by L, and if $\langle u, u_t\rangle = U(t)\langle f, g\rangle$, we have

$$\begin{aligned} u(t) &= (\cos tL) f + (L^{-1} \sin tL) g, \\ u_t(t) &= (-L \sin tL) f + (\cos tL) g. \end{aligned} \tag{4.24}$$

We now begin the proof of Proposition 4.1. Since $U(t)$ is unitary and $E_B(u(t)) \le E(u(t)) = \|\langle u, u_t\rangle\|_{\mathcal{E}}^2$, we see that it suffices to prove the proposition for $\langle f, g\rangle$ in a dense subset of $\mathcal{E}$. In particular, we will take

$$f \in \mathcal{D}(L_1) \subset \mathcal{H}, \quad g \in \mathcal{D}(L_2) \subset L^2(\Omega). \tag{4.25}$$

Lemma 4.2. *If f and g satisfy (4.25), then, as $|t| \to \infty$,*

$$\begin{aligned} u(t) &\longrightarrow 0 \quad \text{weakly in } \mathcal{D}(L_1) \text{ and} \\ u_t(t) &\longrightarrow 0 \quad \text{weakly in } \mathcal{D}(L_2). \end{aligned} \tag{4.26}$$

Proof. Fix $w_0 \in \mathcal{D}(L_1),\ w_1 \in \mathcal{D}(L_2)$. Note that

$$\Phi_\pm f \in L^2(\mathbb{R}^3, |\xi|^2 \langle\xi\rangle^2 d\xi), \tag{4.27}$$

and so on, so using the images under $\Phi_\pm$ to justify the inner-product calculations, and noting that, by (4.27),

$$Lf \in H_0^1(\Omega), \quad L^2 f \in L^2(\Omega), \quad Lg \in L^2(\Omega) \tag{4.28}$$

(and similarly for $w_0,\ w_1$), we obtain

$$\begin{aligned} \big(u(t), w_0\big)_{\mathcal{D}(L_1)} &= \big(Lu(t), Lw_0\big)_{\mathcal{E}} + \big(u(t), w_0\big)_{\mathcal{E}} \\ &= \big(L^2 u(t), L^2 w_0\big)_{L^2} + \big(Lu(t), Lw_0\big)_{L^2}. \end{aligned} \tag{4.29}$$

To examine each term, write (with $j = 1$ or 2)

$$
\begin{aligned}
\big(L^j u(t), L^j w_0\big)_{L^2} &= \big((L^j \cos tL) f + (L^{j-1} \sin tL) g, L^j w_0\big)_{L^2} \\
&= \int_0^\infty (\cos t\lambda)\, d(F_\lambda L^j f, L^j w_0) \\
&\quad + \int_0^\infty (\sin t\lambda)\, d(F_\lambda L^{j-1} g, L^j w_0),
\end{aligned} \tag{4.30}
$$

where F_λ is the spectral measure of L_2. In light of (4.28) and the absolute continuity of F_λ, it follows that $d(F_\lambda L^j f, L^j w_0)$ and $d(F_\lambda L^{j-1} g, L^j w_0)$ are finite measures on $\mathbb{R}$ that are absolutely continuous with respect to Lebesgue measure. Hence (4.30) is the Fourier transform of an L^1-function on $\mathbb{R}$. Thus the Riemann–Lebesgue lemma implies that this tends to 0 as $|t| \to \infty$. Similarly,

$$
\big(u_t(t), w_1\big)_{\mathcal{D}(L_2)} = \big(Lu_t(t), Lw_1\big)_{L^2} + \big(u_t(t), w_1\big)_{L^2}. \tag{4.31}
$$

This time, to examine each term, write (with $j = 0$ or 1)

$$
\begin{aligned}
\big(L^j u_t(t), L^j w_1\big)_{L^2} &= \big((-L^{j+1} \sin tL) f + (L^j \cos tL) g, L^j w_1\big)_{L^2} \\
&= -\int_0^\infty (\sin t\lambda)\, d(F_\lambda L^{j+1} f, L^j w_1) \\
&\quad + \int_0^\infty (\cos t\lambda)\, d(F_\lambda L^j g, L^j w_1).
\end{aligned} \tag{4.32}
$$

Again the Riemann–Lebesgue lemma applies, and the proof of Lemma 4.2 is complete.

To derive local energy decay from this, we reason as follows. For any $R < \infty$, set

$$
\Omega_R = \{x \in \Omega : |x| < R\}. \tag{4.33}
$$

Then, for $f \in \mathcal{H}$, if $\iota_R f = f\big|_{\Omega_R}$, by (4.16) we have

$$
\|\iota_R f\|_{H^1(\Omega_R)} \le C_R \|f\|_{\mathcal{H}}. \tag{4.34}
$$

Similarly, for any $f \in \mathcal{D}(L_1)$,

$$
\|\iota_R f\|_{H^2(\Omega_R)} \le C_R' \|f\|_{\mathcal{D}(L_1)}. \tag{4.35}
$$

Thus, restricted to Ω_R, $u(t)$ is bounded in $H^2(\Omega_R)$ and $u_t(t)$ is bounded in $H^1(\Omega_R)$, for $t \in \mathbb{R}$, given the hypothesis (4.25) on the initial data. Thus these two families of functions on Ω_R are compact in $H^1(\Omega_R)$ and $L^2(\Omega_R)$, respectively, by Rellich's theorem. The weak convergence to zero of (4.26) hence implies the *strong* convergence to zero:

$$u(t) \longrightarrow 0 \text{ in } H^1(\Omega_R), \quad u_t(t) \longrightarrow 0 \text{ in } L^2(\Omega_R), \tag{4.36}$$

as $|t| \to \infty$, whenever f and g satisfy (4.25). Proposition 4.1 is hence proved on the dense set given by (4.25), and as we remarked before, that proves it in general.

Instead of representing $\langle f, g\rangle \in \mathcal{E}$ as a pair of functions, L^2 with respect to different weights, via $\Phi_\pm$, it is often convenient to use the following construction, of Lax–Phillips. Namely, for $f, g \in C_0^\infty(\Omega)$, define $\Psi_\pm\langle f, g\rangle$ on $\mathbb{R} \times S^2$ by

$$\begin{aligned}\Psi_\pm \begin{pmatrix} f \\ g \end{pmatrix}(k,\omega) = \frac{k^2}{4\pi^{3/2}} \int_\Omega f(x)\overline{U_\pm(x,k,\omega)}\, dx \\ + \frac{ik}{4\pi^{3/2}} \int_\Omega g(x)\overline{U_\pm(x,k,\omega)}\, dx.\end{aligned} \tag{4.37}$$

This is the same as the (formally computed) $\mathcal{E}$-inner product

$$\big(\langle f, g\rangle, \langle U_\pm(\cdot,k,\omega), ikU_\pm(\cdot,k,\omega)\rangle\big)_{\mathcal{E}}, \tag{4.38}$$

times $2^{-1/2}(2\pi)^{-3/2}$. Note that $e^{ikt}U_\pm(x,k,\omega)$ solves the wave equation, with Cauchy data $\langle U_\pm(x,k,\omega), ikU_\pm(x,k,\omega)\rangle$. In terms of the operators $\Phi_\pm$, studied before, we can write (4.37) as $1/\sqrt{2}$ times

$$\begin{aligned} &k^2(\Phi_\pm f)(k\omega) + ik(\Phi_\pm g)(k\omega), \qquad \text{for } k > 0, \\ &k^2(\Phi_\mp f)(k\omega) + ik(\Phi_\mp g)(k\omega), \qquad \text{for } k < 0. \end{aligned} \tag{4.39}$$

Note that $f \in \mathcal{H} \Leftrightarrow \Phi_\pm f \in L^2(\mathbb{R}^3, |\xi|^2 d\xi) \Leftrightarrow |\xi|^2\Phi_\pm f \in L^2(\mathbb{R}^3, |\xi|^{-2} d\xi)$, or, switching to polar coordinates,

$$f \in \mathcal{H} \Longleftrightarrow k^2(\Phi_\pm f)(k\omega) \in L^2(\mathbb{R}^+ \times S^2, dk\, d\omega). \tag{4.40}$$

Similarly,

$$g \in L^2(\Omega) \Longleftrightarrow k(\Phi_\pm g)(k\omega) \in L^2(\mathbb{R}^+ \times S^2, dk\, d\omega). \tag{4.41}$$

Therefore, for $\langle f, g\rangle \in \mathcal{E}$, the quantity (4.39) belongs to

$$L^2(\mathbb{R} \times S^2, dk\, d\omega) = L^2(\mathbb{R}, \mathcal{N}), \tag{4.42}$$

with

$$\mathcal{N} = L^2(S^2). \tag{4.43}$$

We can now establish the following.

Proposition 4.3. *For each choice of sign, $\Psi_\pm$ provides a unitary map of $\mathcal{E}$ onto $L^2(\mathcal{N})$.*

Proof. It is clear that the restrictions of $\Psi_\pm$ to $\mathcal{H} \oplus 0$ and to $0 \oplus L^2(\Omega)$ are both isometries, by the arguments leading to (4.40) and (4.41). Also, it is easy to see that the images of these spaces under $\Psi_\pm$ are mutually orthogonal, so $\Psi_\pm$ is an isometry of $\mathcal{E}$ into $L^2(\mathbb{R}, \mathcal{N})$. To show that it is surjective, we show how to solve for $\langle f, g\rangle \in \mathcal{E}$ the pair of equations

$$\Phi_\pm(Lf + ig) = u_0, \quad \Phi_\mp(Lf - ig) = u_1, \tag{4.44}$$

for arbitrary $u_0, u_1 \in L^2(\mathbb{R}^3)$. Inverting the unitary operators $\Phi_\pm$ and $\Phi_\mp$, we reduce this to a trivial system for $Lf + ig$ and $Lf - ig$, easily solved for $f \in \mathcal{H}$, $g \in L^2(\Omega)$, since $L : \mathcal{H} \to L^2(\Omega)$ is an isomorphism. This proves the proposition.

The maps $\Psi_\pm$ intertwine the evolution group $U(t)$ with a simple multiplication operator:

Proposition 4.4. *We have, for $\varphi \in L^2(\mathbb{R}, \mathcal{N})$,*

$$\Psi_\pm \, U(t) \, \Psi_\pm^{-1} \, \varphi(k, \omega) = e^{-ikt} \varphi(k, \omega). \tag{4.45}$$

Proof. This follows directly from the intertwining properties of $\Phi_\pm$, given (4.39) and the following computation:

$$\begin{aligned} & k^2\big(\Phi_\pm u(t)\big)(k\omega) + ik\big(\Phi_\pm u_t(t)\big)(k\omega) \\ &\quad = k^2\big[(\cos kt)\Phi_\pm f + (k^{-1}\sin kt)\Phi_\pm g\big] \\ &\qquad + ik\big[-k(\sin kt)\Phi_\pm f + (\cos kt)\Phi_\pm g\big] \\ &\quad = k^2 \, e^{-ikt} \, \Phi_\pm f + ik \, e^{-ikt} \, \Phi_\pm g, \end{aligned} \tag{4.46}$$

for $k > 0$, with a similar computation for $k < 0$.

The unitary maps discussed above are called "spectral representations" for $U(t)$. In §6 we will study related maps, called "translation representations." Note that in the case $K = \emptyset$, the functions $U_\pm(x, k, \omega)$ become $U_0(x, k, \omega) = e^{-ik\omega\cdot x}$, and both spectral representations coincide. We denote this free-space spectral representation by Ψ_0. It is a unitary map of $\mathcal{E}_0 = \mathcal{H}_0 \oplus L^2(\mathbb{R}^3)$ onto $L^2(\mathbb{R}, \mathcal{N})$, given in terms of the Fourier transform by

$$\Psi_0 \begin{pmatrix} f \\ g \end{pmatrix} (k, \omega) = \frac{k^2}{\sqrt{2}} \hat{f}(k\omega) + \frac{ik}{\sqrt{2}} \hat{g}(k\omega). \tag{4.47}$$

Here, $\mathcal{H}_0$ is the completion of $C_0^\infty(\mathbb{R}^3)$ with respect to the norm $\|\nabla f\|_{L^2(\mathbb{R}^3)}$, mapped unitarily by the Fourier transform onto $L^2(\mathbb{R}^3, |\xi|^2 d\xi)$.

Exercises

Let $\varphi \in \mathcal{S}(\mathbb{R})$ be an even function in the following exercises. Let $\Lambda = \sqrt{-\Delta}$, as in (4.5), and let $K_\varphi(x, y)$ be the Schwartz kernel of $\varphi(\Lambda)$, as in (2.28). Let Δ_0 be the free-space Laplacian on $\mathbb{R}^3$, $\Lambda_0 = \sqrt{-\Delta_0}$, and let $K^0_\varphi(x, y)$ be the Schwartz kernel of $\varphi(\Lambda_0)$, so, parallel to (2.28),

$$K^0_\varphi(x, y) = (2\pi)^{-3} \int\limits_{\mathbb{R}^3} e^{-i\xi\cdot(x-y)} \varphi(|\xi|)\, d\xi.$$

Let $D_\varphi(x, y) = K_\varphi(x, y) - K^0_\varphi(x, y)$, where $K_\varphi(x, y)$ is set equal to 0 if $x \in K$ or $y \in K$.

1. Use the formula

$$\varphi(\Lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{\varphi}(t) \cos t\Lambda \, dt \tag{4.48}$$

together with finite propagation speed to show that

$$\operatorname{supp} \hat{\varphi}(t) \subset \{|t| \le T\} \Longrightarrow \operatorname{supp} D_\varphi(x, y) \subset \{|x|, |y| \le R + T\}$$

if $K \subset B_R(0)$.

2. Use (4.48) to show that, for some $J = J(\alpha, \beta)$,

$$|D^\alpha_x D^\beta_y K_\varphi(x, y)| \le C\big[\|\hat{\varphi}\|_{L^1(\mathbb{R})} + \|D^J_t \hat{\varphi}\|_{L^1(\mathbb{R})}\big],$$

for $x, y \in \Omega$.

3. Use Exercises 1 and 2 to show that when $\varphi \in \mathcal{S}(\mathbb{R})$ is even, then $D_\varphi(x, y)$ is rapidly decreasing and is the Schwartz kernel of a trace class operator on $L^2(\mathbb{R}^3)$.
4. Let $\mathcal{H}^1(\mathbb{R}^n)$ denote the completion of $C^\infty_0(\mathbb{R}^n)$ with respect to the norm in (4.13). Show that if $n \ge 3$, there is a natural injective map

$$\iota : \mathcal{H}^1(\mathbb{R}^n) \longrightarrow \mathcal{S}'(\mathbb{R}^n)$$

and the Fourier transform maps $\mathcal{H}^1(\mathbb{R}^n)$ isomorphically onto

$$\mathcal{F}\mathcal{H}^1(\mathbb{R}^n) = \big\{u \in L^1_{\mathrm{loc}}(\mathbb{R}^n) : |\xi| u(\xi) \in L^2(\mathbb{R}^n)\big\} = L^2(\mathbb{R}^n, |\xi|^2 d\xi).$$

5. Show that, for $n \ge 3$,

$$L^2(\mathbb{R}^n, |\xi|^2 d\xi) \subset L^q_{\mathrm{loc}}(\mathbb{R}^n, d\xi),$$

provided $1 \le q < 2n/(n+2)$. Conclude that if $n \ge 3$, any $\hat{u} \in \mathcal{F}\mathcal{H}^1(\mathbb{R}^n)$ can be written as a sum of an element of $L^2(\mathbb{R}^n)$ and a compactly supported element of $L^q(\mathbb{R}^n)$, given $q \in [1, 2n/(n+2))$.
Show that $L^2(\mathbb{R}^2, |\xi|^2 d\xi)$ is not contained in $L^1_{\mathrm{loc}}(\mathbb{R}^2)$.
6. Let $\psi_\sigma(\xi)$ be the Fourier transform of $\langle x\rangle^{-\sigma}$. Show that if $q \in [1, 2)$, then

$$g \in L^q_{\mathrm{comp}}(\mathbb{R}^n) \Longrightarrow \psi_\sigma * g \in L^2(\mathbb{R}^n),$$

provided $\sigma \ge (2-q)n/2q$. (*Hint*: Interpolate between easy cases.)

7. Show that if $n \ge 3$ and $\sigma > 1$, then

$$\mathcal{H}^1(\mathbb{R}^n) \subset L^2(\mathbb{R}^n, \langle x\rangle^{-2\sigma} dx). \tag{4.49}$$

Note that this extends the estimate (4.15) in several ways.

8. Show that if $n \geq 3$,

$$\mathcal{H}^1(\mathbb{R}^n) \subset L^{2n/(n-2)}(\mathbb{R}^n). \tag{4.50}$$

Show that this result implies (4.49).
Reconstder this problem after reading §2 of Chap. 13.

5. Wave operators

In this section we examine the asymptotic behavior of the unitary group $U(t)$ on $\mathcal{E}$, as $t \to \pm\infty$. More precisely, we show that, as $t \to \pm\infty$, $U(t)M_\varphi U_0(-t)\langle f, g\rangle$ converges to a limit, $W_\pm\langle f, g\rangle$; the operators $W_\pm$ are called *wave operators*, and they are easily seen to be isometries from $\mathcal{E}_0$ into $\mathcal{E}$. Here, $\mathcal{E}$ is the space constructed in §4 for $\Omega = \mathbb{R}^3 \setminus K$, $\mathcal{E}_0$ that for the region $\Omega_0 = \mathbb{R}^3$, and $U_0(t)$ the "free-space" evolution operator for $\mathbb{R}^3$; M_φ is multiplication by a function $\varphi \in C^\infty(\mathbb{R}^3)$, equal to zero in a neighborhood of K, and equal to 1 outside a bounded set. We will show that $W_\pm$ have as right inverses operators $\Omega_\pm = \Psi_0^{-1}\Psi_\pm$, where $\Psi_\pm$ are the unitary operators constructed in §4; Ψ_0 is the corresponding operator constructed for $\Omega_0 = \mathbb{R}^3$. Since $\Omega_\pm$ are unitary, it will follow from this that the wave operators are also unitary.

We begin with the following observation, a simple consequence of Huygens' principle. Suppose f and g are in $C_0^\infty(\mathbb{R}^3)$, supported in $B_R = \{x \in \mathbb{R}^3 : |x| < R\}$. Then, for $|t| > R$,

$$U_0(t)\langle f, g\rangle = 0, \quad \text{for } |x| < |t| - R. \tag{5.1}$$

This follows directly for the formula for the fundamental solution to the wave equation on $\mathbb{R} \times \mathbb{R}^3$, which, recall from Chap. 3, is

$$R(t, x) = \frac{\delta(|x| - |t|)}{4\pi t}. \tag{5.2}$$

Consequently, if $K \subset B_R$ and if f and g are supported in B_{R_0}, then

$$U(s)U_0(-s)\langle f, g\rangle = U(R + R_0)U_0(-R - R_0)\langle f, g\rangle, \quad \text{for } s > R + R_0, \tag{5.3}$$

with a similar identity for $s < -R - R_0$. We can insert an M_φ between the two unitary factors on the left if $\varphi(x) = 1$ for $|x| \geq R$, without altering anything. It follows that

$$W_\pm\langle f, g\rangle = \lim_{t\to\pm\infty} U(-t)M_\varphi U_0(t)\langle f, g\rangle \tag{5.4}$$

exists, for $\langle f, g\rangle$ in the dense subset of $\mathcal{E}_0$ consisting of compactly supported functions. Consequently, the limits exist on all of $\mathcal{E}_0$, and the operators $W_\pm$, called wave operators, are isometries from $\mathcal{E}_0$ into $\mathcal{E}$. A major result, established below, is that these operators are actually unitary, from $\mathcal{E}_0$ *onto* $\mathcal{E}$.

In fact, consider the following operators:

$$\Omega_\pm = \Psi_0^{-1}\Psi_\pm : \mathcal{E} \longrightarrow \mathcal{E}_0. \tag{5.5}$$

By Proposition 4.3 we know $\Omega_\pm$ are unitary. We aim to establish the following result.

Proposition 5.1. *We have*

$$\Omega_+ W_+ = I \ \ \text{and} \ \ \Omega_- W_- = I \ \text{on}\ \mathcal{E}_0. \tag{5.6}$$

In order to prepare to prove this, we introduce the following set of initial data for the wave equation. If R is sufficiently large that $K \subset B_R$, set

$$\begin{aligned} D_0^+(R) = \big\{\langle f, g\rangle \in C_0^\infty(\mathbb{R}^3) \oplus C_0^\infty(\mathbb{R}^3) : U_0(t)\langle f, g\rangle = 0, \\ \text{for } t > 0, |x| < R + t\big\}. \end{aligned} \tag{5.7}$$

In particular, f and g vanish near K, and we can regard $\langle f, g\rangle$ as an element of $\mathcal{E}_0$ or of $\mathcal{E}$, and

$$\langle f, g\rangle \in D_0^+(R) \Longrightarrow U_0(t)\langle f, g\rangle = U(t)\langle f, g\rangle, \ \ \text{for } t > 0. \tag{5.8}$$

Clearly,

$$U_0(t)D_0^+(R) \subset D_0^+(R), \ \ \text{for } t > 0, \tag{5.9}$$

though not for $t < 0$. Also, by the argument involving Huygens' principle discussed above, it is clear that

$$\bigcup_{t<0} U_0(t)D_0^+(R) \text{ is dense in } \mathcal{E}_0. \tag{5.10}$$

Note that (5.8) implies

$$W_+ = I \quad \text{on } D_0^+(R). \tag{5.11}$$

Our first step in establishing Proposition 5.1 is the following.

Lemma 5.2. *We have*

$$\Omega_+ = I \quad \text{on } D_0^+(R). \tag{5.12}$$

Proof. This is equivalent to the identity

$$\Psi_+ = \Psi_0 \quad \text{on } D_0^+(R), \tag{5.13}$$

which in turn follows from the identity

$$\big(\langle f,g\rangle, \langle V_+(\cdot,k,\omega), ikV_+(\cdot,k,\omega)\rangle\big)_{\mathcal{E}} = 0, \tag{5.14}$$

for $k \in \mathbb{R}, \omega \in S^2, \langle f,g\rangle \in D_0^+(R)$. (Here V_+ is the function defined in (1.51).) Note that the left side of (5.14) is equal to e^{ikt} times

$$\big(U_0(t)\langle f,g\rangle, \langle V_+, ikV_+\rangle\big)_{\mathcal{E}}, \tag{5.15}$$

for any $t > 0$. We will show that, for t large, this can be dominated by a small quantity. Indeed, an examination of $\langle u(t), u_t(t)\rangle = U_0(t)\langle f,g\rangle$ via the formula (5.2) for the Riemann function shows that, for t large and positive, $\nabla_x u(t,x)$ is approximately radial, and $u_t(t,x) \sim u_r(t,x)$. Thus (5.15) is equal to

$$\int_\Omega \Big[u_r(t,x)\frac{\partial \overline{V}_+}{\partial r} + (ik)u_t(t,x)\overline{V}_+\Big]\, dx + o(1), \tag{5.16}$$

as $t \to +\infty$. In light of the radiation condition for V_+, the two terms in this integral cancel out, up to a remainder that vanishes as $t \to +\infty$; this proves the lemma.

In view of (5.11), we now know that

$$\Omega_+ W_+ = I \quad \text{on } D_0^+(R). \tag{5.17}$$

Now it follows easily from the definition that

$$W_\pm U_0(t) = U(t)W_\pm, \quad \text{for all } t, \tag{5.18}$$

and from Proposition 4.4 it follows that

$$\Omega_\pm U(t) = U_0(t)\Omega_\pm, \quad \text{for all } t. \tag{5.19}$$

Given that (5.17) holds when applied to $U_0(t)\langle f,g\rangle$, provided this belongs to $D_0^+(R)$, we deduce that

$$\Omega_+ W_+ \langle f,g\rangle = \langle f,g\rangle, \quad \text{for } \langle f,g\rangle \in U_0(-t)D_0^+(R),\ t > 0; \tag{5.20}$$

in other words,

$$\Omega_+ W_+ = I \quad \text{on } \bigcup_{t>0} U_0(-t)D_0^+(R). \tag{5.21}$$

In light of (5.10), this implies that $\Omega_+ W_+ = I$ on $\mathcal{E}_0$, establishing the first identity in (5.6). The second identity is proved in the same fashion, and Proposition 5.1 is done.

The unitarity of $\Omega_\pm$ then gives the following result, known as the completeness of the wave operators.

Corollary 5.3. *The wave operators $W_\pm$ are unitary from $\mathcal{E}_0$ onto $\mathcal{E}$. We have the identities*

$$W_\pm = \Psi_\pm^{-1}\Psi_0. \tag{5.22}$$

Note that (5.6) implies the surjectivity of $\Omega_\pm$, hence of $\Psi_\pm$, since the invertibility of Ψ_0 is obvious (just the Fourier inversion formula). Thus the proof of Proposition 5.1 contains an alternative proof of Proposition 4.3, and hence of Proposition 2.7.

The operator

$$\mathcal{S}_1 = W_+^{-1}W_- = \Psi_0^{-1}\big(\Psi_+\Psi_-^{-1}\big)\Psi_0, \tag{5.23}$$

a unitary operator on $\mathcal{E}_0$, is often called the scattering operator. In view of the simple nature of Ψ_0, it is equally convenient to call the unitary operator on $L^2(\mathbb{R},\mathcal{N})$:

$$\mathcal{S} = \Psi_+\Psi_-^{-1}, \tag{5.24}$$

also a scattering operator. Note that, if we make the identification

$$L^2(\mathbb{R},\mathcal{N}) = L^2(\mathbb{R}^+\times S^2)\oplus L^2(\mathbb{R}^-\times S^2),$$

and follow with the natural unitary map $L^2(\mathbb{R}^\pm\times S^2)\to L^2(\mathbb{R}^3)$ involving polar coordinates, we can write

$$\mathcal{S} = \begin{pmatrix} \Phi_+\Phi_-^{-1} & 0 \\ 0 & \Phi_-\Phi_+^{-1} \end{pmatrix}. \tag{5.25}$$

The operator $S = \Phi_+\Phi_-^{-1}$ is the scattering operator studied in §3; the other operator, $\Phi_-\Phi_+^{-1} = S_-$, appears in Exercise 2 of §3.

Another consequence of the unitarity of the wave operators is the following nontrivial variant of (5.10).

Proposition 5.4. *Pick R so that $K\subset B_R$. Then*

$$\bigcup_{t<0} U(t)D_0^+(R) \text{ is dense in } \mathcal{E}. \tag{5.26}$$

Proof. Any $\langle f,g\rangle\in\mathcal{E}$ can be written in the form

$$\langle f,g\rangle = W_+\langle f_0,g_0\rangle, \tag{5.27}$$

with $\langle f_0, g_0\rangle \in \mathcal{E}_0$. Approximate $\langle f_0, g_0\rangle$ to within ε in the $\mathcal{E}_0$-norm by $\langle f_1, g_1\rangle \in C_0^\infty(\mathbb{R}^3) \oplus C_0^\infty(\mathbb{R}^3)$. Then, for all $t \geq R + R_0$ sufficiently large,

$$W_+\langle f_1, g_1\rangle = U(-t)U_0(t)\langle f_1, g_1\rangle, \tag{5.28}$$

by (5.3) and (5.4), and by the Huygens principle argument given there, for any such $t = t_0$,

$$U_0(t_0)\langle f_1, g_1\rangle = \langle f_2, g_2\rangle \in D_0^+(R). \tag{5.29}$$

Since $\langle f, g\rangle - U(-t_0)\langle f_2, g_2\rangle$ has $\mathcal{E}$-norm less than ε, the proposition is proved.

We can also produce a formula for $W_\pm^{-1}$ of a form similar to (5.4) but not involving an arbitrary choice, for example, of M_φ. Note that there is a natural isometric mapping

$$J : \mathcal{E} \longrightarrow \mathcal{E}_0 \tag{5.30}$$

defined on $\langle f, g\rangle$ by extending these functions to be zero on K. We have tacitly used this before. We now establish the following.

Proposition 5.5. *For any* $\langle f, g\rangle \in \mathcal{E}$,

$$W_\pm^{-1}\langle f, g\rangle = \lim_{t\to\pm\infty} U_0(-t)JU(t)\langle f, g\rangle. \tag{5.31}$$

Proof. For simplicity we analyze W_+^{-1}. By (5.28), for $\langle f_1, g_1\rangle \in C_0^\infty(\mathbb{R}^3) \oplus C_0^\infty(\mathbb{R}^3)$ supported in B_{R_0}, we have

$$JU(t)W_+\langle f_1, g_1\rangle = U_0(t)\langle f_1, g_1\rangle, \tag{5.32}$$

for all $t > R + R_0$. This is equivalent to

$$U_0(-t)JU(t)\langle f_3, g_3\rangle = W_+^{-1}\langle f_3, g_3\rangle, \tag{5.33}$$

for $\langle f_3, g_3\rangle = W_+\langle f_1, g_1\rangle,\ t > R + R_0$. This gives (5.31) on a dense subset of $\mathcal{E}$, hence on all of $\mathcal{E}$ in view of the uniform boundedness of $U(t)$ and $U_0(t)$.

Exercises

The following exercises deal with the existence and completeness of Schrödinger wave operators:

$$W_\pm f = \lim_{t\to\pm\infty} e^{itH}e^{-itH_0}f, \tag{5.34}$$

where $H_0 = -\Delta,\ H = -\Delta + V$, acting on functions on $\mathbb{R}^n$.

1. Show that $W_\pm \in \mathcal{L}(L^2(\mathbb{R}^n))$ exists provided that, for each $f \in C_0^\infty(\mathbb{R}^n)$,

$$\int_0^\infty \|V\, e^{-itH_0} f\|\, dt < \infty. \tag{5.35}$$

   (*Hint*: $e^{itH} e^{-itH_0} f = \int_0^t e^{isH} V e^{-isH_0} f\, ds$.) Note that when $W_\pm$ exist, they are isometries (i.e., $\|W_\pm f\| = \|f\|$ for all $f \in L^2(\mathbb{R}^n)$).
2. Show that $f \in C_0^\infty(\mathbb{R}^n)$ implies $\|e^{itH_0} f\|_{L^\infty} \le C\langle t\rangle^{-n/2}$. Deduce that $W_\pm$ exists if $V \in L^2(\mathbb{R}^n)$. (*Hint*: $e^{it\Delta}\delta(x) = (4\pi i t)^{-n/2} e^{-|x|^2/4it}$.)
3. Show that if $q = 2/(1-\theta) \in [2,\infty)$, then $f \in C_0^\infty(\mathbb{R}^n)$ implies

$$\|e^{itH_0} f\|_{L^q(\mathbb{R}^n)} \le C\langle t\rangle^{-n\theta/2}.$$

   Deduce that $W_\pm$ exists if $V \in L^r(\mathbb{R}^n)$, with $r < n$. In particular, $W_\pm$ exists provided $|V(x)| \le C\langle x\rangle^{-\sigma}$, $\sigma > 1$.
4. Show that, for any $f, g \in L^2(\mathbb{R}^n)$, $(g, e^{-itH_0} f) \to 0$ as $|t| \to \infty$. Use this to show that if g_j is an eigenfunction of H, then $(e^{-itH} g_j, e^{itH_0} f) \to 0$ as $|t| \to \infty$, for all $f \in L^2(\mathbb{R}^n)$. Hence, for $W_\pm$ given by (5.34), $\mathcal{R}(W_\pm) \subset \mathcal{H}_c$.
5. Suppose $V \in C_0^\infty(\mathbb{R}^3)$, so we have $\Phi_\pm$ by Exercise 5 of §2. Let Φ_0 be the inverse Fourier transform. Show that $\Phi_\pm W_\pm \Phi_0^{-1}$ commutes with multiplication by $e^{is|\xi|^2}$, for all $s \in \mathbb{R}$. Hence it commutes with $\varphi(|\xi|)$ for all $\varphi \in C_o(\mathbb{R})$. (*Hint*: $W_\pm = e^{isH} W_\pm e^{-isH_0}$.)
6. When the conditions of Exercise 1 hold, show that

$$W_\pm f = \lim_{\varepsilon\searrow 0} \int_0^{\mp\infty} \varepsilon e^{\pm\varepsilon t} e^{itH} e^{-itH_0} f\, dt$$

   and hence that

$$\big((W_\pm - I) f, g\big) = \lim_{\varepsilon\searrow 0} \int_0^{\mp\infty} i\big(e^{itH} V e^{-itH_0} f, g\big) e^{\pm\varepsilon t}\, dt. \tag{5.36}$$

7. Choosing the $+$ sign, show that the integral on the right side of (5.36) is equal to

$$\begin{aligned}&\int_0^{-\infty} \iint i\,\overline{(\Phi_+ g)(\xi)}\, V(x) \Big[e^{-it(H_0 - |\xi|^2 + i\varepsilon)} f(x)\Big] \overline{u_+(x,\xi)}\, dx\, d\xi\, dt \\ &= \iint \overline{(\Phi_+ g)(\xi)}\, V(x) \Big[(H_0 - |\xi|^2 + i\varepsilon)^{-1} f(x)\Big] \overline{u_+(x,\xi)}\, dx\, d\xi.\end{aligned} \tag{5.37}$$

   (*Hint*: Use Φ_+ to intertwine e^{itH} with e^{itH_0}.)
8. If $V \in C_0^\infty(\mathbb{R}^3)$, show that the limit of (5.37) as $\varepsilon \searrow 0$ is equal to

$$\frac{1}{4\pi} \iiint \overline{(\Phi_+ g)(\xi)}\, V(x)\, \frac{e^{-ik|x-y|}}{|x-y|}\, f(y) \overline{u_+(x,\xi)}\, dy\, dx\, d\xi,$$

   provided $f \in C_0^\infty(\mathbb{R}^3)$ and $(\Phi_+ g)(\xi)$ is supported on $|\xi| \in [a,b] \subset\subset (0,\infty)$. Here $k = |\xi|$. Using (1.58), write this as

$$-\iint \overline{(\Phi_+ g)(\xi)}\; \overline{v_+(y,\xi)}\; f(y)\, dy\, d\xi = -(\Phi_+ f, \Phi_+ g) + (\Phi_0 f, \Phi_+ g).$$

9. Using the previous exercises, show that, given $V \in C_0^\infty(\mathbb{R}^3)$ (real-valued), we have $(W_\pm f, g) = (\Phi_0 f, \Phi_\pm g)$ for all $f, g \in L^2(\mathbb{R}^3)$, hence

$$W_\pm = \Phi_\pm^{-1}\Phi_0.$$

Deduce the completeness of the wave operators: $\mathcal{R}(W_\pm) = \mathcal{H}_c$.

Compare arguments in Chap. 5 of [Si], dealing with a larger class of potentials. Completeness for a nearly maximal class of potentials to which Exercise 3 applies is treated in Chap. 13 of [RS]. Long-range potentials are treated in Chap. 3 of [Ho].

6. Translation representations and the Lax–Phillips semigroup *Z(t)*

From the "spectral representations" $\Psi_\pm : \mathcal{E} \to L^2(\mathbb{R}, \mathcal{N})$ defined in §4, which, as shown in Proposition 4.4, intertwine $U(t)$ with multiplication by e^{-ikt}, we construct "translation representations," unitary operators

$$\mathcal{T}_\pm : \mathcal{E} \longrightarrow L^2(\mathbb{R}, \mathcal{N}), \tag{6.1}$$

by taking the Fourier transform with respect to k:

$$\mathcal{T}_\pm \begin{pmatrix} f \\ g \end{pmatrix}(s, \omega) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{iks}\, \Psi_\pm \begin{pmatrix} f \\ g \end{pmatrix}(k, \omega)\, dk. \tag{6.2}$$

Consequently, Proposition 4.4 implies

$$\mathcal{T}_\pm U(t) \mathcal{T}_\pm^{-1} f(s, \omega) = f(s - t, \omega). \tag{6.3}$$

The operators $\mathcal{T}_\pm$ are useful for exposing various features of $U(t)$, and we explore this in the current section. We begin with a look at the free-space translation representation $\mathcal{T}_0$, a unitary map from $\mathcal{E}_0$ onto $L^2(\mathbb{R}, \mathcal{N})$ given by using Ψ_0 in (6.2).

We can produce an explicit formula for $\mathcal{T}_0$ using the formula (4.47) for Ψ_0, which we recall is

$$2^{1/2}\, \Psi_0 \begin{pmatrix} f \\ g \end{pmatrix}(k, \omega) = k^2 \hat{f}(k\omega) + ik\hat{g}(k\omega). \tag{6.4}$$

The formula for $\mathcal{T}_0$ is expressed naturally in terms of the Radon transform, which is defined (initially for $f \in \mathcal{S}(\mathbb{R}^3)$) by

$$\mathcal{R}f(s, \omega) = \int_{y\cdot\omega = s} f(y)\, dS(y), \tag{6.5}$$

for $s \in \mathbb{R},\ \omega \in S^2$. Note that the Fourier transform can be expressed as

$$\hat{f}(k\omega) = (2\pi)^{-3/2} \int_{-\infty}^{\infty} e^{-iks} \mathcal{R} f(s,\omega)\, ds. \tag{6.6}$$

Thus, taking the inverse Fourier transform in k, we have

$$\mathcal{R} f(s,\omega) = (2\pi)^{1/2} \int_{-\infty}^{\infty} e^{iks} \hat{f}(k\omega)\, dk. \tag{6.7}$$

In light of this, we see that taking the Fourier transform with respect to k of (6.4) gives

$$\mathcal{T}_0 \begin{pmatrix} f \\ g \end{pmatrix} (s,\omega) = \frac{1}{4\pi} \big[-\partial_s^2 \mathcal{R} f(s,\omega) + \partial_s \mathcal{R} g(s,\omega) \big]. \tag{6.8}$$

The unitarity of $\mathcal{T}_0$ gives rise to the inversion formula

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{S^2} k(x\cdot\omega,\omega)\, d\omega, \\ g(x) &= -\frac{1}{2\pi} \int_{S^2} \partial_s k(x\cdot\omega,\omega)\, d\omega, \end{aligned} \tag{6.9}$$

for $\langle f, g\rangle$ in terms of

$$k(s,\omega) = \mathcal{T}_0 \begin{pmatrix} f \\ g \end{pmatrix} (s,\omega). \tag{6.10}$$

This result is related to the Radon inversion formula,

$$f(x) = \frac{1}{8\pi^2} \int_{S^2} \partial_s^2 \mathcal{R} f(x\cdot\omega,\omega)\, d\omega, \tag{6.11}$$

which can be deduced from (6.9), or directly from (6.6) and the Fourier inversion formula.

In view of (6.3), for $\mathcal{T}_0$, we see that the solution to the free-space wave equation $u_{tt} - \Delta u = 0$ with initial data $\langle f, g\rangle$ can be written as

$$u(t,x) = \frac{1}{2\pi} \int_{S^2} k(x\cdot\omega - t,\omega)\, d\omega, \tag{6.12}$$

where $k(s,\omega)$ is given by (6.10). More fully, by (6.8),

$$u(t,x) = \frac{1}{8\pi^2} \int_{S^2} \big[-\partial_s^2 \mathcal{R} f(x\cdot\omega - t,\omega) + \partial_s \mathcal{R} g(x\cdot\omega - t,\omega) \big]\, d\omega. \tag{6.13}$$

Note that if f and g are supported in $B_{R_0} = \{|x| < R_0\}$, then, by (6.5), $\mathcal{R}f(s,\omega)$ and $\mathcal{R}g(s,\omega)$ vanish for $|s| > R_0$. Therefore, $\mathcal{R}f(x\cdot\omega - t, \omega)$ and $\mathcal{R}g(x\cdot\omega - t, \omega)$ vanish for $|x| < |t| - R_0$. Thus from (6.13) we rederive the Huygens principle, that $u(t,x)$ vanishes for $|x| < |t| - R_0$ in this case.

Use of $\mathcal{T}_0$ and $\mathcal{T}_\pm$ will augment arguments involving the Huygens principle made in §5. We introduce the space

$$D^+(R) = \{\langle f, g\rangle \in \mathcal{E}_0 : U_0(t)\langle f, g\rangle = 0 \text{ for } t > 0, |x| < R + t\}. \tag{6.14}$$

Note that $D_0^+(R)$, defined by (5.7), consists of the elements of $D^+(R)$ that are smooth and compactly supported. Similarly, set

$$D^-(R) = \{\langle f, g\rangle \in \mathcal{E}_0 : U_0(t)\langle f, g\rangle = 0 \text{ for } t < 0, |x| < R + |t|\}. \tag{6.15}$$

For $R = 0$, we denote these spaces simply by D^+ and D^-, respectively. From (6.12) it is clear that if $\mathcal{T}_0\langle f, g\rangle(s,\omega)$ is supported in $s \geq R$ (resp., $s \leq -R$), then $\langle f, g\rangle$ belongs to $D^+(R)$ (resp., $D^-(R)$). Furthermore, the converse result is true:

Proposition 6.1. *The transformation $\mathcal{T}_0 : \mathcal{E}_0 \to L^2(\mathbb{R}, \mathcal{N})$ maps $D^+(R)$ (resp. $D^-(R)$) onto the space of functions in $L^2(\mathbb{R}, \mathcal{N})$ supported in $[R, \infty)$ (resp., supported in $(-\infty, -R]$), for any $R \geq 0$. In particular, D^+ and D^- are orthogonal complements of each other in $\mathcal{E}_0$.*

In order to prove this proposition, it suffices to demonstrate that if $\langle f, g\rangle \in \mathcal{E}_0$ belongs to D^+, then $k(s,\omega) = \mathcal{T}_0\langle f, g\rangle$ vanishes for $s < 0$. This comes down to showing that, if $k \in L^2(\mathbb{R}, \mathcal{N})$ and if the integral (6.12) vanishes for $t > |x|$, then $k(s,\omega) = 0$ for $s < 0$. Applying a mollifier, we can suppose $k \in C^\infty(\mathbb{R}, \mathcal{N})$. Since $\mathcal{T}_0$ clearly commutes with rotations, it suffices to prove this for $k(s,\omega)$ of the form $k(s,\omega) = K(s)\varphi(\omega)$, where φ is an eigenfunction of the Laplace operator on S^2. So suppose

$$u(t,x) = \frac{1}{2\pi}\int_{S^2} K(x\cdot\omega - t)\varphi(\omega)\, d\omega \tag{6.16}$$

vanishes for $t > |x|$. Since this implies $D_x^\alpha u(t,0) = 0$ for $t > 0$, for all α, we have

$$0 = \partial_t^{|\alpha|} K(-t) \int_{S^2} \omega^\alpha \varphi(\omega)\, d\omega, \quad t > 0, \tag{6.17}$$

for all α. Since, by the Stone-Weierstrass theorem, $\{\omega^\alpha\}$ has dense linear span in $C(S^2)$, there exists α such that the integral in (6.17) is nonvanishing. This implies that $\partial_t^{|\alpha|} K(-t) = 0$ for $t > 0$, so $K(t)$ coincides with a polynomial in t for $t < 0$. Since $K \in L^2(\mathbb{R})$, this implies $K(t) = 0$ for $t < 0$, and the proposition is proved.

Now we look at the maps $\mathcal{T}_\pm : \mathcal{E} \to L^2(\mathbb{R}, \mathcal{N})$, in the presence of an obstacle K, which we suppose is contained in a ball B_R. Note that $D^\pm(R)$ can be regarded as subspaces both of $\mathcal{E}_0$ and of $\mathcal{E}$. Lemma 5.2 (specifically (5.13)), which was important in the last section, immediately implies the following.

Proposition 6.2. *We have*

$$\mathcal{T}_+ = \mathcal{T}_0 \text{ on } D_0^+(R) \text{ and } \mathcal{T}_- = \mathcal{T}_0 \text{ on } D_0^-(R). \tag{6.18}$$

The potential usefulness of this is indicated by the next result.

Proposition 6.3. *The properties* (6.3) *and* (6.18) *uniquely characterize* $\mathcal{T}_+$ *and* $\mathcal{T}_-$ *as continuous linear maps.*

Proof. Equations (6.3) and (6.18) specify $\mathcal{T}_+$ on $U(t)D_0^+(R)$ for all $t \in \mathbb{R}$. By Proposition 5.4, the union of these spaces is dense in $\mathcal{E}$, so the result follows for $\mathcal{T}_+$. The proof for $\mathcal{T}_-$ is similar.

Note that we can set $\widetilde{\mathcal{T}}_\pm = \mathcal{T}_0$ on $D^\pm(R)$ and since $U(t) = U_0(t)$, for $t \geq 0$ on $D^+(R)$ and for $t \leq 0$ on $D^-(R)$, we can extend $\widetilde{\mathcal{T}}_\pm$ so that (6.3) holds. The uniqueness result above then implies $\widetilde{\mathcal{T}}_\pm = \mathcal{T}_\pm$, so we have

$$\mathcal{T}_+ = \mathcal{T}_0 \text{ on } D^+(R) \text{ and } \mathcal{T}_- = \mathcal{T}_0 \text{ on } D^-(R), \tag{6.19}$$

sharpening (6.18).

If we use the translation representations $\mathcal{T}_\pm$ in place of the spectral representations $\Psi_\pm$, the scattering operator $\mathcal{S}$ defined by (5.24) is replaced by the unitary operator on $L^2(\mathbb{R}, \mathcal{N})$:

$$\hat{\mathcal{S}} = \mathcal{T}_+ \mathcal{T}_-^{-1}. \tag{6.20}$$

The operator $\hat{\mathcal{S}}$ clearly commutes with translations. It also possesses the following important property.

Proposition 6.4. *We have*

$$\hat{\mathcal{S}} : L^2\big((-\infty, -R], \mathcal{N}\big) \longrightarrow L^2\big((-\infty, R], \mathcal{N}\big). \tag{6.21}$$

Proof. $\mathcal{T}_-^{-1}$ maps $L^2((-\infty, -R], \mathcal{N})$ onto $D^-(R)$, which is orthogonal to $D^+(R)$, as a consequence of Proposition 6.1. Since $\mathcal{T}_+$ maps $D^+(R)$ onto $L^2([R, \infty), \mathcal{N})$ and is unitary, it must map $D^-(R)$ into the orthogonal complement of $L^2([R, \infty), \mathcal{N})$; this proves (6.21).

Now the action of $\mathcal{S}$ on $L^2(\mathbb{R}, \mathcal{N})$ is given by multiplication by a unitary operator-valued function $\mathcal{S}(k)$, similar to the action of S in terms of $S(k)$ discussed in §3. The action of $\hat{\mathcal{S}}$ on $L^2(\mathbb{R}, \mathcal{N})$ is then given by convolution by an operator-valued tempered distribution $\hat{\mathcal{S}}(s)$, the Fourier transform of $\mathcal{S}(k)$. From

(6.21) we conclude that $\hat{\mathcal{S}}(s)$ is supported in the half-line $(-\infty, 2R]$. It follows that $\mathcal{S}(k)$ extends to be a holomorphic, operator-valued function in the half-space $\operatorname{Im} k > 0$, a fact that can also be seen directly from an analysis of the scattering amplitude $a(\omega, \theta, k)$, in view of the relation established in §3. We will study the meromorphic continuation of these objects into the lower half-plane in §7.

We now look at a semigroup of operators, introduced by P. Lax and R. Phillips, defined as follows. Fixing R such that $K \subset B_R$, set

$$\mathcal{K} = \big(D^+(R) \oplus D^-(R)\big)^\perp, \tag{6.22}$$

the orthogonal complement in $\mathcal{E}$. For $t \geq 0$, define

$$Z(t) = P_\mathcal{K} U(t) P_\mathcal{K}, \tag{6.23}$$

where $P_\mathcal{K}$ is the orthogonal projection of $\mathcal{E}$ onto $\mathcal{K}$.

Proposition 6.5. *$Z(t)$ is a strongly continuous semigroup of operators on $\mathcal{K}$, so*

$$Z(t+s) = Z(t)Z(s), \quad \textit{for } t, s \geq 0. \tag{6.24}$$

Proof. If $\langle f_j, g_j\rangle \in \mathcal{K}$, then $U(t)\langle f_1, g_1\rangle \in D^+(R)$ for $t \geq 0$, and furthermore $U(-s)\langle f_2, g_2\rangle \in \big(D^-(R)\big)^\perp$ for $s \geq 0$. Hence, for $s, t \geq 0$,

$$\big(U(-s)\langle f_2, g_2\rangle, P_\mathcal{K} U(t)\langle f_1, g_1\rangle\big)_\mathcal{E} = \big(U(-s)\langle f_2, g_2\rangle, U(t)\langle f_1, g_1\rangle\big)_\mathcal{E}. \tag{6.25}$$

Thus $P_\mathcal{K} U(s) P_\mathcal{K} U(t) P_\mathcal{K} = P_\mathcal{K} U(s+t) P_\mathcal{K}$, which implies (6.24). The strong continuity is obvious.

We note that the Lax–Phillips semigroup $Z(t)$ can also be expressed as

$$Z(t) = P_+ U(t) P_- \quad (t \geq 0), \tag{6.26}$$

where $P_\pm$ is the orthogonal projection of $\mathcal{E}$ onto $\big(D^\pm(R)\big)^\perp$. To see this, note that

$$P_\mathcal{K} = P_+ P_- = P_- P_+.$$

Since $U(t)$ leaves $D^+(R)$ invariant, $P_+ U(t) P_+ = P_+ U(t)$, for $t \geq 0$. Similarly, $P_- U(t) P_- = U(t) P_-$, for $t \geq 0$, so

$$\begin{aligned} P_\mathcal{K} U(t) P_\mathcal{K} &= P_- P_+ U(t) P_+ P_- \\ &= P_- P_+ U(t) P_- \\ &= P_+ P_- U(t) P_- \\ &= P_+ U(t) P_-. \end{aligned} \tag{6.27}$$

Since $Z(t)$ is a strongly continuous semigroup on $\mathcal{K}$, it has a generator C, whose resolvent is given by

$$(\lambda - C)^{-1} = \int_0^\infty e^{-\lambda t} Z(t)\, dt, \quad \operatorname{Re} \lambda > 0. \tag{6.28}$$

The following result gives important spectral information on $Z(t)$.

Proposition 6.6. *For any* $T \geq 2R,\ \lambda > 0$,

$$(\lambda - C)^{-1} Z(T) \text{ is compact.} \tag{6.29}$$

We can derive this from the following result, of independent interest. Given $\rho \in C_0^\infty(\mathbb{R}^+)$, let

$$Z(\rho) = \int_0^\infty \rho(t) Z(t)\, dt. \tag{6.30}$$

Define $U(\rho)$ and $U_0(\rho)$ similarly.

Proposition 6.7. *If* $\rho \in C_0^\infty\big((2R, \infty)\big)$, *then*

$$Z(\rho) = P_+ \big[U(\rho) - U_0(\rho)\big] P_-. \tag{6.31}$$

Proof. Since it is easy to see that

$$P_+ U_0(t) P_- = 0, \quad \text{for } t \geq 2R, \tag{6.32}$$

this is clear from the formula (6.26).

Now to prove (6.29), it suffices to show that $Z(\rho)$ is compact for any $\rho \in C_0^\infty\big((2R, \infty)\big)$, since the operator (6.29) is equal to $\int_0^\infty e^{-\lambda t} Z(t+T)\, dt$, which is a norm limit of such $Z(\rho)$. We show that, for such ρ, $U(\rho) - U_0(\rho)$ is compact, from $\mathcal{E}$ to $\mathcal{E}_0$. Indeed, if ρ is supported in $[2R, T]$, then, by finite propagation speed,

$$\big[U(\rho) - U_0(\rho)\big]\langle f, g\rangle \text{ is supported in } |x| \leq 2R + T, \tag{6.33}$$

for any $\langle f, g\rangle \in \mathcal{E}$. Also we have, for such ρ, by integrating by parts, and elliptic regularity,

$$U(\rho) : \mathcal{E} \to C^\infty(\overline{\Omega}), \quad U_0(\rho) : \mathcal{E} \to C^\infty(\mathbb{R}^3). \tag{6.34}$$

The compactness of $U(\rho) - U_0(\rho)$ then follows, by Rellich's theorem. We note that complementing (6.33), we also have, for any $\langle f, g\rangle \in \mathcal{E}$,

$$\big[U(\rho) - U_0(\rho)\big]\langle f, g\rangle \text{ depends only on } \langle f, g\rangle\big|_{B_{R+T}}. \tag{6.35}$$

For any nonzero $\alpha \in \mathbb{C}$ in the spectrum of the operator (6.29) (for a fixed $\lambda > 0,\ T \geq 2R$), this compact operator has an associated finite-dimensional, generalized α-eigenspace V_α. $Z(t)$ clearly preserves V_α, for $t \geq 0$, and the spectrum of $Z(t)\big|_{V_\alpha}$ consists of $e^{\mu_j t}$, where, for each such $\alpha,\ \mu_j$ is a finite set of complex numbers, each satisfying

$$(\lambda - \mu_j)^{-1}\, e^{\mu_j t} = \alpha.$$

We call the set of all such μ_j, as α ranges over the nonzero elements of the spectrum of (6.29), scattering characters. It is a fact that this set coincides precisely with the spectrum of the generator C of $Z(t)$, but we will not make explicit use of this and we do not include a proof. (See [LP1].) By the analysis above, the set of scattering characters μ_j can be characterized as follows:

$$\text{point spec } Z(t) = \{e^{\mu_j t} : \mu_j \text{ scattering character}\}. \tag{6.36}$$

In §7 we relate the set of scattering characters to the set of scattering poles.

We end this section with some comments on the semigroup $Z(t)$ in the translation representation, that is, we look at

$$Z_+(t) = \mathcal{T}_+ Z(t) \mathcal{T}_+^{-1}, \tag{6.37}$$

acting on $\mathcal{K}_+ \subset L^2(\mathbb{R}, \mathcal{N})$, where

$$\mathcal{K}_+ = \mathcal{T}_+(\mathcal{K}). \tag{6.38}$$

Note that $\langle f, g\rangle$ belongs to $\mathcal{K}$ if and only if

$$\text{supp } \mathcal{T}_+\langle f, g\rangle \subset (-\infty, R] \ \text{ and supp } \ \mathcal{T}_-\langle f, g\rangle \subset [-R, \infty), \tag{6.39}$$

in view of Proposition 6.1 and (6.19). Recalling the scattering operator $\hat{\mathcal{S}}$, given by (6.20), we see that

$$\mathcal{K}_+ = \big\{ f \in L^2\big((-\infty, R], \mathcal{N}\big) : \hat{\mathcal{S}}^{-1} f \in L^2\big([-R, \infty), \mathcal{N}\big)\big\}. \tag{6.40}$$

By (6.37) and (6.3) we have, for $f \in \mathcal{K}_+$,

$$\begin{aligned} Z_+(t) f(s, \omega) = f(s - t, \omega), &\quad \text{for } s \leq R, \\ 0, &\quad \text{for } s \geq R. \end{aligned} \tag{6.41}$$

Exercises

1. Prove the Radon inversion formula (6.11) from the definition (6.5) and the Fourier inversion formula.

2. Consider a first-order, constant-coefficient PDE

$$\frac{\partial u}{\partial t} = A(D_x)u, \quad u(0,x) = f(x),$$

where $A(D_x)$ is an $\ell \times \ell$ matrix. Assume the principal symbol $A_1(\xi)$ has ℓ distinct imaginary roots for $\xi \in \mathbb{R}^3 \setminus 0$. Express the solution in terms of the Radon transform. When can you deduce Huygens' principle?

7. Integral equations and scattering poles

In §1 we established results on the existence and uniqueness of solutions to the scattering problem

$$\begin{gathered}(\Delta + k^2)v = 0 \text{ on } \Omega, \quad v = f \text{ on } \partial K, \\ r\Big(\frac{\partial v}{\partial r} - ikv\Big) \longrightarrow 0, \ \text{ as } r \to \infty.\end{gathered} \tag{7.1}$$

As in (1.19), let us denote the solution operator to (7.1) by

$$v = \mathcal{B}(k)f. \tag{7.2}$$

We established the proof that $\mathcal{B}(k)$ is uniquely defined, for $k \in \mathbb{R}$, via the limiting absorption principle in §1; related is the elementary fact that such a solution operator is also uniquely defined for complex k such that $\operatorname{Im} k > 0$, since k^2 belongs to the resolvent set for the Laplace operator on Ω (with Dirichlet boundary condition) for $\operatorname{Im} k > 0$. The limiting absorption principle implies that $\mathcal{B}(k)$ is strongly continuous in $\{k \in \mathbb{C} : \operatorname{Im} k \geq 0\}$; of course, it is holomorphic on $\{k : \operatorname{Im} k > 0\}$.

Here we will show that $v = \mathcal{B}(k)f$ can be obtained as the solution to an integral equation over ∂K. Use of such integral equations is a convenient tool for a number of investigations in scattering theory. We use it here to show that $\mathcal{B}(k)$ has a meromorphic continuation to an operator-valued function on $\mathbb{C}$, with some poles in $\{k : \operatorname{Im} k < 0\}$. These poles are known as scattering poles and provide fundamental objects for study in scattering theory.

The integral equations applying to (7.1) will be obtained from a study of the following operators, called single- and double-layer potentials, respectively:

$$\mathcal{S}\ell(k)f(x) = \int_{\partial K} f(y)\, g(x,y,k)\, dS(y) \tag{7.3}$$

and

$$\mathcal{D}\ell(k)f(x) = \int_{\partial K} f(y)\,\frac{\partial g}{\partial \nu_y}(x,y,k)\,dS(y), \tag{7.4}$$

where, as in §1,

$$g(x,y,k) = \big(4\pi|x-y|\big)^{-1}\,e^{ik|x-y|}. \tag{7.5}$$

For $f \in C^\infty(\partial K)$, or even for $f \in L^1(\partial K)$, the functions (7.3) and (7.4) are well defined and smooth for $x \in \mathbb{R}^3 \setminus \partial K = \Omega \cup \overset{\circ}{K}$, where $\overset{\circ}{K}$ is the interior of K. For such $v,\ x \in \partial K$, we denote by $v_+(x)$ the limit from the exterior region $\Omega,\ v_-(x)$ the limit from the interior region $\overset{\circ}{K}$, and by $\partial v/\partial \nu_+$ and $\partial v/\partial \nu_-$ their normal derivatives, in the direction pointing into Ω, taken as limits from Ω and from $\overset{\circ}{K}$, respectively. By the methods used to treat layer potentials in §11 of Chap. 7, one derives the following results:

$$\begin{aligned} \mathcal{S}\ell(k)f_+(x) &= \mathcal{S}\ell(k)f_-(x) = G(k)f(x), \\ \mathcal{D}\ell(k)f_\pm(x) &= \pm\frac{1}{2}f(x) + \frac{1}{2}N(k)f(x), \end{aligned} \tag{7.6}$$

where, for $x \in \partial K$,

$$G(k)f(x) = \int_{\partial K} f(y)\,g(x,y,k)\,dS(y) \tag{7.7}$$

and

$$N(k)f(x) = 2\int_{\partial K} f(y)\,\frac{\partial g}{\partial \nu_y}(x,y,k)\,dS(y). \tag{7.8}$$

Note that, for $|x-y| \le 1,\ g(x,y,k)$ has an estimate of the form

$$|g(x,y,k)| \le C_k|x-y|^{-1}. \tag{7.9}$$

We have for $\nabla_y g$ the poorer estimate $|\nabla_y g(x,y,k)| \le C_k|x-y|^{-2}$, but the normal derivative $\partial g/\partial \nu_y$ has a weaker singularity on $\partial K \times \partial K$, of the same kind as g:

$$\Big|\frac{\partial g}{\partial \nu_y}(x,y,k)\Big| \le C|x-y|^{-1}, \quad \text{for } x,y \in \partial K. \tag{7.10}$$

It follows that $G(k)$ and $N(k)$ are compact operators on $L^2(\partial K)$, for each $k \in \mathbb{C}$, with holomorphic dependence on k.

We will first consider the possibility of obtaining the solution v to the scattering problem in the form

$$v = \mathcal{B}(k)f = \mathcal{D}\ell(k)g \quad \text{on } \Omega, \tag{7.11}$$

where g (whose dependence on k we suppress) satisfies the identity

$$\big(I + N(k)\big)g = 2f \quad \text{on } \partial K. \tag{7.12}$$

We will establish the following result.

Proposition 7.1. *The operator $I + N(k)$ is invertible on $L^2(\partial K)$ for all $\operatorname{Im} k > 0$, and for all real k, except for $k = \lambda_j$, where $-\lambda_j^2$ is an eigenvalue for Δ on the interior region $\overset{\circ}{K}$, with Neumann boundary condition on ∂K.*

Proof. Since $N(k)$ is compact, it suffices to consider whether $I + N(k)$ is injective. Suppose therefore that

$$\big(I + N(k)\big)g = 0, \tag{7.13}$$

and consider $v = \mathcal{D}\ell(k)g$ on $\mathbb{R}^3 \setminus \partial K$. On $\Omega,\ v$ satisfies (7.1), with $f = 0$ (for real k, and it is also exponentially decaying as $|x| \to \infty$ if $\operatorname{Im} k > 0$), so the uniqueness result implies that $v = 0$ on Ω. Thus $\partial v/\partial \nu_+ = 0$. Now an analysis of the double-layer potential (7.4), parallel to that for (11.39) of Chap. 7, shows that, in general,

$$\frac{\partial \mathcal{D}\ell(k)f}{\partial \nu_+} = \frac{\partial \mathcal{D}(k)f}{\partial \nu_-} \quad \text{on } \partial K. \tag{7.14}$$

Hence, for $v = \mathcal{D}\ell(k)g$, with (7.13) satisfied, we have

$$\frac{\partial v}{\partial \nu_-} = 0 \quad \text{on } \partial K. \tag{7.15}$$

Thus v satisfies the homogeneous Neumann boundary condition, together with the PDE

$$(\Delta + k^2)v = 0 \quad \text{on } \overset{\circ}{K}. \tag{7.16}$$

Since, by (7.7), the jump of v across ∂K is $g(x)$, and since $v_+ = 0$, we deduce that $v_- = -g$, so v is not identically zero in $\overset{\circ}{K}$ if $g \neq 0$. The spectrum of the Laplace operator Δ on K, with Neumann boundary condition, is a discrete subset of $\{\lambda_j^2\}$ of $\mathbb{R}^-$, so the proposition is proved.

The extension of $\mathcal{B}(k)$ to a neighborhood of the real line in $\mathbb{C}$, including the exceptional points λ_j defined above, is neatly accomplished by considering the following alternative integral equation. Namely, we look for a solution v to the scattering problem of the form

$$v = \mathcal{B}(k)f = \mathcal{D}\ell(k)g + i\eta\mathcal{S}\ell(k)g \ \text{ in } \Omega, \tag{7.17}$$

where g is to be determined as a function of f. Here η is a real constant; we can take $\eta = \pm 1$. In this case, we require that g satisfy the identity

$$\big[I + N(k) + 2i\eta G(k)\big]g = 2f. \tag{7.18}$$

Proposition 7.2. *For a given real $\eta \neq 0$, the operator $I + N(k) + 2i\eta G(k)$ is invertible on $L^2(\partial K)$, for all k such that*

$$\mathit{Im}\, k \geq 0 \ \ \mathit{and} \ \ \eta\, \mathit{Re}\, k \geq 0. \tag{7.19}$$

Proof. Again it suffices to check injectivity. Suppose $g \in L^2(\partial K)$ satisfies

$$\big[I + N(k) + 2i\eta G(k)\big]g = 0, \tag{7.20}$$

and let

$$v = \mathcal{D}\ell(k)g + i\eta\mathcal{S}\ell(k)g \ \text{ in } \mathbb{R}^3 \setminus \partial K. \tag{7.21}$$

Then v satisfies (7.1) (for k real, also with exponential decay for $\operatorname{Im} k > 0$) on Ω, with $f = 0$, so our familiar uniqueness result implies $v = 0$ on Ω, hence $v_+ = 0$ and $\partial v/\partial \nu_+ = 0$ on ∂K. Hence, as before, by (7.6)–(7.8),

$$v_- = -g \ \text{ on } \partial K. \tag{7.22}$$

Similarly, $\partial v/\partial \nu_-$ is equal to the jump of $\partial v/\partial \nu$ across ∂K. To calculate this jump, we use (7.14) for $\mathcal{D}\ell(k)g$, and for $i\eta\mathcal{S}\ell(k)g$, we use the identity

$$\frac{\partial \mathcal{S}\ell(k)g}{\partial \nu_\pm}(x) = \frac{1}{2}\big(N^\#(k)g \mp g\big), \tag{7.23}$$

where

$$N^\#(k)g(x) = 2\int_{\partial K} g(y)\,\frac{\partial g}{\partial \nu_x}(x,y,k)\,dS(y), \quad x \in \partial K. \tag{7.24}$$

Consequently, complementing (7.14), we have

$$\frac{\partial \mathcal{S}\ell(k)g}{\partial \nu_+} - \frac{\partial \mathcal{S}\ell(k)g}{\partial \nu_-} = -g \ \text{ on } \partial K. \tag{7.25}$$

Therefore, for v given by (7.21), we have

$$\frac{\partial v}{\partial \nu_-} = -i\eta g \quad \text{on } \partial K. \tag{7.26}$$

Hence, on the interior region, v satisfies

$$(\Delta + k^2)v = 0 \text{ on } \overset{\circ}{K}, \quad \frac{\partial v}{\partial \nu} - i\eta v = 0 \text{ on } \partial K. \tag{7.27}$$

Given that $\eta \neq 0$, we claim that this implies $v = 0$ on $\overset{\circ}{K}$. Indeed, Green's identity implies

$$\|\nabla v\|^2_{L^2(K)} - k^2\|v\|^2_{L^2(K)} = -i\eta\|v\|^2_{L^2(\partial K)}. \tag{7.28}$$

Taking the imaginary part of this identity, we have the following. If $k = \lambda + i\mu$,

$$2\lambda\mu\|v\|^2_{L^2(K)} = -\eta\|v\|^2_{L^2(\partial K)}. \tag{7.29}$$

Under the hypotheses (7.19), the coefficients on the two sides of (7.29) have opposite signs, so $v = 0$ on ∂K. In view of (7.22), this implies $g = 0$, so this proposition is proved.

Taking $\eta = \pm 1$, we have $I + N(k) \pm 2iG(k)$ invertible in the first (resp., second) closed quadrant in $\mathbb{C}$, hence invertible in a neighborhood of such a quadrant. Thus $\mathcal{B}(k)$ is extended to an operator-valued function holomorphic on a neighborhood of the closed upper half-plane $\operatorname{Im} k \geq 0$.

We next show that, in fact, $\mathcal{B}(k)$ has a continuation to a meromorphic operator-valued function on $\mathbb{C}$. This is an immediate consequence of the following result.

Proposition 7.3. *The operator $I + N(k)$ is invertible on $L^2(\partial K)$ for all $k \in \mathbb{C}$ except for a discrete set, and $(I + N(k))^{-1}$ is a meromorphic function on $\mathbb{C}$.*

This result in turn is a special case of the following elementary general result.

Proposition 7.4. *Let $\mathcal{O}$ be a connected open set in $\mathbb{C}$. Suppose $C(z)$ is a compact, operator-valued, holomorphic function of $z \in \mathcal{O}$. Suppose that $I + C(z)$ is invertible at some point $p_0 \in \mathcal{O}$. Then $I + C(z)$ is invertible except at most on a discrete set in $\mathcal{O}$, and $(I + C(z))^{-1}$ is meromorphic on $\mathcal{O}$.*

Proof. The operator $I + C(z)$ fails to be invertible at a point $z \in \mathcal{O}$ if and only if the compact operator $C(z)$ has -1 in its spectrum. For a given $z_0 \in \mathcal{O}$, let γ be a small circle about -1, disjoint from the spectrum of $C(z_0)$. For z in a small neighborhood $\mathcal{U}$ of z_0, we can form the projection-valued function

$$P(z) = \frac{1}{2\pi i}\int_\gamma \left(\lambda - C(z)\right)^{-1} d\lambda. \tag{7.30}$$

For $z \in \mathcal{U}$, this is a projection of finite rank (say ℓ); using $P(z_0)$ we can produce a family of isomorphisms of the range $\mathcal{R}(P(z))$ with $\mathcal{R}(P(z_0))$, and then $C(z)P(z)$ can be treated as a holomorphic family of $\ell \times \ell$ matrices. This proposition in the case of $\ell \times \ell$ matrices is easy, via determinants. By hypothesis, -1 is not identically an eigenvalue for this family, so

$$(I + C(z))^{-1} P(z)$$

is a meromorphic function on $\mathcal{U}$. Clearly,

$$(I + C(z))^{-1}(I - P(z))$$

is a holomorphic function on $\mathcal{U}$, so this establishes the proposition.

Corollary 7.5. *The solution operator $\mathcal{B}(k)$ for (7.1) has a meromorphic continuation to $\mathbb{C}$; all its poles are in the lower half-plane Im $k < 0$.*

This follows from the formula

$$\mathcal{B}(k) = 2\mathcal{D}\ell(k)\big[I + N(k)\big]^{-1}, \tag{7.31}$$

except at the real points $k = \lambda_j$, from Proposition 7.1, together with the formula

$$\mathcal{B}(k) = 2\big[\mathcal{D}\ell(k) + i\eta\mathcal{S}\ell(k)\big]\big[I + N(k) + 2i\eta G(k)\big]^{-1}, \tag{7.32}$$

for $\eta = \pm 1$, which defines $\mathcal{B}(k)$ as holomorphic on a neighborhood of the real axis.

The poles of $\mathcal{B}(k)$ are called scattering poles. It follows immediately from (7.31) that the set of scattering poles is contained in the set of poles of $[I + N(k)]^{-1}$ within the lower half-plane Im $k < 0$. In fact, these two sets coincide; this is a consequence of the following.

Lemma 7.6. *If Im $k \neq 0$, then $\mathcal{D}\ell(k) : L^2(\partial K) \to L^2_{loc}(\Omega)$ is injective.*

Proof. The argument used in the proof of Proposition 7.1 shows that if $g \in L^2(\partial K)$ and $\mathcal{D}\ell(k)g = 0$ on Ω, then $g = v\big|_{\partial K}$ where $v\big|_{\overset{\circ}{K}}$ is an eigenfunction for Δ on $\overset{\circ}{K}$, with Neumann boundary condition on ∂K, and with eigenvalue $-k^2$. Since the spectrum of this elliptic operator is real and nonpositive, the lemma is proved.

Proposition 7.7. *The set of scattering poles is precisely equal to the set of poles k, for $\big[I + N(k)\big]^{-1}$, such that Im $k < 0$.*

Proof. If $[I + N(k)]^{-1}$ has a pole of order m at $k = k_j$, Im $k_j < 0$, then there is an element $h \in L^2(\partial K)$ such that, with nonzero $h_m \in L^2(\partial K)$,

$$\big[I + N(k)\big]^{-1} h = (k - k_j)^{-m}\big[h_m + (k - k_j)h_{m-1} + \cdots\big]. \tag{7.33}$$

Since $\mathcal{D}\ell(k_j)h_m = b_m \neq 0$ in $L^2_{\text{loc}}(\Omega)$, it follows that, for k near k_j,

$$\mathcal{B}(k)h = (k - k_j)^{-m} b_m + O\big((k - k_j)^{-m+1}\big), \quad k \to k_j, \tag{7.34}$$

so $\mathcal{B}(k)$ is singular at k_j.

We also have the following characterization of scattering poles.

Proposition 7.8. *A complex number k_j is a scattering pole if and only if there is a nonzero $v \in C^\infty(\Omega)$ satisfying*

$$(\Delta + k_j^2)v = 0 \text{ on } \Omega, \quad v = 0 \text{ on } \partial K, \tag{7.35}$$

of the form

$$v = \mathcal{D}\ell(k_j)g, \tag{7.36}$$

for some $g \in L^2(\partial K)$.

Proof. We know that, for $\operatorname{Im} k_j \geq 0$, v satisfying (7.35)–(7.36) must vanish on Ω. On the other hand, if $\operatorname{Im} k_j < 0$, we know that k_j is a scattering pole if and only if $I + N(k_j)$ has nonzero kernel. We claim that, for $\operatorname{Im} k_j < 0$,

$$\mathcal{D}\ell(k_j) : \ker\big(I + N(k_j)\big) \longrightarrow \{v \text{ satisfying (7.35)–(7.36)}\}, \tag{7.37}$$

isomorphically. Indeed, surjectivity is obvious, and injectivity follows from Lemma 7.6. This proves Proposition 7.8.

The condition (7.36) can be viewed as an extension of the radiation condition, which we initially defined for real k. A sharper result is given in Proposition 7.13 below.

It is clear that the Green function $G(x, y, k)$, defined in §1 by (1.26)–(1.30), has a meromorphic extension in k, with poles confined to the set of scattering poles defined above. Indeed, we can write

$$G(x, y, k) = g(x, y, k) - \mathcal{B}(k)\gamma_{y,k}(x), \tag{7.38}$$

where $g(x, y, k)$ is given by (1.5) for all $k \in \mathbb{C}$, and $\gamma_{y,k}$ is the restriction of $g(x, y, k)$ to $x \in \partial K$. Similarly, the "eigenfunctions" $u_+(x, k\omega)$, defined by (1.32)–(1.33), have such a meromorphic continuation in k, and so do the scattering amplitude $a(\omega, \theta, k)$ and the scattering operators $S(k)$ and $\mathcal{S}(k)$. We will explore these last objects further at the end of this section. First we consider another integral-equation approach to the scattering problem (7.1).

As another alternative to (7.11), it is of interest to obtain solutions to the scattering problem in the form

$$v = \mathcal{B}(k)f = \mathcal{S}\ell(k)g \text{ on } \Omega, \tag{7.39}$$

where g satisfies the integral equation

$$G(k)g = f \quad \text{on } \partial K. \tag{7.40}$$

The operator-valued function $G(k)$ is defined by (7.6). As we have noted, $G(k)$ is compact on $L^2(\partial K)$. In fact, analysis done in Chap. 7 shows that $G(k)$ is a pseudodifferential operator of order -1 on ∂K, and examination of its symbol shows that it is elliptic. The principal symbol of $G(k)$ is positive on $S^*(\partial K)$. Consequently, for each $k \in \mathbb{C}$, each real s,

$$G(k) : H^s(\partial K) \longrightarrow H^{s+1}(\partial K) \text{ is Fredholm, of index zero.} \tag{7.41}$$

In analogy with Proposition 7.1, we have the following result:

Proposition 7.9. *The operator $G(k) : H^s(\partial K) \to H^{s+1}(\partial K)$ is invertible for all k such that $\operatorname{Im} k > 0$, and for all real k, except for $k = \mu_j$ such that $-\mu_j^2$ is an eigenvalue of Δ on the interior region $\overset{\circ}{K}$, with Dirichlet boundary condition on ∂K.*

Proof. In view of (7.41), it suffices to check the injectivity of $G(k)$. This goes as in the proof of Proposition 7.1. Setting $v = \mathcal{S}\ell(k)g$ on $\mathbb{R}^3 \setminus \partial K$, uniqueness as before yields $v = 0$ on Ω if $g \in \ker G(k)$, $\operatorname{Im} k \geq 0$. Then $v_- = 0$ on ∂K, by (7.6), while by (7.25) $\partial v/\partial \nu_- = g$ on ∂K, so if $g \neq 0$ then $v\big|_{\overset{\circ}{K}} \neq 0$ is an eigenfunction for Δ on $\overset{\circ}{K}$, with Dirichlet boundary condition and with eigenvalue $-k^2$.

In addition to (7.41), we obtain from the analysis of $G(k)$ as a pseudodifferential operator that its principal symbol is independent of k, hence

$$G(k) - G(0) = D(k) : H^s(\partial K) \longrightarrow H^{s+2}(\partial K). \tag{7.42}$$

By Proposition 7.9, $G(0)$ is invertible. Then

$$G(0)^{-1}G(k) = I + G(0)^{-1}D(k) : H^s(\partial K) \longrightarrow H^s(\partial K) \tag{7.43}$$

is holomorphic in k, and

$$G(0)^{-1}D(k) : H^s(\partial K) \longrightarrow H^{s+1}(\partial K); \tag{7.44}$$

in particular, this operator is compact on $H^s(\partial K)$, for each $s \geq 0$. Since Proposition 7.9 implies that the operator (7.43) is invertible for $\operatorname{Im} k > 0$, we can apply the general operator result of Proposition 7.4, to obtain:

Proposition 7.10. *The operator-valued function*

$$G(k)^{-1} : H^{s+1}(\partial K) \longrightarrow H^s(\partial K) \tag{7.45}$$

has a meromorphic continuation to $\mathbb{C}$, with poles contained in Im $k < 0$ together with the set μ_j of real numbers specified in Proposition 7.9.

In view of (7.39), the set of poles of (7.45) satisfying Im $k < 0$ contains the set of scattering poles, and

$$\mathcal{B}(k) = \mathcal{S}\ell(k)\, G(k)^{-1}, \tag{7.46}$$

where $G(k)^{-1}$ is regular. In fact, in parallel with the proofs of Lemma 7.6 and Proposition 7.7, we easily obtain the following:

Proposition 7.11. *If Im $k \neq 0$, then $\mathcal{S}\ell(k) : L^2(\partial K) \to L^2_{loc}(\Omega)$ is injective. Therefore, the set of scattering poles is precisely equal to the set of poles for $G(k)^{-1}$ such that Im $k < 0$. Furthermore, a complex number k_j is a scattering pole if and only if there is a nonzero $v \in C^\infty(\Omega)$ satisfying (7.35), of the form*

$$v = \mathcal{S}\ell(k_j)g, \tag{7.47}$$

for some $g \in L^2(\partial K)$. More precisely, for Im $k_j < 0$,

$$\mathcal{S}\ell(k_j) : \text{ker}\, G(k_j) \longrightarrow \{v \text{ satisfying (7.35) and (7.47)}\}, \tag{7.48}$$

isomorphically.

From the formula (7.7) for $G(k)$, we see that

$$G(k)^* = G(-\overline{k}). \tag{7.49}$$

We therefore have the following:

Corollary 7.12. *The set of scattering poles is symmetric about the imaginary axis.*

We can also obtain a characterization of the set of scattering poles which is more satisfactory than that of Proposition 7.8 or the last part of Proposition 7.11.

Proposition 7.13. *A complex number k is a scattering pole if and only if there is a nonzero $v \in C^\infty(\Omega)$ satisfying (7.35), of the form*

$$v = \mathcal{D}\ell(k)g_1 + \mathcal{S}\ell(k)g_2 \ \text{ on } \Omega, \tag{7.50}$$

for some $g_j \in L^2(\partial K)$.

Proof. For v of the form (7.50), note that

$$v_+ = \frac{1}{2}\big(I + N(k)\big)g_1 + G(k)g_2 \ \text{ on } \partial K. \tag{7.51}$$

In particular, if $v_+ = 0$ on ∂K and k is not a scattering pole, but $\operatorname{Im} k \neq 0$, then $g_1 = -2(I + N(k))^{-1}G(k)g_2$. Now we know that, for $\operatorname{Im} k > 0$, $v_+ = 0$ on ∂K implies $v = 0$ on Ω, so we have the identity

$$2\mathcal{D}\ell(k)\big(I + N(k)\big)^{-1}G(k) - \mathcal{S}\ell(k) = 0, \tag{7.52}$$

for $\operatorname{Im} k > 0$, as a map from $L^2(\partial K)$ to $C^\infty(\Omega)$. This identity continues analytically to the lower half-plane $\operatorname{Im} k < 0$, excluding the scattering poles, and implies that if v is of the form (7.50), $v_+ = 0$ on ∂K, and k is not a scattering pole, then $v = 0$ on Ω. Given the results of Propositions 7.8 and 7.11 when $k = k_j$ is a scattering pole, this finishes the proof.

We can obtain a few more conclusions from (7.52), which we write as

$$\mathcal{D}\ell(k)M(k) = \mathcal{S}\ell(k) \ \text{ on } \Omega, \tag{7.53}$$

valid for all $k \in \mathbb{C}$ at which $I + N(k)$ is invertible, with

$$M(k) = 2\big(I + N(k)\big)^{-1}G(k). \tag{7.54}$$

First, using the injectivity of $\mathcal{D}\ell(k)$ for $\operatorname{Im} k < 0$, as in the proof of Proposition 7.7, we see that $M(k)$ has an analytic continuation to all $\operatorname{Im} k < 0$, including the set of scattering poles. The only poles of $M(k)$ are at the real numbers λ_j of Proposition 7.1. Also, $M(k)$ is invertible, except at the real numbers μ_j of Proposition 7.9; in particular, $M(k)$ is invertible at all the scattering poles. Therefore, when $k = k_j$ is a scattering pole, $M(k_j)$ gives an isomorphism from ker $G(k_j)$, in (7.48), to ker $(I + N(k_j))$, in (7.37). Furthermore, any v of the form (7.50), with $k = k_j$, can be written both in the form (7.36) and in the form (7.47) (with different g's).

Another calculation using the representation of the solution to the scattering problem by a single-layer potential (7.39)–(7.40), produces an analysis of the Neumann operator $\mathcal{N}(k)$, which we define as follows, first for $\operatorname{Im} k \geq 0$. For $f \in C^\infty(\partial K)$, let v be the solution to the scattering problem (7.1), $v = \mathcal{B}(k)f$, and define

$$\mathcal{N}(k)f = \frac{\partial v}{\partial \nu_+} \ \text{ on } \partial K. \tag{7.55}$$

By elliptic regularity estimates, we can deduce that, for $s \geq 1$,

$$\mathcal{N}(k) : H^s(\partial K) \longrightarrow H^{s-1}(\partial K). \tag{7.56}$$

We produce a formula for $\mathcal{N}(k)$ using the representation $v = \mathcal{S}\ell(k)g$, $g = G(k)^{-1}f$, valid for $\operatorname{Im} k > 0$. From the formula (7.23) for $\partial\mathcal{S}\ell(k)g/\partial\nu_\pm$, we see that

$$\mathcal{N}(k) = \frac{1}{2}\big(N^\#(k) - I\big)G(k)^{-1}, \tag{7.57}$$

for Im $k > 0$. This identity continues analytically to the complement of the set of poles of $G(k)^{-1}$ in $\mathbb{C}$. Note that, complementing (7.49),

$$N^{\#}(k) = N(-\overline{k})^*, \tag{7.58}$$

so (7.57) can also be written as

$$\mathcal{N}(-\overline{k}) = \Big(\big[2(N(k) - I)^{-1}G(k)\big]^*\Big)^{-1}. \tag{7.59}$$

By the analysis of the scattering problem for Im $k \geq 0$, we know that $\mathcal{N}(k)$ is a strongly continuous function of k, with values in the Banach space

$$\mathcal{L}(H^s(\partial K), H^{s-1}(\partial K)),$$

for Im $k \geq 0$. Thus $\mathcal{N}(k)$ does not have poles on the real axis; such singularities are therefore removable on the right side of (7.57). The poles of $G(k)^{-1}$ on the real axis must be canceled by a null space of $N^{\#}(k) - I$, for $k = \lambda_j$. The occurrence of these real poles of $G(k)^{-1}$ makes (7.57) a tool of limited value in analyzing the Neumann operator $\mathcal{N}(k)$ for real k.

We can produce another formula for $\mathcal{N}(k)$, first for Im $k > 0$, by using the representation (1.4) for $v = \mathcal{B}(k)f$, that is,

$$\mathcal{B}(k)f = \mathcal{D}\ell(k)f - \mathcal{S}\ell(k)\mathcal{N}(k)f. \tag{7.60}$$

Evaluating this on ∂K, we have

$$f = \frac{1}{2}\big(I + N(k)\big)f - G(k)\mathcal{N}(k)f, \tag{7.61}$$

which implies

$$\mathcal{N}(k) = \frac{1}{2}G(k)^{-1}\big(N(k) - I\big), \tag{7.62}$$

for Im $k > 0$. Of course, this identity also continues analytically to all $k \in \mathbb{C}$ outside the set of poles of $G(k)^{-1}$. Comparing (7.62) with (7.57), we see that $N(k)$ and $N^{\#}(k)$ are related by the identity

$$N(k)G(k) = G(k)N^{\#}(k), \tag{7.63}$$

for all $k \in \mathbb{C}$. Also, comparing (7.62) with (7.59), we see that

$$\mathcal{N}(-\overline{k}) = \mathcal{N}(k)^*; \tag{7.64}$$

in particular, $\mathcal{N}(k)$ is self-adjoint when k is purely imaginary. Furthermore, in view of (7.60) and (7.57), we see that the set of poles of $\mathcal{N}(k)$ coincides exactly with the set of scattering poles.

Note that the factor $(1/2)(N(k) - I)$ in (7.62) arises from evaluating $\mathcal{D}\ell(k)f$ on ∂K as a limit from the *interior* region $\overset{\circ}{K}$, by (7.7). Thus the analogue of the identity (7.53) which is valid on $\overset{\circ}{K}$ is obtained by replacing $M(k)$ by $\mathcal{N}(k)^{-1}$. Equivalently,

$$\mathcal{D}\ell(k) = \mathcal{S}\ell(k)\mathcal{N}(k) \;\; \text{on } \overset{\circ}{K}, \tag{7.65}$$

where $\mathcal{N}(k)$ is the exterior Neumann operator defined above.

So far we have not established that there actually are scattering poles. We will show that in fact there are infinitely many scattering poles on the negative imaginary axis, for any nonempty smooth obstacle K, by a study of $G(k)$. We begin with the following result:

Lemma 7.14. *For real* $s \geq 0$, $G(is)$ *is positive-definite.*

Proof. Given $g \in L^2(\partial K)$, set $v = \mathcal{S}\ell(is)g$ on $\mathbb{R}^3 \setminus \partial K$. Then Green's theorem gives, for $s > 0$,

$$\begin{aligned} (\Delta v, v)_{L^2(\Omega)} + \|dv\|^2_{L^2(\Omega)} &= -\int_{\partial K} G(is)g \, \frac{\partial v}{\partial \nu_+} \, dS, \\ (\Delta v, v)_{L^2(K)} + \|dv\|^2_{L^2(K)} &= \int_{\partial K} G(is)g \, \frac{\partial v}{\partial \nu_-} \, dS. \end{aligned} \tag{7.66}$$

Recall from (7.25) that $\partial v/\partial \nu_- - \partial v/\partial \nu_+ = g$, so adding the identities above gives

$$\begin{aligned} \big(G(is)g, g\big)_{L^2(\partial K)} = s^2 \Big(\|v\|^2_{L^2(\Omega)} + \|v\|^2_{L^2(K)} \Big) \\ + \|\nabla v\|^2_{L^2(\Omega)} + \|\nabla v\|^2_{L^2(K)}, \end{aligned} \tag{7.67}$$

for $s > 0$, which proves the lemma in this case. Since we know that $G(0)$ is invertible, this is also positive-definite.

To proceed with the demonstration that $G(is)^{-1}$ is singular for infinitely many negative real s, we set

$$n(s) = \#\, \text{negative eigenvalues of } G(is), \tag{7.68}$$

for $s < 0$. Our next claim as follows:

Lemma 7.15. *As* $s \searrow -\infty$, $n(s) \to \infty$.

Proof. We will show that $\big(G(is)g_1, g_2\big)$ defines a negative definite inner product on a vector space V whose dimension can be taken large with $|s|$. Then the lemma follows, by the variational characterization of the spectrum of $G(is)$. Pick points $p, q \in K$ such that $|p-q|$ is maximal. Then, for any N, you can pick p_j near p and q_j near q, for $1 \le j \le N$, such that

$$\min \big\{|p_j - q_j| : 1 \le j \le N\big\} > \max \big\{|p_j - q_k| : j \ne k\big\}. \tag{7.69}$$

Put small disjoint disks D_j about p_j, D'_j about q_j, all of the same area, within ∂K, and define functions $g_j \in L^2(\partial K)$ by

$$g_j = 1 \text{ on } D_j, \quad -1 \text{ on } D'_j, \quad 0 \text{ elsewhere.} \tag{7.70}$$

Then $\{g_j : 1 \le j \le N\}$ is a set of orthogonal functions, all of the same norm. Let V be the linear span of these g_j. With V so fixed, of dimension N, a simple calculation gives

$$\begin{aligned} \big(G(is)g_j, g_j\big) &< -\gamma < 0, \\ \big|\big(G(is)g_j, g_k\big)\big| &<< \gamma, \ \text{ for } j \ne k, \end{aligned} \tag{7.71}$$

for s large and negative (because $|x-y|^{-1}e^{|s|\,|x-y|}$ is maximal for x and y distant, if you exclude a small neighborhood of $x = y$), and the lemma follows.

In view of (7.41), only finitely many of the eigenvalues of $G(is)$ (all of which are real) can cross from positive to negative at any point $s = s_k$, so we have the following conclusion from the last two lemmas.

Proposition 7.16. *The operator-valued function $G(k)^{-1}$ has an infinite number of poles on the negative imaginary axis, each of which is a scattering pole.*

As we have already mentioned, the scattering amplitude $a(\omega, \theta, k)$ and also the scattering operators $S(k)$ and $\mathcal{S}(k)$ have meromorphic continuations, with poles confined to the set of scattering poles. Indeed, by (3.28), $a(\omega, \theta, k)$ is the sum of an entire function and $1/4\pi$ times

$$\big(\mathcal{N}(k)e_{k\omega}, e_{k\theta}\big)_{L^2(\partial K)}, \tag{7.72}$$

where $e_\xi(y) = e^{iy\cdot\xi}$ for $y \in \partial K$. We now draw a connection between the set of poles of $\mathcal{S}(k)$ and the set of scattering characters, described in §6 in terms of the spectrum of $Z(t)$.

First note that since $\mathcal{S}(k)$ is unitary for k real, we have

$$\mathcal{S}(k)\mathcal{S}(\overline{k})^* = \mathcal{S}(\overline{k})^*\mathcal{S}(k) = I, \tag{7.73}$$

for k in a neighborhood of the real axis. By continuation, knowing that $\mathcal{S}(k)$ is holomorphic for Im $k \ge 0$, we see that a complex number k such that Im $k < 0$

is a pole of $\mathcal{S}(k)$ if and only if $\mathcal{S}(\overline{k})$ fails to be invertible. Now, by (3.14)–(3.15), and its analogue for $\mathcal{S}(k)$, we see that, for Im $k > 0$,

$$\mathcal{S}(k) = I + \frac{k}{2\pi i}\mathcal{A}(k), \tag{7.74}$$

where, for such k, $\mathcal{A}(k)$ is a compact operator on $L^2(S^2)$, given by a smooth integral kernel. Such $\mathcal{S}(k)$ is Fredholm of index zero. Thus, for Im $k > 0$, $\mathcal{S}(k)$ fails to be invertible if and only if it has a nonzero kernel. Furthermore, this happens if and only if $\mathcal{S}(k)^*$ has a nonzero kernel. We are now prepared to establish the following result.

Proposition 7.17. *A complex number μ is a scattering character if and only if $i\mu$ is a pole of $\mathcal{S}(k)$.*

Proof. μ is a scattering character if and only if there exists a nonzero $f \in \mathcal{K}_+$ such that $Z_+(t)f = e^{\mu t} f$, for $t \ge 0$. By (6.41), this implies

$$\begin{aligned} f(s,\omega) = e^{-\mu s}\varphi(\omega), &\quad \text{for } s \le R, \\ 0, &\quad \text{for } s > R, \end{aligned} \tag{7.75}$$

for some nonzero $\varphi \in L^2(S^2)$. By (6.40), such an f belongs to $\mathcal{K}_+$ if and only if $\hat{\mathcal{S}}^* f$ is supported in $[R,\infty)$. By the Paley–Wiener theorem, we can deduce that this will hold if and only if $\mathcal{S}(\overline{k})^* \hat{f}(k)$ is holomorphic in Im $k < 0$. Now

$$\hat{f}(k) = (2\pi)^{-1/2}\, \frac{\varphi(\omega)}{ik+\mu}, \tag{7.76}$$

which has a pole at $k = i\mu$, so this analyticity holds if and only if φ belongs to the kernel of $\mathcal{S}(\overline{k})^*$, for $k = i\mu$. This establishes the proposition.

Exercises

1. Verify that $G(k)$, defined by (7.7), is an elliptic pseudodifferential operator of order -1 on ∂K. Compute its principal symbol.
2. Justify (7.69).

The following exercises deal with an integral-equation attack on the scattering problem for $H = -\Delta + V$ on $\mathbb{R}^3$. Assume $V \in C_0^\infty$. We use (1.57), that is,

$$(I - \mathcal{V}(k))v = \mathcal{R}(k) f,$$

where $\mathcal{V}(k) = \mathcal{R}(k)(Vv)$ and

$$\mathcal{R}(k)v(x) = -\int v(y) g(x,y,k)\, dy,$$

with $g(x,y,k) = (4\pi|x-y|)^{-1} e^{ik|x-y|}$.

3. Show that, for $\operatorname{Im} k \geq 0,\ \sigma > 1$,

$$\mathcal{R}(k) : L^2_{\mathrm{comp}}(\mathbb{R}^3) \longrightarrow L^2(\mathbb{R}^3, \langle x\rangle^{-2\sigma}\,dx) \text{ is compact.}$$

4. Show that, for $\operatorname{Im} k \geq 0,\ k^2$ not a negative eigenvalue of $-\Delta + V$, and $\sigma > 1$,

$$I - \mathcal{V}(k) : L^2(\mathbb{R}^3, \langle x\rangle^{-2\sigma}\,dx) \longrightarrow L^2(\mathbb{R}^3, \langle x\rangle^{-2\sigma}\,dx)$$

   is injective, hence invertible. (*Hint*: If $u = \mathcal{V}(k)u = \mathcal{R}(k)(Vu)$, show that u satisfies the hypotheses for the uniqueness result of Exercise 6 in §1 when $k \in \mathbb{R}$. When $\operatorname{Im} k > 0$, the argument is easier.)
5. Fix $\kappa \in (0, \infty)$. Show that, for $\operatorname{Im} k > -\kappa$,

$$\mathcal{R}(k) : L^2_{\mathrm{comp}}(\mathbb{R}^3) \longrightarrow L^2(\mathbb{R}^3, e^{-2\kappa|x|}dx) \text{ is compact.}$$

   Also show that

$$I - \mathcal{V}(k) : L^2(\mathbb{R}^3, e^{-2\kappa|x|}dx) \longrightarrow L^2(\mathbb{R}^3, e^{-2\kappa|x|}dx)$$

   is holomorphic in $\{k : \operatorname{Im} k > -\kappa\}$, and invertible for $\operatorname{Im} k \geq 0,\ k^2 \notin$ point spec H. Deduce that its inverse has a meromorphic continuation.

8. Trace formulas; the scattering phase

In Proposition 6.7 we showed that, for any $\rho \in C_0^\infty\big((2R, \infty)\big)$, the operator $Z(\rho) = \int_0^\infty \rho(t)Z(t)\,dt$ is compact. Recall that the proof used the identity

$$Z(\rho) = P_+\big[U(\rho) - U_0(\rho)\big]P_-, \quad \text{for } \rho \in C_0^\infty\big((2R, \infty)\big). \tag{8.1}$$

We then saw that $U(\rho) - U_0(\rho)$ has a smooth, compactly supported integral kernel. It follows that the operator (8.1) is not only compact, but in fact trace class. By a theorem of V. Lidskii, which we will prove in Appendix A at the end of this chapter, it follows that the trace $\operatorname{Tr} Z(\rho)$ is equal to the sum of the eigenvalues of $Z(\rho)$, counted with multiplicity. Thus we have

$$\operatorname{Tr} Z(\rho) = \sum \hat{\rho}(i\mu_j) \tag{8.2}$$

where the sum is over the set of scattering characters, characterized by (6.36). In view of Proposition 7.17, we can write

$$\operatorname{Tr} Z(\rho) = \sum_{\text{poles}} \hat{\rho}(z_j), \tag{8.3}$$

where $\{z_j\}$ is the set of poles of the scattering operator $S(k)$ (counted with multiplicity).

Using (8.1), we will establish the following formula for $\operatorname{Tr} Z(\rho)$, which then sheds light on the right side of (8.3).

Proposition 8.1. *For $\rho \in C_0^\infty\big((2R,\infty)\big)$, we have*

$$\begin{aligned} Tr\, Z(\rho) &= Tr\,\big[U(\rho) - U_0(\rho)\big] \\ &= 2\, Tr \int \rho(t)\big[\cos t\sqrt{-\Delta} - \cos t\sqrt{-\Delta_0}\big]\, dt, \end{aligned} \tag{8.4}$$

where Δ is the Laplacian on $\Omega = \mathbb{R}^3 \setminus K$, with Dirichlet boundary condition, and Δ_0 the Laplacian on $\mathbb{R}^3$.

Proof. Using the facts that Tr AB = Tr BA and that $P_+P_- = P_-P_+$, we see from (8.1) that, for $\rho \in C_0^\infty\big((2R,\infty)\big)$,

$$\text{Tr}\, Z(\rho) = \text{Tr}\, P_-[U(\rho) - U_0(\rho)]P_+.$$

Now for any $t \geq 0$, $U(t) = U_0(t)$ on D^+, so $[U(\rho)-U_0(\rho)]P_+ = U(\rho)-U_0(\rho)$. Similarly, $P_-[U(\rho)-U_0(\rho)] = U(\rho)-U_0(\rho)$, so we have the first identity in (8.4). The second identity is elementary.

Combining (8.3) and (8.4), we have the identity

$$\text{Tr} \int \rho(t)\big[\cos t\sqrt{-\Delta} - \cos t\sqrt{-\Delta_0}\big]\, dt = \frac{1}{2}\sum_{\text{poles}} \hat{\rho}(z_j), \tag{8.5}$$

valid for any $\rho \in C_0^\infty((2R,\infty))$. This identity has been extended to all $\rho \in C_0^\infty(\mathbb{R}^+)$, by R. Melrose [Me1], using a more elaborate argument.

Note that (8.4) is equal to the trace of

$$\varphi\big(\sqrt{-\Delta}\big) - \varphi\big(\sqrt{-\Delta_0}\big), \tag{8.6}$$

with $\varphi(\lambda) = \hat{\rho}(\lambda) + \hat{\rho}(-\lambda)$. It is useful to note that, for any even $\varphi \in \mathcal{S}(\mathbb{R})$, the operator (8.6), given by an integral formula such as in the last line of (8.4) with $\rho = \hat{\varphi}$, has a Schwartz kernel that is smooth and rapidly decreasing at infinity, so that (8.6) is of trace class for this more general class of functions φ. (See Exercises 1–3 from §4.) Recall from (2.7) that if $\varphi \in C_0^\infty(\mathbb{R})$, then

$$\varphi\big(\sqrt{-\Delta}\big)v(x) = (2\pi)^{-3} \int_{\mathbb{R}^3}\int_{\Omega} u_+(x,\xi)\overline{u_+(y,\xi)}v(y)\varphi(|\xi|)\, dy\, d\xi, \tag{8.7}$$

where $u_+(x,\xi)$ are the generalized eigenfunctions of Δ on Ω defined by (1.32)–(1.33). It follows that, for such φ, the trace of (8.6) is equal to

$$\lim_{R\to\infty} (2\pi)^{-3} \int_{\mathbb{R}^3} \varphi(|\xi|)\tau_R(\xi)\, d\xi, \tag{8.8}$$

with

$$\tau_R(\xi) = \int_{B_R} \big(|u_+(x,\xi)|^2 - 1\big)\, dx, \tag{8.9}$$

where we set $u_+(x,\xi) = 0$, for $x \in K = \mathbb{R}^3 \setminus \Omega$, and $B_R = \{x : |x| \le R\}$.

In order to evaluate (8.9), we will calculate $\int |u_+(x,\xi)|^2 dx$ over $\Omega_R = \{x \in \Omega : |x| \le R\}$ via Green's theorem. Note that since $(\Delta + k^2)u_+ = 0$, for $|\xi| = k$, we have

$$(\Delta + k^2)\left(\frac{\partial u_+}{\partial k}\right) = -2ku_+, \quad |\xi| = k. \tag{8.10}$$

Hence, via Green's theorem, we have

$$\begin{aligned} \int_{\Omega_R} |u_+|^2 dx &= -\frac{1}{2k}\int_{\partial\Omega_R} \left(\frac{\partial^2 u_+}{\partial\nu\,\partial k}\overline{u}_+ - \frac{\partial u_+}{\partial k}\frac{\partial \overline{u}_+}{\partial \nu}\right) dS \\ &= \frac{1}{2k}\int_{|x|=R} \left(\frac{\partial^2 u_+}{\partial r\,\partial k}\overline{u}_+ - \frac{\partial u_+}{\partial k}\frac{\partial \overline{u}_+}{\partial r}\right) dS, \end{aligned} \tag{8.11}$$

since $u_+ = 0$ on ∂K. We want to evaluate the limit of (8.11) as $R \to \infty$. Extending (1.41), we can write

$$u_+(r\theta, k\omega) = e^{-ikr(\theta\cdot\omega)} + e^{ikr}B(r,\theta,\omega,k), \tag{8.12}$$

with

$$B \sim r^{-1}a(-\omega,\theta,k) + r^{-2}a_2(-\omega,\theta,k) + \cdots, \quad r \to \infty, \tag{8.13}$$

where $a(\omega,\theta,k)$ is the scattering amplitude and a_j are further coefficients. Differentiating (8.12) yields the following (unfortunately rather long) formula for the integrand in (8.11):

$$\begin{aligned} &\frac{\partial u_+}{\partial k}\frac{\partial \overline{u}_+}{\partial r} - \frac{\partial^2 u_+}{\partial r\,\partial k} \\ &\quad = 2kr(\theta\cdot\omega)^2 + i(\theta\cdot\omega) - 2ki\overline{B}\frac{\partial B}{\partial k} + irB\frac{\partial \overline{B}}{\partial r} \\ &\qquad + \left\{-i\overline{B}\frac{\partial(rB)}{\partial r} + \frac{\partial \overline{B}}{\partial r}\frac{\partial B}{\partial k} - \overline{B}\frac{\partial^2 B}{\partial r\,\partial k}\right\} \\ &\qquad + \Big\{2kr|B|^2 - krBe^{ikr[(\theta\cdot\omega)+1]}[(\theta\cdot\omega) - 1] \\ &\qquad\qquad - kr\overline{B}e^{-ikr[(\theta\cdot\omega)+1]}[(\theta\cdot\omega) - (\theta\cdot\omega)^2]\Big\} \end{aligned} \tag{8.14}$$

$$+ e^{ikr[(\theta\cdot\omega)+1]}\left\{ik\frac{\partial B}{\partial k}[(\theta\cdot\omega)-1] - i\frac{\partial(rB)}{\partial r} - \frac{\partial^2 B}{\partial r\partial k}\right\}$$

$$+ e^{-ikr[(\theta\cdot\omega)+1]}\left\{-ir(\theta\cdot\omega)\frac{\partial\overline{B}}{\partial r} + i(\theta\cdot\omega)\overline{B}\right\}.$$

A primary tool in the analysis of the integral of this quantity over $|x| = R$ will be the stationary phase method, which was established in Appendix B of Chap. 6.

We make some preliminary simplification of (8.14), using the fact that (8.11) is clearly real valued. Also, we can throw out some terms in (8.14) that contribute 0 in the limit $R \to \infty$, after being integrated over $|x| = R$. This includes all the terms in the first set of curly brackets above. Also, a stationary phase evaluation of the last two terms in the third set of curly brackets yields a 0 contribution in the limit $R \to \infty$. Thus, we can replace (8.14) by the real part of

$$\begin{aligned} &2kr(\theta\cdot\omega)^2 - 2ki\overline{B}\frac{\partial B}{\partial k} + irB\frac{\partial\overline{B}}{\partial r} \\ &+ \left\{2kr|B|^2 + krBe^{ikr[\theta\cdot\omega+1]}(1-\theta\cdot\omega)^2\right\} \\ &+ e^{ikr[\theta\cdot\omega+1]}\left\{ik\frac{\partial B}{\partial k}[\theta\cdot\omega-1] + ir\theta\cdot\omega\frac{\partial B}{\partial r} - i\theta\cdot\omega B\right\}. \end{aligned} \tag{8.15}$$

The first term on the right side of (8.15) integrates to $2k$ times $(4/3)\pi R^3$, exactly canceling out $\int_{|x|\le R} dx$. The contribution of the second and third terms to (8.11) is, in the limit $R \to \infty$,

$$-i\overline{a}\,\frac{\partial a}{\partial k} - \frac{i}{2k}|a|^2, \quad \text{integrated with respect to } \theta. \tag{8.16}$$

We can neglect the second term in (8.16), since it is imaginary.

Terms in (8.15) appearing with a factor $e^{\pm ikr[(\theta\cdot\omega)+1]}$ have an asymptotic behavior as $R \to \infty$ given by the stationary phase method, upon integration with respect to θ. The leading part in the terms within the first set of brackets is seen to be (upon taking the real part)

$$\frac{2k}{r}\left\{\int_{S^2} |a(-\omega,\theta,k)|^2 d\theta + \frac{4\pi}{k}\,\mathrm{Im}\,a(-\omega,-\omega,k)\right\}, \tag{8.17}$$

which cancels, by the optical theorem, (3.18). This cancelation is necessary since, if (8.17) were nonzero, one would get an infinite contribution to (8.11) as $R \to \infty$. What gives a finite contribution to (8.11) is the θ-integral of the next leading term in this part of (8.15); the contribution to (8.11) one gets from this, as $R \to \infty$, is (again upon taking the real part)

$$\int_{S^2} (a\overline{a}_2 + \overline{a}a_2)(-\omega, \theta, k)\,d\theta \tag{8.18}$$
$$+ \frac{4\pi}{k} \operatorname{Im} a_2(-\omega, -\omega, k) + \frac{8\pi}{k} \operatorname{Re} a(-\omega, -\omega, k).$$

The rest of the terms in (8.15) also give a finite contribution to (8.11) as $R \to \infty$, via stationary phase, namely $-1/2k$ times

$$4\pi \frac{\partial a}{\partial k} - \frac{4\pi}{k}\overline{a}, \quad \text{at } \theta = -\omega, \tag{8.19}$$

plus a term containing an oscillatory factor e^{-2ikr}, which disappears after integration with respect to ξ. This disappearance is guaranteed, since the limit in (8.8) as $R \to \infty$ does exist. Putting together (8.16)–(8.19), we arrive at a computation of (8.9).

All these contributions are expressed in terms of the scattering amplitude a, except for (8.18), which involves also the coefficient a_2 appearing in (8.13). Now a_2 is related to a in a simple fashion, because $(\Delta + k^2)(e^{ikr}B) = 0$. Expressing Δ in polar coordinates gives a sequence of relations among the coefficients in the expansion of B as $r \to \infty$. In particular, we get

$$2ika_2(\omega, \theta, k) = \Delta_2 a(\omega, \theta, k), \tag{8.20}$$

where Δ_2 denotes the Laplace operator on the sphere $\{|\theta| = 1\}$, applied to the second argument of $a(\omega, \theta, k)$. It follows that $\int a\overline{a}_2(-\omega, \theta, k)d\theta$ is purely imaginary, so the integral in (8.18) vanishes. In concert with the reciprocity formula (3.31), we can deduce that $4ika_2(\omega, \omega, k) = \Delta_2 a(\omega, \omega, k) + \Delta_1 a(-\omega, -\omega, k)$. Hence

$$4ik \int a_2(\omega, \omega, k)\,d\omega = \int (\Delta_1 + \Delta_2)a(\omega, \omega, k)\,d\omega = 0.$$

This disposes of the middle term in (8.18), upon integration with respect to ω. Thus, in addition to (8.16) and (8.19), the last term in (8.18) remains.

Consequently, we have

$$\lim_{R\to\infty} \frac{k^2}{(2\pi)^3} \iint_{B_R \times S^2} \left(|u_+(x, k\omega)|^2 - 1\right) dx\,d\omega$$
$$= \operatorname{Re} \int_{S^2} \left\{ \frac{-ik^2}{(2\pi)^3} \int_{S^2} \overline{a}\, \frac{\partial a}{\partial k}(-\omega, \theta, k)\,d\theta - \frac{1}{(2\pi)^2} a(-\omega, -\omega, k) \right. \tag{8.21}$$
$$\left. - \frac{k}{(2\pi)^2} \frac{\partial a}{\partial k}(-\omega, -\omega, k) \right\} d\omega.$$

On the other hand, $-(1/2\pi i)S(k)(dS^*/dk)$ has integral kernel

$$\begin{aligned}
\frac{-1}{2\pi i}\Big\{&\frac{-1}{2\pi i}\overline{a}(\omega,\theta,k)+\frac{-k}{2\pi i}\frac{\partial\overline{a}}{\partial k}(\theta,\omega,k)\\
&+\frac{k}{4\pi^2}\int\limits_{S^2}a(\omega,\tau,k)\overline{a}(\tau,\theta,k)\,d\tau\\
&+\frac{k^2}{4\pi^2}\int\limits_{S^2}a(\omega,\tau,k)\,\frac{\partial\overline{a}}{\partial k}(\tau,\omega,k)\,d\tau\Big\}.
\end{aligned} \tag{8.22}$$

Noting that the trace of $-(1/2\pi i)S(k)(dS^*/dk)$ must also be real, one sees that (8.21) is equal to the trace of this operator, which proves the following:

Proposition 8.2. *For even* $\varphi\in C_0^\infty(\mathbb{R})$,

$$Tr\,\big[\varphi(\sqrt{-\Delta})-\varphi(\sqrt{-\Delta_0})\big]=-\int_0^\infty\varphi(k)s'(k)\,dk, \tag{8.23}$$

with

$$s'(k)=\frac{1}{2\pi i}\,Tr\,\big(S(k)^*S'(k)\big)=-\frac{1}{2\pi i}\,Tr\,\big(S(k)S'(k)^*\big), \tag{8.24}$$

where $S(k)$ *is the scattering operator* (3.7).

An equivalent characterization of (8.24) is $s'(k)=ds(k)/dk$, with

$$s(k)=\frac{1}{2\pi i}\log\det S(k)=\frac{1}{2\pi}\arg\det S(k). \tag{8.25}$$

The quantity $s(k)$ is called the *scattering phase*. It is real, for $k\in\mathbb{R}$, since the scattering operator is unitary. To give yet another formulation, if we set

$$D(k)=\det S(k), \tag{8.26}$$

then

$$s'(k)=\frac{1}{2\pi i}\,\frac{D'(k)}{D(k)}. \tag{8.27}$$

By both (8.24) and (8.27) it is clear that $s'(k)$ extends from $k\in\mathbb{R}$ to a meromorphic function in the plane, with poles coinciding precisely with the poles of the scattering operator and their complex conjugates. For complex k, one replaces (8.24) by

$$s'(k)=\frac{1}{2\pi i}\,\mathrm{Tr}\,\big(S(\overline{k})^*S'(k)\big).$$

As stated, Propositions 8.1 and 8.2 apply in disjoint situations, but note that the left side of (8.23) is defined for any even $\varphi \in \mathcal{S}(\mathbb{R})$ and defines a continuous linear functional of such φ. Thus the right side of (8.23) is well defined, at least in a distributional sense; in particular, we have $s' \in \mathcal{S}'(\mathbb{R})$. Also, replacing ρ by its even part on the left side of (8.5) leaves this quantity unchanged. We deduce the following.

Proposition 8.3. *Let $\rho \in C_0^\infty\big((2R,\infty)\big)$. Then*

$$\frac{1}{2}\sum_{\text{poles}} \hat{\rho}(z_j) = -\int_0^\infty \varphi(k)s'(k)\,dk,$$

with

$$\varphi(k) = \frac{1}{2}\big[\hat{\rho}(k) + \hat{\rho}(-k)\big].$$

Equivalently, with $s(k) = -s(-k)$ for $k \in \mathbb{R}$,

$$\sum_{\text{poles}} \hat{\rho}(z_j) = \int_{-\infty}^\infty \hat{\rho}'(k)s(k)\,dk, \tag{8.28}$$

the integral interpreted a priori in the sense of tempered distributions.

In view of (8.27), this identity can be thought of as a “formal” consequence of the residue calculus, but a rigorous proof seems to require arguments as described above.

It can be proved that the integral above is actually absolutely convergent. Indeed, it has been shown that $s(k)$ has the asymptotic behavior

$$s(k) = C(\mathrm{vol}\ K)k^3 + O(k^2), \quad \text{as } k \to \infty, \ \text{in } \mathbb{R}. \tag{8.29}$$

This was established for K strictly convex by A. Majda and J. Ralston [MjR], and for K starshaped by A. Jensen and T. Kato [JeK]. We outline a proof for the starshaped case in the exercises (with a weaker remainder estimate).

The result (8.29) was extended to “nontrapping” K by V. Petkov and G. Popov [PP] and finally to general smooth K by Melrose [Me3]. Also, results of Melrose [Me1] extend (8.28) to all $\rho \in C_0^\infty(\mathbb{R}^+)$.

Exercises

1. Use the formula (8.24) to establish the following formula for $s'(k)$:

$$\begin{aligned} s'(k) &= C\int_{\partial K}\int_{S^2} (x\cdot\nu)\Big|\frac{\partial u_+}{\partial \nu}(x,k\theta)\Big|^2\,d\theta\,dS(x) \\ &= C\int_{S^2} \big((x\cdot\nu)\mathcal{N}(k)e_{k\theta}, \mathcal{N}(k)e_{k\theta}\big)_{L^2(\partial K)}\,d\theta. \end{aligned} \tag{8.30}$$

2. Conclude that if K is starshaped, so one can arrange $x\cdot\nu > 0$, then $s(k)$ is monotone.

3. Set $s(k) = -s(-k),\ k \in \mathbb{R}$, as in Proposition 8.3. Show that if K is starshaped, $s'(k)$ is a positive function that defines a tempered distribution on $\mathbb{R}$ and hence that $s(k)$ has a polynomial bound in k:
$$|s(k)| \le C\langle k\rangle^M.$$
4. Write (8.23) in the form
$$\operatorname{Tr}\bigl[\varphi(\sqrt{-\Delta}) - \varphi(\sqrt{-\Delta_0})\bigr] = \frac{1}{2}\int_{-\infty}^{\infty} \varphi'(k)s(k)\,dk,$$
for even $\varphi \in \mathcal{S}(\mathbb{R})$. If K is starshaped, Exercise 3 implies that the integral on the right is absolutely convergent. Use $\varphi(k) = \varphi_t(k) = e^{-tk^2}$ and the results on heat kernel asymptotics of Chap. 7 to deduce that
$$t\int_{-\infty}^{\infty} e^{-tk^2}\, ks(k)\,dk = (4\pi t)^{-3/2}\operatorname{vol} K + o(t^{-3/2}), \tag{8.31}$$
as $t \searrow 0$.
5. Show that Karamata's Tauberian theorem (established in §3 of Chap. 8) applies to (8.31) to yield
$$s(k) = C(\operatorname{vol} K)k^3 + o(k^3), \quad k \to \infty.$$
Evaluate C.

9. Scattering by a sphere

In this section we analyze solutions to problems of scattering by the unit sphere $S^2 \subset \mathbb{R}^3$, starting with the scattering problem
$$(\Delta + k^2)v = 0 \text{ on } \Omega,\ v = f \text{ on } S^2,\ r(\partial_r v - ikv) \to 0, \text{ as } r \to \infty, \tag{9.1}$$
where $\Omega = \{x \in \mathbb{R}^3 : |x| > 1\}$, the complement of the unit ball. We start by considering real k. This problem can be solved by writing the Laplace operator Δ on $\mathbb{R}^3$ in polar coordinates,
$$\Delta = \partial_r^2 + 2r^{-1}\partial_r + r^{-2}\Delta_S, \tag{9.2}$$
where Δ_S is the Laplace operator on the sphere S^2. Thus v in (9.1) satisfies
$$r^2\partial_r^2 v + 2r\partial_r v + (k^2r^2 + \Delta_S)v = 0, \tag{9.3}$$
for $r > 1$. In particular, if $\{\varphi_j\}$ is an orthonormal basis of $L^2(S^2)$ consisting of eigenfunctions of Δ_S, with eigenvalue $-\lambda_j^2$, and we write
$$v(r\omega) = \sum_j v_j(r)\varphi_j(\omega), \quad r \ge 1, \tag{9.4}$$

then the functions $v_j(r)$ satisfy

$$r^2 v_j''(r) + 2r v_j'(r) + (k^2 r^2 - \lambda_j^2) v_j(r) = 0, \quad r > 1. \tag{9.5}$$

As in (1.14), this is a modified Bessel equation, and the solution satisfying the radiation condition $r(v_j'(r) - ikv_j(r)) \to 0$ as $r \to \infty$ is of the form

$$v_j(r) = a_j r^{-1/2} H_{\nu_j}^{(1)}(kr), \tag{9.6}$$

where $H_\nu^{(1)}(\lambda)$ is the Hankel function, which arose in the proof of Lemma 1.2. We recall from (6.33) of Chap. 3 the integral formula

$$H_\nu^{(1)}(z) = \Bigl(\frac{2}{\pi z}\Bigr)^{1/2} \frac{e^{i(z-\pi\nu/2-\pi/4)}}{\Gamma(\nu+\frac{1}{2})} \int_0^\infty e^{-s} s^{\nu-1/2} \Bigl(1 - \frac{s}{2iz}\Bigr)^{\nu-1/2} ds. \tag{9.7}$$

This is valid for $\operatorname{Re} \nu > -1/2$ and $-\pi/2 < \arg z < \pi$. Also, in (9.6), ν_j is given by

$$\nu_j = \Bigl(\lambda_j^2 + \frac{1}{4}\Bigr)^{1/2}. \tag{9.8}$$

The coefficients a_j in (9.6) are determined by the boundary condition $v_j(1) = (f, \varphi_j)$, so

$$a_j = \frac{(f, \varphi_j)}{H_{\nu_j}^{(1)}(k)}. \tag{9.9}$$

Using these calculations, we can write the solution operator $\mathcal{B}(k)$ to (9.1), $v = \mathcal{B}(k) f$, as follows. Introduce the self-adjoint operator

$$A = \Bigl(-\Delta_S + \frac{1}{4}\Bigr)^{1/2}, \tag{9.10}$$

so

$$A\varphi_j = \nu_j \varphi_j. \tag{9.11}$$

Then

$$\mathcal{B}(k) f(r\theta) = r^{-1/2} \varkappa(A, k, kr) f(\theta), \tag{9.12}$$

where $\varkappa(\nu, k, kr) = H_\nu^{(1)}(kr)/H_\nu^{(1)}(k)$ and, for each k, r, $\varkappa(A, k, kr)$ is regarded as a function of the self-adjoint operator A. For convenience, we use the notation

$$\mathcal{B}(k) f(r\theta) = r^{-1/2} \frac{H_A^{(1)}(kr)}{H_A^{(1)}(k)} f(\theta), \quad \theta \in S^2. \tag{9.13}$$

Similar families of functions of the operator A will arise below.

Taking the r-derivative of (9.13), we have the following formula for the Neumann operator:

$$\mathcal{N}(k)f(\theta) = \left[k \frac{H_A^{(1)\prime}(k)}{H_A^{(1)}(k)} - \frac{1}{2} \right] f(\theta). \tag{9.14}$$

We also denote the operator on the right by $kQ(A,k) - 1/2$, with

$$Q(\nu, k) = \frac{H_\nu^{(1)\prime}(k)}{H_\nu^{(1)}(k)}. \tag{9.15}$$

We will want to look at the Green function and scattering amplitude, but first we derive some properties of the operators (9.13) and (9.14) which follow from the special nature of the operator defined by (9.10). The analysis of the spectrum of the Laplace operator on S^2 given in Chap. 8 shows that

$$\text{spec } A = \left\{ m + \frac{1}{2} : m = 0, 1, 2, \dots \right\}. \tag{9.16}$$

Now, as shown in Chap. 3, $H_{m+1/2}^{(1)}(\lambda)$ and the other Bessel functions of order $m + 1/2$ are all elementary functions of λ. We have

$$H_{m+1/2}^{(1)}(\lambda) = \left(\frac{2\lambda}{\pi} \right)^{1/2} h_m(\lambda), \tag{9.17}$$

where

$$\begin{aligned} h_m(\lambda) &= -i(-1)^m \left(\frac{1}{\lambda} \frac{d}{d\lambda} \right)^m \left(\frac{e^{i\lambda}}{\lambda} \right) \\ &= \lambda^{-m-1} p_m(\lambda) e^{i\lambda} \end{aligned} \tag{9.18}$$

and $p_m(\lambda)$ is a polynomial of order m in λ, given by

$$\begin{aligned} p_m(\lambda) &= i^{-m-1} \sum_{k=0}^{m} \left(\frac{i}{2} \right)^k \frac{(m+k)!}{k!(m-k)!} \lambda^{m-k} \\ &= i^{m-1} \lambda^m + \cdots + \frac{1}{2^m i} \frac{(2m)!}{m!}. \end{aligned} \tag{9.19}$$

Consequently,

$$r^{-\frac{1}{2}} \varkappa\left(m + \frac{1}{2}, k, kr\right) = \frac{h_m(kr)}{h_m(k)} = r^{-m-1} e^{ik(r-1)} \frac{p_m(kr)}{p_m(k)} \tag{9.20}$$

and

$$kQ\left(m+\frac{1}{2},k\right) = ik - \left(m+\frac{1}{2}\right) + k\,\frac{p_m'(k)}{p_m(k)}. \tag{9.21}$$

Each polynomial $p_m(\lambda)$ has m complex zeros $\{\zeta_{m1},\dots,\zeta_{mm}\}$, by the fundamental theorem of algebra, and the collection of all these ζ_{mj} is clearly the set of scattering poles for S^2. Note that (9.21) can be written as

$$kQ\left(m+\frac{1}{2},k\right) = ik - \left(m+\frac{1}{2}\right) + k\sum_{j=1}^{m}(k-\zeta_{mj})^{-1}. \tag{9.22}$$

We now look at the expression for the Green kernel $G(x,y,k)$ for the operator $(\Delta+k^2)^{-1}$, for k real. Thus we look for a solution to

$$(\Delta+k^2)u = f \text{ on } \Omega, \quad u = 0 \text{ on } \partial K, \tag{9.23}$$

satisfying the radiation condition at infinity, given $f \in C_0^\infty(\Omega)$. If we write

$$f(r\theta) = \sum_j f_j(r)\varphi_j(\theta), \tag{9.24}$$

using the eigenfunctions φ_j as before, and

$$u(r\theta) = \sum_j u_j(r)\varphi_j(\theta), \tag{9.25}$$

then the functions $u_j(r)$ satisfy

$$r^2u_j''(r) + 2ru_j'(r) + (k^2r^2-\lambda_j^2)u_j(r) = r^2 f_j(r), \quad r > 1, \tag{9.26}$$

together with the boundary condition $u_j(1) = 0$ and, as a consequence of the radiation condition, $r(u_j'(r) - iku_j(r)) \to 0$ as $r \to \infty$. We will write the solution in the form

$$u_j(r) = \int_1^\infty G_{\nu_j}(r,s,k) f_j(s)s^2\,ds, \tag{9.27}$$

where the kernel $G_\nu(r,s,k)$ remains to be constructed, as the Green kernel for the ordinary differential operator

$$L_\nu = \frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} + \left(k^2 - \frac{\lambda^2}{r^2}\right), \quad \nu = \left(\lambda^2+\frac{1}{4}\right)^{1/2}, \tag{9.28}$$

that is,

$$L_\nu g_\nu(\cdot,s,k) = s^{-2}\delta_s \quad \text{on } (1,\infty), \tag{9.29}$$

satisfying the boundary condition of vanishing at $r = 1$, together with the radiation condition as $r \to \infty$. This operator is self-adjoint on the space $L^2([1,\infty), r^2 dr)$, and $G_\nu(r,s,k)$ satisfies the symmetry condition

$$G_\nu(r,s,k) = G_\nu(s,r,k). \tag{9.30}$$

Thus it suffices to specify $G_\nu(r,s,k)$ for $r > s$. Since $G_\nu(\cdot,s,k)$ is annihilated by L_ν for $r > s$ and satisfies the radiation condition, we must have

$$G_\nu(r,s,k) = c_\nu(s,k) r^{-1/2} H_\nu^{(1)}(rk), \quad \text{for } r > s, \tag{9.31}$$

for some coefficient $c_\nu(s,k)$ that remains to be determined. In view of the symmetry (9.30), $c_\nu(\cdot,k)$ satisfies the same sort of modified Bessel equation, and so is a linear combination of $s^{-1/2} J_\nu(sk)$ and $s^{-1/2} H_\nu^{(1)}(sk)$. The boundary condition gives $c_\nu(s,k) = 0$ at $s = 1$, so we can write

$$c_\nu(s,k) = b_\nu(k) s^{-1/2} \Big(J_\nu(sk) - \frac{J_\nu(k)}{H_\nu^{(1)}(k)} H_\nu^{(1)}(sk) \Big), \tag{9.32}$$

where the coefficient $b_\nu(k)$ remains to be determined. This can be done by plugging (9.32) into (9.31), using (9.30) to write $G_\nu(r,s,k)$ for $r < s$, and examining the jump in the first derivative of g_ν with respect to r across $r = s$. Achieving (9.29) then specifies $b_\nu(k)$ uniquely. A straightforward calculation shows that $b_\nu(k)$ is the following constant, independent of ν and k, in view of the Wronskian relation:

$$b_\nu(k) = b = \frac{sk}{J_\nu(sk) H_\nu^{(1)\prime}(sk) - J_\nu'(sk) H_\nu^{(1)}(sk)} = \frac{\pi}{2i}. \tag{9.33}$$

To summarize, $G_\nu(r,s,k)$ is given by

$$\begin{aligned} &b(rs)^{-1/2} \Big(J_\nu(sk) - \frac{J_\nu(k)}{H_\nu^{(1)}(k)} H_\nu^{(1)}(sk) \Big) H_\nu^{(1)}(rk), \quad r \ge s, \\ &b(rs)^{-1/2} \Big(J_\nu(rk) - \frac{J_\nu(k)}{H_\nu^{(1)}(k)} H_\nu^{(1)}(rk) \Big) H_\nu^{(1)}(sk), \quad r \le s. \end{aligned} \tag{9.34}$$

In light of this, we can represent the Green kernel for the solution to (9.23) satisfying the radiation condition as follows. Using the Schwartz kernel theorem, we can identify an operator on functions on Ω with a (generalized) function of r, s with values in the space of operators on functions on the sphere S^2. With this identification, we have

$$G(x,y,k) = \frac{1}{4\pi} G_A(r,s,k), \tag{9.35}$$

where $|x| = r$, $|y| = s$, and A is given by (9.10). This is also the formula for the resolvent kernel of $(\Delta + k^2)^{-1}$, for $\operatorname{Im} k > 0$.

The formula (9.34) for $G_\nu(r, s, k)$, as a sum of two terms, corresponds to the decomposition (1.30) for $G(x, y, k)$, that is,

$$G(x, y, k) = g(x, y, k) + h(x, y, k), \tag{9.36}$$

where, as in (1.5),

$$g(x, y, k) = \frac{e^{ik|x-y|}}{4\pi|x-y|}. \tag{9.37}$$

Now recall from Proposition 1.6 how we can obtain the eigenfunctions

$$u(x, \xi) = e^{-ix\cdot\xi} + v(x, \xi) \tag{9.38}$$

from the asymptotic behavior of $G(x, y, k)$ as $|y| \to \infty$, via

$$h(x, r\omega, k) = \frac{e^{ikr}}{4\pi r} v(x, k\omega) + O(r^{-2}), \quad r \to \infty, \tag{9.39}$$

proved in (1.37). We therefore have

$$v(r\theta, k\omega) = \lim_{s\to\infty} s e^{-iks} h_A(r, s, k), \tag{9.40}$$

where we set

$$h_\nu(r, s, k) = 4\pi\, b(rs)^{-1/2} \frac{J_\nu(k)}{H_\nu^{(1)}(k)} H_\nu^{(1)}(sk) H_\nu^{(1)}(rk). \tag{9.41}$$

As before we identify a function of (θ, ω) with an operator on $C^\infty(S^2)$, with A acting on functions of θ. To evaluate the limit in (9.40), we can use

$$H_\nu^{(1)}(\lambda) = \left(\frac{2}{\pi\lambda}\right)^{1/2} e^{i(\lambda - \pi\nu/2 - \pi/4)} + o(\lambda^{-1/2}), \quad \lambda \to \infty, \tag{9.42}$$

which can be deduced from the integral formula (9.7). We obtain

$$v(r\theta, k\omega) = \mathcal{V}(A, r, k), \tag{9.43}$$

where

$$\mathcal{V}(\nu, r, k) = 2\pi^2 i \left(\frac{2}{\pi r k}\right)^{1/2} e^{-(1/2)\pi i(\nu+1/2)} \frac{J_\nu(k)}{H_\nu^{(1)}(k)} H_\nu^{(1)}(rk). \tag{9.44}$$

We can now evaluate the scattering amplitude, which satisfies

$$a(-\omega,\theta,k)=\lim_{r\to\infty} re^{-ikr}v(r\theta,k\omega), \tag{9.45}$$

according to (1.41). Using (9.42)–(9.44), we have

$$a(-\omega,\theta,k)=\frac{4\pi}{k}\frac{J_A(k)}{H_A^{(1)}(k)}e^{-\pi iA}. \tag{9.46}$$

In other words, if the right side is $\Xi(A)$, then

$$\Xi(A)f(\theta)=\int a(-\omega,\theta,k)f(\omega)\,d\omega.$$

Now, as shown in the study of harmonic analysis on spheres, in (4.44) of Chap. 8,

$$e^{-\pi iA}f(\omega)=-i\ f(-\omega),\quad f\in L^2(S^2), \tag{9.47}$$

so we can write

$$a(\omega,\theta,k)=-\frac{4\pi i}{k}\frac{J_A(k)}{H_A^{(1)}(k)}. \tag{9.48}$$

Recall that the scattering amplitude $a(\omega,\theta,k)$ is related to the scattering operator $S(k)$ by

$$S(k)=I+\frac{k}{2\pi i}A(k), \tag{9.49}$$

where $a(\omega,\theta,k)$ is the kernel of $A(k)$, by (3.14)–(3.15). In other words, $A(k)$ is the operator on the right side of (9.48). Therefore, the scattering operator itself has the form

$$S(k)=-\frac{H_A^{(2)}(k)}{H_A^{(1)}(k)} \tag{9.50}$$

in view of the identity

$$H_\nu^{(1)}(\lambda)+H_\nu^{(2)}(\lambda)=2\,J_\nu(\lambda). \tag{9.51}$$

We also note that

$$H_\nu^{(2)}(k)=\overline{H_\nu^{(1)}(k)}, \tag{9.52}$$

for ν and k real, so (9.50) explicitly displays the unitarity of the scattering operator, for real k.

The investigation of scattering by a sphere can be carried further, based on these formulas. For example, qualitative information on the zeros of $H_\nu^{(1)}(\lambda)$ yields qualitative information on the scattering poles. Some of the most delicate results on such scattering make use of the *uniform* asymptotic behavior of $H_\nu^{(1)}(\lambda)$ as ν and λ both tend to ∞. A treatment of this in a modern spirit, touching on more general approaches to diffraction problems, is given in [T2] and [MT1], and, in more detail, in Appendix C of [MT2]. Also, [Nus] gives a lengthy analysis of scattering by a sphere, from a more classical perspective.

Exercises

1. Derive from (9.7) that $H^{(1)}_{m+1/2}(z) = (2z/\pi)^{1/2} z^{-m-1} p_m(z) e^{iz}$, with

$$p_m(z) = \frac{(-i)^{m+1}}{m!} \int_0^\infty e^{-s} s^m \Big(z - \frac{s}{2i}\Big)^m \, ds.$$

   Show that this yields (9.19).

2. From the material on Bessel functions developed in Chap. 3, show that there is the Wronskian identity

$$H_\nu^{(1)\prime}(\lambda) H_\nu^{(2)}(\lambda) - H_\nu^{(1)}(\lambda) H_\nu^{(2)\prime}(\lambda) = \frac{C}{\lambda},$$

   and evaluate C. Using this, prove that $H_\nu^{(1)}(\lambda)$ is not zero for any $\lambda \in (0,\infty)$, $\nu \in (0,\infty)$.

3. Use results on the location of scattering poles from §7 to show that (9.13) and (9.14) imply $H_\nu^{(1)}(z)$ has zeros only in $\operatorname{Im} z < 0$, for $\nu = m + 1/2,\ m = 0, 1, 2, \dots$.
   It is known that this property holds for all $\nu \in [0,\infty)$. See [Wat], p. 511. There it is stated in terms of the zeros of $K_\nu(z)$, which is related to the Hankel function by $K_\nu(z) = (\pi i/2) e^{\pi i \nu/2} H_\nu^{(1)}(iz)$.

4. A formula of Nicholson (see [Olv], p. 340, or [Wat], p. 444) implies

$$J_\nu(z)^2 + Y_\nu(z)^2 = \frac{8}{\pi^2} \int_0^\infty K_0(2z \sinh t) \cosh 2\nu t \, dt,$$

   for $\operatorname{Re} z > 0$. Here $K_0(r)$ is Macdonald's function, the $\nu = 0$ case of the function mentioned in Exercise 3; cf. (6.50)–(6.54) in Chap. 3. $K_0(r)$ is a decreasing function of $r \in (0,\infty)$, and hence, for fixed $\nu > 0$, $J_\nu(x)^2 + Y_\nu(x)^2$ is a decreasing function of $x \in \mathbb{R}^+$. Show that this implies that

$$\mathcal{B}(k,r) : L^2(S^2) \longrightarrow L^2(S^2),$$

   defined by $\mathcal{B}(k,r) f(\theta) = \mathcal{B}(k) f(r\theta)$, has operator norm $\le r^{-1/2}$, for $r \ge 1$. Consequently,

$$\|\mathcal{B}(k) f\|_{L^2(\Omega, |x|^{-4} dx)} \le \|f\|_{L^2(S^2)}. \tag{9.53}$$

Using the integral formula (6.50) of Chap. 3, show that $rK_0(r)$ is decreasing on $\mathbb{R}^+$, hence that $|r^{1/2}H_\nu^{(1)}(r)|$ is decreasing on $\mathbb{R}^+$, for fixed $\nu > 0$. Use this to show that $\|\mathcal{B}(k,r)\| \le r^{-1}$ for $r \ge 1$, and sharpen (9.53).

5. Let $\mathfrak{A} = \{x \in \mathbb{R}^3 : 1 < |x| < 2\}$. With $u = \mathcal{B}(k)f$, use $\Delta u = -k^2 u$ and estimates derivable from Chap. 5, in concert with Exercise 4, to show that

$$\|\mathcal{B}(k)f\|_{H^2(\mathfrak{A})} \le C\|f\|_{H^{3/2}(S^2)} + Ck^2\|f\|_{L^2(S^2)}. \tag{9.54}$$

Deduce that

$$\|\mathcal{N}(k)f\|_{H^{1/2}(S^2)} \le C\|f\|_{H^{3/2}(S^2)} + k^2\|f\|_{L^2(S^2)}. \tag{9.55}$$

6. Show that

$$\big|k\, Q(m+1/2,k)\big| \le C\big(|k|+m+1\big),$$

for $k \in \mathbb{R},\ m \ge 0$. Deduce that, for $s \in \mathbb{R},\ k \in \mathbb{R}$,

$$\|\mathcal{N}(k)f\|_{H^s(S^2)} \le C_s\|f\|_{H^{s+1}(S^2)} + C_s|k|\cdot\|f\|_{H^s(S^2)}. \tag{9.56}$$

Compare this with the bound on $\mathcal{N}(k)$ derived in the previous exercise.
(*Hint*: Consider uniform asymptotic expansions of Bessel and Hankel functions, discussed in [Erd] and in Chap. 11 of [Olv]. Compare a related analysis in [T2].)

7. Suppose an obstacle K is contained in the unit ball $B_1 = \{|x| < 1\}$. Show that the solution to the scattering problem (1.1)–(1.3) is uniquely characterized on $\Omega_1 = (\mathbb{R}^3 \setminus K) \cap B_1$ as the solution to

$$(\Delta + k^2)v = 0 \text{ on } \Omega_1, \quad v = f \text{ on } \partial K, \quad \frac{\partial v}{\partial r} = \mathcal{N}(k)v \text{ on } S^2, \tag{9.57}$$

where $\mathcal{N}(k)$ is given by (9.14).

8. Derive the formulas of this section, particularly the formula analogous to (9.50) for $S(k)$, in the case of scattering by a sphere of radius R, centered at $p \in \mathbb{R}^3$, displaying explicitly the dependence of the various quantities on R and p.

9. It follows from (9.46)–(9.48) that the scattering amplitude for S^2 satisfies

$$a(\omega,\theta,k) = a(\theta,\omega,k) \quad \text{and} \quad a(\omega,\theta,k) = a(-\omega,-\theta,k). \tag{9.58}$$

Demonstrate these identities directly, for $\partial K = S^2$. How much more generally do they hold? Compare (3.31).

10. Suppose $v \in C^\infty(\mathbb{R}^3)$ solves $(\Delta + k^2)v = 0$. Show that $v(r\theta)$ has the form

$$v(r\theta) = \sum_j v_j(r)\varphi_j(\theta),$$

where φ_j is an eigenfunction of Δ_S, as in (9.4), and $v_j(r) = b_j\ j_{\nu_j - 1/2}(kr)$.
(*Hint*: $v_j(r)$ solves (9.5) and does not blow up as $r \to 0$.)
Deduce that, for some coefficients β_ℓ,

$$e^{ikr\omega\cdot\theta} = \sum_{\ell=0}^{\infty} \beta_\ell\ j_\ell(kr)\ P_\ell(\omega\cdot\theta), \tag{9.59}$$

where $P_\ell(t)$ are the Legendre polynomials, defined in (4.36) of Chap. 8.

As shown in (4.49) of Chap. 8, this formula holds with $\beta_\ell = (2\ell+1)i^\ell$, so

$$e^{ist} = \sum_{\ell=0}^{\infty} (2\ell+1)\, i^\ell\, j_\ell(s)\, P_\ell(t), \quad s \in \mathbb{R}, \quad t \in [-1,1]. \tag{9.60}$$

11. As noted in §4 of Chap. 8, the identity (9.59), with $\beta_\ell = (2\ell+1)i^\ell$, is equivalent to the assertion that $e^{ikr\omega\cdot\theta}$ is the integral kernel of

$$\Xi_{kr}(A) = 4\pi\, e^{(1/2)\pi i(A-1/2)}\, j_{A-1/2}(kr). \tag{9.61}$$

Show that this is in turn equivalent to the $r = 1$ case of (9.43)–(9.44). (*Hint*: Use (9.47).)

12. Derive explicit formulas for scattering objects (e.g., $S(k)$), in the case of the quantum scattering problem for $H = -\Delta + V$, when

$$V(x) = b, \quad \text{for } |x| \le R,$$
$$0, \quad \text{otherwise.}$$

Keep track of the dependence on b and R. If you fix $R = 1$ and let b decrease from $b = 0$ to the first value $b = -\beta_0$, below which $-\Delta + V$ has a negative eigenvalue, what happens to some of the scattering poles?

10. Inverse problems I

By "inverse problems" we mean problems of determining a scatterer ∂K in terms of information on the scattered waves. These problems are of practical interest. One might be given observations of the scattered wave $v(x, k\omega)$, for x in a region not far from ∂K, k belonging to some restricted set of frequencies (maybe a single frequency). Or one might have only the far field behavior, defined by the scattering amplitude $a(\omega, \theta, k)$, which we recall is related to $v(x, k\omega)$ by

$$v(r\theta, k\omega) \sim \frac{e^{ikr}}{r} a(-\omega, \theta, k), \quad r \to \infty. \tag{10.1}$$

In this section we examine the question of what scattering data are guaranteed to specify ∂K uniquely, at least if the data are measured perfectly.

It is useful to begin with the following explicit connection between the scattered wave $v(x, k\omega)$ and the scattering amplitude.

Proposition 10.1. *If $K \subset B_R(0)$, then, for $r \ge R$,*

$$a(-\omega, \theta, k) = -ik^{-1}\, e^{(1/2)\pi i(A-1/2)}\, h_{A-1/2}(kr)^{-1} g(\theta), \tag{10.2}$$

where $g = g_{r,\omega,k}$ is given by

$$g(\theta) = v(r\theta, k\omega). \tag{10.3}$$

As in §9, $h_{A-1/2}(kr)^{-1}$ is regarded as a family of functions of the self-adjoint operator A defined by (9.10). Recall that $h_m(\lambda)$ is given by (9.17)–(9.18).

Proof. This result follows easily from (9.12), which implies

$$\mathcal{B}(k)f(r\theta) = \frac{h_{A-1/2}(kr)}{h_{A-1/2}(k)} f(\theta), \tag{10.4}$$

for $f \in C^\infty(S^2)$. To prove (10.2)–(10.3), we can suppose without loss of generality that $R = 1$ and apply (10.4) with $f(\theta) = v_+(\theta, k\omega)$ to get

$$v_+(r\theta, k\omega) = \frac{h_{A-1/2}(kr)}{h_{A-1/2}(k)} f(\theta); \quad f(\theta) = v(\theta, k\omega).$$

Now compare the asymptotic behavior of both sides as $r \to \infty$. For the left side we have (10.1), while the behavior of the right side is governed by

$$h_m(kr) \sim i^{m-1} \frac{e^{ikr}}{kr} \tag{10.5}$$

by (9.18)–(9.19), so (10.2) follows.

Now we can invert the operator in (10.2), to write

$$v(r\theta, k\omega) = ik\, e^{-(1/2)\pi i(A-1/2)}\, h_{A-1/2}(kr) a(-\omega, \theta, k), \tag{10.6}$$

where the operator acts on functions of θ. The operator $h_{A-1/2}(kr)$ is an unbounded operator on $L^2(S^2)$; indeed, it is not continuous from $C^\infty(S^2)$ to $\mathcal{D}'(S^2)$, which has consequences for the inverse problem, as we will see in §11.

Suppose now that K_1 and K_2 are two compact obstacles in $\mathbb{R}^3$ giving rise to scattered waves which both agree with $v(x, k\omega)$ in some open set $\mathcal{O}$ in $\mathbb{R}^3 \setminus (K_1 \cup K_2)$. In other words $v_j(x, k\omega) = v(x, k\omega)$ for $x \in \mathcal{O}$, where the functions v_j are solutions to

$$(\Delta + k^2)v_j = 0 \text{ on } \mathbb{R}^3 \setminus K_j, \quad v_j = -e^{-ikx\cdot\omega} \text{ on } \partial K_j, \tag{10.7}$$

satisfying the radiation condition. We suppose the sets K_j have no "cavities"; that is, each $\Omega_j = \mathbb{R}^3 \setminus K_j$ has just one connected component. In this case, possibly the complement of $K_1 \cup K_2$ is not connected; cf. Fig. 10.1. We will let $\mathcal{U}$ denote the unbounded, connected component of this complement, and consider $\mathbb{R}^3 \setminus \mathcal{U}$, which we denote by $\widetilde{K}_2$, so $K_1 \subset \widetilde{K}_2$. This is illustrated in Figs. 10.1 and 10.2. We assume $\mathcal{O} \subset \mathcal{U}$. Let $\mathcal{R}$ be any connected component of the interior of $\widetilde{K}_2 \setminus K_1$. (Switch indices if $K_2 \subset K_1$.)

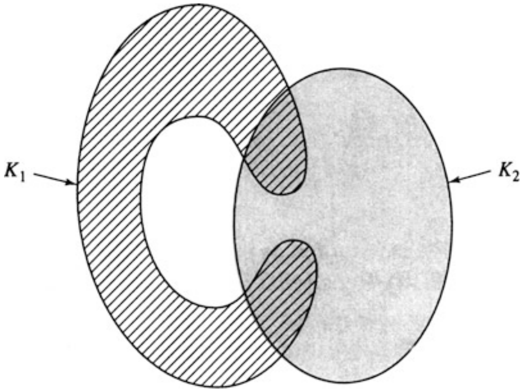

FIGURE 10.1 Two Obstacles

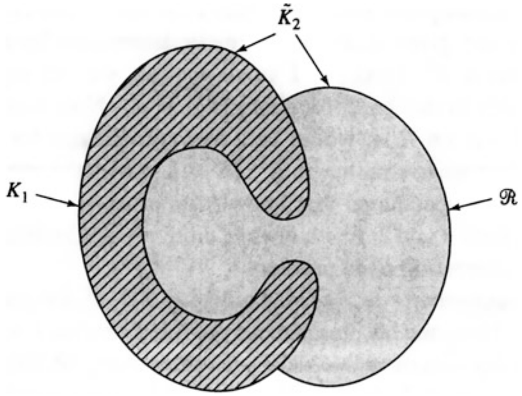

FIGURE 10.2 Filled Obstacle

The functions v_1 and v_2 described above agree on $\mathcal{U}$, since they are real analytic and agree on $\mathcal{O}$. Thus u_1 and u_2 agree on $\mathcal{U}$, where $u_j(x) = v_j(x) + e^{-ikx\cdot\omega}$. Since each u_j vanishes on ∂K_j, it follows that $u = u_1\big|_{\mathcal{R}}$ vanishes on $\partial\mathcal{R}$, so

$$(\Delta + k^2)u = 0 \text{ on } \mathcal{R}, \quad u = 0 \text{ on } \partial\mathcal{R}. \tag{10.8}$$

In fact, $u \in H_0^1(\mathcal{R})$. However, u does not vanish identically on $\mathcal{R}$. In particular, u provides an eigenfunction of Δ on each connected component $\mathcal{R}$ of the interior of $\widetilde{K}_2 \setminus K_1$, with Dirichlet boundary condition (and with eigenvalue $-k^2$) if u is not identically zero, and if the symmetric difference $K_1 \,\triangle\, K_2$ has nonempty interior. Now, there are circumstances where we can obtain bounds on

$$\dim \ker\, (\Delta + k^2)\big|_{H_0^1(\mathcal{R})} = d(k); \tag{10.9}$$

for example, if we know the obstacle is contained in a ball B_R. We then have the following uniqueness result:

Proposition 10.2. *Let $k \in (0, \infty)$ be fixed. Suppose $\Sigma = \{\omega_\ell\}$ is a subset of S^2 whose cardinality is known to be greater than $d(k)/2$. (If ω_ℓ and $-\omega_\ell$ both belong to Σ, do not count them separately.) Then knowledge of $v_+(x, k\omega_\ell)$ for x in an open set $\mathcal{O}$ uniquely determines the obstacle K. Hence knowledge of $a(-\omega_\ell, \theta, k)$ for $\theta \in S^2$ uniquely determines K.*

Proof. If K were not uniquely determined, there would be a nonempty set $\mathcal{R}$ such as described above. The corresponding $u_\ell(x) = v(x, k\omega_\ell) + e^{ikx\cdot\omega_\ell}$, together with their complex conjugates, which are all eigenfunctions on $\mathcal{R}$, must be linearly independent. Indeed, any linear dependence relation valid on $\mathcal{R}$ must continue on all of $\mathbb{R}^3 \setminus (K_1 \cap K_2)$; but near infinity, $u_\ell(x) = e^{ikx\cdot\omega_\ell} + O(|x|^{-1})$ guarantees independence.

We make a few complementary remarks. First, $a(-\omega, \theta, k)$ is analytic in its arguments, so for any given ω, k, it is uniquely determined by its behavior for θ in any open subset of S^2. Next, for k small enough, we can say that $d(k) = 0$, so uniqueness holds in that case, for a single $\omega = \omega_\ell$. Note that even when k^2 is an eigenvalue of $-\Delta$ on $\mathcal{R}$, it would be a real coincidence for a corresponding eigenfunction to happen to continue to $\mathbb{R}^3 \setminus (K_1 \cap K_2)$ with the appropriate behavior at infinity. It is often speculated that knowledge of $a(-\omega, \theta, k)$, for $\theta \in S^2$ (or an open set) and both k and ω fixed, always uniquely determines the obstacle K. This remains an interesting open problem.

Furthermore, suppose $a(-\omega, \theta, k)$ is known on $\theta \in S^2$, for a set $\{\omega_\ell\} \subset S^2$ and a set $\{k_m\} \subset \mathbb{R}^+$. Then one has uniqueness provided $\text{card}\{\omega_\ell\} > \min d(k_m)/2$. In particular, if $\{k_m\}$ consists of an interval I (of nonzero length), then $\min d(k_m) = 0$, so knowledge of $a(-\omega, \theta, k)$ for $\theta \in S^2$, $k \in I$, and a single ω uniquely determines K.

All of these considerations are subject to the standing assumption made throughout this chapter on the smoothness of ∂K. There are interesting cases of non-smooth obstacles, not equal to the closure of their interiors, to which the proof of Proposition 10.2 would not apply. We will discuss this further in §12.

We also mention that the method used to prove Proposition 10.2 is ineffective when one has the Neumann boundary condition. A uniqueness result in that case, using a different technique, can be found in [CK2]; see also [Isa].

One study that sheds light on the inverse problem is the linearized inverse problem. Here, given an obstacle K, denote by $\mathcal{B}_K(k)$ the solution operator (7.1)–(7.2) and by $S_K(k)$ the scattering operator (3.7), with corresponding scattering amplitude $a_K(\omega, \theta, k)$, as in (3.14)–(3.15). We want to compute the "derivative" with respect to K of these objects, and study their inverses.

More precisely, if K is given, ∂K smooth, we can parameterize nearby smooth obstacles by a neighborhood of 0 in $C^\infty(\partial K)$, via the correspondence that, to $\psi \in C^\infty(\partial K)$ (real-valued), we associate the image ∂K_ψ of ∂K under the map

$$F_\psi(x) = x + \psi(x)N(x), \quad x \in \partial K, \tag{10.10}$$

where $N(x)$ is the unit outward-pointing normal to ∂K, at x. Then, denote $\mathcal{B}_{K_\psi}(k)$ and $a_{K_\psi}(\omega,\theta,k)$ by $\mathcal{B}_\psi(k)$ and $a_\psi(\omega,\theta,k)$. We want to compute

$$D_\psi\, \mathcal{B}_K(k)f = \frac{\partial}{\partial s}\mathcal{B}_{s\psi}(k)f\big|_{s=0} \tag{10.11}$$

and $D_\psi\, a_K(\omega,\theta,k) = \partial_s a_{s\psi}(\omega,\theta,k)\big|_{s=0}$. The following is a straightforward exercise.

Proposition 10.3. *If f is smooth near ∂K and $v_\psi(x) = \partial_s \mathcal{B}_{s\psi}(k)f\big|_{s=0}$, for $x \in \mathbb{R}^3 \setminus K$, then $v_\psi(x)$ is uniquely characterized by*

$$\begin{aligned} (\Delta + k^2)v_\psi &= 0 \ \ \text{on } \mathbb{R}^3 \setminus K, \\ r\Big(\frac{\partial v_\psi}{\partial r} - ikv_\psi\Big) &\to 0, \ \ \text{as } r \to \infty, \\ v_\psi &= \psi(x)\Big(\mathcal{N}(k)f - \frac{\partial f}{\partial \nu}\Big) \ \ \text{on } \partial K. \end{aligned} \tag{10.12}$$

Here, $\mathcal{N}(k)$ is the Neumann operator, defined by (7.55). In other words,

$$D_\psi\, \mathcal{B}_K(k)f = \mathcal{B}_K(k)\Big\{\psi(x)\Big(\mathcal{N}(k)f - \frac{\partial f}{\partial \nu}\Big)\Big\}. \tag{10.13}$$

The linearized problem is to find ψ.

Therefore, for a given smooth obstacle K, granted that the operators $\mathcal{B}_K(k)$ and $\mathcal{N}(k)$ have been constructed (e.g., by methods of §7), we can to some degree reduce the linearized inverse problem for ψ to the following linear inverse problem:

Problem. Given (an approximation to) $w = \mathcal{B}(k)g(x)$ on $|x| = R_1$ (and assuming that $K \subset \{x : |x| < R_1\}$), find (an approximation to) g on ∂K.

As for finding $\mathcal{B}_K(k)$ and $\mathcal{N}(k)$ via an integral-equation method, we mention that an integral equation of the form (7.18) is preferable to one of the form (7.12), since it is very inconvenient to deal with the set of values of k for which (7.12) is not solvable. This point is made in many expositions on the subject, such as, [Co]. Solving (7.18) leads to the formula (7.32) for $\mathcal{B}_K(k)$.

We note that when we take $f = e^{-ik\omega\cdot x}$, the solution to the linearized inverse problem is unique:

Proposition 10.4. *Given K nonempty, smooth, and compact, such that $\mathbb{R}^3 \setminus K$ is connected, define*

$$\mathcal{L}_K(k,\omega) : H^s(\partial K) \to C^\infty(\mathbb{R}^3 \setminus K) \tag{10.14}$$

by

$$\mathcal{L}_K(k,\omega)\psi = D_\psi \mathcal{B}_K(k) f, \quad f(x) = e^{-ik\omega\cdot x}. \tag{10.15}$$

Then $\mathcal{L}_K(k,\omega)$ is always injective.

Proof. By (10.13), our claim is that

$$\mathcal{B}_K(k)\Big\{\psi(x)\Big(\mathcal{N}(k)f - \frac{\partial f}{\partial \nu}\Big)\Big\} = 0 \ \text{ on } \mathbb{R}^3 \setminus K \Longrightarrow \psi = 0 \text{ on } \partial K. \tag{10.16}$$

Since $\mathcal{B}_K(k)g\big|_{\partial K} = g$, the hypothesis in (10.16) implies $\psi(x)\big(\mathcal{N}(k)f - \partial_\nu f\big) = 0$ on ∂K, so it suffices to show that

$$\mathcal{N}(k)f - \frac{\partial f}{\partial \nu} \text{ vanishes on no open subset } \mathcal{O} \text{ of } \partial K \tag{10.17}$$

when $f(x) = e^{-ik\omega\cdot x}$. To see this, consider $w = \mathcal{B}(k)f - e^{-ik\omega\cdot x}$, which satisfies

$$(\Delta + k^2)w = 0 \text{ on } \mathbb{R}^3 \setminus K, \quad w = 0 \text{ on } \partial K. \tag{10.18}$$

If $\mathcal{N}(k)f - \partial_\nu f = 0$ on $\mathcal{O}$, then

$$\frac{\partial w}{\partial \nu} = 0 \ \text{ on } \mathcal{O}. \tag{10.19}$$

But if $\mathcal{O}$ is a nonempty, open subset of ∂K, then (10.18)–(10.19) imply that w is identically zero, by uniqueness in the Cauchy problem for $\Delta + k^2$. This is impossible, so the proof is complete.

Parallel to (10.13), we have

$$\begin{aligned} D_\psi a_K(-\omega,\theta,k) &= \mathcal{A}_K(k)\Big\{\psi(x)\Big(\mathcal{N}(k)f - \frac{\partial f}{\partial \nu}\Big)\Big\}(\theta), \\ f(x) &= e^{-ik\omega\cdot x}, \end{aligned} \tag{10.20}$$

where $\mathcal{A}_K(k)$ is as in (3.33)–(3.34). Note that (10.2) extends readily to the identity

$$\mathcal{A}_K(k)f = -ik^{-1}\, e^{(1/2)\pi i(A-1/2)}\, h_{A-1/2}(kr)^{-1}\, \mathcal{B}_{Kr}(k)f. \tag{10.21}$$

In view of the injectivity of the operator acting on $\mathcal{B}_{Kr}(k)f$, on the right side of (10.21) we see that, under the hypotheses of Proposition 10.4, we have

$$\widetilde{\mathcal{L}}_K(k,\omega) : H^s(\partial K) \longrightarrow C^\infty(S^2) \quad \text{injective,} \tag{10.22}$$

for each $k \in \mathbb{R},\ \omega \in S^2$, where

$$\widetilde{\mathcal{L}}_K(k,\omega)\psi = D_\psi \mathcal{A}_K(k) f, \quad f = e^{-ik\omega\cdot x}. \tag{10.23}$$

Exercises

1. Fix $k \in (0,\infty)$. Show that a given obstacle K is a ball centered at 0 if and only if
$$a(\omega,\theta,k) = a(R(\omega), R(\theta), k),$$
for every rotation $R : S^2 \to S^2$. (*Hint*: For the "if" part, make use of Proposition 10.2 to compare K and its image under a rotation.)
2. Suppose you are given that K is a ball, but you are not given its radius or its center. How can you determine these quantities from the scattering amplitude? How little information on $a(\omega,\theta,k)$ will suffice?
3. The set $\mathcal{R}$ arising in the proof of Proposition 10.2 might not have smooth boundary, so how do you know that $u = u_1\big|_{\mathcal{R}}$, which vanishes on $\partial\mathcal{R}$, belongs to $H_0^1(\mathcal{R})$?
4. Suppose K is known to be contained in the unit cube $Q = \{x \in \mathbb{R}^3 : 0 \le x_j \le 1\}$. Let $\omega \in S^2,\ k \in \mathbb{R}$ be fixed. Show that exact knowledge of $a(-\omega,\theta,k)$, for $\theta \in S^2$, uniquely determines K, as long as
$$|k| < \sqrt{6}\,\pi.$$
Given $\omega_1,\omega_2 \in S^2$, such that $\omega_1,\omega_2,-\omega_1$, and $-\omega_2$ are distinct, show that $a(-\omega_j,\theta,k)$, for k fixed, $\theta \in S^2,\ j = 1,2$, uniquely determines K, as long as
$$|k| < 3\pi.$$
Can you improve these results?
5. Give a detailed proof of Proposition 10.3.

11. Inverse problems II

In this section we describe some of the methods used to determine an obstacle K (approximately) when given a measurement of scattering data, and we deal with some aspects of the "ill-posed" nature of such an inverse problem.

For simplicity, suppose you know that $B_1 \subset K \subset B_{R_0}$, where $B_r = \{x \in \mathbb{R}^3 : |x| \le r\}$. Suppose you have a measurement of $a(-\omega,\theta,k)$ on $\theta \in S^2$, with k fixed and ω fixed. One strategy is to minimize

$$\begin{aligned}\Phi(f,K) = \|\mathcal{A}_1(k)f - a(-\omega,\cdot,k)\|^2_{L^2(S^2)} \\ + \|\mathcal{B}_{1K}(k)f + e^{-ik\omega\cdot x}\|^2_{L^2(\partial K)},\end{aligned} \tag{11.1}$$

with f and K varying over certain compact sets, determined by a priori hypotheses on the scatterer. This is close to some methods of Angell and Kleinman, Kirsch and Kress, as described at the end of [Co2].

Here we use the following notation: $\mathcal{A}_r(k) = \mathcal{A}_{B_r}(k)$, where $\mathcal{A}_K(k)$ is as in (10.20), that is,

$$\mathcal{A}_K(k)f(\theta) = \lim_{r\to\infty} re^{-ikr}\,\mathcal{B}_K(k)f(r\theta). \tag{11.2}$$

Also, if K_1 is contained in the interior of K_2, then $g = \mathcal{B}_{K_1}(k)f\big|_{\partial K_2}$ defines a bounded operator

$$\mathcal{B}_{K_1K_2}(k) : H^s(\partial K_1) \longrightarrow C^\infty(\partial K_2), \tag{11.3}$$

and if either $K_1 = B_r$ or $K_2 = B_\rho$, we use the notation $\mathcal{B}_{rK_2}(k)$ or $\mathcal{B}_{K_1\rho}(k)$; if both K_1 and K_2 are such balls, we use the notation $\mathcal{B}_{r\rho}(k)$.

More generally, we might have measurements of $a(-\omega_j, \theta, k)$ on $\theta \in S^2$, for a sequence of directions ω_j. Then one might take

$$\begin{aligned}\Phi = &\sum_{j=1}^N \big\|\mathcal{A}_1(k)f_j - a(-\omega_j, \cdot, k)\big\|^2_{L^2(S^2)} \\ &+ \sum_{j=1}^N \big\|\mathcal{B}_{1K}(k)f_j + e^{-ik\omega_j\cdot x}\big\|^2_{L^2(\partial K)}\end{aligned} \tag{11.4}$$

and minimize over $(f_1, \dots, f_N; K)$. One might also consider weighted sums, and perhaps stronger norms.

Note that

$$\mathcal{A}_1(k)f_j = \mathcal{A}_K(k)\mathcal{B}_{1K}(k)f_j. \tag{11.5}$$

The feasibility of approximating the actual scattered wave by such a function follows from the next lemma, provided K is connected and has connected complement.

Lemma 11.1. *If K_j are compact sets in $\mathbb{R}^3$ (with connected complement) such that $K_1 \subset \overset{\circ}{K}_2$, then for any $k \in \mathbb{R}$ the map $\mathcal{B}_{K_1K_2}(k)$ is injective. If also K_2 is connected, this map has dense range.*

Proof. If $u = \mathcal{B}_{K_1}(k)f$ vanishes on ∂K_2, then u restricted to $\mathbb{R}^3 \setminus K_2$ is an outgoing solution to the basic scattering problem (1.1), with $K = K_2$, so by the uniqueness of solutions to (1.1)–(1.3), we have $u = 0$ on $\mathbb{R}^3 \setminus K_2$. Then unique continuation forces $u = 0$ on $\mathbb{R}^3 \setminus K_1$, so the injectivity of (11.3) is established.

As for the second claim, note that if $y \in \overset{\circ}{K}_1$, then

$$|x-y|^{-1}e^{ik|x-y|} = g_y(x) \in \text{Range}\,\mathcal{B}_{K_1}(k). \tag{11.6}$$

Thus if $f \in L^2(\partial K_2)$ is in the orthogonal complement of the range of $\mathcal{B}_{K_1 K_2}(k)$, we deduce that

$$F(x) = \int_{\partial K_2} f(y) g_x(y)\, dS(y) \tag{11.7}$$

is zero for $x \in \overset{\circ}{K}_1$, hence for $x \in \overset{\circ}{K}_2$ (if $\overset{\circ}{K}_2$ is connected). Also material in §7 implies that F is continuous across ∂K_2 and is an outgoing solution of (1.1) on $\mathbb{R}^3 \setminus K_2$. The uniqueness of solutions to (1.1)–(1.3) forces $F = 0$ on $\mathbb{R}^3 \setminus K_2$. Since, by (7.25), the jump of $\partial_\nu F$ across ∂K_2 is proportional to f, this implies $f = 0$, proving denseness.

If K is not known to be connected, one could use several spherical bodies as domains of f_j in (11.4), provided it is known that each connected component of K contains one of them, as can be seen by a variant of the proof of Lemma 11.1.

Instead of minimizing (11.4) over $(f_1, \dots, f_N; K)$, an alternative is first to minimize the first term of (11.4), thus choosing f_j, within some compact set of functions, and then to pick K to minimize the second term, within some compact set of obstacles.

An attack pursued in [Rog] and [MTW] takes a guess K_μ of K, solves (approximately) a linearized inverse problem, given K_μ, and applies an iteration, provided by Newton's method, to approximate K. See also [Kir].

If one has a measurement of the scattered wave $v(x, k\omega)$ on the sphere $|x| = r$, say for k fixed and $\omega = \omega_1, \dots, \omega_N$, instead of a measurement of $a(-\omega, \theta, k)$, then parallel to (11.4) one might take

$$\begin{aligned} \Phi = &\sum_{j=1}^{N} \left\| \mathcal{B}_{1r}(k) f_j - v(\cdot, k\omega_j) \right\|^2_{L^2(S_r^2)} \\ &+ \sum_{j=1}^{N} \left\| \mathcal{B}_{1K}(k) f_j + e^{-ik\omega_j \cdot x} \right\|^2_{L^2(\partial K)} \end{aligned} \tag{11.8}$$

and minimize over $(f_1, \dots, f_N; K)$. Alternatively, first minimize the first sum over f_j (in some compact set of functions) and then minimize the second sum over K (in some compact set of obstacles).

In fact, a number of approaches to the inverse problem, when measurements of the scattering amplitude $a(-\omega, \theta, k)$ are given, start by first constructing an approximation to $v(x, k\omega)$ on some sphere $|x| = r$, such that $K \subset B_r$, and then proceed from there to tackle the problem of approximating K. Recall the relation established in §10:

$$v(r\theta, k\omega) = ik\, e^{-(1/2)\pi i (A - 1/2)}\, h_{A-1/2}(kr)\, a(-\omega, \theta, k), \tag{11.9}$$

where the operator acts on functions of θ. As noted there, the operator $h_{A-1/2}(kr)$ is a seriously unbounded operator on $L^2(S^2)$. In fact, this phenomenon is behind the ill-posed nature of recovering the near field behavior $v(x, k\omega)$ from the far field behavior defined by the scattering amplitude $a(-\omega, \theta, k)$. As this is one of the simplest examples of an ill-posed problem, we will discuss the following problem. Suppose you know that an obstacle K is contained in a ball B_{R_0}. Fix $k \in \mathbb{R}$, $\omega \in S^2$.

Problem A. Given an approximation $b(-\omega, \theta, k)$ to the scattering amplitude, with

$$\|a(-\omega, \cdot, k) - b(-\omega, \cdot, k)\|_{L^2(S^2)} \le \varepsilon, \tag{11.10}$$

how well can you approximate the scattered wave $v(x, k\omega)$, for x on the shell $|x| = R_1$, given $R_1 > R_0$?

As we have said, what makes this problem difficult is the failure of the operator $h_{A-1/2}(kr)$ appearing in (11.9) to be bounded, even from $C^\infty(S^2)$ to $\mathcal{D}'(S^2)$. Indeed, for fixed $s \in (0, \infty)$, one has the asymptotic behavior, as $\nu \to +\infty$,

$$H_\nu^{(1)}(s) \sim -i \left(\frac{2}{\pi\nu}\right)^{1/2} \left(\frac{2\nu}{es}\right)^\nu \tag{11.11}$$

(see the exercises) and hence

$$h_{\nu-1/2}(s) \sim -i\,(s\nu)^{-1/2} \left(\frac{2\nu}{es}\right)^\nu. \tag{11.12}$$

Consequently, an attempt to approximate $v_+(x, k\omega)$ for $x = R_1\theta$ by

$$v_0(\theta) = ik\,e^{-(1/2)\pi i(A-1/2)}\,h_{A-1/2}(kR_1) b(-\omega, \theta, k) \tag{11.13}$$

could lead to nonsense. We will describe a method below that is well behaved. But first we look further into the question of how well can we possibly hope to approximate $v(x, k\omega)$ on the shell $|x| = R_1$ with the data given.

In fact, it is necessary to have some further a priori bound on v_+ to make progress here. We will work under the hypothesis that a bound on $v(x, k\omega)$ is known on the sphere $|x| = R_0$:

$$\|v(R_0\theta, k\omega)\|_{L^2(S^2_\theta)} \le E. \tag{11.14}$$

Now, if we are given that (11.10) and (11.14) are both true and we have $b(-\omega, \theta, k)$ in hand (for ω, k fixed, $\theta \in S^2$), then we can consider the set $\mathcal{F}$ of functions $f(\theta)$ such that

$$\|f - b(-\omega, \cdot, k)\|_{L^2(S^2)} \le \varepsilon \tag{11.15}$$

and such that

$$\|k\, h_{A-1/2}(kR_0) f\|_{L^2(S^2)} \le E, \tag{11.16}$$

knowing that $\mathcal{F}$ is nonempty. We know that $a(\theta) = a(-\omega, \theta, k)$ belongs to $\mathcal{F}$, and that is all we know about $a(-\omega, \theta, k)$, in the absence of further data. The greatest accuracy of an approximation $v_1(\theta)$ to $v(R_1\theta, k\omega)$ that we can count on, measured in the $L^2(S^2)$ norm, is

$$\|v_1(\theta) - v(R_1\theta, k\omega)\|_{L^2(S^2_\theta)} \le 2\, M(\varepsilon, E), \tag{11.17}$$

where $M(\varepsilon, E)$ is defined as follows. Denote by

$$T_j : \mathcal{F} \longrightarrow L^2(S^2), \quad j = 0, 1, \tag{11.18}$$

the maps

$$T_j f(\theta) = ik\, e^{-(1/2)\pi i (A-1/2)}\, h_{A-1/2}(kR_j) f(\theta). \tag{11.19}$$

Then we set

$$2M(\varepsilon, E) = \sup \{\|T_1 f - T_1 g\|_{L^2(S^2)} : f, g \in \mathcal{F}\}, \tag{11.20}$$

that is,

$$M(\varepsilon, E) = \sup \{\|T_1 f\|_{L^2} : \|f\|_{L^2} \le \varepsilon \text{ and } \|T_0 f\|_{L^2} \le E\}. \tag{11.21}$$

One way to obtain as accurate as possible an approximation to v on $|x| = R_1$ would be to pick any $f \in \mathcal{F}$ and evaluate $T_1 f$. However, it might not be straightforward to obtain elements of $\mathcal{F}$. We describe a method, from [Mr2] and [MrV], which is effective in producing a "nearly best possible" approximation.

We formulate a more general problem. We have a linear equation

$$Sv = a, \tag{11.22}$$

where S is a bounded operator on a Hilbert space H, which is injective, but S^{-1} is unbounded (with domain a proper linear subspace of H). Given an approximate measurement b of a, we want to find an approximation to the solution v. This is a typical ill-posed linear problem. As a priori given information, we assume that

$$\|b - a\|_H \le \varepsilon, \quad \|T_0 a\|_H \le E. \tag{11.23}$$

T_0 is an auxiliary operator. In the example above, $H = L^2(S^2_\theta)$, and T_0 and T_1 are given by (11.19). Generalizing (11.20)–(11.21), we have a basic measurement of error:

$$M(\varepsilon, E) = \sup \{ \|T_1 f\|_H : \|f\|_H \le \varepsilon \text{ and } \|T_0 f\|_H \le E \} \tag{11.24}$$
$$= \frac{1}{2} \sup \{ \|T_1 f - T_1 g\|_H : f, g \in \mathcal{F} \},$$

where

$$\mathcal{F} = \{ f \in H : \|f - b\|_H \le \varepsilon, \|T_0 f\|_H \le E \}. \tag{11.25}$$

Now, if all one knows about a in (11.22) is that it belongs to $\mathcal{F}$, then the greatest accuracy of an approximation v_1 to the solution v of (11.22) one can count on is

$$\|v_1 - v\|_H \le 2M(\varepsilon, E). \tag{11.26}$$

This recaps the estimates in (11.14)–(11.21). Now we proceed. An approximation method is called nearly best possible (up to a factor γ) if it yields a $v_1 \in H$ such that

$$\|v_1 - v\|_H \le 2\gamma M(\varepsilon, E). \tag{11.27}$$

We now describe one nearly best possible method for approximating v, in cases where T_0 is a self-adjoint operator, with discrete spectrum accumulating only at $+\infty$. Then, pick an orthonormal basis $\{u_j : 1 \le j < \infty\}$ of H, consisting of eigenvectors, such that

$$T_0 u_j = \alpha_j u_j, \quad \alpha_j \nearrow +\infty. \tag{11.28}$$

When T_0 is given by (11.19), this holds as a consequence of (11.12). It is essential that the α_j be monotonic, so the eigenvectors need to be ordered correctly. Now let

$$f_\ell = P_\ell b, \quad P_\ell u = \sum_{j=1}^{\ell} (u, u_j) u_j. \tag{11.29}$$

Now let N be the *first* ℓ such that

$$\|f_\ell - b\|_H \le 2\varepsilon. \tag{11.30}$$

We then claim that

$$\|T_0 f_N\|_H \le 2E. \tag{11.31}$$

This can be deduced from:

Lemma 11.2. *If the set $\mathcal{F}$ defined by* (11.25) *is nonempty, and if $M+1$ is the first j such that $\alpha_j > E/\varepsilon$, then*

$$\|f_M - b\|_H \le 2\varepsilon \ \text{ and } \ \|T_0 f_M\|_H \le 2E. \tag{11.32}$$

Proof of lemma. The key facts about M are the following:

$$\begin{gathered}\|P_M g\| \le \varepsilon \Longrightarrow \|T_0 P_M g\| \le E,\\ \|(T_0(1-P_M)h\| \le E \Longrightarrow \|(1-P_M)h\| \le \varepsilon.\end{gathered} \tag{11.33}$$

We are given that there exists f such that

$$\|f - b\| \le \varepsilon \ \text{ and } \ \|T_0 f\| \le E. \tag{11.34}$$

The first part of (11.34) implies $\|P_M f - f_M\| \le \varepsilon$, which via the first part of (11.33) yields the second part of (11.32). The second part of (11.34) implies $\|T_0(1-P_M)f\| \le E$, which by the second part of (11.33) gives $\|(1-P_M)f\| \le \varepsilon$. Since $\|f-b\| \le \varepsilon$, this yields the first part of (11.32).

Having the lemma, we see that $N \le M$, so $\|T_0 f_N\| \le \|T_0 f_M\|$, giving (11.31). Then (11.30) and (11.31), together with (11.23), yield

$$\|f_N - a\|_H \le 3\varepsilon \ \text{ and } \ \|T_0(f_N - a)\|_H \le 3E. \tag{11.35}$$

We have established:

Proposition 11.3. *Under the hypotheses* (11.23)*, if we set $v_N = T_1 f_N$, where N is the smallest ℓ such that (11.30) holds, we have*

$$\|v_N - v\|_H \le 3\, M(\varepsilon, E). \tag{11.36}$$

Hence this method of approximating v is nearly best possible. Note that the value of the estimate E of (11.14) does not play an explicit role in the method described above for producing the approximation v_N; it plays a role in estimating the error $v_N - v$.

The method described above provides a technique for solving a certain class of ill-posed problems. Other related problems involve the analytic continuation of functions and solving backwards heat equations. Further discussions of this and other techniques, can be found in papers of K. Miller [Mr1], [Mr2], and references given there.

We now turn to the task of estimating $M(\varepsilon, E)$, for our specific problem, defined by (11.18)–(11.21). Thus, with $R_0 < R_1$, we want to estimate $\|h_{A-1/2}(kR_1)\|_{L^2}$, given that $\|f\|_{L^2} \le \varepsilon$ and $\|h_{A-1/2}(kR_0)f\|_{L^2} \le E$. If we also assume that k lies in a bounded interval, this is basically equivalent to estimating $\|A^{-1/2}e^{-\beta A}A^A f\|_{L^2}$, given that

$$\|f\|_{L^2} \le \varepsilon, \quad \|A^{-1/2}e^{-\alpha A}A^A f\|_{L^2} \le E,$$

where $e^{-\alpha} = 2/eR_0$ and $e^{-\beta} = 2/eR_1$, so $\alpha < \beta$. We can get a hold on this using the inequality

$$\nu^{-1/2}e^{-\beta\nu}\nu^\nu \le e^{-(\beta-\alpha)x}(\nu^{-1/2}e^{-\alpha\nu}\nu^\nu) + \sqrt{2}x^x, \tag{11.37}$$

valid for $\nu \ge 1/2,\ 0 < x < \infty$, to write

$$\|T_1 f\|_{L^2} \le C \inf_{x\in\mathbb{R}+} \Big(e^{-\gamma x}E + x^x\varepsilon\Big), \tag{11.38}$$

where $\gamma = \beta - \alpha$, given $\|f\| \le \varepsilon,\ \|T_0 f\| \le E$. While picking x to minimize the last quantity is not easy, we can obtain a reasonable estimate by picking

$$x = \frac{\log E/\varepsilon}{\log\log E/\varepsilon}, \tag{11.39}$$

in which case

$$e^{-\gamma x} = \Big(\frac{\varepsilon}{E}\Big)^{\gamma\alpha(E/\varepsilon)}, \quad \varepsilon x^x = E\Big(\frac{\varepsilon}{E}\Big)^{\beta(E/\varepsilon)},$$

with $\alpha(E/\varepsilon) = 1/(\log\log E/\varepsilon)$ and $\beta(E/\varepsilon) = (\log\log\log E/\varepsilon)/(\log\log E/\varepsilon)$. Consequently,

$$M(\varepsilon, E) \le CE\left[\Big(\frac{\varepsilon}{E}\Big)^{\gamma\alpha(E/\varepsilon)} + \Big(\frac{\varepsilon}{E}\Big)^{\beta(E/\varepsilon)}\right]. \tag{11.40}$$

As for the exponents that appear in (11.40), note the following values (to three digits):

ε/E	$\alpha(E/\varepsilon)$	$\beta(E/\varepsilon)$
10^{-2}	.655	.277
10^{-3}	.517	.341
10^{-4}	.450	.359
10^{-5}	.409	.366
10^{-6}	.381	.368
10^{-7}	.360	.368

The close agreement of the last two figures in the right column is due to the fact that $f(y) = (\log\log\log\ y)/(\log\log\ y)$ achieves its maximum value of $1/e$ at $y = e^{e^e} \approx 3.81 \times 10^6$ and is very slowly varying in this region. As for the close agreement of the two figures corresponding to $\varepsilon = 10^{-7}$, note that log log log $e^{e^e} = 1$. An estimate similar to (11.40) is also given in [Isa].

Even though the analysis in (11.13)–(11.39) does not directly deal with the problem of describing ∂K given an approximation $b(\omega, \theta, k)$ to the scattering amplitude $a(\omega,\theta,k)$, to some degree it reduces this problem to that of describing ∂K, given the solution $u = \mathcal{B}(k)f$ to the scattering problem (1.1)–(1.3) (i.e., to (7.1)), with $u(x)$ evaluated near $|x| = R_1$, for a certain class of boundary data, namely $f(x) = e^{-k\omega\cdot x}\big|_{\partial K}$ (where k and ω belong to a specified subset of $\mathbb{R}$ and S^2, respectively). One assumes it given that $K \subset \{x : |x| < R_0\}$, where $R_0 < R_1$. This reduction is an intermediate step in many studies of inverse problems. Thus Problem A is complemented by:

Problem B. Approximate $v = \mathcal{B}(k)f$ on $|x| = R_0$, given (an approximation to) v on $|x| = R_1$ and having some a priori estimate of v on $|x| = R_0$, but not on a smaller sphere.

Rescaling, we can consider the case $R_0 = 1,\ R_1 = R > 1$. By (10.4), we have

$$g = v(\theta) \text{ and } w = v(R\theta) \Longrightarrow g = \frac{h_{A-1/2}(k)}{h_{A-1/2}(kR)} w = \mathcal{C}_R(k)w, \tag{11.41}$$

where the last identity is the definition of the unbounded operator $\mathcal{C}_R(k)$ on $L^2(S^2)$. In view of (11.12), we have, for fixed $k \in (0,\infty),\ R > 1$,

$$\frac{h_{\nu-1/2}(k)}{h_{\nu-1/2}(kR)} \sim R^{\nu+1/2} = C\, e^{\gamma\nu}, \quad \nu \to +\infty, \tag{11.42}$$

where $\gamma = \log\ R > 0$.

Parallel to (11.14)–(11.21), we consider the problem of estimating $\mathcal{C}_R(k)w$ in $L^2(S^2)$, given a small bound on $\|w\|_{L^2(S^2)}$ (estimate on observational error) and an a priori bound on $\mathcal{C}_R(k)w$ in $H^\ell(S^2)$, for some $\ell > 0$. That is, we want to estimate

$$M(\varepsilon, E) = \sup\big\{\|\mathcal{C}_R(k)w\|_{L^2} : \|w\|_{L^2} \le \varepsilon,\ \|\mathcal{C}_R(k)w\|_{H^\ell} \le E\big\}. \tag{11.43}$$

Parallel to (11.37)–(11.38), we can attack this by writing

$$e^{\gamma\nu} \le (\nu x^{-1})^\ell e^{\gamma\nu} + e^{\gamma x}, \tag{11.44}$$

valid for $\nu, x \in (0,\infty)$. Thus, if $\|g\|_{H^\ell} = \|A^\ell g\|_{L^2}$, we have

$$\|\mathcal{C}_R(k)w\|_{L^2} \le C_k \inf_{x>0} \big(x^{-\ell}E + e^{\gamma x}\varepsilon\big). \tag{11.45}$$

We get a decent upper bound by setting $x = (1/2\gamma)\log(\ell E/\varepsilon\gamma)$. This yields

$$M(\varepsilon, E) \le C_k(2\gamma)^\ell E\Big(\log\frac{\ell E}{\gamma\varepsilon}\Big)^{-\ell} + C_k\Big(\frac{\ell E\varepsilon}{\gamma}\Big)^{1/2}. \tag{11.46}$$

This is bad news; ε would have to be terribly tiny for $M(\varepsilon, E)$ to be small. Fortunately, this is not the end of the story.

As a preliminary to deriving a more satisfactory estimate, we produce a variant of the "bad" estimate (11.46). Fix $\psi \in C_0^\infty(\mathbb{R})$, supported on $[-1, 1]$, such that $\psi(0) = 1$. Instead of (11.43), we estimate

$$M_\delta(\varepsilon, E) = \sup\left\{\|\psi(\delta A)\mathcal{C}_R(k)w\|_{L^2} : \|w\|_{L^2} \le \varepsilon, \|\mathcal{C}_R(k)w\|_{L^2} \le E\right\}. \tag{11.47}$$

We proceed via

$$\psi(\delta\nu)e^{\gamma\nu} \le e^{-\gamma x}\left[\psi(\delta\nu)e^{\gamma\nu}\right]e^{\gamma\nu} + e^{\gamma x}\psi(\delta\nu), \tag{11.48}$$

to get

$$\|\psi(\delta A)e^{\gamma A}w\|_{L^2} \le \inf_{x>0}\left[C(\gamma,\delta)e^{-\gamma x}C_k' E + e^{\gamma x}\varepsilon\right], \tag{11.49}$$

where

$$C(\gamma,\delta) = \sup_{\nu>0}\ \psi(\delta\nu)e^{\gamma\nu} \le e^{\gamma/\delta} = R^{1/\delta}. \tag{11.50}$$

Using (11.42) again, we have the estimate

$$M_\delta(\varepsilon, E) \le C_k\ R^{1/2\delta}\ \sqrt{E\varepsilon}. \tag{11.51}$$

Now we do want to be able to take δ small, to make $\psi(\delta A)f$ close to f, but $R^{1/2\delta} = e^{\gamma/2\delta}$ blows up very rapidly as $\delta \searrow 0$, so this gives no real improvement over (11.46). Compare (11.51) with the estimate

$$\|\psi(\delta A)\mathcal{C}_R(k)w\|_{L^2} \le C_k R^{1/\delta}\ \varepsilon, \tag{11.52}$$

when $\|w\|_{L^2} \le \varepsilon$, involving no use of the a priori estimate $\|\mathcal{C}_R(k)w\|_{L^2} \le E$.

We now show that a different technique yields a useful bound on the quantity $M_\delta(\varepsilon, E)$, when δ lies in the range $\delta > 1/k$.

Proposition 11.4. *Let $R > 1$ and $\alpha > 1$ be fixed. Then there is an estimate*

$$\|\psi(\delta A)\mathcal{C}_R(k)w\|_{L^2(S^2)} \le C\,\|w\|_{L^2(S^2)},\ \ \textit{for}\ \frac{\alpha}{k} \le \delta. \tag{11.53}$$

In particular, C is independent of k.

Proof. Since $\psi(\delta A)$ and $\mathcal{C}_R(k)$ commute, it suffices to show that

$$\|\mathcal{C}_R(k)w\|_{L^2} \le C\,\|w\|_{L^2},\ \ \text{for } w \in \text{Range } \chi(\delta A), \tag{11.54}$$

given $\alpha/k \le \delta$, where $\chi(\lambda)$ is the characteristic function of [0, 1]. Thus

$$0 \le A \le (k\alpha^{-1})I \quad \text{on Range } \chi(\delta A). \tag{11.55}$$

Equivalently, we claim that an outgoing solution $u(r,\omega)$ to the reduced wave equation $(\Delta + k^2)u = 0$ satisfies

$$\|u(1,\cdot)\|_{L^2(S^2)} \le C\|u(R,\cdot)\|_{L^2(S^2)}, \quad \text{for } u(R,\cdot) \in \text{Range } \chi(\delta A), \tag{11.56}$$

given $\alpha/k \le \delta$.

Now u satisfies the equation

$$\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} + \bigl(k^2 - r^{-2}L\bigr)u = 0, \quad L = A^2 - \frac{1}{4} = -\Delta_S. \tag{11.57}$$

We can replace $u(r,\omega)$ by $v(r,\omega) = ru(r,\omega)$, satisfying

$$\frac{\partial^2 v}{\partial r^2} + \bigl(k^2 - r^{-2}L\bigr)v = 0, \tag{11.58}$$

and it suffices to establish

$$\|v(1,\cdot)\|_{L^2(S^2)} \le C\|v(R,\cdot)\|_{L^2(S^2)}, \tag{11.59}$$

given $v(R,\cdot) \in \text{Range } \chi(\delta A)$, and assuming that $v = ru$, u an outgoing solution to (11.57); let us denote by $\mathcal{V}_{k\delta}$ the vector space of such functions v.

It will be convenient to use a family of norms, depending on r and k, given by

$$N_{kr}(v)^2 = \Bigl(\bigl(1 - (kr)^{-2}L\bigr)v, v\Bigr)_{L^2(S^2)} + k^{-2}\Bigl\|\frac{\partial v}{\partial r}\Bigr\|^2_{L^2(S^2)}, \tag{11.60}$$

where $v = v(r,\cdot) \in \mathcal{V}_{k\delta}$. Note that $\partial v/\partial r = [\mathcal{N}_r(k) + r^{-1}]v(r,\cdot)$, where $\mathcal{N}_r(k)$ is the Neumann operator (7.55), for the obstacle $\{|x| = r\}$. By (9.56), extended to treat balls of radius $r \in [1, R]$,

$$\begin{aligned} \|\mathcal{N}_r(k)f\|^2_{L^2(S^2)} &\le C\|f\|^2_{H^1(S^2)} + Ck^2\|f\|^2_{L^2(S^2)} \\ &\le Ck^2\Bigl[\bigl(k^{-2}Lf, f\bigr)_{L^2} + \|f\|^2_{L^2}\Bigr]. \end{aligned} \tag{11.61}$$

Now, by (11.55),

$$0 \le (kr)^{-2}L \le \alpha^{-2}I \quad \text{on Range } \chi(\delta A), \tag{11.62}$$

given $\alpha/k \le \delta$ and $r \ge 1$. Consequently, if $\alpha > 1$, we have constants $C_j \in (0,\infty)$, independent of k, such that

$$C_0 \|v(r)\|^2_{L^2(S^2)} \le N_{kr}(v(r))^2 \le C_1 \|v(r)\|^2_{L^2(S^2)}, \tag{11.63}$$

for all $v \in \mathcal{V}_{k\delta}$.

We now show that, for $v \in \mathcal{V}_{k\delta}$, $N_{kr}(v(r))^2 = E(r)$ is a monotonically increasing function of $r \in [1, R]$; this will establish the estimate (11.59) and hence complete the proof of Proposition 11.4. To see this, write

$$\begin{aligned}\frac{dE}{dr} &= 2\,\mathrm{Re}\Big(\big(1-(kr)^{-2}L\big)\frac{\partial v}{\partial r}, v\Big) \\ &\quad + \frac{2}{k^2 r^3}\,(Lv, v) + 2\,\mathrm{Re}\left(k^{-2}\frac{\partial^2 v}{\partial r^2}, \frac{\partial v}{\partial r}\right),\end{aligned} \tag{11.64}$$

and use (11.58) to replace $k^{-2}\,\partial^2 v/\partial r^2$ by $-\big(1-(kr)^{-2}L\big)v$. We obtain

$$\frac{dE}{dr} = \frac{2}{r}\big((kr)^{-2}Lv, v\big) \ge 0, \tag{11.65}$$

and the proof is complete.

We can place the analysis in (11.60)–(11.65) in the following more general context. Suppose

$$\frac{\partial^2 v}{\partial r^2} + A(r)v = 0, \tag{11.66}$$

where each $A(r)$ is positive-definite, all having the same domain. If we set

$$Q_r(v) = \big(A(r)v, v\big) + \|\partial_r v\|^2, \tag{11.67}$$

then

$$\begin{aligned}\frac{d}{dr}\,Q_r(v) &= 2\,\mathrm{Re}\big(A(r)v, \partial_r v\big) + 2\,\mathrm{Re}\big(\partial_r^2 v, \partial_r v\big) \\ &\quad + \big(A'(r)v, v\big) = \big(A'(r)v, v\big).\end{aligned} \tag{11.68}$$

If $A'(r)$ can be bounded by $A(r)$, then we have an estimate

$$\left|\frac{d}{dr}\,Q_r(v)\right| \le C\;Q_r(v). \tag{11.69}$$

Of course, if $A'(r)$ is positive-semidefinite, we have monotonicity of $Q_r(v)$, as in (11.65).

This result indicates that, using signals of wavelength $\lambda = 1/k$, one can expect to "regularize" inverse problems, to perceive details in an unknown obstacle on a length scale $\sim \lambda$. Further analytical estimates with the goal of making this precise are given in [T4]. This idea is very much consistent with intuition and experience. For example, a well-known statement on the limitations of an optical microscope is that if it has perfect optics, one can use it to examine microscopic detail on a length scale approximately equal to, but not smaller than, the wavelength of visible light.

We emphasize that this limitation applies to discerning detail on an obstacle whose diameter is much larger than $1/k$. If one has a single obstacle whose diameter is $\sim 1/k$, then one is said to be dealing with an inverse problem in the "resonance region," and, given some a priori hypotheses on the obstacle, one can hope to make out some details of its structure to a higher precision than one wavelength. This sort of problem is discussed in a number of papers on inverse problems, such as [ACK], [AKR], [JM], and [MTW].

Exercises

1. Using the power series for $J_\nu(z)$ given as (6.19) of Chap. 3, show that, for fixed $s \in (0, \infty)$, as $\nu \to +\infty$,
$$J_\nu(s) \sim (2\pi\nu)^{-1/2} \Big(\frac{es}{2\nu}\Big)^\nu.$$
Modify this argument to establish (11.11).
2. Generalize Proposition 11.3 to the case where different Hilbert spaces (or even different Banach spaces) H_j are involved, that is, $S : H_1 \to H_2$ and $T_j : V_j \to H_j,\ j = 0, 1$, where $V_j \subset H_2,\ a \in V_0 \cap V_1$.

12. Scattering by rough obstacles

In the previous sections we have restricted attention to scattering of waves by compact obstacles in $\mathbb{R}^3$ with *smooth* boundary. Here we extend some of this material to the case of compact $K \subset \mathbb{R}^3$, which is not assumed to have smooth boundary. We do assume that $\Omega = \mathbb{R}^3 \setminus K$ is connected.

The first order of business is to construct the solution (in a suitable sense) to the problem

$$(\Delta + k^2)v = 0 \text{ on } \Omega = \mathbb{R}^3 \setminus K, \quad v = f \text{ on } \partial K, \tag{12.1}$$

satisfying the radiation condition

$$|rv(x)| \le C, \quad r\Big(\frac{\partial v}{\partial r} - ikv\Big) \to 0, \ \text{ as } r \to \infty, \tag{12.2}$$

given $k > 0$. Our analysis will use a method from §5 of Chap. 5; we take compact K_j with smooth boundary such that

$$\overset{\circ}{K}_1 \supset\supset \overset{\circ}{K}_2 \supset\supset \cdots \supset\supset \overset{\circ}{K}_j \searrow K. \tag{12.3}$$

Set $\Omega_j = \mathbb{R}^3 \setminus K_j$, so $\Omega_j \nearrow \Omega$.

Let us assume $f\big|_{\partial K}$ is the restriction to ∂K of some $f \in C_0^2(\mathbb{R}^3)$. We will extend the method of proof of Theorem 1.3. For $\varepsilon > 0$, $j \in \mathbb{Z}^+$, let $w_{\varepsilon j} \in L^2(\Omega_j)$ be the solution to

$$(\Delta + (k+i\varepsilon)^2)w_{\varepsilon j} = h_\varepsilon \text{ on } \Omega_j, \quad w_{\varepsilon j}\big|_{\partial K_j} = 0, \tag{12.4}$$

where

$$h_\varepsilon = -(\Delta + (k+i\varepsilon)^2)f, \text{ on } \Omega. \tag{12.5}$$

Set $w_{\varepsilon j} = 0$ on K_j. Then set $v_{\varepsilon j} = f + w_{\varepsilon j}$, so

$$(\Delta + (k+i\varepsilon)^2)v_{\varepsilon j} = 0 \text{ on } \Omega_j, \quad v_{\varepsilon j}\big|_{\partial K_j} = f. \tag{12.6}$$

By methods of Chap. 5, §5, for fixed $\varepsilon > 0$, as $j \to \infty$, $w_{\varepsilon j} \to w_\varepsilon$ in $H_0^1(\Omega)$, the domain of $(-\Delta)^{1/2}$, when Δ is the self-adjoint operator on $L^2(\Omega)$ with the Dirichlet boundary condition on $\partial\Omega = \partial K$, and

$$w_\varepsilon = (\Delta + (k+i\varepsilon)^2)^{-1}h_\varepsilon \in H_0^1(\Omega). \tag{12.7}$$

We have $w_\varepsilon\big|_{\partial K} = 0$ in a generalized sense. It follows that $v_{\varepsilon j} \to v_\varepsilon = w_\varepsilon + f$ in $H^1(\Omega)$, and

$$(\Delta + (k+i\varepsilon)^2)v_\varepsilon = 0 \text{ on } \Omega, \quad v_\varepsilon\big|_{\partial K} = f, \tag{12.8}$$

the boundary condition holding in a generalized sense. Furthermore, v_ε is the unique solution to (12.8) with the property that $v_\varepsilon - f \in H_0^1(\Omega)$.

If $f \in C_0^2(\mathbb{R}^3)$ is supported in $B_A = \{|x| \le A\}$, then elliptic estimates imply $w_{\varepsilon j} \to w_\varepsilon$ in $C^\infty(\mathbb{R}^3 \setminus B_A)$, hence $v_{\varepsilon j} \to v_\varepsilon$ in $C^\infty(\mathbb{R}^3 \setminus B_A)$. It follows that, for any fixed $A_1 > A$, if $\Sigma = \{|x| = A_1\}$, then

$$v_\varepsilon(x) = \int_\Sigma \Big[v_\varepsilon(y)\frac{\partial g_\varepsilon}{\partial \nu} - g_\varepsilon \frac{\partial v_\varepsilon(y)}{\partial \nu}\Big]\, dS(y), \quad |x| > A_1, \tag{12.9}$$

where $g_\varepsilon = g(x, y, k+i\varepsilon)$ is given by (1.6). Compare with the identity (1.24).

We now state a result parallel to Theorem 1.3.

Theorem 12.1. *For v_ε constructed above, we have, as $\varepsilon \searrow 0$, a unique limit*

$$v_\varepsilon \to v = \mathcal{B}(k)f, \tag{12.10}$$

satisfying (12.1)–(12.2). Convergence occurs in the norm topology of the space $L^2(\Omega, \langle x\rangle^{-1-\delta}dx)$, for any $\delta > 0$, as well as weakly in $H^1(\Omega \cap \{|x| < R\})$, for any $R < \infty$, and the limit v satisfies the identity

$$v(x) = \int_{\Sigma} \Big[v(y) \frac{\partial g}{\partial \nu} - g \frac{\partial v(y)}{\partial \nu} \Big] \, dS(y), \quad |x| > A_1. \tag{12.11}$$

More generally, we can replace the boundary condition on v_ε *in* (12.8) *by* $v_\varepsilon = f_\varepsilon$ *on* ∂K*, with* $f_\varepsilon \to f$ *in* $C_0^2(\mathbb{R}^3)$*.*

As in §1, we begin the proof by establishing a uniqueness result.

Lemma 12.2. *Given* $k > 0$*, if* v *satisfies* (12.1)–(12.2) *with* $f = 0$*, then* $v = 0$*.*

Proof. Here, to say $v = 0$ on ∂K means $\chi v \in H_0^1(\Omega)$, for some $\chi \in C_0^\infty(\mathbb{R}^3)$, chosen so $\chi(x) = 1$ for $|x| \le A$. The proof of Proposition 1.1 works here, with a minor change in the identity (1.8). Namely, write the equation

$$(\Delta + k^2) v = 0 \text{ on } B_R \setminus K, \quad v\big|_{\partial K} = 0, \quad v\big|_{S_R} = v \tag{12.12}$$

in the weak form

$$\int_{B_R \setminus K} \big[-\langle dv, d\varphi \rangle + k^2 v \varphi \big] \, dx = \int_{S_R} v \frac{\partial \varphi}{\partial \nu} \, dS, \tag{12.13}$$

for all $\varphi \in H^1(B_R \setminus K)$ such that $\varphi\big|_{\partial K} = 0$ and φ is smooth near S_R. This applies to $\varphi = \overline{v}$, yielding

$$\int_{B_R \setminus K} \big[-\langle dv, d\overline{v} \rangle + k^2 v \overline{v} \big] \, dx = \int_{S_R} v \overline{v}_r \, dS. \tag{12.14}$$

Also, we can interchange the roles of v and $\overline{v}$ and subtract the resulting identity from (12.14), obtaining

$$\int_{S_R} (v \overline{v}_r - \overline{v} v_r) \, dS = 0, \tag{12.15}$$

as in (1.8). The rest of the proof proceeds exactly as in the proof of Proposition 1.1.

We continue with the proof of Theorem 12.1. Pick $R > A_1$ and set $\mathcal{O}_R = \Omega \cap \{|x| < R\}$. Parallel to Lemma 1.4, we have

Lemma 12.3. *Assume* $v_\varepsilon\big|_{\mathcal{O}_R}$ *is bounded in* $L^2(\mathcal{O}_R)$ *as* $\varepsilon \searrow 0$*. Then the conclusions of Theorem 12.1 hold.*

Proof. Fix $S \in (A_1, R)$. Elliptic estimates imply that if $\|v_\varepsilon\|_{L^2(\mathcal{O}_R)}$ is bounded, then

$$\|v_\varepsilon\|_{H^1(\mathcal{O}_S)} \le C. \tag{12.16}$$

Passing to a subsequence, which we continue to denote by v_ε, we have

$$v_\varepsilon \to v \text{ weakly in } H^1(\mathcal{O}_S). \tag{12.17}$$

Also, $w_\varepsilon = v_\varepsilon - f_\varepsilon \to w = v - f$, and for $\chi \in C_0^\infty(|x| < S)$ such that $\chi = 1$ on a neighborhood of K, we have $\chi w_\varepsilon \to \chi w$ in $H_0^1(\mathcal{O}_S)$. Thus $v\big|_{\partial K} = f$, in our current sense.

Since $\big(\Delta + (k + i\varepsilon)^2\big)v_\varepsilon = 0$ on $\mathcal{O}_R$, elliptic estimates imply that if $v_\varepsilon\big|_{\mathcal{O}_R}$ is bounded in $L^2(\mathcal{O}_R)$, then v_ε is bounded in $C^\infty(A < |x| < R)$. Thus we obtain (12.11) from (12.9), and (12.11) implies the radiation condition (12.2) and also the convergence $v_\varepsilon \to v$ in $L^2(\Omega, \langle x\rangle^{-1-\delta}dx)$.

So far we have convergence for subsequences, but by Lemma 12.2 this limit is unique, so Lemma 12.3 is proved. The proof of Theorem 12.1 is completed by:

Lemma 12.4. *The hypotheses of Theorem 12.1 imply that the family $\{v_\varepsilon\}$ is bounded in the space $L^2(\Omega, \langle x\rangle^{-1-\delta}dx)$, for any $\delta > 0$.*

The proof is the same as that of Lemma 1.5.

The fact that, for each $j \in \mathbb{Z}^+$, $v_{\varepsilon j}$ converges as $\varepsilon \to 0$ to a limit v_j solving the scattering problem

$$(\Delta + k^2)v_j = 0 \text{ on } \Omega_j, \quad v_j = f \text{ on } \partial K_j, \tag{12.18}$$

plus the radiation condition, is a consequence of Theorem 1.3. The following approximation result is useful. Extend v_j to be equal to f on $\Omega \setminus \Omega_j$.

Proposition 12.5. *For any $R \in (A, \infty)$, $\delta > 0$, we have*

$$v_j \to v \ \text{ in } C^\infty(\Omega_1) \cap L^2\big(\Omega, \langle x\rangle^{-1-\delta}dx\big). \tag{12.19}$$

More generally, we can replace the boundary condition by $v_j = f_j$ on ∂K_j, where $f_j \to f$ in $C_0^2(\mathbb{R}^3)$. Furthermore,

$$v_j \to v \ \text{ in } H^1(\mathcal{O}_R), \ \text{ in norm.} \tag{12.20}$$

Proof. To establish (12.19), an argument parallel to that used for Lemmas 1.4 and 12.3 shows that it suffices to demonstrate that $\{v_j\}$ is bounded in the space $L^2(\mathcal{O}_R)$, and then an argument parallel to that used for Lemmas 1.5 and 12.4 shows that indeed $\{v_j\}$ is bounded in $L^2(\Omega, \langle x\rangle^{-1-\delta}\, dxdx)$.

Arguments such as those used to prove Theorems 1.3 and 12.1 also show that $v_j \to v$ weakly in $H^1(\mathcal{O}_R)$. To get the norm convergence stated in (12.20), note that, parallel to (12.14),

$$\int_{\mathcal{O}_R} \big[-\langle dv_j, d\overline{v}_j\rangle + k^2|v_j|^2\big]\, dx = \int_{S_R} v_j(\partial_r \overline{v}_j)\, dS. \tag{12.21}$$

Since $v_j \to v$ in $L^2(\mathcal{O}_R)$ and $v_j \to v$ in C^∞ on a neighborhood of $S_R = \{|x| = R\}$, we have $\int_{\mathcal{O}_R} k^2|v_j|^2\, dx \to \int_{\mathcal{O}_R} k^2|v|^2\, dx$ and $\int_{S_R} v_j(\partial_r\overline{v}_j)\, dS \to \int_{S_R} v(\partial_r\overline{v})\, dS$. Consequently,

$$\int\limits_{\mathcal{O}_R} |dv_j|^2\, dx \longrightarrow \int\limits_{\mathcal{O}_R} |dv|^2\, dx,$$

so

$$\|v_j\|_{H^1(\mathcal{O}_R)} \longrightarrow \|v\|_{H^1(\mathcal{O}_R)}. \tag{12.22}$$

This, together with weak convergence, yields (12.20).

It is useful to note that if we extend v_j to be f_j on K_j and extend v to be f on K, then (12.20) can be sharpened to

$$v_j \longrightarrow v \text{ in } H^1(B_R), \quad \text{in norm.} \tag{12.23}$$

Now, we have a well-defined operator

$$\mathcal{B}_K(k) : C^2(K) \longrightarrow C^\infty(\mathbb{R}^3 \setminus K), \tag{12.24}$$

for any $k > 0$, any compact $K \subset \mathbb{R}^3$, extending (1.19). By (12.11), we have asymptotic results on $\mathcal{B}_K(k)f(x)$, as $|x| \to \infty$, of the same nature as derived in §1. In particular, the scattering amplitude $a_K(-\omega, \theta, k)$ is defined as before, in terms of the asymptotic behavior of $\mathcal{B}_K(k)f(r\theta)$, when $f(x) = -e^{-ikx\cdot\omega}$ on ∂K.

We next want to discuss the uniqueness of the scatterer, when K is not required to be smooth. A special case of Proposition 10.2 is that if $K \subset B_R$ is assumed to be smooth and one has fixed $k \in (0, \infty)$ and sufficiently many $\omega_\ell \in S^2$, then the knowledge that $a_K(-\omega_\ell, \theta, k) = 0,\ \forall\, \theta \in S^2$, implies K is empty. The appropriate statement of this result when K is not required to have any smoothness is the following:

Proposition 12.6. *Given compact $K \subset B_R$, fixed $k \in (0, \infty)$, and $\theta \in S^2$, then if*

$$a_K(-\omega_\ell, \theta, k) = 0, \quad \forall\, \theta \in S^2, \tag{12.25}$$

for a single $\omega_\ell \in S^2$, it follows that

$$\mathit{cap}\ K = 0. \tag{12.26}$$

Here "cap K" is the Newtonian capacity of K, which is discussed in detail in §6 of Chap. 11. One characterization is

$$\text{cap } K = \inf \left\{ \int |\nabla f(x)|^2\, dx : f \in C_0^\infty(\mathbb{R}^3),\ f = 1 \text{ on nbd of } K \right\}. \tag{12.27}$$

One can derive the estimate that if $f \in \operatorname{Lip}(\mathbb{R}^3)$ has compact support and $\lambda > 0$, then

$$\operatorname{cap}\big(\{x \in \mathbb{R}^3 : |f(x)| \geq \lambda\}\big) \leq \lambda^{-2} \|\nabla f\|_{L^2}^2. \tag{12.28}$$

See (6.64)–(6.65) of Chap. 11.

To prove the proposition, first note that, as in the proof of Proposition 10.2, the hypothesis (12.25) implies

$$u(x, k\omega_\ell) = e^{-ik\omega_\ell \cdot x}, \tag{12.29}$$

for $x \in \mathbb{R}^3 \setminus \widehat{K}$, the unbounded, connected component of $\mathbb{R}^3 \setminus K$; we may as well suppose that $\widehat{K} = K$. Fix $\varphi \in C_0^\infty(\mathbb{R}^3)$ so that $\varphi = 1$ on a neighborhood of K. Then (12.29) implies

$$\varphi(x) e^{-ik\omega_\ell \cdot x} \in H_0^1(\mathbb{R}^3 \setminus K). \tag{12.30}$$

Hence

$$\varphi \in H_0^1(\mathbb{R}^3 \setminus K). \tag{12.31}$$

We claim this implies $\operatorname{cap} K = 0$. Indeed, take $\varphi_\nu \in C_0^\infty(\mathbb{R}^3 \setminus K)$ so that $\varphi_\nu \to \varphi$ in H^1-norm. Then $f_\nu = \varphi - \varphi_\nu \in C_0^\infty(\mathbb{R}^3)$, $f_\nu = 1$ on a neighborhood of K, and $\int |\nabla f_\nu(x)|^2\, dx \to 0$. By (12.27), this implies $\operatorname{cap} K = 0$, so the proposition is proved.

We now want to compare two nonempty obstacles, K_1 and K_2, with identical scattering data $a(-\omega, \theta, k)$, perhaps for (ω, k) running over some set. Our next step is to push the arguments used in the proof of Proposition 10.2 to show that under certain conditions the symmetric difference $K_1 \,\triangle\, K_2$ has empty interior. After doing that, we will take up the question of whether $\operatorname{cap}(K_1 \,\triangle\, K_2) = 0$.

So, as in the proof of Proposition 10.2, suppose K_1 and K_2 are two compact obstacles in $\mathbb{R}^3$ giving rise to scattered waves v_j that agree on an open set $\mathcal{O}$ in the unbounded, connected component of $\mathbb{R}^3 \setminus (K_1 \cup K_2)$. In other words, $u_j = e^{-ik\omega\cdot x} + v_j(x, k\omega)$ has the properties

$$(\Delta + k^2) u_j = 0 \ \text{ on } \ \mathbb{R}^3 \setminus K_j, \quad \varphi u_j \in H_0^1(\mathbb{R}^3 \setminus K_j), \tag{12.32}$$

and $v_j = u_j - e^{-ik\omega\cdot x}$ satisfies the radiation condition. Here, we fix $\varphi \in C_0^\infty(\mathbb{R}^3)$ such that $\varphi = 1$ on a ball containing $K_1 \cup K_2$ in its interior. We suppose the sets K_j have no cavities, so each $\Omega_j = \mathbb{R}^3 \setminus K_j$ has just one connected component. As before, $\mathbb{R}^3 \setminus (K_1 \cup K_2)$ might not be connected, so let $\mathcal{U}$ be its unbounded component, and consider $\widetilde{K} = \mathbb{R}^3 \setminus \mathcal{U}$. If $K_1 \neq K_2$, then either K_1 or K_2 is a proper subset of $\widetilde{K}$. Let us suppose K_1 is.

Note that the functions u_1 and u_2 agree on $\mathcal{U}$, since they are real analytic and agree on $\mathcal{O}$.

Proposition 12.7. *For any* (ω, k) *for which* K_1 *and* K_2 *have identical scattering data (for all* θ*), so do* K_1 *and* $\widetilde{K}$.

Proof. By the uniqueness result, Lemma 12.2, it suffices to show that

$$u = u_1 = u_2 \quad \text{on } \mathcal{U} \tag{12.33}$$

has the property that $\varphi u \in H_0^1(\mathcal{U})$, for any $\varphi \in C_0^\infty(\mathbb{R}^3)$ equal to 1 on a neighborhood of $\widetilde{K}$. In turn, this is a consequence of the following general result.

Lemma 12.8. *Let* Ω_j *be open in* $\mathbb{R}^n$, $f_j \in H_0^1(\Omega_j)$. *Let* $\mathcal{O}$ *be a connected component of* $\Omega_1 \cap \Omega_2$. *Then*

$$f_1 = f_2 = f \quad \text{on } \mathcal{O} \Longrightarrow f \in H_0^1(\mathcal{O}). \tag{12.34}$$

Proof. It suffices to assume that the functions f_j are real-valued. The hypotheses imply $f_j^+ \in H_0^1(\Omega_j)$ and $f_1^+ = f_2^+ = f^+$ on $\mathcal{O}$. Thus it suffices to assume in addition that $f_j \geq 0$ on Ω_j. Now we can find $g_\nu \in C_0^\infty(\Omega_1)$ and $h_\nu \in C_0^\infty(\Omega_2)$ such that $g_\nu \to f_1$ in $H^1(\Omega_1)$ and $h_\nu \to f_2$ in $H^1(\Omega_2)$. Hence $g_\nu^+ \to f_1$ and $h_\nu^+ \to f_2$ in H^1-norm. Now

$$\varphi_\nu = \min\,(g_\nu^+, h_\nu^+)\big|_{\mathcal{O}}$$

has the properties

$$\varphi_\nu \in H_0^1(\mathcal{O}), \quad \varphi_\nu \to f \ \text{ in } \ H^1\text{-norm},$$

so (12.34) is proved.

Thus we replace K_2 by $\widetilde{K}$ (which we relabel as K_2), and we investigate whether $K_1 \subset K_2$ can have identical scattering data, for (k, ω) belonging to some set. We return to the considerations of the functions u_j, as in (12.32). (Now, $\mathcal{U} = \Omega_2$.)

Suppose $K_2 \setminus K_1$ has nonempty interior $\mathcal{R}$. Note that $\partial\mathcal{R} \setminus \partial\Omega_1 \subset \Omega_1 \cap \partial\Omega_2$. We claim that $w = u_1\big|_{\mathcal{R}}$ has the property

$$w \in H_0^1(\mathcal{R}). \tag{12.35}$$

This is a consequence of the following general result.

Lemma 12.9. *Let* $\mathcal{R} \subset \Omega$ *be open. Then*

$$f \in H_0^1(\Omega) \cap C(\Omega), \ f = 0 \ \text{on } \partial\mathcal{R} \setminus \partial\Omega \Longrightarrow f\big|_{\mathcal{R}} \in H_0^1(\mathcal{R}). \tag{12.36}$$

Proof. It suffices to assume f is real-valued. Then the hypotheses apply to f^+ and f^-, so it suffices to assume $f \geq 0$ on Ω. Take $f_\nu \in C_0^\infty(\Omega)$, $f_\nu \to f$ in H^1-norm. Then $f_\nu^+ \to f$ in H^1-norm. Also, if we define $\eta_\varepsilon(s)$ for $s \geq 0$ to be

$$\eta_\varepsilon(s) = \begin{array}{ll} 0 & \text{if } 0 \le s \le \varepsilon, \\ s-\varepsilon & \text{if } s \ge \varepsilon, \end{array} \tag{12.37}$$

and extend η_ε to be an odd function, we have $\eta_\varepsilon(f) \to f$ in H^1-norm. Now set

$$g_\nu = \min\bigl(f_\nu^+, \eta_{1/\nu}(f)\bigr)\big|_{\mathcal{R}}. \tag{12.38}$$

We see that $g_\nu \in H_0^1(\mathcal{R})$ and $g_\nu \to f$ in H^1-norm, so we have (12.36).

Now return to $w = u_1\big|_{\mathcal{R}}$. We claim this function satisfies the hypotheses of (12.36), with $\Omega = \Omega_1$. To see that w vanishes on $\partial\mathcal{R} \setminus \partial\Omega_1$, we use the fact that $u_1 = u_2$ on Ω_2 and argue that u_2 vanishes pointwise on a dense subset of $\partial\mathcal{R} \setminus \partial\Omega_1$. In fact, a dense subset satisfies an exterior sphere condition (with respect to Ω_2). Hence, a barrier construction (applied to the harmonic function $e^{ky}u_2$) gives this fact. Thus we have (12.35). Also,

$$(\Delta + k^2)w = 0 \quad \text{on } \mathcal{R}. \tag{12.39}$$

Of course, w is not identically zero on $\mathcal{R}$, so k^2 must be an eigenvalue of $-\Delta$ on $\mathcal{R}$. Hence we have the following parallel to Proposition 10.2. Suppose we have a bound on

$$\dim \ker(\Delta + k^2)\big|_{H_0^1(\mathcal{R})} = d(k). \tag{12.40}$$

Proposition 12.10. *Let K_1 and K_2 be arbitrary compact obstacles, with no cavities. Let $k \in (0,\infty)$ be fixed. Suppose $\Sigma = \{\omega_\ell\}$ is a subset of S^2 whose cardinality is known to be greater than $d(k)/2$. (If ω_ℓ and $-\omega_\ell$ both belong to Σ, do not count them separately.) Then*

$$a_{K_1}(-\omega_\ell,\theta,k) = a_{K_2}(-\omega_\ell,\theta,k), \quad \forall\, \omega_\ell \in \Sigma,\ \theta \in S^2 \tag{12.41}$$

implies that $K_1 \triangle K_2$ has empty interior.

We next show that under stronger hypotheses we can draw a stronger conclusion.

Proposition 12.11. *If $K_1 \subset K_2$ are compact sets without cavities in $\mathbb{R}^3$ and*

$$a_{K_1}(-\omega,\theta,k) = a_{K_2}(-\omega,\theta,k), \quad \forall\, \omega,\theta \in S^2,\ k \in (0,\infty), \tag{12.42}$$

then every compact subset of $K_2 \setminus K_1$ is negligible.

Proof. What we need to show is that if L is a compact subset of $K_2 \setminus K_1$, and if $\beta \in H^{-1}(\mathbb{R}^3)$ is supported on L, then $\beta = 0$.

By Proposition 12.10, the current hypotheses imply that $K_2 \setminus K_1$ has empty interior. Hence $K_2 \setminus K_1 \subset \partial\Omega_2$, so $K_2 \setminus K_1 = \Omega_1 \cap \partial\Omega_2$. Also, as in the considerations above, we have $u_1(x, k\omega) = u_2(x, k\omega)$ for all $x \in \Omega_2$; this time, for all $k\omega \in \mathbb{R}^3 \setminus 0$. We claim that this implies, for all compact $L \subset \Omega_1 \cap \partial\Omega_2$,

$$\beta \in H^{-1}(\mathbb{R}^3), \quad \operatorname{supp} \beta \subset L \Longrightarrow \langle u_1(\cdot, k\omega), \beta \rangle = 0. \tag{12.43}$$

To see this, we argue as follows (suppressing the parameters k, ω): Pick $\varphi \in C_0^\infty(\Omega_1)$, equal to 1 on a neighborhood of L. Then $\varphi u_2 \in H_0^1(\Omega_2)$, so we can take a sequence $f_\nu \in C_0^\infty(\Omega_2)$ such that $f_\nu \to \varphi u_2$ in $H^1(\Omega_2)$-norm. We can also regard f_ν as an element of $C_0^\infty(\Omega_1)$, and of course (f_ν) is Cauchy in $H_0^1(\Omega_1)$. We claim that

$$f_\nu \to \varphi u_1 \ \text{ in } \ H^1(\Omega_1). \tag{12.44}$$

Indeed, we have $f_\nu \to w$ for some $w \in H_0^1(\Omega_1)$, and hence $w = \varphi u_2$ on Ω_2. We want to show that $w = \varphi u_1$ on Ω_1. Since $u_1 = u_2$ on Ω_2, we have $w = \varphi u_1$ on Ω_2, so

$$\operatorname{supp}\, (w - \varphi u_1) \subset \Omega_1 \cap K_2, \tag{12.45}$$

a set that, in the current setting, is equal to $\Omega_1 \cap \partial\Omega_2$. Of course, if $\Omega_1 \cap \partial\Omega_2$ has three-dimensional Lebesgue measure 0, we can deduce $w = \varphi u_1$. If it has positive measure, we argue as follows. First, the characterization $w = \lim f_\nu$ clearly implies $w = 0$ on $\Omega_1 \cap K_2$. Furthermore, material on regular points discussed in Chap. 11, §6, applied to the harmonic functions $e^{ky} u_j(x)$, implies that $\lim_{x \to x_0} u_2(x, k\omega) = 0$ for all $x_0 \in \partial\Omega_2$ except for a set of interior capacity zero; in the current situation this implies $u_1 = 0$ a.e. on $\Omega_1 \cap \partial\Omega_2$. Hence we again have $w = \varphi u_1$, so (12.44) holds. On the other hand, (12.43) follows from (12.44).

Having (12.43), for all $\xi = k\omega \in \mathbb{R}^3 \setminus 0$, we deduce that, given $F \in C_0^\infty(\mathbb{R}^3 \setminus 0)$, if we set

$$g(x) = \int u_1(x, \xi) F(\xi)\, d\xi, \tag{12.46}$$

then

$$\beta \in H^{-1}(\mathbb{R}^3), \ \operatorname{supp} \beta \subset L \Longrightarrow \langle g, \beta \rangle = 0. \tag{12.47}$$

However, the set of functions of the form (12.46) is dense in $H_0^1(\Omega_1)$, by the isomorphism (4.10), which continues to hold in this setting. Thus

$$\beta \in H^{-1}(\mathbb{R}^3), \ \operatorname{supp} \beta \subset L \Longrightarrow \beta = 0. \tag{12.48}$$

As discussed in Chap. 5, this means L is negligible.

A consequence of material in §6 of Chap. 11 is that if a compact set is negligible, then its capacity is zero. Thus, by Proposition 12.11 (in conjunction with Proposition 12.7), when (12.43) holds, $K_1 \triangle K_2$ has "inner capacity" zero; see §6 of Chap. 11 for further discussion of inner capacity.

Exercises

1. Extend results of §§1–6 to obstacles considered here, with particular attention to results needed in the proof of Proposition 12.11.
2. Show that, for any open $\Omega \subset \mathbb{R}^n$, the map $u \mapsto |u|$ is continuous on $H^1(\Omega)$. Use this to justify the limiting arguments made in the proofs of Lemmas 12.8 and 12.9.

A. Lidskii's trace theorem

The purpose of this appendix is to prove the following result of V. Lidskii, which is used for (8.2):

Theorem A.1. *If A is a trace class operator on a Hilbert space H, then*

$$Tr\ A = \sum (dim\ V_j)\lambda_j, \tag{A.1}$$

where $\{\lambda_j : j \geq 1\} = Spec\ A \setminus \{0\}$ and V_j is the generalized λ_j-eigenspace of A.

We will make use of elementary results about trace class operators, established in §6 of Appendix A, Functional Analysis. In particular, if $\{u_j\}$ is any orthonormal basis of H, then

$$\text{Tr}\ A = \sum (Au_j, u_j), \tag{A.2}$$

the result being independent of the choice of orthonormal basis, provided A is trace class.

To begin the proof, let $E_\ell = \bigoplus_{j\leq \ell} V_j$, and let $P_\ell = Q_1 + \cdots + Q_\ell$ denote the orthogonal projection of H onto E_ℓ. Thus

$$AP_\ell = P_\ell A P_\ell,$$

restricted to E_ℓ, has spectrum $\{\lambda_j : 1 \leq j \leq \ell\}$. We will choose an orthonormal set $\{u_j : j \geq 1\}$ according to the following prescription: $\{u_j : 1 + \dim E_{\ell-1} \leq j \leq \dim E_\ell\}$ will be an orthonormal basis of $\mathcal{R}(Q_\ell)$, with the property that $Q_\ell A Q_\ell$ (restricted to $\mathcal{R}(Q_\ell)$) is upper triangular. That this can be done is proved in Theorem 4.7 of Chap. 1. Note that $\{u_j : 1 \leq j \leq \dim E_\ell\}$ is then an orthonormal basis of E_ℓ, with respect to which $AP_\ell = P_\ell A P_\ell$ is upper triangular. It follows that the diagonal entries of $P_\ell A P_\ell\big|_{E_\ell}$ with respect to this basis are precisely λ_j, $1 \le j \le \ell$, counted with multiplicity dim V_j. Inductively, we conclude that the diagonal entries of each block $Q_\ell A Q_\ell$ consist of dim V_ℓ copies of λ_ℓ.

Let H_0 denote the closed linear span of $\{u_j : j \ge 1\}$, and H_1 the orthogonal complement of H_0 in H, and let R_ν be the orthogonal projection of H on H_ν. We can write A in block form

$$A = \begin{pmatrix} A_0 & B \\ 0 & A_1 \end{pmatrix}, \tag{A.3}$$

where $A_\nu = R_\nu A R_\nu$, restricted to H_ν. Clearly, A_0 and A_1 are trace class and, by the construction above plus (A.2), we have

$$\operatorname{Tr} A_0 = \sum (\dim V_j)\lambda_j. \tag{A.4}$$

Thus (A.1) will follow if we can show that Tr $A_1 = 0$. If $H_1 = 0$, there is no problem.

Lemma A.2. *If $H_1 \neq 0$, then Spec $A_1 = \{0\}$.*

Proof. Suppose Spec A_1 contains an element $\mu \neq 0$. Since A_1 is compact on H_1, there must exist a unit vector $v \in H_1$ such that $A_1 v = \mu v$. Let $\mathcal{H} = H_0 + (v)$. Note that

$$Av = \mu v + w, \quad w \in H_0.$$

Hence $\mathcal{H}$ is invariant under A; let $\mathcal{A}$ denote A restricted to $\mathcal{H}$. Of course, H_0 is invariant under $\mathcal{A}$, and $\mathcal{A}$ restricted to H_0 is A_0.

Note that both $T_\mu = A_0 - \mu I$ (on H_0) and $\mathcal{T}_\mu = \mathcal{A} - \mu I$ (on $\mathcal{H}$) are Fredholm operators of index zero, and that

$$\operatorname{Codim} \mathcal{T}_\mu(\mathcal{H}) = 1 + \operatorname{Codim} T_\mu(H_0).$$

Hence

$$\operatorname{Dim} \operatorname{Ker}(\mathcal{A} - \mu I) = 1 + \operatorname{Dim} \operatorname{Ker}(A_0 - \mu I).$$

It follows that the μ-eigenspace of $\mathcal{A}$ is bigger than the μ-eigenspace of A_0. But this is impossible, since by construction, for any $\mu \neq 0$, the μ-eigenspace of A_0 is the entire μ-eigenspace of A. Thus the lemma is proved.

A linear operator K is said to be *quasi-nilpotent* provided Spec $K = \{0\}$. If this holds, then $(I + zK)^{-1}$ is an entire holomorphic function of z. The convergence of its power series implies

$$\sup_j |z|^j \|K^j\| < \infty, \quad \forall\, z \in \mathbb{C}, \tag{A.5}$$

a condition that is in fact equivalent to Spec $K = \{0\}$. To prove Theorem A.1, it suffices to demonstrate the following.

Lemma A.3. *If K is a trace-class operator on a Hilbert space and K is quasi-nilpotent, then Tr $K = 0$.*

To prove Lemma A.3, we use results on the determinant established in §6 of Appendix A, Functional Analysis. Thus, we consider the entire holomorphic function

$$\varphi(z) = \det(I + zK), \tag{A.6}$$

which is well defined for trace class K. By (6.45) of Appendix A,

$$|\varphi(z)| \le C_\varepsilon e^{\varepsilon|z|}, \quad \forall\, \varepsilon > 0. \tag{A.7}$$

Also, by Proposition 6.16 of Appendix A, $\varphi(z) \neq 0$ whenever $I + zK$ is invertible. Now, if K is quasi-nilpotent, then, as remarked above, $I + zK$ is invertible for all $z \in \mathbb{C}$. Hence $\varphi(z)$ is nowhere vanishing, so we can write

$$\varphi(z) = e^{f(z)}, \tag{A.8}$$

with $f(z)$ holomorphic on $\mathbb{C}$. Now (A.7) implies Re $f(z) \le C_\varepsilon + \varepsilon|z|$ for all $\varepsilon > 0$, and a Harnack inequality argument applied to this gives

$$|\text{Re } f(z)| \le C'_\varepsilon + \varepsilon|z|, \quad \forall\, \varepsilon > 0, \tag{A.9}$$

See Chap. 3, §2, Exercises 13–16. The estimate (A.9) in turn (e.g., by Proposition 4.6 of Chap. 3) implies that Re f is constant, so f is constant, and hence φ is constant. But, by (6.41) of Appendix A, we have

$$\text{Tr } K = \varphi'(0), \tag{A.10}$$

so the lemma is proved. Hence the proof of Theorem A.1 is complete.

A proof of Lidskii's theorem—avoiding the first part of the argument given above, and simply using determinants, but making heavier use of complex function theory—is given in [Si2].

References

[Ag] S. Agmon, *Lectures on Elliptic Boundary Problems*, Van Nostrand, New York, 1964.

[ACK] T. Angell, D. Colton, and A. Kirsch, The three dimensional inverse scattering problem for acoustic waves, *J. Diff. Eq.* 46(1982), 46–58.

[AKR] T. Angell, R. Kleinman, and G. Roach, An inverse transmission problem for the Helmholtz equation, *Inverse Probl.* 3(1987), 149–180.

[Ber] A. M. Berthier, *Spectral Theory and Wave Operators for the Schrödinger Equation*, Pitman, Boston, 1982.

[Bs] V. Buslaev, Scattered plane waves, spectral asymptotics, and trace formulas in exterior problems, *Dokl. Akad. Nauk. SSSR* 197(1971), 1067–1070.

[Co] D. Colton, The inverse scattering problem for time-harmonic accoustic waves, *SIAM Rev.* 26(1984), 323–350.

[Co2] D. Colton, *Partial Differential Equations, an Introduction,* Random House, New York, 1988.

[CK] D. Colton and R. Kress, *Integral Equation Methods in Scattering Theory*, Wiley, New York, 1983.

[CK2] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer, New York, 1992.

[CH] R. Courant and D. Hilbert, *Methods of Mathematical Physics II*, J. Wiley, New York, 1966.

[Ei] D. Eidus, On the principle of limiting absorption, *Mat. Sb.* 57(1962), 13–44.

[Erd] A. Erdélyi, *Asymptotic Expansions*, Dover, New York, 1956.

[HeR] J. Helton and J. Ralston, The first variation of the scattering matrix, *J. Diff. Equations* 21(1976), 378–394.

[HiP] E. Hille and R. Phillips, *Functional Analysis and Semigroups*, Colloq. Publ. Vol. 31, AMS, Providence, RI, 1957.

[Ho] L. Hörmander, *The Analysis of Linear Partial Differential Operators*, Vols. 3 and 4, Springer, New York, 1985.

[Isa] V. Isakov, Uniqueness and stability in multi-dimensional inverse problems, *Inverse Probl.* 9(1993), 579–621.

[JeK] A. Jensen and T. Kato, Asymptotic behavior of the scattering phase for exterior domains, *Comm. PDE* 3(1978), 1165–1195.

[JM] D. Jones and X. Mao, The inverse problem in hard acoustic scattering, *Inverse Probl.* 5(1989), 731–748.

[Kt] T. Kato, *Perturbation Theory for Linear Operators*, Springer, New York, 1966.

[Kir] A. Kirsch, The domain derivative and two applications in inverse scattering theory, *Inverse Probl.* 9(1993), 81–96.

[KZ] A. Kirsch, R. Kress, P. Monk, and A. Zinn, Two methods for solving the inverse scattering problem, *Inverse Probl.* 4(1988), 749–770.

[KV] G. Kristensson and C. Vogel, Inverse problems for acoustic waves using the penalised liklihood method, *Inverse Probl.* 2(1984), 461–479.

[LP1] P. Lax and R. Phillips, *Scattering Theory*, Academic, New York, 1967.

[LP2] P. Lax and R. Philips, Scattering theory, *Rocky Mountain J. Math.* 1(1971), 173–223.

[LP3] P. Lax and R. Phillips, Scattering theory for the acoustic equation in an even number of space dimensions, *Indiana University Math. J.* 22(1972), 101–134.

[LP4] P. Lax and R. Phillips, The time delay operator and a related trace formula, in *Topics in Func. Anal.*, I. Gohberg and M. Kac, eds., Academic, New York, 1978, pp. 197–215.

[LP5] P. Lax and R. Phillips, *Scattering Theory for Automorphic Functions*, Princeton University Press, Princeton, NJ, 1976.

[MjR] A. Majda and J. Ralston, An analogue of Weyl's theorem for unbounded domains, *Duke Math. J.* 45(1978), 513–536.

[MjT] A. Majda and M. Taylor, Inverse scattering problems for transparent obstacles, electromagnetic waves, and hyperbolic systems, *Comm. PDE* 2(1977), 395–438.

[Me1] R. Melrose, Scattering theory and the trace of the wave group, *J. Func. Anal.* 45(1982), 29–40.

[Me2] R. Melrose, Polynomial bound on the number of scattering poles, *J. Func. Anal.* 53(1983), 287–303.

[Me3] R. Melrose, Weyl asymptotics for the phase in obstacle scattering, *Comm. PDE* 11(1988), 1431–1439.

[Me4] R. Melrose, *Geometric Scattering Theory*, Cambridge University Press, Cambridge, 1995.

[MT1] R. Melrose and M. Taylor, Near peak scattering and the corrected Kirchhoff approximation for convex bodies, *Adv. Math.* 55(1985), 242–315.

[MT2] R. Melrose and M. Taylor, *Boundary Problems for Wave Equations with Grazing and Gliding Rays*. Monograph, Preprint, 2010.

[Mr1] K. Miller, Least squares methods for ill-posed problems with a prescribed bound, *SIAM J. Math. Anal.* 1(1970), 52–74.

[Mr2] K. Miller, Efficient numerical methods for backward solutions of parabolic problems with variable coefficients, in *Improperly Posed Boundary Value Problems*, A.Carasso and A.Stone, eds., Pitman, London, 1975.

[MrV] K. Miller and G. Viano, On the necessity of nearly-best possible methods for analytic continuation of scattering data, *J. Math Phys.* 14(1973), 1037–1048.

[Mw] C. Morawetz, The limiting amplitude principle, *Comm. Pure Appl. Math.* 15(1962), 349–361.

[Mw2] C. Morawetz, Exponential decay of solutions to the wave equation, *Comm. Pure Appl. Math.* 19(1966), 439–444.

[MRS] C. Morawetz, J. Ralston, and W. Strauss, Decay of solutions of the wave equation outside nontrapping obstacles, *Comm. Pure Appl. Math.* 30(1977), 447–508.

[MTW] R. Murch, D. Tan, and D. Wall, Newton-Kantorovich method applied to two-dimensional inverse scattering for an exterior Helmholtz problem, *Inverse Probl.* 4(1988), 1117–1128.

[New] R. G. Newton, Scattering Theory of Waves and Particles, Springer, New York, 1966.

[Nus] H. Nussensweig, High frequency scattering by an impenetrable sphere, *Ann. Phys.* 34(1965), 23–95.

[Olv] F. Olver, *Asymptotics and Special Functions*, Academic, New York, 1974.

[PP] V. Petkov and G. Popov, Asymptotic behavior of the scattering phase for nontrapping obstacles, *Ann. Inst. Fourier (Grenoble)* 32(1982), 111–149.

[Rl] J. Ralston, Propagation of singularities and the scattering matrix, in *Singularities in Boundary Value Problems*, H.Garnir, ed., D.Reidel, Dordrecht, 1981, pp. 169–184.

[Rm] A. Ramm, *Scattering by Obstacles*, D.Reidel, Dordrecht, 1986.

[Rau] J. Rauch, *Partial Differential Equations*, Springer, New York, 1991.

[RT] J. Rauch and M. Taylor, Potential and scattering theory on wildly perturbed domains, *J. Func. Anal.* 18(1975), 27–59.

[RS] M. Reed and B. Simon, *Methods of Mathematical Physics*, Academic, New York, Vols. 1,2, 1975; Vols. 3,4, 1978.

[Rog] A. Roger, Newton-Kantorovich algorithm applied to an electromagnetic inverse problem, *IEEE Trans. Antennas Propagat* 29(1981), 232–238.

[Si] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms*, Princeton University Press, Princeton, NJ, 1971.

[Si2] B. Simon, *Trace Ideals and Their Applications*, London Math. Soc. Lecture Notes no. 35, Cambridge University Press, Cambridge, 1979.

[Stk] I. Stakgold, *Boundary Value Problems of Mathematical Physics*, Macmillan, New York, 1968.

[T1] M. Taylor, Propagation, reflection, and diffraction of singularities of solutions to wave equations, *Bull. AMS* 84(1978), 589–611.
[T2] M. Taylor, Fourier integral operators and harmonic analysis on compact manifolds, *Proc. Symp. Pure Math*. 35(pt. 2)(1979), 113–136.
[T3] M. Taylor, *Pseudodifferential Operators*, Princeton University Press, Princeton, NJ, 1981.
[T4] M. Taylor, Estimates for approximate solutions to acoustic inverse scattering problems, IMA Preprint #1303, 1995.
[VV] V. K. Varadan and V. V. Varadan (eds.), *Acoustic, Electromagnetic, and Elastic Wave Scattering – Focus on the T matrix Approach*, Pergammon, New York, 1980.
[Wat] G. Watson, *Theory of Bessel Functions*, Cambridge University Press, Cambridge, 1945.
[Wil] C. Wilcox, *Scattering Theory for the d'Alembert Equation in Exterior Domains*, LNM no. 442, Springer, New York, 1975.
[Yo] K. Yosida, *Functional Analysis*, Springer, New York, 1965.

10

Dirac Operators and Index Theory

Introduction

The physicist P. A. M. Dirac constructed first-order differential operators whose squares were Laplace operators, or more generally wave operators, for the purpose of extending the Schrodinger–Heisenberg quantum mechanics to the relativistic setting. Related operators have been perceived to have central importance in the interface between PDE and differential geometry, and we discuss some of this here.

We define various classes of "Dirac operators," some arising on arbitrary Riemannian manifolds, some requiring some special geometrical structure, such as a spin structure, discussed in §3, or a spin^c structure, discussed in §8.

Dirac operators on compact Riemannian manifolds are elliptic and have an index. Evaluating this index, in terms of an integrated "curvature," is the essence of the famous Atiyah–Singer index theorem. We present a proof of this index formula here, using a "heat-equation" method of proof. Such a proof was first suggested in [McS], but it seemed difficult to carry out, as it required understanding of a coefficient in the asymptotic expansion of the traces of e^{-tL_j}, for a pair of positive, second-order, elliptic operators L_j, well below the principal term. Ingenious arguments, beginning with V. Patodi [Pt1, Pt2], led to a proof in Atiyah–Bott–Patodi [ABP]. Later, physicists, motivated by ideas from "supersymmetry," proposed more direct heat-equation proofs. Such proposals were first made by E. Witten [Wit]; particularly elegant mathematical treatments were given by E. Getzler in [Gt1] and [Gt2]. We present a heat-equation proof in §6, using Getzler's method of exploiting an analogue of Mehler's formula for the exponential of the harmonic oscillator Hamiltonian. Our analytical details differ from Getzler's; instead of introducing a noncommutative symbol calculus as in [Gt1], or the dilation argument of [Gt2], we fit the analysis more into a "classical" examination of heat-equation asymptotics, such as dealt with in Chap. 7. One major achievement of Getzler's approach is to make the appearance of the (rather subtle) $\hat{A}$-genus of M in the index formula arise quite naturally.

We present two specific examples of the index formula here. In §7 we derive the Chern–Gauss–Bonnet formula, giving the Euler characteristic $\chi(M)$

M.E. Taylor, *Partial Differential Equations II: Qualitative Studies of Linear Equations*, Applied Mathematical Sciences 116, DOI 10.1007/978-1-4419-7052-7_4,

of a compact, orientable Riemannian manifold in terms of an integral of the "Pfaffian" applied to its curvature tensor. In §9 (following a discussion in §8 of spin^c structures) we discuss the Riemann–Roch formula, a tool for understanding holomorphic (and meromorphic) sections of line bundles over Riemann surfaces (of real dimension 2), which is important in the study of the structure and function theory of Riemann surfaces. These are the simplest applications of the Atiyah–Singer formula; both were established well before the general formula. From a technical point of view, both have in common that the $\hat{A}$-genus of M is effectively discarded.

Other examples of the index formula include higher-dimensional Riemann–Roch formulas and signature formulas. Further material on these can be found in [Pal] and [Gil]. There is also an operator associated with "self-dual" connections on bundles over 4-manifolds, whose index plays an important role in the study of the Yang–Mills equations; see [AHS] and [FU]. For a recent variant, arising from the Seiberg–Witten equations, see [D] and [Mor].

The heat-equation proof of the Chern–Gauss–Bonnet theorem was Patodi's [Pt1] first step in this circle of results. An exposition of the heat-equation proof of this result due to B. Simon is given in the last chapter of [CS]. Another proof of the Chern–Gauss–Bonnet theorem, celebrating closely physicists' ideas about supersymmetry, is given in [Rog].

Due to the low dimension, one can give a direct proof of the Riemann–Roch theorem, using techniques of [McS]; such a proof is given in [Kot]. Such a direct approach, with a good bit more effort, could be expected to be effective in other low dimensions (e.g., complex dimension 2); in a sense, the sort of analysis required to accomplish this is what was done in [Ko1]. In §10, we give a direct proof of an index formula for first-order, elliptic differential operators of Dirac type on a 2-manifold M, in terms of a direct calculation of the second term in the expansion of the heat kernel, carried out in §14 of Chap. 7, using the Weyl calculus. We show how this formula yields the Gauss–Bonnet formula and the Riemann–Roch formula.

There are also other heat-equation proofs of the index theorem, particularly [Bi1] and [BV]. In [Bi2] the heat equation method is applied to families of operators; see also [Don] and [BiC]. There are several recent books devoted to expositions of heat-equation proofs of the index theorem, including [BGV,Gil,Mel], and [Roe].

A systematic "blow up" of the original proof of the Atiyah–Singer index theorem has led to the interesting subject of operator K-theory. An introductory exposition is given in Blackadar [Bl]. Further developments are described in [Con].

In §11 we change course, and produce an index formula for a class of elliptic $k \times k$ systems on Euclidean space $\mathbb{R}^n$. We do this for the class of pseudodifferential operators of harmonic oscillator type, introduced in §15 of Chap. 7. The proof here makes no use of heat-equation techniques. It uses some results from topology, particularly results on the homotopy groups of the unitary groups $U(k)$, including the Bott periodicity theorem, results for which we refer to [Mil] for proof. Section 11 can be read independently of the other sections of this chapter.

1. Operators of Dirac type

Let M be a Riemannian manifold, $E_j \to M$ vector bundles with Hermitian metrics. A first-order, elliptic differential operator

$$D : C^\infty(M, E_0) \longrightarrow C^\infty(M, E_1) \tag{1.1}$$

is said to be of Dirac type provided D^*D has scalar principal symbol. This implies

$$\sigma_{D^*D}(x, \xi) = g(x, \xi)I : E_{0x} \longrightarrow E_{0x}, \tag{1.2}$$

where $g(x, \xi)$ is a positive quadratic form on T_x^*M. Thus g itself arises from a Riemannian metric on M. Now the calculation of (1.2) is independent of the choice of Riemannian metric on M. We will suppose M is endowed with the Riemannian metric inducing the form $g(x, \xi)$ on T^*M.

If $E_0 = E_1$ and $D = D^*$, we say D is a symmetric Dirac-type operator. Given a general operator D of Dirac type, if we set $E = E_0 \oplus E_1$ and define $\tilde{D}$ on $C^\infty(M, E)$ as

$$\tilde{D} = \begin{pmatrix} 0 & D^* \\ D & 0 \end{pmatrix}, \tag{1.3}$$

then D is a symmetric Dirac-type operator.

Let $\vartheta(x, \xi)$ denote the principal symbol of a symmetric Dirac-type operator. With $x \in M$ fixed, set $\vartheta(\xi) = \vartheta(x, \xi)$. Thus ϑ is a linear map from $T_x^*M = \{\xi\}$ into $\mathrm{End}(E_x)$, satisfying

$$\vartheta(\xi) = \vartheta(\xi)^* \tag{1.4}$$

and

$$\vartheta(\xi)^2 = \langle \xi, \xi \rangle I. \tag{1.5}$$

Here, $\langle \ , \ \rangle$ is the inner product on T_x^*M; let us denote this vector space by V. We will show how ϑ extends from V to an algebra homomorphism, defined on a Clifford algebra $Cl(V, g)$, which we now proceed to define.

Let V be a finite-dimensional, real vector space, g a quadratic form on V. We allow g to be definite or indefinite if nondegenerate; we even allow g to be degenerate. The Clifford algebra $Cl(V, g)$ is the quotient algebra of the tensor algebra

$$\bigotimes V = \mathbb{R} \oplus V \oplus (V \otimes V) \oplus (V \otimes V \otimes V) \oplus \cdots \tag{1.6}$$

by the ideal $\mathcal{I} \subset \bigotimes V$ generated by

$$\{v \otimes w + w \otimes v - 2\langle v, w \rangle \cdot 1 : v, w \in V\}, \tag{1.7}$$

where $\langle\ ,\ \rangle$ is the symmetric bilinear form on V arising from g. Thus, in $Cl(V,g)$, V occurs naturally as a linear subspace, and there is the anti-commutation relation

$$vw + wv = 2\langle v, w\rangle \cdot 1 \text{ in } Cl(V,g), \quad v, w \in V. \tag{1.8}$$

We will look more closely at the structure of Clifford algebras in the next section.

Now if $\vartheta : V \to \text{End}(E)$ is a linear map of the V into the space of endomorphisms of a vector space E, satisfying (1.5), i.e.,

$$\vartheta(v)^2 = \langle v, v\rangle I, \quad v \in V, \tag{1.9}$$

it follows from expanding $\vartheta(v+w)^2 = [\vartheta(v) + \vartheta(w)]^2$ that

$$\vartheta(v)\vartheta(w) + \vartheta(w)\vartheta(v) = 2\langle v, w\rangle I, \quad v, w \in V. \tag{1.10}$$

Then, from the construction of $Cl(V,g)$, it follows that ϑ extends uniquely to an algebra homomorphism

$$\vartheta : Cl(V,g) \longrightarrow \text{End}(E), \quad \vartheta(1) = I. \tag{1.11}$$

This gives E the structure of a module over $Cl(V,g)$, or a Clifford module. If E has a Hermitian metric and (1.4) also holds, that is,

$$\vartheta(v) = \vartheta(v)^*, \quad v \in V, \tag{1.12}$$

we call E a Hermitian Clifford module. For this notion to be useful, we need g to be positive-definite.

In the case where $E = E_0 \oplus E_1$ is a direct sum of Hermitian vector spaces, we say a homomorphism $\vartheta : Cl(V,g) \to \text{End}(E)$ gives E the structure of a graded Clifford module provided $\vartheta(v)$ interchanges E_0 and E_1, for $v \in V$, in addition to the hypotheses above. The principal symbol of (1.3) has this property if D is of Dirac type.

Let us give some examples of operators of Dirac type. If M is a Riemannian manifold, the exterior derivative operator

$$d : \Lambda^j M \longrightarrow \Lambda^{j+1} M \tag{1.13}$$

has a formal adjoint

$$\delta = d^* : \Lambda^{j+1} M \longrightarrow \Lambda^j M, \tag{1.14}$$

discussed in Chap. 2, §10, and in Chap. 5, §§8 and 9. Thus we have

$$d + \delta : \Lambda^* M \longrightarrow \Lambda^* M, \tag{1.15}$$

where, with $n = \dim M$,

$$\Lambda^* M = \bigoplus_{j=0}^{n} \Lambda^j M.$$

As was shown in Chap. 2, $(d+\delta)^*(d+\delta) = d^*d + dd^*$ is the negative of the Hodge Laplacian on each $\Lambda^j M$, so (1.15) is a symmetric Dirac-type operator. There is more structure. Indeed, we have

$$d+\delta : \Lambda^{\text{even}} M \longrightarrow \Lambda^{\text{odd}} M. \tag{1.16}$$

If D is this operator, then $D^* = d+\delta : \Lambda^{\text{odd}} M \to \Lambda^{\text{even}} M$, and an operator of type (1.3) arises. If M is compact, the operator (1.16) is Fredholm, with index equal to the Euler characteristic of M, in view of the Hodge decomposition. A calculation of this index in terms of an integrated curvature gives rise to the generalized Gauss–Bonnet formula, as will be seen in §7.

Computations implying that (1.15) is of Dirac type were done in §10 of Chap. 2, leading to (10.22) there. If we define

$$\wedge_v : \Lambda^j V \longrightarrow \Lambda^{j+1} V, \quad \wedge_v(v_1 \wedge \cdots \wedge v_j) = v \wedge v_1 \wedge \cdots \wedge v_j, \tag{1.17}$$

on a vector space V with a positive-definite inner product, and then define

$$\iota_v : \Lambda^{j+1} V \longrightarrow \Lambda^j V \tag{1.18}$$

to be its adjoint, then the principal symbol of $d+\delta$ on $V = T_x^* M$ is $1/i$ times $\wedge_\xi - \iota_\xi$. That is to say,

$$iM(v) = \wedge_v - \iota_v \tag{1.19}$$

defines a linear map from V into $\text{End}(\Lambda_{\mathbb{C}}^* V)$ which extends to an algebra homomorphism

$$M : Cl(V,g) \longrightarrow \text{End}(\Lambda_{\mathbb{C}}^* V).$$

Given $\wedge_v \wedge_w = -\wedge_w \wedge_v$ and its analogue for ι, the anticommutation relation

$$M(v)M(w) + M(w)M(v) = 2\langle v, w\rangle I \tag{1.20}$$

follows from the identity

$$\wedge_v \iota_w + \iota_w \wedge_v = \langle v, w\rangle I. \tag{1.21}$$

In this context we note the role that (1.21) played as the algebraic identity behind Cartan's formula for the Lie derivative of a differential form:

$$\mathcal{L}_X \alpha = d(\alpha \rfloor X) + (d\alpha)\rfloor X; \tag{1.22}$$

cf. Chap. 1, Proposition 13.1, and especially (13.51).

Another Dirac-type operator arises from (1.15) as follows. Suppose $\dim M = n = 2k$ is even. Recall from Chap. 5, §8, that $d^* = \delta$ is given in terms of the Hodge star operator on $\Lambda^j M$ by

$$\begin{aligned} d^* &= (-1)^{j(n-j)+j} * d * \\ &= *d* \qquad \text{if } n = 2k. \end{aligned} \tag{1.23}$$

Also recall that, on $\Lambda^j M$,

$$*^2 = (-1)^{j(n-j)} = (-1)^j \quad \text{if } n = 2k. \tag{1.24}$$

Now, on the complexification $\Lambda^*_{\mathbb{C}} M$ of the real vector bundle $\Lambda^* M$, define

$$\alpha : \Lambda^j_{\mathbb{C}} M \longrightarrow \Lambda^{n-j}_{\mathbb{C}} M \tag{1.25}$$

by

$$\alpha = i^{j(j-1)+k} * \quad \text{on } \Lambda^j_{\mathbb{C}} M. \tag{1.26}$$

It follows that

$$\alpha^2 = 1 \tag{1.27}$$

and

$$\alpha(d + \delta) = -(d + \delta)\alpha. \tag{1.28}$$

Thus we can write

$$\Lambda^*_{\mathbb{C}} M = \Lambda^+ M \oplus \Lambda^- M, \ \text{ with } \alpha = \pm I \text{ on } \Lambda^\pm M, \tag{1.29}$$

and we have

$$D^\pm_H = d + \delta : C^\infty(M, \Lambda^\pm) \longrightarrow C^\infty(M, \Lambda^\mp). \tag{1.30}$$

Thus D^+_H is an operator of Dirac type, with adjoint D^-_H. This operator is called the Hirzebruch signature operator, and its index is called the Hirzebruch signature of M.

Other examples of operators of Dirac type will be considered in the following sections.

Both of the examples just discussed give rise to Hermitian Clifford modules. We now show conversely that generally such modules produce operators of Dirac type. More precisely, if M is a Riemannian manifold, $T^*_x M$ has an induced inner product, giving rise to a bundle $Cl(M) \to M$ of Clifford algebras. We suppose

$E \to M$ is a Hermitian vector bundle such that each fiber is a Hermitian $Cl_x(M)$-module (in a smooth fashion). Let $E \to M$ have a connection ∇, so

$$\nabla : C^\infty(M, E) \longrightarrow C^\infty(M, T^* \otimes E). \tag{1.31}$$

Now if E_x is a $Cl_x(M)$-module, the inclusion $T_x^* \hookrightarrow Cl_x$ gives rise to a linear map

$$m : C^\infty(M, T^* \otimes E) \longrightarrow C^\infty(M, E), \tag{1.32}$$

called "Clifford multiplication." We compose these two operators; set

$$D = i\, m \circ \nabla : C^\infty(M, E) \longrightarrow C^\infty(M, E). \tag{1.33}$$

We see that, for $v \in E_x$,

$$\sigma_D(x, \xi)v = m(\xi \otimes v) = \xi \cdot v, \tag{1.34}$$

so $\sigma_D(x, \xi)$ is $|\xi|_x$ times an isometry on E_x. Hence D is of Dirac type.

If U is an open subset of M, on which we have an orthonormal frame $\{e_j\}$ of smooth vector fields, with dual orthonormal frame $\{v_j\}$ of 1-forms, then, for a section φ of E,

$$D\varphi = i \sum v_j \cdot \nabla_{e_j} \varphi \quad \text{on } U. \tag{1.35}$$

Note that $\sigma_D(x, \xi)^* = \sigma_D(x, \xi)$, so D can be made symmetric by altering it at most by a zero-order term. Given a little more structure, we have more. We say ∇ is a "Clifford connection" on E if ∇ is a metric connection that is also compatible with Clifford multiplication, in that

$$\nabla_X(v \cdot \varphi) = (\nabla_X v) \cdot \varphi + v \cdot \nabla_X \varphi, \tag{1.36}$$

for a vector field X, a 1-form v, and a section φ of E. Here, of course, $\nabla_X v$ arises from the Levi–Civita connection on M.

Proposition 1.1. *If ∇ is a Clifford connection on E, then D is symmetric.*

Proof. Let $\varphi, \psi \in C_0^\infty(M, E)$. We want to show that

$$\int_M \Big[\langle D\varphi, \psi \rangle - \langle \varphi, D\psi \rangle \Big] \, dV = 0. \tag{1.37}$$

We can suppose φ, ψ have compact support in a set U on which local orthonormal frames e_j, v_j as above are given. Define a vector field X on U by

$$\langle X, v \rangle = \langle \varphi, v \cdot \psi \rangle, \quad v \in \Lambda^1 U.$$

If we show that, pointwise in U,

$$i \text{ div } X = \langle D\varphi, \psi\rangle - \langle \varphi, D\psi\rangle, \tag{1.38}$$

then (1.37) will follow from the divergence theorem. Indeed, starting with

$$\text{div } X = \sum \langle \nabla_{e_j} X, v_j\rangle, \tag{1.39}$$

and using the metric and derivation properties of ∇, we have

$$\begin{aligned} \text{div } X &= \sum \Big[e_j \cdot \langle X, v_j\rangle - \langle X, \nabla_{e_j} v_j\rangle \Big] \\ &= \sum \Big[e_j \langle \varphi, v_j \cdot \psi\rangle - \langle \varphi, (\nabla_{e_j} v_j) \cdot \psi\rangle \Big]. \end{aligned}$$

Looking at the last quantity, we expand the first part into a sum of three terms, one of which cancels the last part, and obtain

$$\text{div } X = \sum \Big[\langle \nabla_{e_j} \varphi, v_j \cdot \psi\rangle + \langle \varphi, v_j \cdot \nabla_{e_j} \psi\rangle \Big], \tag{1.40}$$

which gives (1.38) and completes the proof.

If $E = E_0 \oplus E_1$ is a graded Hermitian $Cl(M)$-module, if E_0 and E_1 are each provided with metric connections, and if (1.36) holds, then the construction above gives an operator of Dirac type, of the form (1.3).

The examples in (1.15) and (1.30) described above can be obtained from Hermitian Clifford modules via Clifford connections. The Clifford module is $\Lambda^* M \to M$, with natural inner product on each factor $\Lambda^k M$ and $Cl(M)$-module structure given by (1.19). The connection is the natural connection on $\Lambda^* M$, extending that on $T^* M$, so that the derivation identity

$$\nabla_X (\varphi \wedge \psi) = (\nabla_X \varphi) \wedge \psi + \varphi \wedge (\nabla_X \psi) \tag{1.41}$$

holds for a j-form φ and a k-form ψ. In this case it is routine to verify the compatibility condition (1.36) and to see that the construction (1.33) gives rise to the operator $d + d^*$ on differential forms.

We remark that it is common to use Clifford algebras associated to negative-definite forms rather than positive-definite ones. The two types of algebras are simply related. If a linear map $\vartheta : V \to \text{End}(E)$ extends to an algebra homomorphism $Cl(V, g) \to \text{End}(E)$, then $i\vartheta$ extends to an algebra homomorphism $Cl(V, -g) \to \text{End}(E)$. If one uses a negative form, the condition (1.12) that E be a Hermitian Clifford module should be changed to $\vartheta(v) = -\vartheta(v)^*, \ v \in V$. In such a case, we should drop the factor of i in (1.33) to associate the Dirac-type operator D to a $Cl(M)$-module E. In fact, getting rid of this factor of i in (1.33) and (1.35) is perhaps the principal reason some people use the negative-definite quadratic form to construct Clifford algebras.

Exercises

1. Let E be a $Cl(M)$-module with connection ∇. If φ is a section of E and f is a scalar function, show that
$$D(f\varphi) = f\, D\varphi + i(df)\cdot\varphi,$$
where the last term involves a Clifford multiplication.
2. If ∇ is a Clifford connection on E and u is a 1-form, show that
$$D(u\cdot\varphi) = -u\cdot D\varphi + 2i\nabla_U\varphi + i(\mathcal{D}u)\cdot\varphi,$$
where U is the vector field corresponding to u via the metric tensor on M, and
$$\mathcal{D} : C^\infty(M,\Lambda^1) \longrightarrow C^\infty(M, Cl)$$
is given by
$$\mathcal{D}u = i\sum v_j\cdot\nabla_{e_j}u,$$
with respect to local dual orthonormal frames e_j, v_j, and ∇ arising from the Levi–Civita connection.
3. Show that $\mathcal{D}(df) = i\Delta f$.
Note: Compare with Exercise 6 of §2.
4. If D arises from a Clifford connection on E, show that
$$D^2(f\varphi) = f\, D^2\varphi - 2\nabla_{\mathrm{grad}\, f}\varphi - (\Delta f)\varphi.$$

2. Clifford algebras

In this section we discuss some further results about the structure of Clifford algebras, which were defined in §1.

First we note that, by construction, $Cl(V,g)$ has the following universal property. Let A_0 be any associative algebra over $\mathbb{R}$, with unit, containing V as a linear subset, generated by V, such that the anticommutation relation (1.8) holds in A_0, for all $v,w\in V$; that is, $vw+wv = 2\langle v,w\rangle\cdot 1$ in A_0. Then there is a natural surjective homomorphism

$$\alpha : Cl(V,g) \longrightarrow A_0. \tag{2.1}$$

If $\{e_1,\dots,e_n\}$ is a basis of V, any element of $Cl(V,g)$ can be written as a polynomial in the e_j. Since $e_je_k = -e_ke_j + 2\langle e_j,e_k\rangle\cdot 1$ and in particular $e_j^2 = \langle e_j,e_j\rangle\cdot 1$, we can, starting with terms of highest order, rearrange each monomial in such a polynomial so the e_j appear with j in ascending order, and no exponent greater than 1 occurs on any e_j. In other words, each element $w\in Cl(V,g)$ can be written in the form

$$w = \sum_{i_\nu = 0 \text{ or } 1} a_{i_1\cdots i_n}\, e_1^{i_1}\cdots e_n^{i_n}, \tag{2.2}$$

with real coefficients $a_{i_1\cdots i_n}$.

Denote by A the set of formal expressions of the form (2.2), a real vector space of dimension 2^n; we have a natural inclusion $V \subset A$. We can define a "product" $A \otimes A \to A$ in which a product of monomials $(e_1^{i_1} \cdots e_n^{i_n}) \cdot (e_1^{j_1} \cdots e_n^{j_n})$, with each i_ν and each j_μ equal to either 0 or 1, is a linear combination of monomials of such a form, by pushing each $e_\mu^{j_\mu}$ past the $e_\nu^{i_\nu}$ for $\nu > \mu$, invoking the anticommutation relations. It is routine to verify that this gives A the structure of an associative algebra, generated by V. The universal property mentioned above implies that A is isomorphic to $Cl(V, g)$. Thus each $w \in Cl(V, g)$ has a unique representation in the form (2.2), and $\dim Cl(V, g) = 2^n$ if $\dim V = n$.

Recall from §1 the algebra homomorphism $M : Cl(V, g) \to \text{End}(\Lambda^* V)$, defined there provided g is positive-definite (which can be extended to include general g). Then, we can define a linear map

$$\tilde{M} : Cl(V, g) \longrightarrow \Lambda^* V; \quad \tilde{M}(w) = M(w)(1), \tag{2.3}$$

for $w \in Cl(V, g)$. Note that if $v \in V \subset Cl(V, g)$, then $\tilde{M}(v) = v$. Comparing the anticommutation relations of $Cl(V, g)$ with those of $\Lambda^* V$, we see that if $w \in Cl(V, g)$ is one of the monomials in (2.2), say $w = e_1^{j_1} \cdots e_n^{j_n}$, all j_ν either 0 or 1, $k = j_1 + \cdots + j_n$, then

$$\tilde{M}(e_1^{j_1} \cdots e_n^{j_n}) - e_1^{j_1} \wedge \cdots \wedge e_n^{j_n} \in \Lambda^{k-1} V. \tag{2.4}$$

It follows easily that (2.3) is an isomorphism of vector spaces. This observation also shows that the representation of an element of $Cl(V, g)$ in the form (2.2) is unique. If g is positive-definite and e_j is an orthonormal basis of V, the difference in (2.4) vanishes.

In the case $g = 0$, the anticommutation relation (1.8) becomes $vw = -wv$, for $v, w \in V$, and we have the exterior algebra

$$Cl(V, 0) = \Lambda^* V.$$

Through the remainder of this section we will restrict attention to the case where g is positive-definite. We denote $\langle v, v \rangle$ by $|v|^2$. For $V = \mathbb{R}^n$ with g its standard Euclidean inner product, we denote $Cl(V, g)$ by $Cl(n)$.

It is useful to consider the complexified Clifford algebra

$$\mathbb{C}l(n) = \mathbb{C} \otimes Cl(n),$$

as it has a relatively simple structure, specified as follows.

Proposition 2.1. *There are isomorphisms of complex algebras*

$$\mathbb{C}l(1) \approx \mathbb{C} \oplus \mathbb{C}, \quad \mathbb{C}l(2) \approx \mathit{End}(\mathbb{C}^2), \tag{2.5}$$

and

$$\mathbb{C}l(n+2) \approx \mathbb{C}l(n) \otimes \mathbb{C}l(2); \tag{2.6}$$

hence, with $\kappa = 2^k$,

$$\mathbb{C}l(2k) \approx End(\mathbb{C}^\kappa), \quad \mathbb{C}l(2k+1) \approx End(\mathbb{C}^\kappa) \oplus End(\mathbb{C}^\kappa). \tag{2.7}$$

Proof. The isomorphisms (2.5) are simple exercises. To prove (2.6), imbed $\mathbb{R}^{n+2}$ into $\mathbb{C}l(n) \otimes \mathbb{C}l(2)$ by picking an orthonormal basis $\{e_1, \dots, e_{n+2}\}$ and taking

$$\begin{aligned} e_j &\mapsto i\, e_j \otimes e_{n+1}e_{n+2}, \quad \text{for } 1 \le j \le n, \\ e_j &\mapsto 1 \otimes e_j, \quad \text{for } j = n+1 \text{ or } n+2. \end{aligned} \tag{2.8}$$

Then the universal property of $\mathbb{C}l(n+2)$ leads to the isomorphism (2.6). Given (2.5) and (2.6), (2.7) follows by induction.

While, parallel to (2.5), one has $Cl(1) = \mathbb{R} \oplus \mathbb{R}$ and $Cl(2) = \text{End}(\mathbb{R}^2)$, other algebras $Cl(n)$ are more complicated than their complex analogues; in place of (2.6) one has a form of periodicity with period 8. We refer to [LM] for more on this.

It follows from Proposition 2.1 that $\mathbb{C}^{2^k}$ has the structure of an irreducible $Cl(2k)$-module, though making the identification (2.7) explicit involves some untangling, in a way that depends strongly on a choice of basis. It is worthwhile to note the following explicit, invariant construction, for V, a vector space of real dimension $2k$, with a positive inner product $\langle\ ,\ \rangle$, endowed with one other piece of structure, namely a complex structure J. Assume J is an isometry for $\langle\ ,\ \rangle$. Denote the complex vector space (V, J) by $\mathcal{V}$, which has complex dimension k. On $\mathcal{V}$ we have a positive Hermitian form

$$(u, v) = \langle u, v\rangle + i\langle u, Jv\rangle. \tag{2.9}$$

Form the complex exterior algebra

$$\Lambda^*_{\mathbb{C}}\mathcal{V} = \bigoplus_{j=0}^{k} \Lambda^j_{\mathbb{C}}\mathcal{V}, \tag{2.10}$$

with its natural Hermitian form. For $v \in \mathcal{V}$, one has the exterior product $v\wedge : \Lambda^j_{\mathbb{C}}\mathcal{V} \to \Lambda^{j+1}_{\mathbb{C}}\mathcal{V}$; denote its adjoint, the interior product, by $j_v : \Lambda^{j+1}_{\mathbb{C}}\mathcal{V} \to \Lambda^j_{\mathbb{C}}\mathcal{V}$. Set

$$i\, \mu(v)\varphi = v \wedge \varphi - j_v\varphi, \quad v \in \mathcal{V}, \quad \varphi \in \Lambda^*_{\mathbb{C}}\mathcal{V}. \tag{2.11}$$

Note that $v \wedge \varphi$ is $\mathbb{C}$-linear in v and $j_v\varphi$ is conjugate linear in v, so $\mu(v)$ is only $\mathbb{R}$-linear in v. As in (1.20), we obtain

$$\mu(u)\mu(v) + \mu(v)\mu(u) = 2\langle u, v\rangle \cdot I, \tag{2.12}$$

so $\mu : V \to \operatorname{End}(\Lambda^*_{\mathbb{C}}\mathcal{V})$ extends to a homomorphism of algebras

$$\mu : Cl(V, g) \longrightarrow \operatorname{End}(\Lambda^*_{\mathbb{C}}\mathcal{V}), \tag{2.13}$$

hence to a homomorphism of $\mathbb{C}$-algebras

$$\mu : \mathbb{C}l(V, g) \longrightarrow \operatorname{End}(\Lambda^*_{\mathbb{C}}\mathcal{V}), \tag{2.14}$$

where $\mathbb{C}l(V, g)$ denotes $\mathbb{C} \otimes Cl(V, g)$.

Proposition 2.2. *The homomorphism (2.14) is an isomorphism when V is a real vector space of dimension $2k$, with complex structure J, $\mathcal{V}$ the associated complex vector space.*

Proof. We already know that both $\mathbb{C}l(V, g)$ and $\operatorname{End}(\Lambda^*_{\mathbb{C}}\mathcal{V})$ are isomorphic to $\operatorname{End}(\mathbb{C}^\kappa)$, $\kappa = 2^k$. We will make use of the algebraic fact that this is a complex algebra with no proper two-sided ideals. Now the kernel of μ in (2.14) would have to be a two-sided ideal, so either $\mu = 0$ or μ is an isomorphism. But for $v \in V$, $\mu(v) \cdot 1 = v$, so $\mu \neq 0$; thus μ is an isomorphism.

We next mention that a grading can be put on $Cl(V, g)$. Namely, let $Cl^0(V, g)$ denote the set of sums of the form (2.2) with $i_1 + \cdots + i_n$ even, and let $Cl^1(V, g)$ denote the set of sums of that form with $i_1 + \cdots + i_n$ odd. It is easy to see that this specification is independent of the choice of basis $\{e_j\}$. Also we clearly have

$$u \in Cl^j(V, g), \ w \in Cl^k(V, g) \Longrightarrow uw \in Cl^{j+k}(V, g), \tag{2.15}$$

where j and k are each 0 or 1, and we compute $j + k$ mod 2. If (V, g) is $\mathbb{R}^n$ with its standard Euclidean metric, we denote $Cl^j(V, g)$ by $Cl^j(n)$, $j = 0$ or 1.

We note that there is an isomorphism

$$j : Cl(2k - 1) \longrightarrow Cl^0(2k) \tag{2.16}$$

uniquely specified by the property that, for $v \in \mathbb{R}^{2k-1}$, $j(v) = ve_{2k}$, where $\{e_1, \dots, e_{2k-1}\}$ denotes the standard basis of $\mathbb{R}^{2k-1}$, with e_{2k} added to form a basis of $\mathbb{R}^{2k}$. This will be useful in the next section for constructing spinors on odd-dimensional spaces.

We can construct a finer grading on $Cl(V, g)$. Namely, set

$$Cl^{[k]}(V, g) = \text{ set of sums of the form (2.2), with } i_1 + \cdots + i_n = k. \tag{2.17}$$

Thus $Cl^{[0]}(V, g)$ is the set of scalars and $Cl^{[1]}(V, g)$ is V. If we insist that $\{e_j\}$ be an orthonormal basis of V, then $Cl^{[k]}(V, g)$ is invariantly defined, for all k. In fact, using the isomorphism (2.3), we have

$$Cl^{[k]}(V, g) = \tilde{M}^{-1}\big(\Lambda^k V\big). \tag{2.18}$$

Note that

$$Cl^0(V,g) = \bigoplus_{k \text{ even}} Cl^{[k]}(V,g) \ \text{ and } \ Cl^1(V,g) = \bigoplus_{k \text{ odd}} Cl^{[k]}(V,g).$$

Let us also note that $Cl^{[2]}(V,g)$ has a natural Lie algebra structure. In fact, if $\{e_j\}$ is orthonormal,

$$\begin{aligned} \big[e_ie_j, e_ke_\ell\big] &= e_ie_je_ke_\ell - e_ke_\ell e_ie_j \\ &= 2(\delta_{jk}e_ie_\ell - \delta_{\ell j}e_ie_k + \delta_{ik}e_\ell e_j - \delta_{\ell i}e_ke_j). \end{aligned} \tag{2.19}$$

The construction (2.17) makes $Cl(V,g)$ a graded vector space, but not a graded algebra, since typically $Cl^{[j]}(V,g)\cdot Cl^{[k]}\,(V,g)$ is not contained in $Cl^{[j+k]}(V,g)$, as (2.19) illustrates. We can set

$$Cl^{(k)}(V,g) = \bigoplus \big\{Cl^{[j]}(V,g) : j \le k,\ j = k \bmod 2\big\}, \tag{2.20}$$

and then $Cl^{(j)}(V,g)\cdot Cl^{(k)}(V,g) \subset Cl^{(j+k)}(V,g)$. As k ranges over the even or the odd integers, the spaces (2.20) provide *filtrations* of $Cl^0(V,g)$ and $Cl^1(V,g)$.

Exercises

1. Let V have an oriented orthonormal basis $\{e_1, \cdots, e_n\}$. Set

$$\nu = e_1\cdots e_n \in Cl(V,g). \tag{2.21}$$

   Show that ν is independent of the choice of such a basis.
   Note: $\tilde{M}(\nu) = e_1\wedge\cdots\wedge e_n \in \Lambda^nV$, with $\tilde{M}$ as in (2.3).
2. Show that $\nu^2 = (-1)^{n(n-1)/2}$.
3. Show that, for all $u \in V,\ \nu u = (-1)^{n-1}u\nu$.
4. With μ as in (2.11)–(2.14), show that

$$\mu(\nu)^* = (-1)^{n(n-1)/2}\mu(\nu) \ \text{ and } \ \mu(\nu)^*\mu(\nu) = I.$$

5. Show that

$$\tilde{M}(\nu w) = c_{nk} * \tilde{M}(w),$$

   for $w \in Cl^{[k]}(V,g)$, where $* : \Lambda^kV \to \Lambda^{n-k}V$ is the Hodge star operator. Find the constants c_{nk}.
6. Let $\mathcal{D} : C^\infty(M,T^*) \to C^\infty(M,Cl)$ be as in Exercise 2 of §1, namely,

$$\mathcal{D}u = i\sum v_j\cdot\nabla_{e_j}u,$$

   where $\{e_j\}$ is a local orthonormal frame of vector fields, $\{v_j\}$ the dual frame. Show that

$$\tilde{M}(\mathcal{D}v) = -i(d+d^*)v.$$

7. Show that $\mathrm{End}(\mathbb{C}^m)$ has no proper two-sided ideals. (*Hint*: Suppose $M_0 \neq 0$ belongs to such an ideal $\mathcal{I}$ and $v_0 \neq 0$ belongs to the range of M_0. Show that every $v \in \mathbb{C}^m$ belongs to the range of some $M \in \mathcal{I}$, and hence that every one-dimensional projection belongs to $\mathcal{I}$.)

3. Spinors

We define the spinor groups $\mathrm{Pin}(V, g)$ and $\mathrm{Spin}(V, g)$, for a vector space V with a positive-definite quadratic form g; set $|v|^2 = g(v, v) = \langle v, v\rangle$. We set

$$\mathrm{Pin}(V, g) = \{v_1 \cdots v_k \in Cl(V, g) : v_j \in V,\ |v_j| = 1\}, \tag{3.1}$$

with the induced multiplication. Since $(v_1 \cdots v_k)(v_k \cdots v_1) = 1$, it follows that $\mathrm{Pin}(V, g)$ is a group. We can define an action of $\mathrm{Pin}(V, g)$ on V as follows. If $u \in V$ and $x \in V$, then $ux + xu = 2\langle x, u\rangle \cdot 1$ implies

$$uxu = -xuu + 2\langle x, u\rangle u = -|u|^2 x + 2\langle x, u\rangle u. \tag{3.2}$$

If also $y \in V$,

$$\langle uxu, uyu\rangle = |u|^2\langle x, y\rangle = \langle x, y\rangle \quad \text{if } |u| = 1. \tag{3.3}$$

Thus if $u = v_1 \cdots v_k \in \mathrm{Pin}(V, g)$ and if we define a conjugation on $Cl(V, g)$ by

$$u^* = v_k \cdots v_1, \quad v_j \in V, \tag{3.4}$$

it follows that

$$x \mapsto uxu^*, \quad x \in V, \tag{3.5}$$

is an isometry on V for each $u \in \mathrm{Pin}(V, g)$. It will be more convenient to use

$$u^\# = (-1)^k u^*, \quad u = v_1 \cdots v_k. \tag{3.6}$$

Then we have a group homomorphism

$$\tau : \mathrm{Pin}(V, g) \longrightarrow O(V, g), \tag{3.7}$$

defined by

$$\tau(u)x = uxu^\#, \quad x \in V,\ u \in \mathrm{Pin}(V, g). \tag{3.8}$$

Note that if $v \in V,\ |v| = 1$, then, by (3.2),

$$\tau(v)x = x - 2\langle x, v\rangle v \tag{3.9}$$

is the reflection across the hyperplane in V orthogonal to v. It is easy to show that any orthogonal transformation $T \in O(V, g)$ is a product of a finite number of such reflections, so the group homomorphism (3.7) is surjective.

Note that each isometry (3.9) is orientation reversing. Thus, if we define

$$\text{(3.10)}\qquad \begin{aligned} \mathrm{Spin}(V, g) &= \{v_1 \cdots v_k \in Cl(V, g) : v_j \in V,\ |v_j| = 1,\ k \text{ even}\} \\ &= \mathrm{Pin}(V, g) \cap Cl^0(V, g), \end{aligned}$$

then

$$\text{(3.11)}\qquad \tau : \mathrm{Spin}(V, g) \longrightarrow \mathrm{SO}(V, g)$$

and in fact $\mathrm{Spin}(V, g)$ is the inverse image of $\mathrm{SO}(V, g)$ under τ in (3.7). We now show that τ is a 2-fold covering map.

Proposition 3.1. *τ is a 2-fold covering map. In fact, ker $\tau = \{\pm 1\}$.*

Proof. Note that $\pm 1 \in \mathrm{Spin}(V, g) \subset Cl(V, g)$ and ± 1 acts trivially on V, via (3.8). Now, if $u = v_1 \cdots v_k \in \ker \tau$, k must be even, since $\tau(u)$ must preserve orientation, so $u^{\#} = u^*$. Since $uxu^* = x$ for all $x \in V$, we have $ux = xu$, so $uxu = |u|^2 x,\ x \in V$. If we pick an orthonormal basis $\{e_1, \dots, e_n\}$ of V and write $u \in \ker \tau$ in the form (2.2), each $i_1 + \cdots + i_n$ even, since $e_j u e_j = u$ for each j, we deduce that, for each j,

$$u = \sum (-1)^{i_j}\, a_{i_1 \cdots i_n}\, e^{i_1 \cdots i_n} \quad \text{if } u \in \ker \tau.$$

Hence $i_j = 0$ for all j, so u is a scalar; hence $u = \pm 1$.

We next consider the connectivity properties of $\mathrm{Spin}(V, g)$.

Proposition 3.2. $\mathrm{Spin}(V, g)$ *is the connected 2-fold cover of SO(V, g), provided g is positive-definite and dim $V \geq 2$.*

Proof. It suffices to connect $-1 \in \mathrm{Spin}(V, g)$ to the identity element 1 via a continuous curve in $\mathrm{Spin}(V, g)$. In fact, pick orthogonal $e_1,\ e_2$, and set

$$\gamma(t) = e_1 \cdot \big[-(\cos t) e_1 + (\sin t) e_2\big], \quad 0 \leq t \leq \pi.$$

If $V = \mathbb{R}^n$ with its standard Euclidean inner product g, denote $\mathrm{Spin}(V, g)$ by $\mathrm{Spin}(n)$. It is a known topological fact that $\mathrm{SO}(n)$ has fundamental group $\mathbb{Z}_2$, and $\mathrm{Spin}(n)$ is simply connected, for $n \geq 3$. Though we make no use of this result, we mention that one route to it is via the “homotopy exact sequence” (see [BTu]) for $S^n = \mathrm{SO}(n+1)/\mathrm{SO}(n)$. This leads to $\pi_1\big(\mathrm{SO}(n+1)\big) \approx \pi_1\big(\mathrm{SO}(n)\big)$ for $n \geq 3$. Meanwhile, one sees directly that SU(2) is a double cover of SO(3), and it is homeomorphic to S^3.

We next produce representations of $\mathrm{Pin}(V,g)$ and $\mathrm{Spin}(V,g)$, arising from the homomorphism (2.13). First assume V has real dimension $2k$, with complex structure J; let $\mathcal{V} = (V,J)$ be the associated complex vector space, of complex dimension k, and set

$$S(V,g,J) = \Lambda^*_{\mathbb{C}}\mathcal{V}, \tag{3.12}$$

with its induced Hermitian metric, arising from the metric (2.9) on $\mathcal{V}$. The inclusion $\mathrm{Pin}(V,g) \subset Cl(V,g) \subset \mathbb{C}l(V,g)$ followed by (2.14) gives the representation

$$\rho : \mathrm{Pin}(V,g) \longrightarrow \mathrm{Aut}\big(S(V,g,J)\big). \tag{3.13}$$

Proposition 3.3. *The representation* ρ *of* $\mathrm{Pin}(V,g)$ *is irreducible and unitary.*

Proof. Since the $\mathbb{C}$-subalgebra of $\mathbb{C}l(V,g)$ generated by $\mathrm{Pin}(V,g)$ is all of $\mathbb{C}l(V,g)$, the irreducibility follows from the fact that μ in (2.14) is an isomorphism. For unitarity, it follows from (2.11) that $\mu(v)$ is self-adjoint for $v \in V$; by (2.12), $\mu(v)^2 = |v|^2 I$, so $v \in V,\ |v| = 1$ implies that $\rho(v)$ is unitary, and unitarity of ρ on $\mathrm{Pin}(V,g)$ follows.

The restriction of ρ to $\mathrm{Spin}(V,g)$ is not irreducible. In fact, set

$$S_+(V,g,J) = \Lambda^{\mathrm{even}}_{\mathbb{C}}\mathcal{V}, \quad S_-(V,g,J) = \Lambda^{\mathrm{odd}}_{\mathbb{C}}\mathcal{V}. \tag{3.14}$$

Under ρ, the action of $\mathrm{Spin}(V,g)$ preserves both S_+ and S_-. In fact, we have (2.14) restricting to

$$\mu : \mathbb{C}l^0(V,g) \longrightarrow \mathrm{End}_{\mathbb{C}}\big(S_+(V,g,J)\big) \oplus \mathrm{End}_{\mathbb{C}}\big(S_-(V,g,J)\big), \tag{3.15}$$

this map being an isomorphism. On the other hand,

$$z \in \mathbb{C}l^1(V,g) \Longrightarrow \mu(z) : S_\pm \longrightarrow S_\mp. \tag{3.16}$$

From (3.15) we get representations

$$D^\pm_{1/2} : \mathrm{Spin}(V,g) \longrightarrow \mathrm{Aut}\big(S_\pm(V,g,J)\big), \tag{3.17}$$

which are irreducible and unitary.

If $V = \mathbb{R}^{2k}$ with its standard Euclidean metric, standard orthonormal basis $e_1,\dots,e_{2k}$, we impose the complex structure $Je_i = e_{i+k},\ Je_{i+k} = -e_i,\ 1 \le i \le k$, and set

$$S(2k) = S(\mathbb{R}^{2k}, |\ |^2, J), \quad S_\pm(2k) = S_\pm(\mathbb{R}^{2k}, |\ |^2, J). \tag{3.18}$$

Then (3.17) defines representations

$$D^{\pm}_{1/2} : \mathrm{Spin}(2k) \longrightarrow \mathrm{Aut}\big(S_{\pm}(2k)\big). \tag{3.19}$$

We now consider the odd dimensional case. If $V = \mathbb{R}^{2k-1}$, we use the isomorphism

$$Cl(2k-1) \longrightarrow Cl^0(2k) \tag{3.20}$$

produced by the map

$$v \mapsto ve_{2k}, \quad v \in \mathbb{R}^{2k-1}. \tag{3.21}$$

Then the inclusion $\mathrm{Spin}(2k-1) \subset Cl(2k-1)$ composed with (3.20) gives an inclusion

$$\mathrm{Spin}(2k-1) \hookrightarrow \mathrm{Spin}(2k). \tag{3.22}$$

Composing with $D^{+}_{1/2}$ from (3.19) gives a representation

$$D^{+}_{1/2} : \mathrm{Spin}(2k-1) \longrightarrow \mathrm{Aut}\, S_{+}(2k). \tag{3.23}$$

We also have a representation $D^{-}_{1/2}$ of $\mathrm{Spin}(2k-1)$ on $S_{-}(2k)$, but these two representations are equivalent. They are intertwined by the map

$$\mu(e_{2k}) : S_{+}(2k) \to S_{-}(2k). \tag{3.24}$$

We now study spinor bundles on an oriented Riemannian manifold M, with metric tensor g. Over M lies the bundle of oriented orthonormal frames,

$$P \longrightarrow M, \tag{3.25}$$

a principal $\mathrm{SO}(n)$-bundle, $n = \dim M$. A spin structure on M is a "lift,"

$$\tilde{P} \longrightarrow M, \tag{3.26}$$

a principal $\mathrm{Spin}(n)$-bundle, such that $\tilde{P}$ is a double covering of P in such a way that the action of $\mathrm{Spin}(n)$ on the fibers of $\tilde{P}$ is compatible with the action of $\mathrm{SO}(n)$ on the fibers of P, via the covering homomorphism $\tau : \mathrm{Spin}(n) \to \mathrm{SO}(n)$. Endowed with such a spin structure, M is called a spin manifold. There are topological obstructions to the existence of a spin structure, which we will not discuss here (see [LM]). It turns out that there is a naturally defined element of $\mathcal{H}^2(M, \mathbb{Z}_2)$ whose vanishing guarantees the existence of a lift, and when such lifts exist, equivalence classes of such lifts are parameterized by elements of $\mathcal{H}^1(M, \mathbb{Z}_2)$.

Given a spin structure as in (3.26), spinor bundles are constructed via the representations of Spin(n) described above. Two cases arise, depending on whether $n = \dim M$ is even or odd. If $n = 2k$, we form the bundle of spinors

$$S(\tilde{P}) = \tilde{P} \times_\rho S(2k), \tag{3.27}$$

where $\rho = D^+_{1/2} \oplus D^-_{1/2}$ is the sum of the representations in (3.19); this is a sum of the two vector bundles

$$S_\pm(\tilde{P}) = \tilde{P} \times_{D^\pm_{1/2}} S_\pm(2k). \tag{3.28}$$

Recall that, as in §6 of Appendix C, on Connections and Curvature, the sections of $S(\tilde{P})$ are in natural correspondence with the functions f on $\tilde{P}$, taking values in the vector space $S(2k)$, which satisfy the compatibility conditions

$$f(p \cdot g) = \rho(g)^{-1} f(p), \quad p \in \tilde{P},\ g \in \text{Spin}(2k), \tag{3.29}$$

where we write the Spin(n)-action on $\tilde{P}$ as a right action.

Recall that $S(2k)$ is a $Cl(2k)$-module, via (2.13). This result extends to the bundle level.

Proposition 3.4. *The spinor bundle $S(\tilde{P})$ is a natural $Cl(M)$-module.*

Proof. Given a section u of $Cl(M)$ and a section φ of $S(\tilde{P})$, we need to define $u \cdot \varphi$ as a section of $S(\tilde{P})$. We regard u as a function on $\tilde{P}$ with values in $Cl(n)$ and φ as a function on $\tilde{P}$ with values in $S(n)$. Then $u \cdot \varphi$ is a function on $\tilde{P}$ with values in $S(n)$; we need to verify the compatibility condition (3.29). Indeed, for $p \in \tilde{P},\ g \in \text{Spin}(2k)$,

$$\begin{aligned} u \cdot \varphi(p \cdot g^{-1}) &= \tau(g)u(p) \cdot \rho(g)\varphi(p) \\ &= gu(p)g^{\#}g\varphi(p) \\ &= gu(p) \cdot \varphi(p), \end{aligned} \tag{3.30}$$

since $gg^{\#} = 1$ for $g \in \text{Spin}(n)$. This completes the proof.

Whenever (M, g) is an oriented Riemannian manifold, the Levi–Civita connection provides a connection on the principal SO(n)-bundle of frames P. If M has a spin structure, this choice of horizontal space for P lifts in a unique natural fashion to provide a connection on $\tilde{P}$. Thus the spinor bundle constructed above has a natural connection, which we will call the Dirac–Levi–Civita connection.

Proposition 3.5. *The Dirac–Levi–Civita connection ∇ on $S(\tilde{P})$ is a Clifford connection.*

Proof. Clearly, ∇ is a metric connection, since the representation ρ of $\mathrm{Spin}(2k)$ on $S(2k)$ is unitary. It remains to verify the compatibility condition (1.36), namely,

$$\nabla_X(v\cdot\varphi) = (\nabla_X v)\cdot\varphi + v\cdot\nabla_X\varphi, \tag{3.31}$$

for a vector field X, a 1-form v, and a section φ of $S(\tilde{P})$. To see this, we first note that as stated in (3.30), the bundle $Cl(M)$ can be obtained from $\tilde{P}\to M$ as $\tilde{P}\times_\kappa Cl(2k)$, where κ is the representation of $\mathrm{Spin}(2k)$ on $Cl(2k)$ given by $\kappa(g)w = gwg^{\#}$. Furthermore, T^*M can be regarded as a subbundle of $Cl(M)$, obtained from $\tilde{P}\times_\kappa \mathbb{R}^{2k}$ with the same formula for κ. The connection on T^*M obtained from that on $\tilde{P}$ is identical to the usual connection on T^*M defined via the Levi–Civita formula. Given this, (3.31) is a straightforward derivation identity.

Using the prescription (1.31)–(1.33), we can define the Dirac operator on a Riemannian manifold of dimension $2k$, with a spin structure:

$$D : C^\infty(M, S(\tilde{P})) \longrightarrow C^\infty(M, S(\tilde{P})). \tag{3.32}$$

We see that Proposition 1.1 applies; D is symmetric. Note also the grading:

$$D : C^\infty(M, S_\pm(\tilde{P})) \longrightarrow C^\infty(M, S_\mp(\tilde{P})). \tag{3.33}$$

In other words, this Dirac operator is of the form (1.3).

On a Riemannian manifold of dimension $2k$ with a spin structure $\tilde{P}\to M$, let $F\to M$ be another vector bundle. Then the tensor product $E = S(\tilde{P})\otimes F$ is a $Cl(M)$-module in a natural fashion. If F has a connection, then E gets a natural product connection. Then the construction (1.31)–(1.33) yields an operator D_F of Dirac type on sections of E; in fact

$$D_F : C^\infty(M, E_\pm) \longrightarrow C^\infty(M, E_\mp), \quad E_\pm = S_\pm(\tilde{P})\otimes F. \tag{3.34}$$

If F has a metric connection, then E gets a Clifford connection. The operator D_F is called a *twisted* Dirac operator. Sometimes it will be convenient to distinguish notationally the two pieces of D_F; we write

$$\begin{aligned} D_F^+ &: C^\infty(M, E_+) \longrightarrow C^\infty(M, E_-),\\ D_F^- &: C^\infty(M, E_-) \longrightarrow C^\infty(M, E_+). \end{aligned} \tag{3.35}$$

When $\dim M = 2k-1$ is odd, we use the representation (3.23) to form the bundle of spinors

$$S_+(\tilde{P}) = \tilde{P}\times_{D^+_{1/2}} S_+(2k).$$

The inclusion $Cl(2k-1)\hookrightarrow Cl^0(2k)$ defined by (3.20)–(3.21) makes $S_+(2k)$ a $Cl(2k-1)$-module, and analogues of Propositions 3.4 and 3.5 hold. Hence there arises a Dirac operator, $D : C^\infty(M, S_+(\tilde{P})) \to C^\infty(M, S_+(\tilde{P}))$. Twisted Dirac operators also arise; however, in place of (3.34), we have $D_F : C^\infty(M, E_+) \to C^\infty(M, E_+)$, with $E_+ = S_+(\tilde{P}) \otimes F$.

Exercises

1. Verify that the map (3.15) is an isomorphism and that the representations (3.17) of $\mathrm{Spin}(V, g)$ are irreducible when $\dim V = 2k$.
2. Let ν be as in Exercises 1–4 of §2, with $n = 2k$. Show that

   a) the center of $\mathrm{Spin}(V, g)$ consists of $\{1, -1, \nu, -\nu\}$,
   b) $\mu(\nu)$ leaves S_+ and S_- invariant,
   c) $\mu(\nu)$ commutes with the action of $Cl^0(V, g)$ under μ, hence with the representations $D^\pm_{1/2}$ of $\mathrm{Spin}(V, g)$,
   d) $\mu(\nu)$ acts as a pair of scalars on S_+ and S_-, respectively. These scalars are the two square roots of $(-1)^k$.

3. Calculate $\mu(\nu) \cdot 1$ directly, making use of the definition (2.11). Hence match the scalars in exercise 2d) to S_+ and S_-. (*Hint*: $\mu(e_{k+1} \cdots e_{2k}) \cdot 1 = (-i)^k e_{k+1} \wedge \cdots \wedge e_{2k}$ in $\Lambda^k_{\mathbb{C}} \mathcal{V}$. Using $e_{j+k} = i\, e_j$ in $\mathcal{V}$, for $1 \le j \le k$, we have

$$\mu(\nu) \cdot 1 = \mu(e_1 \cdots e_k)(e_1 \wedge \cdots \wedge e_k),$$

   and there are k interior products to compute.)
4. Show that $Cl^{[2]}(V, g)$, with the Lie algebra structure (2.19), is naturally isomorphic to the Lie algebra of $\mathrm{Spin}(V, g)$. In fact, if (a_{jk}) is a real, antisymmetric matrix, in the Lie algebra of $\mathrm{SO}(n)$, which is the same as that of $\mathrm{Spin}(n)$, show that there is the correspondence

$$A = (a_{jk}) \mapsto \frac{1}{4} \sum a_{jk}\, e_j\, e_k = \kappa(A).$$

   In particular, show that $\kappa(A_1 A_2 - A_2 A_1) = \kappa(A_1)\kappa(A_2) - \kappa(A_2)\kappa(A_1)$.
5. If X is a spin manifold and $M \subset X$ is an oriented submanifold of codimension 1, show that M has a spin structure. Deduce that an oriented hypersurface in $\mathbb{R}^n$ has a spin structure.

4. Weitzenbock formulas

Let $E \to M$ be a Hermitian vector bundle with a metric connection ∇. Suppose E is also a $Cl(M)$-module and that ∇ is a Clifford connection. If we consider the Dirac-type operator $D : C^\infty(M, E) \to C^\infty(M, E)$ and the covariant derivative $\nabla : C^\infty(M, E) \to C^\infty(M, T^* \otimes E)$, then D^2 and $\nabla^*\nabla$ are operators on $C^\infty(M, E)$ with the same principal symbol. It is of interest to examine their difference, clearly a differential operator of order ≤ 1. In fact, the difference has order 0. This can be seen in principle from the following considerations. From Exercise 4 of §1, we have

$$D^2(f\varphi) = f\, D^2\varphi - 2\nabla_{\mathrm{grad}\, f}\varphi - (\Delta f)\varphi \tag{4.1}$$

when $\varphi \in C^\infty(M, E)$, f a scalar function. Similarly, we compute $\nabla^*\nabla(f\varphi)$. The derivation property of ∇ implies

$$\nabla(f\varphi) = f\nabla\varphi + df \otimes \varphi. \tag{4.2}$$

To apply ∇^* to this, first a short calculation gives

$$\nabla^* f(u \otimes \varphi) = f\nabla^*(u \otimes \varphi) - \langle df, u\rangle\varphi, \tag{4.3}$$

for $u \in C^\infty(M, T^*)$, $\varphi \in C^\infty(M, E)$, and hence

$$\nabla^*(f\nabla\varphi) = f\nabla^*\nabla\varphi - \nabla_{\mathrm{grad}\, f}\varphi. \tag{4.4}$$

This gives ∇^* applied to the first term on the right side of (4.2). To apply ∇^* to the other term, we can use the identity (see Appendix C, (1.35))

$$\nabla^*(u \otimes \varphi) = -\nabla_U\varphi - (\mathrm{div}\, U)\varphi, \tag{4.5}$$

where U is the vector field corresponding to u via the metric on M. Hence

$$\nabla^*(df \otimes \varphi) = -\nabla_{\mathrm{grad}\, f}\varphi - (\Delta f)\varphi. \tag{4.6}$$

Then (4.6) and (4.4) applied to (4.2) gives

$$\nabla^*\nabla(f\varphi) = f\nabla^*\nabla\varphi - 2\nabla_{\mathrm{grad}\, f}\varphi - (\Delta f)\varphi. \tag{4.7}$$

Comparing (4.1) and (4.7), we have

$$(D^2 - \nabla^*\nabla)(f\varphi) = f(D^2 - \nabla^*\nabla)\varphi, \tag{4.8}$$

which implies $D^2 - \nabla^*\nabla$ has order zero, hence is given by a bundle map on E. We now derive the Weitzenbock formula for what this difference is.

Proposition 4.1. *If $E \to M$ is a $Cl(M)$-module with Clifford connection and associated Dirac-type operator D, then, for $\varphi \in C^\infty(M, E)$,*

$$D^2\varphi = \nabla^*\nabla\varphi - \sum_{j>k} v_k v_j K(e_k, e_j)\varphi, \tag{4.9}$$

where $\{e_j\}$ is a local orthonormal frame of vector fields, with dual frame field $\{v_j\}$, and K is the curvature tensor of (E, ∇).

Proof. Starting with $D\varphi = i\sum v_j \nabla_{e_j}\varphi$, we obtain

$$
\begin{aligned}
D^2\varphi &= -\sum_{j,k} v_k \nabla_{e_k}\big(v_j \nabla_{e_j}\varphi\big) \\
&= -\sum_{j,k} v_k\Big[v_j \nabla_{e_k}\nabla_{e_j}\varphi + \big(\nabla_{e_k} v_j\big)\nabla_{e_j}\varphi\Big],
\end{aligned}
\tag{4.10}
$$

using the compatibility condition (1.36). We replace $\nabla_{e_k}\nabla_{e_j}$ by the Hessian, using the identity

$$
\nabla^2_{e_k,e_j}\varphi = \nabla_{e_k}\nabla_{e_j}\varphi - \nabla_{\nabla_{e_k}e_j}\varphi;
\tag{4.11}
$$

cf. (2.4) of Appendix C. We obtain

$$
\begin{aligned}
D^2\varphi = &-\sum_{j,k} v_k v_j \nabla^2_{e_k,e_j}\varphi \\
&-\sum_{j,k} v_k\Big[v_j \nabla_{\nabla_{e_k}e_j}\varphi + \big(\nabla_{e_k} v_j\big)\nabla_{e_j}\varphi\Big].
\end{aligned}
\tag{4.12}
$$

Let us look at each of the two double sums on the right. Using $v_j^2 = 1$ and the anticommutator property $v_k v_j = -v_j v_k$ for $k \neq j$, we see that the first double sum becomes

$$
-\sum_j \nabla^2_{e_j,e_j}\varphi - \sum_{j>k} v_k v_j K(e_k, e_j)\varphi,
\tag{4.13}
$$

since the antisymmetric part of the Hessian is the curvature. This is equal to the right side of (4.9), in light of the formula for $\nabla^*\nabla$ established in Proposition 2.1 of Appendix C. As for the remaining double sum in (4.12), for any $p \in M$, we can choose a local orthonormal frame field $\{e_j\}$ such that $\nabla_{e_j} e_k = 0$ at p, and then this term vanishes at p. This proves (4.9).

We denote the difference $D^2 - \nabla^*\nabla$ by $\mathcal{K}$, so

$$
(D^2 - \nabla^*\nabla)\varphi = \mathcal{K}\varphi, \quad \mathcal{K} \in C^\infty(M, \operatorname{End} E).
\tag{4.14}
$$

The formula for $\mathcal{K}$ in (4.9) can also be written as

$$
\mathcal{K}\varphi = -\frac{1}{2}\sum_{j,k} v_k v_j K(e_k, e_j)\varphi.
\tag{4.15}
$$

Since a number of formulas that follow will involve multiple summation, we will use the summation convention.

This general formula for $\mathcal{K}$ simplifies further in some important special cases. The first simple example of this will be useful for further calculations.

Proposition 4.2. *Let $E = \Lambda^* M$, with $Cl(M)$-module structure and connection described in §1, so $\mathcal{K} \in C^\infty(M, End\, \Lambda^*)$. In this case,*

$$u \in \Lambda^1 M \Longrightarrow \mathcal{K}u = Ric(u). \tag{4.16}$$

Proof. The curvature of $\Lambda^* M$ is a sum of curvatures of each factor $\Lambda^k M$. In particular, if $\{e_j, v_j\}$ is a local dual pair of frame fields,

$$K(e_i, e_j)v_k = -R^k{}_{\ell ij} v_\ell, \tag{4.17}$$

where $R^k{}_{\ell ij}$ are the components of the Riemann tensor, with respect to these frame fields, and we use the summation convention. In light of (4.15), the desired identity (4.16), will hold provided

$$\frac{1}{2} v_i v_j v_\ell R^k{}_{\ell ij} = \text{Ric}(v_k), \tag{4.18}$$

so it remains to establish this identity. Since, if (i, j, ℓ) are distinct, $v_i v_j v_\ell = v_\ell v_i v_j = v_j v_\ell v_i$, and since by Bianchi's first identity

$$R^k{}_{\ell ij} + R^k{}_{j\ell i} + R^k{}_{ij\ell} = 0,$$

it follows that in summing the left side of (4.18), the sum over (i, j, ℓ) distinct vanishes. By antisymmetry of $R^k{}_{\ell ij}$, the terms with $i = j$ vanish. Thus the only contributions arise from $i = \ell \neq j$ and $i \neq \ell = j$. Therefore, the left side of (4.18) is equal to

$$\frac{1}{2}\big(-v_j R^k{}_{iij} + v_i R^k{}_{jij}\big) = v_i R^k{}_{jij} = \text{Ric}(v_k), \tag{4.19}$$

which completes the proof.

We next derive Lichnerowicz's calculation of $\mathcal{K}$ when $E = S(\tilde{P})$, the spinor bundle of a manifold M with spin structure. First we need an expression for the curvature of $S(\tilde{P})$.

Lemma 4.3. *The curvature tensor of the spinor bundle $S(\tilde{P})$ is given by*

$$K(e_i, e_j)\varphi = \frac{1}{4} R^k{}_{\ell ij} v_k v_\ell \varphi. \tag{4.20}$$

Proof. This follows from the relation between curvatures on vector bundles and on principal bundles established in Appendix C, §6, together with the identification of the Lie algebra of Spin(n) with $Cl^{[2]}(n)$ given in Exercise 4 of §3.

Proposition 4.4. *For the spin bundle $S(\tilde{P})$, $\mathcal{K} \in C^\infty(M, End\, S(\tilde{P}))$ is given by*

$$\mathcal{K}\varphi = \frac{1}{4} S\varphi, \tag{4.21}$$

where S is the scalar curvature of M.

Proof. Using (4.20), the general formula (4.15) yields

$$\mathcal{K}\varphi = -\frac{1}{8} R^k{}_{\ell ij}\; v_i v_j v_k v_\ell \varphi = \frac{1}{8} v_i v_j v_\ell R^k{}_{\ell ij}\; v_k \varphi, \tag{4.22}$$

the last identity holding by the anticommutation relations; note that only the sum over $k \neq \ell$ counts. Now, by (4.18), this becomes

$$\begin{aligned} \mathcal{K}\varphi &= \frac{1}{4} v_i v_k R^k{}_{jij} \varphi \\ &= \frac{1}{4}\, \mathrm{Ric}_{ii} \varphi \quad \text{(by symmetry)} \\ &= \frac{1}{4} S\varphi, \end{aligned} \tag{4.23}$$

completing the proof.

We record the generalization of Proposition 4.4 to the case of twisted Dirac operators. We mention that one often sees a different sign before the sum, due to a different sign convention for Clifford algebras.

Proposition 4.5. *Let $E \to M$ have a metric connection ∇, with curvature R^E. For the twisted Dirac operator on sections of $F = S(\tilde{P}) \otimes E$, the section $\mathcal{K}$ of End F has the form*

$$\mathcal{K}\varphi = \frac{1}{4} S\varphi - \frac{1}{2} \sum_{i,j} v_i v_j R^E(e_i, e_j)\varphi. \tag{4.24}$$

Proof. Here $R^E(e_i, e_j)$ is shorthand for $I \otimes R^E(e_i, e_j)$ acting on $S(\tilde{P}) \otimes E$. This formula is a consequence of the general formula (4.15) and the argument proving Proposition 4.4, since the curvature of $S(\tilde{P}) \otimes E$ is $K \otimes I + I \otimes R^E$, K being the curvature of $S(\tilde{P})$, given by (4.20).

These Weitzenbock formulas will be of use in the following sections. Here we draw some interesting conclusions, due to Bochner and Lichnerowitz.

Proposition 4.6. *If M is compact and connected, and the section $\mathcal{K}$ in (4.14)–(4.15) has the property that $\mathcal{K} \geq 0$ on M and $\mathcal{K} > 0$ at some point, then ker $D = 0$.*

Proof. This is immediate from

$$(D^2\varphi,\varphi)=(\mathcal{K}\varphi,\varphi)+\|\nabla\varphi\|_{L^2}^2.$$

Proposition 4.7. *If M is a compact Riemannian manifold with positive Ricci tensor, then $b_1(M)=0$, that is, the deRham cohomology group $\mathcal{H}^1(M,\mathbb{R})=0$.*

Proof. Via Hodge theory, we want to show that if $u\in\Lambda^1(M)$ and $du=d^*u=0$, then $u=0$. This hypothesis implies $Du=0$, where D is the Dirac-type operator dealt with in Proposition 4.2. Consequently we have, for a 1-form u on M,

$$\|Du\|_{L^2}^2=\big(\mathrm{Ric}(u),u\big)+\|\nabla u\|_{L^2}^2, \tag{4.25}$$

so the result follows.

Proposition 4.8. *If M is a compact, connected Riemannian manifold with a spin structure whose scalar curvature is ≥ 0 on M and >0 at some point, then M has no nonzero harmonic spinors, that is, ker $D=0$ in $C^\infty(M,S(\tilde{P}))$.*

Proof. In light of (4.21), this is a special case of Proposition 4.6.

Exercises

1. Let Δ be the Laplace operator on functions (0-forms) on a compact Riemannian manifold M, Δ_k the Hodge Laplacian on k-forms. If Spec$(-\Delta)$ consists of $0=\lambda_0<\lambda_1\leq\lambda_2\leq\cdots$, show that $\lambda_1\in\mathrm{Spec}(-\Delta_1)$.
2. If $\mathrm{Ric}\geq c_0I$ on M, show that $\lambda_1\geq c_0$.
3. Recall the deformation tensor of a vector field u:

$$\mathrm{Def}\,u=\frac{1}{2}\mathcal{L}_ug=\frac{1}{2}(\nabla u+\nabla u^t),\quad \mathrm{Def}:C^\infty(M,T)\to C^\infty(M,S^2).$$

   Show that

$$\mathrm{Def}^*v=-\,\mathrm{div}\,v,$$

   where $(\mathrm{div}\,v)^j=v^{jk}{}_{;k}$. Establish the Weitzenbock formula

$$2\,\mathrm{div}\,\mathrm{Def}\,u=-\nabla^*\nabla u+\ \mathrm{grad}\,\mathrm{div}\,u+\ \mathrm{Ric}(u). \tag{4.26}$$

   The operator div on the right is the usual divergence operator on vector fields. (This formula will appear again in Chap. 17, in the study of the Navier–Stokes equation.)
4. Suppose M is a compact, connected Riemannian manifold, whose Ricci tensor satisfies

$$\mathrm{Ric}(x)\leq 0 \text{ on } M,\quad \mathrm{Ric}(x_0)<0,\ \text{for some } x_0\in M. \tag{4.27}$$

   Show that the operator Def is injective, so there are no nontrivial Killing fields on M, hence no nontrivial one-parameter groups of isometries. (*Hint*: From (4.26), we have

$$2\|\mathrm{Def}\,u\|_{L^2}^2=\|\nabla u\|_{L^2}^2+\|\mathrm{div}\,u\|^2-\big(\mathrm{Ric}(u),u\big)_{L^2}.) \tag{4.28}$$

5. As shown in (3.39) of Chap. 2, the equation of a conformal Killing field on an n-dimensional Riemannian manifold M is

$$\text{Def } X - \frac{1}{n}\big(\text{div } X\big)g = 0. \tag{4.29}$$

Note that the left side is the trace-free part of Def $X \in C^\infty(M, S^2T^*)$. Denote it by $\mathcal{D}_{TF}X$. Show that

$$\mathcal{D}_{TF}^* = -\,\text{div}\big|_{S_0^2T^*}, \quad \mathcal{D}_{TF}^*\mathcal{D}_{TF}X = -\,\text{div Def } X + \frac{1}{n}\big(\text{grad div } X\big), \tag{4.30}$$

where $S_0^2T^*$ is the trace-free part of S^2T^*. Show that

$$\|\mathcal{D}_{TF}X\|_{L^2}^2 = \frac{1}{2}\|\nabla X\|_{L^2}^2 + \left(\frac{1}{2} - \frac{1}{n}\right)\|\text{div } X\|_{L^2}^2 - \frac{1}{2}\big(\text{Ric}(X), X\big)_{L^2}. \tag{4.31}$$

Deduce that if M is compact and satisfies (4.27), then M has no nontrivial one-parameter group of conformal diffeomorphisms.

6. Show that if M is a compact Riemannian manifold which is Ricci flat (i.e., Ric $= 0$), then every conformal Killing field is a Killing field, and the dimension of the space of Killing fields is given by

$$\dim_{\mathbb{R}} \ker \text{Def} \;=\; \dim \mathcal{H}^1(M, \mathbb{R}). \tag{4.32}$$

(*Hint*: Combine (4.25) and (4.28).)

7. Suppose $\dim M = 2$ and M is compact and connected. Show that, for $u \in C^\infty(M, S_0^2T^*)$,

$$\|\mathcal{D}_{TF}^* u\|_{L^2}^2 = \frac{1}{2}\|\nabla u\|_{L^2}^2 + \int_M K|u|^2\, dV,$$

where K is the Gauss curvature. Deduce that if $K \ge 0$ on M, and $K(x_0) > 0$ for some $x_0 \in M$, then Ker $\mathcal{D}_{TF}^* = 0$. Compare with Exercises 6–8 of §10.

8. If u and v are vector fields on a Riemannian manifold M, show that

$$\text{div } \nabla_u v = \nabla_u(\text{div } v) + \text{Tr}\big((\nabla u)(\nabla v)\big) - \text{Ric}(u, v). \tag{4.33}$$

Compare with formula (3.17) in Chap. 17, on the Euler equation. Relate this identity to the Weitzenbock formula for Δ on 1-forms (a special case of Proposition 4.2).

5. Index of Dirac operators

If $D : C^\infty(M, E_0) \to C^\infty(M, E_1)$ is an elliptic, first-order differential operator between sections of vector bundles E_0 and E_1 over a compact manifold M, then, as we have seen, $D : H^{k+1}(M, E_0) \to H^k(M, E_1)$ is Fredholm, for any real k. Furthermore, ker D is a finite-dimensional subspace of $C^\infty(M, E_0)$, independent of k, and $D^* : H^{-k}(M, E_1) \to H^{-k-1}(M, E_0)$ has the same properties. A quantity of substantial importance is the *index* of D:

$$\text{Index } D = \dim\ker D - \dim\ker D^*. \tag{5.1}$$

In this section and the next we derive a formula for this index, due to Atiyah and Singer. Later sections will consider a few applications of this formula.

One basic case for such index theorems is that of twisted Dirac operators. Thus, let $F \to M$ be a vector bundle with metric connection, over a compact Riemannian manifold M with a spin structure. Assume $\dim M = n = 2k$ is even. The twisted Dirac operator constructed in §3 in particular gives an elliptic operator

$$D_F : C^\infty(M, S_+(\tilde{P}) \otimes F) \longrightarrow C^\infty(M, S_-(\tilde{P}) \otimes F). \tag{5.2}$$

The Atiyah–Singer formula for the index of this operator is given as follows.

Theorem 5.1. *If M is a compact Riemannian manifold of dimension $n = 2k$ with spin structure and D_F the twisted Dirac operator (5.2), then*

$$\textit{Index } D_F = \langle \hat{A}(M)\, Ch(F), [M] \rangle. \tag{5.3}$$

What is meant by the right side of (5.3) is the following. $\hat{A}(M)$ and $\mathrm{Ch}(F)$ are certain *characteristic classes*; each is a sum of even-order differential forms on M, computed from the curvatures of $S(\tilde{P})$ and F, respectively. We will derive explicit formulas for what these are in the course of the proof of this theorem, in the next section, so we will not give the formulas here. The pairing with M indicated in (5.3) is the integration over M of the form of degree $2k = n$ arising in the product $\hat{A}(M)\mathrm{Ch}(F)$.

The choice of notation in $\hat{A}(M)$ and $\mathrm{Ch}(F)$ indicates an independence of such particulars as the choice of Riemannian metric on M and of connection on F. This is part of the nature of characteristic classes, at least after integration is performed; for a discussion of this, see §7 of Appendix C. There is also a simple direct reason why Index D_F does not depend on such choices. Namely, any two Riemannian metrics on M can be deformed to each other, and any two connections on F can be deformed to each other. The invariance of the index of D_F is thus a special case of the following.

Proposition 5.2. *If D_s, $0 \le s \le 1$, is a continuous family of elliptic differential operators $D_s : C^\infty(M, E_0) \to C^\infty(M, E_1)$ of first order, then Index D_s is independent of s.*

Proof. We have a norm-continuous family of Fredholm operators $D_s : H^1(M, E_0) \to L^2(M, E_1)$; the constancy of the index of any continuous family of Fredholm operators is proved in Appendix A, Proposition 7.4.

The proof of Theorem 5.1 will be via the heat-equation method, involving a comparison of the spectra of D^*D and DD^*, self-adjoint operators on $L^2(M, E_0)$ and $L^2(M, E_1)$, respectively. As we know, since $D^*D = L_0$ and $DD^* = L_1$ are both elliptic and self-adjoint, they have discrete spectra, with eigenspaces of finite dimension, contained in $C^\infty(M, E_j)$, say

$$\mathrm{Eigen}(L_j, \lambda) = \{u \in C^\infty(M, E_j) : L_j u = \lambda u\}. \tag{5.4}$$

We have the following result:

Proposition 5.3. *The spectra of* L_0 *and* L_1 *are discrete subsets of* $[0,\infty)$ *which coincide, except perhaps at* 0*. All nonnero eigenvalues have the same finite multiplicity.*

Proof. It is easy to see that for each $\lambda \in [0,\infty)$, $D : \text{Eigen}(L_0,\lambda) \to \text{Eigen}(L_1,\lambda)$ and $D^* : \text{Eigen}(L_1,\lambda) \to \text{Eigen}(L_0,\lambda)$. For $\lambda \neq 0$, D and $\lambda^{-1}D^*$ are inverses of each other on these spaces.

We know from the spectral theory of Chap. 8 that $\varphi(L_0)$ and $\varphi(L_1)$ are trace class for any $\varphi \in \mathcal{S}(\mathbb{R})$. We hence have the following.

Proposition 5.4. *For any* $\varphi \in \mathcal{S}(\mathbb{R})$*, with* $\varphi(0) = 1$*,*

$$\text{Index } D = \text{Tr}\,\varphi(D^*D) - \text{Tr}\,\varphi(DD^*). \tag{5.5}$$

In particular, for any $t > 0$*,*

$$\text{Index } D = \text{Tr}\,e^{-tD^*D} - \text{Tr}\,e^{-tDD^*}. \tag{5.6}$$

Now, whenever D is of Dirac type, so $D^*D = L_0$ and $DD^* = L_1$ have scalar principal symbol, results of Chap. 7 show that

$$e^{-tL_j}u(x) = \int_M k_j(t,x,y)\,u(y)\,dV(y), \tag{5.7}$$

with

$$k_j(t,x,x) \sim t^{-n/2}\Big[a_{j0}(x) + a_{j1}(x)t + \cdots + a_{j\ell}(x)t^\ell + \cdots\Big], \tag{5.8}$$

as $t \searrow 0$, with $a_j \in C^\infty(M, \text{End } E_j)$, so

$$\text{Tr}\,e^{-tL_j} \sim t^{-n/2}\Big(b_{j0} + b_{j1}t + \cdots + b_{j\ell}t^\ell + \cdots\Big), \tag{5.9}$$

with

$$b_{j\ell} = \int_M \text{Tr}\,a_{j\ell}(x)\,dV(x). \tag{5.10}$$

In light of (5.6), we have the following result:

Proposition 5.5. *If D is of Dirac type on M, of dimension $n = 2k$, then*

$$\text{(5.11)} \qquad \textit{Index } D = b_{0k} - b_{1k} = \int_M Tr\left[a_{0k}(x) - a_{1k}(x)\right] dV(x),$$

where $a_{j\ell}$ are the coefficients in (5.8).

We remark that these calculations are valid for dim $M = n$ odd. In that case, there is no coefficient of t^0 in (5.8) or (5.9), so the identity (5.6) implies Index $D = 0$ for dim M odd. In fact, this holds for any elliptic differential operator, not necessarily of Dirac type. On the other hand, if dim M is odd, there exist elliptic pseudodifferential operators on M with nonzero index.

We will establish the Atiyah–Singer formula (5.3) in the next section by showing that, for a twisted Dirac operator D_F, the $2k$-form part of the right side of the formula (5.3), with $\hat{A}(M)$ and $Ch(F)$ given by curvatures in an appropriate fashion, is equal *pointwise* on M to the integrand in (5.11). Such an identity is called a *local index formula*.

6. Proof of the local index formula

Let D_F be a twisted Dirac operator on a compact spin manifold, as in (5.2). If $L_0 = D_F^* D_F$ and $L_1 = D_F D_F^*$, we saw in §5 that, for all $t > 0$,

$$\text{(6.1)} \qquad \text{Index } D_F = \int_M \left[\text{Tr}\, k_0(t,x,x) - \text{Tr}\, k_1(t,x,x)\right] dV(x),$$

where $k_j(t,x,y)$ are the Schwartz kernels of the operators e^{-tL_j}. In the index formula stated in (5.3), $\hat{A}(M)$ and $Ch(F)$ are to be regarded as differential forms on M, arising in a fashion we will specify later in this section, from curvature forms given by the spin structure on M and a connection on F; the product is the wedge product of forms. The following is the local index formula, which refines (5.3).

Theorem 6.1. *For the twisted Dirac operator D_F, we have the pointwise identity*

$$\text{(6.2)} \qquad \lim_{t\to 0}\left[Tr\, k_0(t,x,x) - Tr\, k_1(t,x,x)\right] dV = \left\{\hat{A}(M) \wedge Ch(F)\right\}_n,$$

where $\{\beta\}_n$ denotes the component of degree $n = \dim M$ of a differential form β, and dV denotes the volume form of the oriented manifold M.

We first obtain a formula for the difference in the traces of $k_0(t,x,x)$ and of $k_1(t,x,x)$, which are elements of $\text{End}((S_\pm)_x \otimes F_x)$. It is convenient to put these together, and consider

$$\text{(6.3)} \qquad K = \begin{pmatrix} k_0 & 0 \\ 0 & k_1 \end{pmatrix} \in \text{End}(S \otimes F),$$

where $S = S_+ \oplus S_-$, and we have dropped x and t. Using the isomorphism (2.14), $\mu : \mathbb{C}l(2k) \to \text{End}\, S$, we can write

$$\text{End}(S \otimes F) = \mathbb{C}l(2k) \otimes \text{End}(F). \tag{6.4}$$

We will suppose $\dim M = n = 2k$. In other words, we can think of an element of $\text{End}(S \otimes F)$ as a combination of elements of the Clifford algebra, whose coefficients are linear transformations on F. Since (6.3) preserves $S_+ \otimes F$ and $S_- \otimes F$, we have

$$K \in \mathbb{C}l^0(2k) \otimes \text{End}(F). \tag{6.5}$$

For K of the form (6.3), the difference $\text{Tr}\, k_0 - \text{Tr}\, k_1$ is called the "supertrace" of K, written

$$\text{Str}\, K = \text{Tr}(\varepsilon K), \quad \text{with } \varepsilon = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{6.6}$$

The first key step in establishing (6.2) is the following identity, which arose in the work of F. Berezin [Ber] and V. Patodi [Pt1]. Define the map

$$\tau : \mathbb{C}l(2k) \longrightarrow \mathbb{C} \tag{6.7}$$

to be the evaluation of the coefficient of the "volume element" $\nu = e_1 \cdots e_{2k}$, introduced in Exercises 1–4 of §2. Similarly define

$$\tau_F : \mathbb{C}l(2k) \otimes \text{End}(F) \longrightarrow \text{End}\, F, \quad \tilde{\tau} : \mathbb{C}l(2k) \otimes \text{End}(F) \longrightarrow \mathbb{C} \tag{6.8}$$

to be

$$\tau_F = \tau \otimes I, \quad \tilde{\tau} = \text{Tr} \circ \tau_F, \tag{6.9}$$

where the last trace is $\text{Tr} : \text{End}\, F \to \mathbb{C}$.

Lemma 6.2. *The supertrace is given by*

$$Str\, K = (-2i)^k \tilde{\tau}(K), \tag{6.10}$$

using the identification (6.4).

Proof. If this is established for the case $F = \mathbb{C}$, the general case follows easily. We note that, with $\nu = e_1 \cdots e_{2k}$,

$$S_\pm = \{x \in S : \mu(i^k \nu)x = \pm x\}.$$

Thus, for $K \in \mathbb{C}l(2k)$,

$$\text{Str}\, K = \text{Tr}(i^k \nu K). \tag{6.11}$$

Thus (6.10) is equivalent to

$$\operatorname{Tr} w = 2^k\, w_0, \tag{6.12}$$

for $w \in \mathbb{C}l(2k) \approx \operatorname{End} S$, where w_0 is the scalar term in the expansion (2.2) for w. This in turn follows from

$$\operatorname{Tr} 1 = 2^k \tag{6.13}$$

and

$$\operatorname{Tr} e_1^{i_1} \cdots e_n^{i_n} = 0 \quad \text{if } i_1 + \cdots + i_n > 0, \quad i_\nu = 0 \text{ or } 1. \tag{6.14}$$

To verify these identities, note that 1 acts on S as the identity, so (6.13) holds by the computation of dim S. As for (6.14), using $S \otimes S' \approx \mathbb{C}l(2k)$, we see that (6.14) is a multiple of the trace of $e_1^{i_1} \cdots e_n^{i_n}$ acting on $\mathbb{C}l(2k)$ by Clifford multiplication, which is clearly zero. The proof is complete.

Thus we want to analyze the $Cl^{[2k]}(2k) \otimes \operatorname{End} F$ component of $K(t,x,x)$, the value on the diagonal of $K(t,x,y)$, the Schwartz kernel of

$$e^{-tD_F^2} = \begin{pmatrix} e^{-tL_0} & 0 \\ 0 & e^{-tL_1} \end{pmatrix}.$$

We recall that a construction of $K(t,x,y)$ was made in Chap. 7, §13. It was shown that, in local coordinates and with a local choice of trivializations of $S(\tilde{P})$ and of F, we could write, modulo a negligible error,

$$e^{-tD_F^2} u(x) = (2\pi)^{-n/2} \int a(t,x,\xi) \hat{u}(\xi) e^{ix\cdot\xi}\, d\xi, \tag{6.15}$$

where the amplitude $a(t,x,\xi)$ has an asymptotic expansion

$$a(t,x,\xi) \sim \sum_{j \geq 0} a_j(t,x,\xi). \tag{6.16}$$

The terms $a_j(t,x,\xi)$ were defined recursively in the following manner. If, with such local coordinates and trivializations,

$$D_F^2 = L(x, D_x), \tag{6.17}$$

then, by the Leibniz formula, write

(6.18)

$$\begin{aligned} L(a\, e^{ix\cdot\xi}) &= e^{ix\cdot\xi} \sum_{|\alpha| \leq 2} \frac{i^{|\alpha|}}{\alpha!} L^{(\alpha)}(x,\xi)\; D_x^\alpha a(t,x,\xi) \\ &= e^{ix\cdot\xi} \left[L_2(x,\xi) a(t,x,\xi) + \sum_{\ell=1}^{2} B_{2-\ell}(x,\xi,D_x) a(t,x,\xi) \right], \end{aligned}$$

where $B_{2-\ell}(x, \xi, D_x)$ is a differential operator (of order ℓ) whose coefficients are polynomials in ξ, homogeneous of degree $2-\ell$ in ξ. $L_2(x, \xi)$ is the principal symbol of $L = D_F^2$.

Thus we want the amplitude $a(t, x, \xi)$ in (6.15) to satisfy formally

$$\frac{\partial a}{\partial t} \sim -L_2 a - \sum_{\ell=1}^{2} B_{2-\ell}(x, \xi, D_x) a. \tag{6.19}$$

If a is taken to have the form (6.16), we produce the following transport equations for a_j:

$$\frac{\partial a_0}{\partial t} = -L_2(x, \xi) a_0(t, x, \xi) \tag{6.20}$$

and, for $j \geq 1$,

$$\frac{\partial a_j}{\partial t} = -L_2(x, \xi) a_j + \Omega_j(t, x, \xi), \tag{6.21}$$

where

$$\Omega_j(t, x, \xi) = -\sum_{\ell=1}^{2} B_{2-\ell}(x, \xi, D_x) a_{j-\ell}(t, x, \xi). \tag{6.22}$$

By convention, we set $a_{-1} = 0$. So that (6.15) reduces to Fourier inversion at $t = 0$, we set

$$a_0(0, x, \xi) = 1, \quad a_j(0, x, \xi) = 0, \text{ for } j \geq 1. \tag{6.23}$$

Then we have

$$a_0(t, x, \xi) = e^{-tL_2(x,\xi)}. \tag{6.24}$$

The solution to (6.21) is

$$a_j(t, x, \xi) = \int_0^t e^{(s-t)L_2(x,\xi)}\, \Omega_j(s, x, \xi)\, ds. \tag{6.25}$$

Now, as shown in Chap. 7, we have

$$\operatorname{Tr} e^{-tL} \sim \sum_{j \geq 0} \operatorname{Tr} \iint a_j(t, x, \xi)\, d\xi\, dx, \tag{6.26}$$

with

$$\int a_j(t, x, \xi)\, d\xi = t^{(-n+j)/2} b_j(x). \tag{6.27}$$

Furthermore, the integral (6.27) *vanishes* for j odd. Thus we have the expansion

$$K(t,x,x) \sim t^{-n/2}\Big[a_0(x) + a_1(x)t + \cdots + a_\ell(x)t^\ell + \cdots\Big], \tag{6.28}$$

with $a_j(x) = b_{2j}(x)$.

Our goal is to analyze the $\mathbb{C}l^{[2k]} \otimes \mathrm{End}\, F$ component of $a_k(x)$, with $n = 2k$. In fact, the way the local index formula (6.2) is stated, the claim is made that $a_\ell(x)$ has zero component in this space, for $\ell < k$. The next lemma gives a more precise result. Its proof will also put us in a better position to evaluate the treasured $\mathbb{C}l^{[2]} \otimes \mathrm{End}\, F$ component of $a_k(x)$. Recall the filtration (2.20) of $Cl^0(2k)$; complexification gives a similar filtration of $\mathbb{C}l(2k)$.

Lemma 6.3. *In the expansion (6.28), we have*

$$a_j(x) \in \mathbb{C}l^{(2j)}(2k) \otimes \mathit{End}\, F, \quad 0 \le j \le k. \tag{6.29}$$

In order to prove this, we examine the expression for $L = D_F^2$, in local coordinates, with respect to convenient local trivializations of $S(\tilde{P})$ and F. Fix $x_0 \in M$. Use geodesic normal coordinates centered at x_0; in these coordinates, $x_0 = 0$. Let $\{e_\alpha\}$ denote an orthonormal frame of tangent vectors, obtained by parallel translation along geodesics from x_0 of an orthonormal basis of $T_{x_0}M$; let $\{v_\alpha\}$ denote the dual frame. The frame $\{e_\alpha\}$ gives rise to a local trivialization of the spinor bundle $S(\tilde{P})$. Finally, choose an orthonormal frame $\{\varphi_\mu\}$ of F, obtained by parallel translation along geodesics from x_0 of an orthonormal basis of F_{x_0}. The connection coefficients for the Levi–Civita connection will be denoted as $\Gamma^k{}_{\ell j}$ for the coordinate frame, $\Gamma^\alpha{}_{\beta j}$ for the frame $\{e_\alpha\}$; both sets of connection coefficients vanish at 0, their first derivatives at 0 being given in terms of the Riemann curvature tensor. Similarly, denote by $\theta_j = \big(\theta^\mu{}_{\nu j}\big)$ the connection coefficients for F, with respect to the frame $\{\varphi_\mu\}$. Denote by $\Phi_{\alpha\beta}$ the curvature of F, with respect to the frame $\{e_\alpha\}$.

With respect to these choices, we write down a local coordinate expression for D_F^2, using the Weitzenbock formula

$$D_F^2 = \nabla^*\nabla + \frac{1}{4}S - \frac{1}{2}v_\alpha v_\beta \Phi_{\alpha\beta},$$

together with the identity $\nabla^*\nabla = -\gamma \circ \nabla^2$, proved in Proposition 2.1 of Appendix C. We obtain

$$\begin{aligned} D_F^2 = {} & -g^{j\ell}\Big(\partial_j + \frac{1}{4}\Gamma^\beta{}_{\alpha j}\, v_\alpha v_\beta + \theta_j\Big)\Big(\partial_\ell + \frac{1}{4}\Gamma^\delta{}_{\gamma\ell}\, v_\gamma v_\delta + \theta_\ell\Big) \\ & + g^{j\ell}\Gamma^i{}_{\ell j}\Big(\partial_i + \frac{1}{4}\Gamma^\beta{}_{\alpha i} v_\alpha v_\beta + \theta_i\Big) + \frac{1}{4}S - \frac{1}{2}\Phi_{\alpha\beta} v_\alpha v_\beta. \end{aligned} \tag{6.30}$$

This has scalar second-order part. The coefficients of ∂_j are products of elements of $Cl^{(2)}(2k)$ with connection coefficients, which vanish at 0. Terms involving no derivatives include products of elements of $Cl^{(2)}(2k)$ with curvatures, which may not vanish, and products of elements of $Cl^{(4)}(2k)$ with coefficients that vanish to *second* order at 0.

Hence we can say the following about the operators $B_{2-\ell}(x, \xi, D_x)$, which arise in (6.18) and which enter into the recursive formulas for $a_j(t, x, \xi)$. First, $B_0(x, \xi, D_x)$, a differential operator of order 2 that is homogeneous of degree 0 in ξ (thus actually independent of ξ), can be written as

$$B_0(x, \xi, D_x) = \sum_{|\alpha| \le 2} B_{0\alpha}(x, \xi) D_x^\alpha,$$

where $B_{00}(x, \xi)$ has coefficients in $Cl^{(2)}(2k)$, and also coefficients that are $O(|x|^2)$ in $Cl^{(4)}(2k)$; $B_{0\alpha}(x, \xi)$ for $|\alpha| = 1$ has some coefficients that are $O(|x|)$ in $Cl^{(2)}(2k)$. Each $B_{0\alpha}(x, \xi)$ for $|\alpha| = 2$ is scalar. Note that $B_0(x, \xi, D_x)$ acts on $a_{j-2}(t, x, \xi)$ in the recursive formula (6.21)–(6.22) for $a_j(t, x, \xi)$.

The operator $B_1(x, \xi, D_x)$, a differential operator of order 1 that is homogeneous of degree 1 in ξ, can be written as

$$B_1(x, \xi, D_x) = \sum_{|\alpha| \le 1} B_{1\alpha}(x, \xi) D_x^\alpha,$$

and among the coefficients are terms that are $O(|x|)$ in $Cl^{(2)}(2k)$. The operator $B_1(x, \xi, D_x)$ acts on $a_{j-1}(t, x, \xi)$ in (6.21)–(6.22).

We see that while the coefficients in $Cl^{[\ell]}(2k)$ in $a_j(t, x, \xi)$ give rise to coefficients in $Cl^{[\ell+2]}(2k)$ in $a_{j+1}(t, x, \xi)$ and in $Cl^{[\ell+4]}(2k)$ in $a_{j+2}(t, x, \xi)$, the degree of vanishing described above leads exactly to the sort of increase in "Clifford order" stated in Lemma 6.3, which is consequently proved.

The proof of Lemma 6.3 gives more. Namely, the $Cl^{[2j]}$-components of $a_j(x_0)$, for $0 \le j \le k$, are *unchanged* if we replace D_F^2 by the following:

$$\tilde{L} = -\sum_{j=1}^{n} \left(\frac{\partial}{\partial x_j} - \frac{1}{8} \Omega_{j\ell} x_\ell \right)^2 - \frac{1}{2} \Phi_{\alpha\beta} v_\alpha v_\beta, \tag{6.31}$$

where $\Omega_{j\ell}$ denotes the Riemann curvature tensor, acting on sections of $S(\tilde{P})$ as

$$\Omega_{j\ell} = R_{j\ell\alpha\beta}\, v_\alpha v_\beta. \tag{6.32}$$

In (6.31), summation over ℓ is understood. At this point, we can exploit a key observation of Getzler—that the Schwartz kernel $\tilde{K}(t, x, y)$ of $e^{-t\tilde{L}}$ can be evaluated in closed form at $y = 0$—by exploiting the similarity of (6.31) with the harmonic oscillator Hamiltonian, whose exponential is given by Mehler's formula, provided we modify $\tilde{L}$ in the following fashion.

Namely, for the purpose of picking out the $Cl^{[2j]}$-components of $a_j(x_0)$, we might as well let $\tilde{L}$ act on sections of $Cl(2k) \otimes F$ rather than $S(\tilde{P}) \otimes F$, and then we use the linear isomorphism $Cl(2k) \approx \Lambda^*\mathbb{R}^n$, and let the products involving v_α and v_β in (6.31) and (6.32) be *wedge* products, which, after all, for the purpose of our calculation are the principal parts of the Clifford products.

We can then separate $\tilde{L}$ into two commuting parts. Let $\tilde{L}_0$ denote the sum over j in (6.31), and let $\tilde{K}_0(t,x,y)$ be the Schwartz kernel of $e^{-t\tilde{L}_0}$. We can evaluate $\tilde{K}_0(t,x,0)$ using Mehler's formula, established in §6 of Chap. 8 (see particularly Exercises 6 and 7 at the end of that section), which implies that whenever $(\Omega_{j\ell})$ is an antisymmetric matrix of imaginary numbers (hence a self-adjoint matrix), then

$$\tilde{K}_0(t,x,0) = (4\pi t)^{-n/2} \det\Big(\frac{\Omega t/4}{\sinh(\Omega t/4)}\Big)^{1/2} e^{-(f(\Omega t/4)x,x)/4t}, \tag{6.33}$$

where $f(s) = 2s \coth 2s$. Now it is straightforward to verify that this formula is also valid whenever Ω is a nilpotent element of any commutative ring (assumed to be an algebra over $\mathbb{C}$), as in the case (6.32), where Ω is an $\mathrm{End}(T_{x_0}M)$-valued 2-form. Evaluating (6.33) at $x = 0$ gives

$$\tilde{K}_0(t,0,0) = (4\pi t)^{-n/2} \det\Big(\frac{\Omega t/4}{\sinh(\Omega t/4)}\Big)^{1/2}. \tag{6.34}$$

When Ω is the curvature 2-form of M, with its Riemannian metric, this is to be interpreted in the same way as the characteristic classes discussed in §7 of Appendix C. The $\hat{A}$-genus of M is defined to be this determinant, at $t = 1/2\pi i$:

$$\hat{A}(M) = \det\Big(\frac{\Omega/8\pi i}{\sinh(\Omega/8\pi i)}\Big)^{1/2}. \tag{6.35}$$

The $Cl^{[2k]}$-component of the t^0-coefficient in the expansion of $e^{-t\tilde{L}}$ is $(-2i)^{-k}$ times the $2k$-form part of the product of (6.35) with $\mathrm{Tr}\, e^{-\Phi/2\pi i}$, where Φ is the End F-valued curvature 2-form of the connection on F. This is also a characteristic class; we have the Chern character:

$$\mathrm{Ch}(F) = \mathrm{Tr}\, e^{-\Phi/2\pi i}. \tag{6.36}$$

This completes the proof of Theorem 6.1.

Exercises

1. Write out the first few terms in the expansion of the formula (6.35) for $\hat{A}(M)$, such as forms of degree 0, 4, 8.
2. If M is a compact, oriented, four-dimensional manifold, show that

(6.37)
$$\langle \hat{A}(M), [M] \rangle = -\frac{1}{24} \int_M p_1(TM),$$

where p_1 is the first Pontrjagin class, defined in §7 of Appendix C.

3. If $M = \mathbb{C}P^2$, show that $\langle \hat{A}(M), [M] \rangle = -1/8$. Deduce that $\mathbb{C}P^2$ has no spin structure.
4. If M is a spin manifold with positive scalar curvature, to which Proposition 4.8 applies, show that $\langle \hat{A}(M), [M] \rangle = 0$. What can you deduce about the right side of (5.3) in such a case? Consider particularly the case where $\dim M = 4$.
5. Let $F_j \to M$ be complex vector bundles. Show that
$$\mathrm{Ch}(F_1 \oplus F_2) = \mathrm{Ch}(F_1) + \mathrm{Ch}(F_2),$$
$$\mathrm{Ch}(F_1 \otimes F_2) = \mathrm{Ch}(F_1) \wedge \mathrm{Ch}(F_2).$$
6. If $F \to M$ is a complex *line* bundle, relate $\mathrm{Ch}(F)$ to the first Chern class $c_1(F)$, defined in §7 of Appendix C.

7. The Chern–Gauss–Bonnet theorem

Here we deduce from the Atiyah–Singer formula (5.3) the generalized Gauss–Bonnet formula expressing as an integrated curvature the Euler characteristic $\chi(M)$ of a compact, oriented Riemannian manifold M, of dimension $n = 2k$. As we know from Hodge theory, $\chi(M)$ is the index of

(7.1)
$$d + d^* : \Lambda^{\mathrm{even}} M \longrightarrow \Lambda^{\mathrm{odd}} M.$$

This is an operator of Dirac type, but it is not actually a twisted Dirac operator of the form (3.34), even when M has a spin structure. Rather, a further twist in the twisting procedure is required. Until near the end of this section, we assume that M has a spin structure.

With $V = \mathbb{R}^{2k}$, we can identify $\mathbb{C}\Lambda^* V$, both as a linear space and as a Clifford module, with $\mathbb{C}l(2k)$. Recall the isomorphism (2.14):

(7.2)
$$\mu : \mathbb{C}l(2k) \longrightarrow \mathrm{End}\, S,$$

where $S = S(2k) = S_+(2k) \oplus S_-(2k)$. This can be rewritten as

$$\mathbb{C}l(2k) \approx S \otimes S'.$$

Now if $\mathbb{C}l(2k)$ acts on the left factor of this tensor product, then there is a twisted Dirac operator

(7.3)
$$\begin{pmatrix} 0 & D^-_{S'} \\ D^+_{S'} & 0 \end{pmatrix},$$

produced from the grading $S \otimes S' = (S_+ \otimes S') \oplus (S_- \otimes S')$, but this is not the operator (7.1). Rather, it is the signature operator. To produce (7.1), we use the identities $\mathbb{C}\Lambda^{\text{even}}V = \mathbb{C}l^0(2k)$ and $\mathbb{C}\Lambda^{\text{odd}}V = \mathbb{C}l^1(2k)$. Recall the isomorphism (3.15):

$$\mu : \mathbb{C}l^0(2k) \longrightarrow \text{ End } S_+ \oplus \text{ End } S_-. \tag{7.4}$$

We rewrite this as

$$\mathbb{C}l^0(2k) \approx (S_+ \otimes S'_+) \oplus (S_- \otimes S'_-). \tag{7.5}$$

Similarly, we have an isomorphism

$$\mu : \mathbb{C}l^1(2k) \longrightarrow \text{ Hom}(S_+, S_-) \oplus \text{ Hom}(S_-, S_+), \tag{7.6}$$

which we rewrite as

$$\mathbb{C}l^1(2k) \approx (S_- \otimes S'_+) \oplus (S_+ \otimes S'_-). \tag{7.7}$$

It follows from this that the operator (7.1) is a "twisted" Dirac operator of the form

$$D = \begin{pmatrix} 0 & D^-_{S'_+} \oplus D^+_{S'_-} \\ D^+_{S'_+} \oplus D^-_{S'_-} & 0 \end{pmatrix}. \tag{7.8}$$

In other words, the index $\chi(M)$ of (7.1) is a difference:

$$\text{Index } D^+_{S'_+} - \text{ Index } D^+_{S'_-},$$

since Index $D^-_{S'_-} = -\text{Index } D^+_{S'_-}$. Furthermore, this difference is respected in the local index formula, an observation that will be useful later when we remove the hypothesis that M have a spin structure.

The Atiyah–Singer formula (5.3) thus yields

$$\chi(M) = \big\langle \hat{A}(M)[\text{Ch}(S'_+) - \text{Ch}(S'_-)], [M]\big\rangle. \tag{7.9}$$

The major step from here to the Chern–Gauss–Bonnet theorem is to produce a $2k$-form on M expressing $\text{Ch}(S'_+) - \text{Ch}(S'_-)$ in purely differential geometric terms, independent of a spin structure.

If $\pi_\pm$ are the representations of $\text{Spin}(2k)$ on $S_\pm$, $d\pi_\pm$ the derived representations of $\text{spin}(2k)$, and $\tilde{\Omega}$ the $\text{spin}(2k)$-valued curvature form on $\tilde{P}$, then

$$\text{Ch}(S_\pm) = \text{ Tr } e^{-d\pi_\pm(\tilde{\Omega})/2\pi i}, \tag{7.10}$$

a sum of even-order forms formally related to the characters of $\pi_\pm$,

(7.11) $$\chi_\pm(g) = \operatorname{Tr} \pi_\pm(g), \quad g \in \operatorname{Spin}(2k).$$

Note that $\dim S_+ = \dim S_-$ implies $\chi_+(e) - \chi_-(e) = 0$. It is a fact of great significance that the difference $\chi_+(g) - \chi_-(g)$ vanishes to order k at the identity element $e \in \operatorname{Spin}(2k)$. More precisely, we have the following. Take $X \in \operatorname{spin}(2k) \approx \operatorname{so}(2k)$, identified with a real, skew-symmetric matrix, $X = (X_{ij})$; there is the exponential map $\operatorname{Exp} : \operatorname{spin}(2k) \to \operatorname{Spin}(2k)$. The key formula is given as follows:

Lemma 7.1. *For $X \in so(2k)$,*

(7.12) $$\lim_{t\to 0} t^{-k}\big[\chi_+(Exp\, tX) - \chi_-(Exp\, tX)\big] = (-i)^k\, Pf(X).$$

Here, $\operatorname{Pf} : \operatorname{so}(2k) \to \mathbb{R}$ is the Pfaffian, defined as follows. Associate to $X \in \operatorname{so}(2k)$ the 2-form

(7.13) $$\xi = \xi(X) = \frac{1}{2} \sum X_{ij} e_i \wedge e_j,$$

$e_1, \dots, e_{2k}$ denoting an oriented orthonormal basis of $\mathbb{R}^{2k}$. Then

(7.14) $$k!\, (\operatorname{Pf} X) e_1 \wedge \cdots \wedge e_{2k} = \xi \wedge \cdots \wedge \xi \quad (k \text{ factors}).$$

It follows from this definition that if $T : \mathbb{R}^{2k} \to \mathbb{R}^{2k}$ is linear, then $T^*\xi(X) = \xi(T^t X T)$, so

(7.15) $$\operatorname{Pf}(T^t X T) = (\det T) \operatorname{Pf}(X).$$

Now any $X \in \operatorname{so}(n)$ can be written as $X = T^t A T$, where $T \in \operatorname{SO}(n)$, and A is a sum of 2×2, skew-symmetric blocks, of the form

$$A_\nu = \begin{pmatrix} 0 & a_\nu \\ -a_\nu & 0 \end{pmatrix}, \quad a_\nu \in \mathbb{R}.$$

Thus $\xi(A) = a_1 e_1 \wedge e_2 + \cdots + a_k e_{2k-1} \wedge e_{2k}$, so

(7.16) $$\operatorname{Pf}(X) = \operatorname{Pf}(A) = a_1 \cdots a_k.$$

It follows that

(7.17) $$\operatorname{Pf}(X)^2 = \det X.$$

We also note that, if one uses Clifford multiplication rather than exterior multiplication, on k factors of $\xi(X)$, then the result has as its highest-order term $k!(\text{Pf } X)e_1 \cdots e_{2k}$. In other words, in terms of the map $\tau : Cl(2k) \to \mathbb{C}$ of (7.7),

$$k!(\text{Pf } X) = \tau(\xi \cdots \xi), \tag{7.18}$$

with k factors of ξ.

To prove Lemma 7.1, note that the representation $\pi = \pi_+ \oplus \pi_-$ of $\text{Spin}(2k)$ on $S = S_+ \oplus S_-$ is the restriction to $\text{Spin}(2k)$ of the representation μ of $Cl(2k)$ on S characterized by (2.11). Consequently, in view of Exercise 4 in §3,

$$\text{Tr } \pi_+(\text{Exp } tX) - \text{Tr } \pi_-(\text{Exp } tX) = \text{Str } \mu\Big(e^{t \sum X_{ij} e_i e_j / 4}\Big), \tag{7.19}$$

where Str stands for the supertrace, as in (6.6). This can be evaluated by Berezin's formula, (6.10), as $(-2i)^k$ times the coefficient of $\nu = e_1 \cdots e_{2k}$ in $e^{t \sum X_{ij} e_i e_j / 4}$. Now the lowest power of t in the power-series expansion of this quantity, which has a multiple of ν as coefficient, is the kth power; the corresponding term is

$$\frac{1}{k!} \frac{t^k}{4^k} \Big(\sum X_{ij} e_i e_j\Big)^k = \frac{t^k}{2^k} (\text{Pf } X)\nu + \cdots, \tag{7.20}$$

by (7.18). Thus, by (6.10), the leading term in the expansion in powers of t of (7.19) is $(-it)^k (\text{Pf } X)$, which proves (7.12).

We remark that the formula (7.12) plays a central role in the proof of the index formula for (twisted) Dirac operators, in the papers of Bismut [Bi] and of Berline–Vergne [BV].

In §8 of Appendix C, it is shown that the Pfaffian arises directly for the generalized Gauss–Bonnet formula for a hypersurface $M \subset \mathbb{R}^{2k+1}$ when one expresses the degree of the Gauss map $M \to S^{2k}$ as an integral of the Jacobian determinant of the Gauss map and evaluates this Jacobian determinant using the Weingarten formula and Gauss' *Theorema Egregium.*

From (7.12) it follows that

$$\text{Ch}(S'_+) - \text{Ch}(S'_-) = (2\pi)^{-k} \, \text{Pf}\,(\Omega). \tag{7.21}$$

This is defined independently of any spin structure on M. Since locally any manifold has spin structures, the local index formula of §6 provides us with the following conclusion.

Theorem 7.2. *If M is a compact, oriented Riemannian manifold of dimension $n = 2k$, then the Euler characteristic $\chi(M)$ satisfies the identity*

$$\chi(M) = (2\pi)^{-k} \int_M Pf(\Omega). \tag{7.22}$$

Proof. It remains only to note that in the formula

$$(2\pi)^{-k}\langle \hat{A}(M)\mathrm{Pf}\,(\Omega), [M]\rangle = \chi(M),$$

since the factor $\mathrm{Pf}(\Omega)$ is a pure form of degree $2k = n$, only the leading term 1 in $\hat{A}(M)$ contributes to this product.

Exercises

1. Verify that when $\dim M = 2$, the formula (7.22) coincides with the classical Gauss–Bonnet formula:

$$\int_M K\, dV = 2\pi\chi(M). \tag{7.23}$$

2. Work out "more explicitly" the formula (7.22) when $\dim M = 4$. Show that

$$\chi(M) = \frac{1}{8\pi^2}\int_M \Big(|R|^2 - 4|\mathrm{Ric}|^2 + S^2\Big)\, dV, \tag{7.24}$$

   where R is the Riemann curvature tensor, Ric the Ricci tensor, and S the scalar curvature. For some applications, see [An].
3. Evaluate (7.19); show that

$$\mathrm{Str}\,\mu\Big(e^{t\sum X_{ij}e_ie_j/4}\Big) = (-it)^k\,\det\Big(\frac{\sinh tX/2}{tX/2}\Big)^{1/2}\,\mathrm{Pf}\,X. \tag{7.25}$$

   (*Hint*: Reduce to the case where X is a sum of 2×2 blocks.)
4. Apply Theorem 6.1 to give a formula for the index of the signature operator D_H^+, using the representation (7.3) of $D_H^+ \oplus D_H^-$ as a twisted Dirac operator. Justify the formula when M has no spin structure. Show that, if M is a compact, oriented 4-manifold, then

$$\mathrm{Index}\ D_H^+ = -8\langle \hat{A}(M), [M]\rangle. \tag{7.26}$$

   (*Hint*: Take a peek in [Roe].)

8. Spin^c manifolds

Here we consider a structure that arises more frequently than a spin structure, namely a *spin^c structure*. Let M be an oriented Riemannian manifold of dimension n, $P \to M$ the principal $SO(n)$-bundle of oriented orthonormal frames. A spin^c structure on M is a principal bundle $Q \to M$ with structure group

$$\mathrm{Spin}^c(n) = \mathrm{Spin}(n)\times S^1/\{(1,1),(-1,-1)\} = G. \tag{8.1}$$

Note that $\{-1, 1\} \subset \mathrm{Spin}(n)$ is the pre-image of the identity element of $\mathrm{SO}(n)$. For this principal bundle Q, we require that there be a bundle map $\rho : Q \to P$, commuting with the natural $\mathrm{Spin}(n)$ actions on Q and P.

There is a natural injection $\mathrm{Spin}(n) \hookrightarrow \mathrm{Spin}^c(n)$, as a normal subgroup. Note that taking the quotient $R = Q/\,\mathrm{Spin}(n)$ produces a principal S^1-bundle, over which Q projects. We display the various principal bundles:

$$\begin{array}{ccc} Q & \longrightarrow & R \\ \downarrow & & \downarrow \\ P & \longrightarrow & M \end{array} \tag{8.2}$$

There is a topological obstruction to the existence of a spin^c structure on M, though it is weaker than the obstruction to the existence of a spin structure. We refer to [LM] for these topological considerations; we will give some examples of spin^c-manifolds later in this section.

The standard representation of S^1 on $\mathbb{C}$ produces a complex line bundle

$$L \longrightarrow M. \tag{8.3}$$

Suppose $n = 2k$. Recall the representation $D^+_{1/2} \oplus D^-_{1/2}$ of $\mathrm{Spin}(n)$ on $S(2k)$ from (3.19). If we take the product with the standard representation of S^1 on $\mathbb{C}$, this is trivial on the factor group appearing in (8.1), so we get a representation of $\mathrm{Spin}^c(n)$ on $S(2k)$, which we continue to denote $D^+_{1/2} \oplus D^-_{1/2}$. This representation produces a vector bundle over M, which we continue to call a spinor bundle:

$$S(Q) = S_+(Q) \oplus S_-(Q); \quad S_\pm(Q) = Q \times_{D^\pm_{1/2}} S_\pm(2k). \tag{8.4}$$

In case n is odd, we have instead the bundle of spinors constructed from the representation (3.24) of $\mathrm{Spin}(n)$, via the same sort of procedure.

As in §3, we will be able to define a Dirac operator on $C^\infty(M, S(Q))$ in terms of a connection on Q, which we now construct. The Levi–Civita connection on M defines an $\mathrm{so}(n)$-valued form θ_0 on P, which pulls back to an $\mathrm{so}(n)$-valued form θ_0 on Q. Endow the bundle $R = Q/\,\mathrm{Spin}(n) \to M$ with a connection θ_1, so $L \to M$ gets a metric connection. Then θ_1 pulls back to an $i\mathbb{R}$-valued form θ_1 on Q, and

$$\theta = \theta_0 + \theta_1 \tag{8.5}$$

defines a spin$^c(n)$-valued form on Q, which gives rise to a connection on Q. This leads to a connection on the spinor bundle $S(Q) \to M$, and the analogues of Propositions 3.4 and 3.5 hold. Thus we produce the Dirac operator

$$D = i\, m \circ \nabla : C^\infty(M, S) \longrightarrow C^\infty(M, S). \tag{8.6}$$

More generally, if $E \to M$ is a vector bundle with a metric connection, one gets a Clifford connection on $S(Q) \otimes E$ and hence a twisted Dirac operator

$$D_E : C^\infty(M, S \otimes E) \longrightarrow C^\infty(M, S \otimes E). \tag{8.7}$$

If dim M is even, D_E maps sections of $S_\pm \otimes E$ to sections of $S_\mp \otimes E$.

We consider some ways in which spin^c structures arise. First, a spin structure gives rise to a spin^c structure. Indeed, if the frame bundle $P \to M$ lifts to a principal Spin(n)-bundle $\tilde{P} \to M$, then Q can be taken to be the quotient of the product bundle $\tilde{P} \times S^1 \to M$ by the natural $\mathbb{Z}_2$-action on the fibers. The canonical flat connection on $S^1 \times M \to M$ is used, to provide $Q \to M$ with a connection, and then the Dirac operator (8.6) defined by $Q \to M$ coincides with that defined by $\tilde{P} \to M$.

Another family of examples of spin^c structures of considerable importance arises as follows. Suppose M is a manifold of dimension $n = 2k$ with an almost complex structure, $J : T_x M \to T_x M,\ J^2 = -I$. Endow M with a Riemannian metric such that J is an isometry. $\mathcal{T}M$, which is (TM, J) regarded as a *complex* vector bundle of fiber dimension k, then acquires a natural Hermitian metric, as in (2.9). The associated frame bundle $F \to M$ is a principal $U(k)$ bundle. Note that

$$\mathrm{U}(k) \approx \mathrm{SU}(k) \times S^1 / \Gamma, \tag{8.8}$$

where $\Gamma = \{(I, 1), (-I, -1)\}$. Since SU($k$) is simply connected, the inclusion $\mathrm{U}(k) \hookrightarrow \mathrm{SO}(n)$ yields a uniquely defined homomorphism

$$\mathrm{SU}(k) \longrightarrow \mathrm{Spin}(n), \tag{8.9}$$

and hence a homomorphism

$$\mathrm{U}(k) \approx \mathrm{SU}(k) \times S^1 / \Gamma \longrightarrow \mathrm{Spin}^c(n). \tag{8.10}$$

From the bundle $F \to M$, this gives rise to a principal Spin$^c(n)$ bundle $Q \to M$.

In this case, the map $\mathrm{U}(k) \to \mathrm{Spin}^c(n) \to S^1$ is given by the determinant, $\det : \mathrm{U}(k) \to S^1$. The principal S^1-bundle $R \to M$ is obtained by taking the quotient of the principal U(k)-bundle F by the action of SU(k). The associated line bundle $L \to M$ is seen to be

$$L = \Lambda^k_{\mathbb{C}} \mathcal{T}. \tag{8.11}$$

Other geometrical structures give rise to spin^c structures; we refer to [LM] for more on this. We mention the following: namely, any oriented hypersurface in a spin^c manifold inherits a natural spin^c structure. In this fashion the sphere bundle S^*M over a Riemannian manifold gets a spin^c structure, as a hypersurface of T^*M, which can be given an almost complex structure.

Though a spin^c structure is more general than a spin structure, it is a very significant fact that a spin^c structure in turn gives rise to a spin structure, in the following circumstance. Namely, suppose the principal S^1-bundle $R \to M$ lifts to a double cover

$$\tilde{R} \longrightarrow M, \tag{8.12}$$

corresponding to the natural two-to-one surjective homomorphism $\mathrm{sq} : S^1 \to S^1$. This is equivalent to the hypothesis that the line bundle $L \to M$ possess a "square root" $\lambda \to M$:

$$\lambda \otimes \lambda = L. \tag{8.13}$$

In such a case, the quotient of $Q \times \tilde{R} \to M$ by the natural action of S^1 on each factor gives a lift of Q to a principal $\mathrm{Spin}(n) \times S^1$-bundle

$$\tilde{Q} \longrightarrow M. \tag{8.14}$$

Then the quotient

$$\tilde{P} = \tilde{Q}/S^1 \longrightarrow M \tag{8.15}$$

defines a spin structure on M. The vector bundles $S(Q)$ and $S(\tilde{P})$ are related by

$$S(Q) = S(\tilde{P}) \otimes \lambda. \tag{8.16}$$

Furthermore, the connection on $S(Q)$ defined above coincides with the product connection on $S(\tilde{P}) \otimes \lambda$ arising from the natural connections on each factor. Therefore, if D_E and D_E^0 are respectively the twisted Dirac operator associated with a vector bundle $E \to M$ (given a metric connection) via the spin^c and spin structures described above, then

$$D_E = D^0_{\lambda \otimes E}. \tag{8.17}$$

This holds, we recall, provided L has a square root λ.

One consequence of this is the following extension of the Weitzenbock formula (4.24). Namely, if D_E is the twisted Dirac operator on $S(Q) \otimes E$ described there, then applying (4.24) to the right side of (8.17) gives

$$D_E^2 = \nabla^* \nabla + \mathcal{K}, \tag{8.18}$$

with

$$\mathcal{K}\varphi = \frac{1}{4} S\varphi - \frac{1}{2} \sum_{i,j} v_i v_j \omega^\lambda(e_i, e_j)\varphi - \frac{1}{2} \sum_{i,j} v_i v_j R^E(e_i, e_j)\varphi, \tag{8.19}$$

or equivalently

$$\mathcal{K}\varphi = \frac{1}{4}S\varphi - \frac{1}{4}\sum_{i,j} v_i v_j \omega^L(e_i, e_j)\varphi - \frac{1}{2}\sum_{i,j} v_i v_j R^E(e_i, e_j)\varphi, \tag{8.20}$$

where, as before, $\{e_j\}$ is a local orthonormal frame of vector fields on M, with dual frame field $\{v_j\}$. Here ω^λ is the curvature form of the line bundle λ and ω^L that of L.

Now locally there is no topological obstruction to the existence of the lift (8.12). Consequently, the identity (8.20) holds regardless of whether L possesses a global square root. Therefore, the proof of the local index formula given in §6 extends to this case. Furthermore, we have the pointwise identity of forms:

$$\mathrm{Ch}(\lambda \otimes E) = e^{c_1(\lambda)}\,\mathrm{Ch}(E), \quad c_1(\lambda) = \frac{1}{2}c_1(L), \tag{8.21}$$

where c_1 is the first Chern class, defined in §7 of Appendix C. Therefore, we have the following extension of Theorem 5.1:

Theorem 8.1. *If M is a compact Riemannian manifold of dimension $n = 2k$ with spin^c structure and $D_E : C^\infty(M, S_+ \otimes E) \to C^\infty(M, S_- \otimes E)$ is a twisted Dirac operator, then*

$$\mathit{Index}\ D_E = \left\langle e^{c_1(L)/2}\, Ch(E)\hat{A}(M), [M]\right\rangle, \tag{8.22}$$

where L is the line bundle (8.3), and $c_1(L)$ is its first Chern class.

The index formula for twisted Dirac operators on spin^c manifolds furnishes a tool with which one can evaluate the index of general elliptic pseudodifferential operators. Indeed, let P be any elliptic pseudodifferential operator (of order m),

$$P : C^\infty(M, E_0) \longrightarrow C^\infty(M, E_1), \tag{8.23}$$

$E_j \to M$ being vector bundles. Then, as seen in Chap. 7, we have the principal symbol

$$\sigma_P \in C^\infty(S^*M, \mathrm{Hom}(\tilde{E}_0, \tilde{E}_1)), \tag{8.24}$$

$\tilde{E}_j \to S^*M$ being the pull-backs of $E_j \to M$. The ellipticity of P is equivalent to σ_P being an isomorphism at each point of S^*M. Now, we can construct a new vector bundle E over $\widehat{B}M$, the double of the ball bundle B^*M, as follows. We let $\tilde{E}_j$ also denote the pull-back of E_j to B^*M, and, when the two copies of B^*M are glued together along S^*M to form $\widehat{B}M$, we also glue together $\tilde{E}_0$ and $\tilde{E}_1$, over S^*M, using the isomorphism (8.24). The construction of $E \to \widehat{B}M$ by this process is known as the "clutching construction." Now $\widehat{B}M$ can be given a

Riemannian metric, and also a spin^c structure, arising from the almost complex structure on B^*M. If E is endowed with a connection, one obtains a twisted Dirac operator D_E on $\widehat{B}M$. The following result, together with the formula for Index D_E given by Theorem 8.1, provides the general Atiyah–Singer index formula.

Theorem 8.2. *If P is an elliptic pseudodifferential operator, giving rise to a twisted Dirac operator D_E by the clutching construction described above, then*

$$\text{Index } P = \text{Index } D_E. \tag{8.25}$$

The proof of this result will not be given here; it involves use of the Bott periodicity theorem. Related approaches, computing Index P from a knowledge of the index of twisted signature operators, are discussed in [Pal] and [ABP]. A refinement of (8.25), involving an identity in K-homology is established in [BDT].

Exercises

1. Consider the following zero-order pseudodifferential operator on $L^2(S^1)$:

$$Q = M_f P + M_g(I - P),$$

where P is the projection

$$P\left(\sum_{-\infty}^{\infty} c_n e^{in\theta}\right) = \sum_{0}^{\infty} c_n e^{in\theta}.$$

We assume f and g are smooth, complex-valued functions; $M_f u = fu$. If f and g are nowhere vanishing on S^1, Q is elliptic. A formula for its index is produced in Exercises 1–5 of Chap. 4, §3.

Construct the associated twisted Dirac operator D_E, acting on sections of a vector bundle over the manifold $\widehat{B}S^1 \approx \mathbb{T}^2$. Evaluate the index of D_E using Theorem 8.1, and verify the identity (8.25) in this case.

9. The Riemann–Roch theorem

In this section we will show how the index formula (8.22) implies the classical Riemann–Roch formula on compact Riemann surfaces, and we also discuss some of the implications of that formula. For implications of generalizations of the Riemann–Roch formula to higher-dimensional, compact, complex manifolds, which also follows from (8.22), see [Har] and [Hir].

Let M be a compact two-dimensional manifold, with a complex structure, defined by $J : T_xM \to T_xM,\ J^2 = -I$. As shown in Chap. 5, §10, this a priori "almost complex" structure automatically gives rise to holomorphic charts on M in this dimension. We can put a Riemannian metric and an orientation on M such that J is an isometry on each tangent space, counterclockwise rotation by 90°. Then TM gets the structure of a complex line bundle, which we denote $\mathcal{T}M$, with a Hermitian metric. We have the dual line bundle $\mathcal{T}'M$. Note that the Hermitian metric on $\mathcal{T}M$ yields a Hermitian metric on $\mathcal{T}'M$ and also produces a conjugate linear bundle isomorphism of $\mathcal{T}M$ with $\mathcal{T}'M$. We also define the complex line bundle $\overline{\mathcal{T}}M$ to be the tangent bundle TM with complex structure given by $-J$ and $\overline{\mathcal{T}}'M$ to be its dual.

A function $u \in C^\infty(M)$ is holomorphic if $\partial u/\partial \bar{z} = 0$ in any local holomorphic coordinate system and is antiholomorphic if $\partial u/\partial z = 0$. We denote the space of holomorphic functions on an open set $U \subset M$ by $\mathcal{O}_U$, and antiholomorphic functions by $\overline{\mathcal{O}}_U$. There are invariantly defined operators

$$\partial : C^\infty(M) \longrightarrow C^\infty(M, \mathcal{T}'), \quad \bar{\partial} : C^\infty(M) \longrightarrow C^\infty(M, \overline{\mathcal{T}}'), \tag{9.1}$$

given as follows. If X is a real vector field, namely, a section of TM, set

$$\partial_X u = \frac{1}{2}\big(Xu - i(JX)u\big), \quad \bar{\partial}_X u = \frac{1}{2}\big(Xu + i(JX)u\big). \tag{9.2}$$

Note that

$$\partial_{JX} u = i\,\partial_X u, \quad \bar{\partial}_{JX} u = -i\,\bar{\partial}_X u, \tag{9.3}$$

which justifies (9.1).

In addition to holomorphic functions, we also have the notion of a holomorphic line bundle over M. Given a complex line bundle $L \to M$, let $\{U_j\}$ be a covering of M by geodesically convex sets. A holomorphic structure on L is a choice of nowhere-vanishing sections s_j of L over U_j such that $s_j = \sigma_{jk} s_k$ on $U_{jk} = U_j \cap U_k$, with σ_{jk} holomorphic complex-valued functions. Similarly, a choice of nowhere-vanishing sections t_j of L over U_j such that $t_j = \tau_{jk} t_k$ on U_{jk}, τ_{jk} antiholomorphic, gives L the structure of an antiholomorphic line bundle.

The bundle $\mathcal{T}M$ has a natural structure of a holomorphic line bundle; in a local holomorphic coordinate system $\{U_j\}$, let $s_j = \partial/\partial x$. $\mathcal{T}'$ is a holomorphic line bundle with $s_j = dx$. To see this, note that if $\psi : U \to V$ is a holomorphic map relating two local coordinate charts on M, $\psi = u + iv$, then $(D\psi)(\partial/\partial x)$ is equal to

$$\frac{\partial u}{\partial x}\frac{\partial}{\partial x} + \frac{\partial v}{\partial x}\frac{\partial}{\partial y} = \frac{\partial u}{\partial x}\frac{\partial}{\partial x} + \frac{\partial v}{\partial x}\, J\, \frac{\partial}{\partial x} = \Big(\frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}\Big)\frac{\partial}{\partial x} = \frac{\partial \psi}{\partial x}\,\frac{\partial}{\partial x}.$$

Here, the first two quantities are regarded as local sections of TM, the last two as local sections of $\mathcal{T}M$. Similarly, $\overline{\mathcal{T}}$ and $\overline{\mathcal{T}}'$ have natural structures as antiholomorphic line bundles, using the same choices of local sections as above.

It is also common to identify $\mathcal{T}M$ and $\overline{\mathcal{T}}M$ with complementary subbundles of the complexified tangent bundle $T_{\mathbb{C}}M = \mathbb{C} \otimes TM$, a complex vector bundle whose fibers are two-dimensional complex vector spaces. Namely, the local section $\partial/\partial x$ of TM is identified with $(1/2)(\partial/\partial x - i\partial/\partial y) = \partial/\partial z$ to yield $\mathcal{T}M \hookrightarrow T_{\mathbb{C}}M$ and it is identified with $(1/2)(\partial/\partial x + i\partial/\partial y) = \partial/\partial \overline{z}$ to yield $\overline{\mathcal{T}}M \hookrightarrow T_{\mathbb{C}}M$. More generally, these two maps are given respectively by $X \mapsto (1/2)(X - iJX)$ and $X \mapsto (1/2)(X + iJX)$. Identifying $\mathcal{T}M$ and $\overline{\mathcal{T}}M$ with their images in $T_{\mathbb{C}}M$, we have

$$T_{\mathbb{C}}M = \mathcal{T}M \oplus \overline{\mathcal{T}}M.$$

Similarly, we have the complexified cotangent bundle $T^*_{\mathbb{C}}M = \mathbb{C} \otimes T^*M$, and natural injections $\mathcal{T}'M \hookrightarrow T^*_{\mathbb{C}}M$, $\overline{\mathcal{T}}'M \hookrightarrow T^*_{\mathbb{C}}M$, so that

$$T^*_{\mathbb{C}}M = \mathcal{T}'M \oplus \overline{\mathcal{T}}'M.$$

In this case, dx is mapped respectively to $(dx + i\,dy)/2 = dz/2$ and to $(dx - i\,dy)/2 = d\overline{z}/2$.

We use the following common notation for these line bundles equipped with these extra structures:

$$\mathcal{T} = \kappa^{-1}, \quad \mathcal{T}' = \kappa, \quad \overline{\mathcal{T}} = \overline{\kappa}^{-1}, \quad \overline{\mathcal{T}}' = \overline{\kappa}. \tag{9.4}$$

We can rewrite (9.1) as

$$\partial : C^\infty(M) \longrightarrow C^\infty(M, \kappa), \quad \overline{\partial} : C^\infty(M) \longrightarrow C^\infty(M, \overline{\kappa}).$$

We note that κ^{-1} and $\overline{\kappa}$ are isomorphic as C^∞-line bundles; κ is called the *canonical bundle*.

More generally, if $L \to M$ is any holomorphic line bundle, we have a naturally defined operator

$$\overline{\partial} : C^\infty(M, L) \longrightarrow C^\infty(M, L \otimes \overline{\kappa}), \tag{9.5}$$

defined as follows. Pick any local (nowhere-vanishing) holomorphic section S of L, for example, $S = s_j$ on U_j, used in the definition above of holomorphic structure. Then an arbitrary section u is of the form $u = vS$, v complex-valued, and we set

$$\overline{\partial} u = \frac{\partial v}{\partial \overline{z}}\, S \otimes d\overline{z}. \tag{9.6}$$

It is easy to see that this is independent of the choice of holomorphic section S or of local holomorphic coordinate system. Sometimes, to emphasize the dependence of (9.5) on L, we denote this operator by $\overline{\partial}_L$. The operator (9.5) is a first-order, elliptic differential operator, and the Riemann–Roch formula is a formula for its index.

The kernel of $\overline{\partial}_L$ in (9.5) consists of holomorphic sections of L; namely, sections u such that, with respect to the defining sections s_j on U_j, $u = v_j s_j$ with v_j holomorphic. We denote this space of holomorphic sections by

$$\mathcal{O}(L) = \ker \overline{\partial}_L. \tag{9.7}$$

The significance of the Riemann–Roch formula lies largely in its use as a tool for understanding as much as possible about the spaces (9.7).

The cokernel of $\overline{\partial}_L$ in (9.5) can be interpreted as follows. The Hermitian metric on $\mathcal{T}$ gives rise to a trivialization of $\kappa \otimes \overline{\kappa}$ and to a duality of $L^2(M, L \otimes \overline{\kappa})$ with $L^2(M, L^{-1} \otimes \kappa)$. With respect to this duality, the adjoint of $\overline{\partial}_L$ is

$$-\overline{\partial} : C^\infty(M, L^{-1} \otimes \kappa) \longrightarrow C^\infty(M, L^{-1} \otimes \kappa \otimes \overline{\kappa}). \tag{9.8}$$

Consequently,

$$\text{Index}\, \overline{\partial}_L = \dim \mathcal{O}(L) - \dim \mathcal{O}(L^{-1} \otimes \kappa). \tag{9.9}$$

The Riemann–Roch theorem will produce a formula for (9.9) in terms of topological information, specifically, in terms of $c_1(L)$ and $c_1(\kappa)$.

Recall that M has a natural spin^c structure, arising from its complex structure. We will produce a twisted Dirac operator on M whose index is the same as that of $\overline{\partial}_L$. In fact, when the construction of the spinor bundle made in §8 is specialized to the case at hand, we get

$$S_+ = 1, \quad S_- = \mathcal{T} \approx \overline{\kappa}, \tag{9.10}$$

where 1 denotes the trivial line bundle over M. Furthermore, the line bundle denoted as L in (8.11) is $\mathcal{T} \approx \kappa^{-1}$. If L is a (holomorphic) line bundle over M, we give L a Hermitian metric and metric connection ∇. Then the twisted Dirac operator

$$D_L : C^\infty(M, L) \longrightarrow C^\infty(M, L \otimes \overline{\kappa}) \tag{9.11}$$

is given by

$$\langle D_L u, X \rangle = \frac{1}{2}\big(\nabla_X u + i \nabla_{JX} u\big), \tag{9.12}$$

for X a section of TM, identified with $\overline{\mathcal{T}M} \approx \overline{\kappa}'$, noting that

$$\langle D_L u, JX \rangle = -\langle D_L u, X \rangle. \tag{9.13}$$

It is easy to see that $\overline{\partial}_L$ and D_L are differential operators with the same principal symbol. Disregarding the question of whether one can pick a connection on L making these operators equal, we clearly have

$$\text{Index}\, \overline{\partial}_L = \text{Index}\, D_L. \tag{9.14}$$

Now applying the index formula (8.22) to the right side of (9.14) gives

$$\text{Index } D_L = \left\langle e^{-c_1(\kappa)/2}\,\mathrm{Ch}(L)\hat{A}(M), [M]\right\rangle. \tag{9.15}$$

Since $\hat{A}(M)$ is 1 plus a formal sum of forms of degree $4, 8, \dots$, we obtain

$$\text{Index } D_L = c_1(L)[M] - \frac{1}{2}c_1(\kappa)[M]. \tag{9.16}$$

Putting together (9.9), (9.14), and (9.16) gives the Riemann–Roch formula:

Theorem 9.1. *If L is a holomorphic line bundle over a compact Riemann surface M, with canonical bundle κ, then*

$$\dim \mathcal{O}(L) - \dim \mathcal{O}(L^{-1}\otimes\kappa) = c_1(L)[M] - \frac{1}{2}c_1(\kappa)[M]. \tag{9.17}$$

According to the characterization of the Chern classes given in §7 of Appendix C, if L has a connection with curvature 2-form ω_L, then

$$c_1(L)[M] = -\frac{1}{2\pi i}\int_M \omega_L. \tag{9.18}$$

In particular, $c_1(\kappa)[M]$ is given by the Gauss–Bonnet formula:

$$c_1(\kappa)[M] = -\chi(M) = 2g-2, \tag{9.19}$$

where $\chi(M)$ is the Euler characteristic and g is the genus of M.

We begin to draw some conclusions from the Riemann–Roch formula (9.17). First, for the trivial line bundle 1 we clearly have

$$\dim \mathcal{O}(1) = 1, \tag{9.20}$$

assuming M is connected, since holomorphic functions on M must be constant. If we apply (9.17) to $L=\kappa$, using $\kappa^{-1}\otimes\kappa = 1$ and the formula (9.19), we obtain

$$\dim \mathcal{O}(\kappa) = g. \tag{9.21}$$

The space $\mathcal{O}(\kappa)$ is called the space of holomorphic 1-forms, or "Abelian differentials." We claim there is a decomposition

$$\mathcal{H}_1(M) = \mathcal{O}(\kappa)\oplus\overline{\mathcal{O}}(\overline{\kappa}), \tag{9.22}$$

of the space $\mathcal{H}_1(M)$ of (complex) harmonic 1-forms on M into a direct sum of $\mathcal{O}(\kappa)$ and the space $\overline{\mathcal{O}}(\overline{\kappa})$ of antiholomorphic sections of $\overline{\kappa}$. In fact, the Hodge star operator $* : \Lambda^1 M \to \Lambda^1 M$, extended to be $\mathbb{C}$-linear on $\mathbb{C} \otimes \Lambda^1 M$, acts on $\mathcal{H}_1(M)$, with $** = -1$, and $\mathcal{O}(\kappa)$ and $\overline{\mathcal{O}}(\overline{\kappa})$ are easily seen to be the i and $-i$ eigenspaces of $*$ in $\mathcal{H}_1(M)$. Furthermore, there is a conjugate linear isomorphism

$$C : \mathcal{O}(\kappa) \longrightarrow \overline{\mathcal{O}}(\overline{\kappa}) \tag{9.23}$$

given in local holomorphic coordinates by

$$C\big(u(z)\,dz\big) = \overline{u(z)}\,d\overline{z}.$$

Now (9.22) and (9.23) imply

$$\dim \mathcal{O}(\kappa) = \frac{1}{2}\dim \mathcal{H}_1(M) = \frac{1}{2}\dim \mathcal{H}^1(M,\mathbb{C}), \tag{9.24}$$

where $\mathcal{H}^1(M,\mathbb{C})$ is a deRham cohomology group, and the last identity is by Hodge theory. Granted that $\dim \mathcal{H}^1(M,\mathbb{C}) = 2g$, this gives an alternative derivation of (9.21), not using the Riemann–Roch theorem.

The Hodge theory used to get the last identity in (9.24) is contained in Proposition 8.3 of Chap. 5. Actually, in §8 of Chap. 5, $\mathcal{H}_1$ denoted the space of real harmonic 1-forms, which was shown to be isomorphic to the real deRham cohomology group $\mathcal{H}^1(M,\mathbb{R})$, which in turn was denoted $\mathcal{H}^1(M)$ there.

Just for fun, we note the following. Suppose that instead of (9.17) one had in hand the weaker result

$$\dim \mathcal{O}(L) - \dim \mathcal{O}(L^{-1}\otimes\kappa) = Ac_1(L)[M] + Bc_1(\kappa)[M], \tag{9.25}$$

with constants A and B that had not been calculated. Then using the results (9.19) and (9.21), one can determine A and B. Indeed, substituting $L = 1$ into (9.25) gives $1 - g = B(2g-2)$, while substituting $L = \kappa$ in (9.25) gives $g - 1 = (A+B)(2g-2)$. As long as $g \neq 1$, this forces $A = 1,\ B = -1/2$. The $g = 1$ case would also follow if one knew that (9.25) held with constants independent of M.

Before continuing to develop implications of the Riemann–Roch formula, we note that, in addition to $\mathcal{O}(L)$, it is also of interest to study $\mathcal{M}(L)$, the space of meromorphic sections of a holomorphic line bundle. The following is a fundamental existence result.

Proposition 9.2. *If $L \to M$ is a holomorphic line bundle, there exist nontrivial elements of $\mathcal{M}(L)$.*

Proof. The operator (9.5) extends to

$$\overline{\partial} : H^{s+1}(M,L) \longrightarrow H^s(M, L\otimes\overline{\kappa}), \tag{9.26}$$

which is Fredholm. There are elements $v_1, \dots, v_K \in C^\infty(M, L^{-1} \otimes \kappa)$ such that, for all $s \in \mathbb{R}$, if $f \in H^s(M, L \otimes \overline{\kappa})$ and $\langle f, v_j \rangle = 0$ for $j = 1, \dots, K$, then there exists $u \in H^{s+1}(M, L)$ such that $\overline{\partial} u = f$. Now, for $s < -1$, there is a finite linear combination of "delta functions," in $H^s(M, L \otimes \overline{\kappa})$, orthogonal to these v_j. Denote such an f by $f = \sum a_j \delta_{p_j}$. Then let $u \in H^{s+1}(M, L)$ satisfy $\overline{\partial} u = f$. In particular, $\overline{\partial} u = 0$ on the complement of a finite set of points. Near each $p \in \operatorname{supp} f$, u looks like the Cauchy kernel, so u is a nontrivial meromorphic section of L.

Such an existence result need not hold for $\mathcal{O}(L)$; in Corollary 9.4 we will see a condition that guarantees $\mathcal{O}(L) = 0$. Such a result should not be regarded in a negative light; indeed knowing that $\mathcal{O}(L) = 0$ for some line bundles can give important information on $\mathcal{O}(L_1)$ for certain other line bundles, as we will see.

Any nontrivial $u \in \mathcal{M}(L)$ will have a finite number of zeros and poles. If p is a zero of u, let $\nu_u(p)$ be the order of the zero; if p is a pole of u, let $-\nu_u(p)$ be the order of the pole. We define the "divisor" of $u \in \mathcal{M}(L)$ to be the formal finite sum

$$\vartheta(u) = \sum_p \nu_u(p) \cdot p \tag{9.27}$$

over the set of zeros and poles of u. It is a simple exercise in complex analysis that if u is a nontrivial meromorphic function on M (i.e., an element of $\mathcal{M}(1)$), then $\sum_p \nu_u(p) = 0$. The following is a significant generalization of that.

Proposition 9.3. *If $L \to M$ is a holomorphic line bundle and $u \in \mathcal{M}(L)$ is nontrivial, then*

$$c_1(L)[M] = \sum_p \nu_u(p). \tag{9.28}$$

Proof. The left side of (9.28) is given by (9.18), where ω_L is the curvature 2-form associated to any connection on L. We will use the formula

$$-\frac{1}{2\pi i} \int_M \omega_L = \text{Index } X, \tag{9.29}$$

for any $X \in C^\infty(M, L)$ with nondegenerate zeros, proved in Appendix C, Proposition 5.4, as a variant of the Gauss–Bonnet theorem. The section X will be constructed from $u \in \mathcal{M}(L)$ as follows. Except on the union of small neighborhoods of the poles of u, we take $X = u$. Near the poles of u, write $u = vS$, S a nonvanishing holomorphic section of L defined on a neighborhood of such poles, v meromorphic. Pick $R > 0$ sufficiently large, and replace u by $(R^2/\overline{v})S$, where $|v| \geq R$. Smooth out X near the loci $|v| = R$. Then the formula (9.29) for X is equivalent to the desired formula, (9.28).

The following is an immediate consequence.

Corollary 9.4. *If $L \to M$ is a holomorphic line bundle with $c_1(L)[M] < 0$, then every nontrivial $u \in \mathcal{M}(L)$ has poles; hence $\mathcal{O}(L) = 0$.*

Note that if $c_1(L)[M] = 0$ and $\mathcal{O}(L) \neq 0$, by (9.28) we have that any $u \in \mathcal{O}(L)$ not identically zero is nowhere vanishing. Thus we have

$$c_1(L)[M] = 0,\ \mathcal{O}(L) \neq 0 \Longrightarrow L \text{ is trivial holomorphic line bundle.} \tag{9.30}$$

To relate Corollary 10.4 to the Riemann–Roch formula (9.17), we note that since $\dim \mathcal{O}(L^{-1} \otimes \kappa) \geq 0$, (9.17) yields Riemann's inequality:

$$\dim \mathcal{O}(L) \geq c_1(L)[M] - g + 1. \tag{9.31}$$

In view of the identities

$$\begin{aligned} c_1(L_1 \otimes L_2)[M] &= c_1(L_1)[M] + c_1(L_2)[M], \\ c_1(L^{-1})[M] &= -c_1(L)[M], \end{aligned} \tag{9.32}$$

we see that

$$c_1(L)[M] > 2g - 2 \Longrightarrow \mathcal{O}(L^{-1} \otimes \kappa) = 0. \tag{9.33}$$

Thus we have the following sharpening of Riemann's inequality:

Proposition 9.5. *If M has genus g and $c_1(L)[M] > 2g - 2$, then*

$$\mathit{dim}\ \mathcal{O}(L) = c_1(L)[M] - g + 1. \tag{9.34}$$

Generalizing (9.27), we say a divisor on M is a finite formal sum

$$\vartheta = \sum_p \nu(p) \cdot p, \tag{9.35}$$

$\nu(p)$ taking values in $\mathbb{Z}$. One defines $-\vartheta$ and the sum of two divisors in the obvious fashion. To any divisor ϑ we can associate a holomorphic line bundle, denoted E_ϑ; one calls E_ϑ a divisor bundle. To construct E_ϑ, it is most convenient to use the method of transition functions. Cover M with holomorphic coordinate sets U_j, pick $\psi_j \in \mathcal{M}_{U_j}$, having a pole of order exactly $|\nu(p)|$ at p, if $\nu(p) < 0$, a zero of order exactly $\nu(p)$ if $\nu(p) > 0$ (provided $p \in U_j$), and no other poles or zeros. The transition functions

$$\varphi_{jk} = \psi_k^{-1} \psi_j \tag{9.36}$$

define a holomorphic line bundle E_ϑ. The collection $\{\psi_j, U_j\}$ defines a meromorphic section

$$\psi \in \mathcal{M}(E_\vartheta) \tag{9.37}$$

and

$$-\vartheta(\psi) = \vartheta. \tag{9.38}$$

Thus Proposition 9.3 implies

$$c_1(E_\vartheta) = -\sum_p \nu(p) = \langle\vartheta\rangle, \tag{9.39}$$

where the last identity defines $\langle\vartheta\rangle$.

Divisor bundles help one study meromorphic sections of one line bundle in terms of holomorphic sections of another. A basic question in Riemann surface theory is when can one construct a meromorphic function on M (more generally, a meromorphic section of L) with prescribed poles and zeros. A closely related question is the following. Given a divisor ϑ on M, describe the space

$$\mathcal{M}(L,\vartheta) = \{u \in \mathcal{M}(L) : \vartheta(u) \geq \vartheta\}, \tag{9.40}$$

where $\vartheta_1 \geq \vartheta$ means $\vartheta_1 - \vartheta \geq 0$, that is, all integers $\mu(p)$ in $\vartheta_1 - \vartheta = \sum \mu(p)\cdot p$ are ≥ 0. When $L = 1$, we simply write $\mathcal{M}(\vartheta)$ for the space (9.40). A straightforward consequence of the construction of E_ϑ is the following:

Proposition 9.6. *There is a natural isomorphism*

$$\mathcal{M}(L,\vartheta) \approx \mathcal{O}(L \otimes E_\vartheta). \tag{9.41}$$

Proof. The isomorphism takes $u \in \mathcal{M}(L,\vartheta)$ to $u\psi$, where ψ is described by (9.36)–(9.37).

We can hence draw some conclusions about the dimension of $\mathcal{M}(L,\vartheta)$. From the identity (9.34) we have

$$c_1(L)[M] + \langle\vartheta\rangle > 2g-2 \Longrightarrow \dim \mathcal{M}(L,\vartheta) = c_1(L)[M] + \langle\vartheta\rangle - g + 1, \tag{9.42}$$

and, in particular,

$$\langle\vartheta\rangle > 2g-2 \Longrightarrow \dim \mathcal{M}(\vartheta) = \langle\vartheta\rangle - g + 1. \tag{9.43}$$

Also one has general inequalities, as a consequence of (9.31).

Now Corollary 9.4 and Proposition 9.5 specify precisely $\dim \mathcal{O}(L)$ provided either $c_1(L)[M] < 0$ or $c_1(L)[M] > 2g-2$, but (9.31) gives weaker information if $0 \le c_1(L)[M] \le 2g-2$; in fact, for $c_1(L)[M] \le g-1$, it gives no information at all. In this range the lower bound (9.31) can be complemented by an upper bound. For example, (9.30) implies

$$c_1(L)[M] = 0 \Longrightarrow \dim \mathcal{O}(L) = 0 \text{ or } 1. \tag{9.44}$$

We will show later that both possibilities can occur. We now establish the following generalization of (9.44).

Proposition 9.7. *Let $k = 0, 1, \dots, g-1$. Then, for a holomorphic line bundle $L \to M$,*

$$c_1(L)[M] = g - 1 - k \Longrightarrow 0 \le \ dim\, \mathcal{O}(L) \le g - k \tag{9.45}$$

and

$$c_1(L)[M] = g - 1 + k \Longrightarrow k \le \ dim\, \mathcal{O}(L) \le g. \tag{9.46}$$

Proof. First we establish (9.46). The lower estimate follows from (9.31). For the upper estimate, pick any divisor $\vartheta \le 0$ with $\langle \vartheta \rangle = k$. Then $\dim \mathcal{O}(L) \le \dim \mathcal{M}(L, \vartheta) = \dim \mathcal{O}(L \otimes E_\vartheta)$, which is equal to g since $c_1(L \otimes E_\vartheta)[M] = 2g-1$ and Proposition 9.5 applies. The upper estimate in (9.45) follows by interchanging L and $L^{-1} \otimes \kappa$ in the Riemann–Roch identity.

To illustrate (9.46), we note the following complement to (9.44):

$$c_1(L)[M] = 2g - 2 \Longrightarrow \dim \mathcal{O}(L) = g - 1 \text{ or } g. \tag{9.47}$$

On the other hand, the closer $c_1(L)[M]$ gets to $g-1$, the greater the uncertainty in $\dim \mathcal{O}(L)$, except of course when $g = 0$; then Corollary 9.4 and Proposition 9.5 cover all possibilities. It turns out that, for "typical" L, the *minimum* value of $\dim \mathcal{O}(L)$ in (9.45)–(9.46) is achieved; see [Gu].

We now use some of the results derived above to obtain strong results on the structure of compact Riemann surfaces of genus $g = 0$ and 1.

Proposition 9.8. *If M is a compact Riemann surface of genus $g = 0$, then M is holomorphically diffeomorphic to the Riemann sphere S^2.*

Proof. Pick $p \in M$; with $\vartheta = -p$, so $\langle \vartheta \rangle = 1$, (9.43) implies $\dim \mathcal{M}(\vartheta) = 2$. Of course, the constants form a one-dimensional subspace of $\mathcal{M}(\vartheta)$; thus we know that there is a nonconstant $u \in \mathcal{M}(\vartheta)$; u cannot be holomorphic, so it must have a simple pole at p. The proof thus follows from the next result.

Proposition 9.9. *If there exists a meromorphic function u on a compact Riemann surface M, regular except at a single point, where it has a simple pole, then M is holomorphically diffeomorphic to S^2.*

Proof. By the simple argument mentioned above (9.28), u must have precisely one zero, a simple zero. By the same reasoning, for any $\lambda \in \mathbb{C}$, $u - \lambda$ must have precisely one simple zero, so $u : M \to \mathbb{C} \cup \{\infty\} = S^2$ is a holomorphic diffeomorphism.

Proposition 9.10. *If M is a compact Riemann surface of genus $g = 1$, then there exists a lattice $\Gamma \subset \mathbb{C}$ such that M is holomorphically diffeomorphic to $\mathbb{C}/\Gamma$.*

Proof. By (9.21), or alternatively by (9.24), $\dim \mathcal{O}(\kappa) = 1$ in this case. Pick a nontrivial section ξ. By (9.28), $\sum \nu_\xi(p) = 2g - 2 = 0$. Since ξ has no poles, it also has no zeros, that is, κ is holomorphically trivial if $g = 1$. (Compare with (9.30).)

We use a topological fact. Namely, since $\dim \mathcal{H}^1(M, \mathbb{C}) = 2$ if $g = 1$, by deRham's theorem there exist closed curves γ_1, γ_2 in M such that, for *any* closed curve γ in M, there are integers m_1, m_2 such that

$$\int_\gamma v = m_1 \int_{\gamma_1} v + m_2 \int_{\gamma_2} v,$$

for any closed 1-form v on M. Granted this, it follows that if we pick $p_0 \in M$, the map

$$M \ni z \mapsto \int_{p_0}^{z} \xi \tag{9.48}$$

defines a holomorphic map

$$\Phi : M \longrightarrow \mathbb{C}/\Gamma', \tag{9.49}$$

where Γ' is the lattice in $\mathbb{C}$ generated by $\zeta_j = \int_{\gamma_j} \xi,\ j = 0, 1$. Since ξ is nowhere vanishing, the map (9.49) is a covering map. It follows that there is a holomorphic covering map $\Psi : \mathbb{C} \to M$, and the covering transformations form a group of translations of $\mathbb{C}$ (a subgroup of Γ', call it Γ). This gives the holomorphic diffeomorphism $M \approx \mathbb{C}/\Gamma$. We remark that, with a little extra argument, one can verify that (9.49) is already a diffeomorphism.

Propositions 9.8 and 9.10 are special cases of the *uniformization theorem* for compact Riemann surfaces. The $g \geq 2$ case will be established in Chap. 14 as a consequence of solving a certain nonlinear PDE. Also in that chapter, an alternative proof of Proposition 9.10 will be presented; in that case the PDE becomes linear. Also in Chap. 14 we present a linear PDE proof that treats the case $g = 0$. We note that in the treatment of the $g = 1$ case given above, the Riemann–Roch theorem is not essential; the analysis giving (9.22)–(9.24) suffices.

We return to the study of $\dim \mathcal{O}(L)$, for $L = E_\vartheta$. We illustrate how the first possibility can occur in (9.44). In fact, pick distinct points $p, q \in M$, and consider

$\vartheta = p - q$. Clearly, $c_1(E_{p-q})[M] = 0$. Now $\mathcal{O}(E_{p-q}) \approx \mathcal{M}(p-q)$, and it follows from Proposition 9.9 that if there is a nontrivial member of $\mathcal{M}(p-q)$, then M must be the sphere S^2. We thus have

$$\mathcal{O}(E_{p-q}) = 0 \quad \text{if } p \neq q \in M, \text{ of genus } g \geq 1. \tag{9.50}$$

On the other hand, if $p, q, r \in M$ are distinct, then $c_1(E_{-p-q+r})[M] = 1$, and (9.34) applies for $g = 1$; hence

$$g = 1 \Longrightarrow \dim \mathcal{M}(-p-q+r) = 1. \tag{9.51}$$

By the discussion above, a nontrivial $u \in \mathcal{M}(-p-q+r)$ cannot have just a simple pole; it must have poles at p and q. This proves the next result:

Proposition 9.11. *If $p, q,$ and r are distinct points in M, of genus 1, there is a meromorphic function on M with simple poles at p and q, and a zero at r, unique up to a multiplicative constant. Similarly, if $p = q \neq r \in M$, one has a meromorphic u with a double pole at p, and a zero at r.*

Given that $M = \mathbb{C}/\Gamma$, these meromorphic functions are the elliptic functions of Weierstrass, and they can be constructed explicitly. The uniqueness statement can also be established on elementary grounds. Note that, with p, q, and r as in Proposition 9.11, the corresponding elliptic function u vanishes at one other uniquely determined point s (or perhaps has a double zero at r, so $s = r$). In other words, if we set $\vartheta = -p - q + r + s$, for M of genus 1, the line bundle E_ϑ is trivial for a unique $s \in M$, given $p, q, r \in M$, r different from p or q. Actually, this last qualification can be dispensed with; $r = p$ forces $s = q$. It is a basic general question in Riemann surface theory to specify conditions on a divisor ϑ (in addition to $\langle \vartheta \rangle = 0$) necessary and sufficient for E_ϑ to be a trivial holomorphic line bundle over M. The question of whether E_ϑ is trivial is equivalent to the question of whether there exists a nontrivial meromorphic function on M, with poles at p of order exactly $|\nu(p)|$, where $\nu(p) < 0$, in the representation (9.35) for ϑ, and zeros of order exactly $\nu(p)$, where $\nu(p) > 0$. This question is answered by a theorem of Abel; see [Gu] for a discussion. The answer is essentially equivalent to a classification of holomorphic line bundles over M.

Exercises

1. Show that the conjugate linear map C in (9.23) is indeed well defined, independently of a choice of local holomorphic coordinates.
2. Show that if M is a compact Riemann surface, then the complex line bundle κ has a square root, i.e., a line bundle λ such that $\kappa \approx \lambda \otimes \lambda$. Show that λ can even be taken to be a holomorphic square root. Thus M actually has a spin structure. (Note also Exercise 5 of §3.)
3. Deduce the index formula (9.15), which leads to the Riemann–Roch formula, directly from Theorem 5.1, for twisted Dirac operators on spin manifolds.

4. Is it possible to choose a connection on L such that the operators $\overline{\partial}_L$ and D_L in (9.14) are actually equal?
5. Sections of the line bundle $\kappa \otimes \kappa$ are called *quadratic differentials*. Compute the dimension of $\mathcal{O}(\kappa \otimes \kappa)$. Given a divisor $\vartheta \le 0$, compute $\dim \mathcal{M}(\kappa \otimes \kappa, \vartheta)$.
6. Extend Theorem 9.1 to the case where $L \to M$ is a holomorphic *vector bundle*.
7. Formulate a version of the Riemann–Roch theorem for a compact, complex manifold M of higher dimension, and prove it, using Theorem 8.1.
8. Show that (9.41)–(9.42) provide an alternative proof of the existence result, Proposition 9.2.
9. Deduce from Proposition 9.2 that every holomorphic line bundle L over a Riemann surface is isomorphic to a divisor bundle E_ϑ.

A nonconstant meromorphic function $f : M \to \mathbb{C} \cup \{\infty\}$ can be regarded as a holomorphic map $f : M \to S^2$, which is onto. It is called a branched covering of S^2 by the Riemann surface M. A branch point of M is a point $p \in M$ such that $df(p) = 0$. The *order* $o(p)$ is the order to which $df(p)$ vanishes at p.

10. If $f : M \to S^2$ is a holomorphic map with branch points p_j, show that

$$\sum_j o(p_j) = 2\deg(f) + 2g - 2. \tag{9.52}$$

(*Hint*: Reduce to the case where all poles of f are simple, so (counting multiplicity)

$$\#\text{ poles of } f' = 2 \times \#\text{ poles of } f,$$

while the left side of (9.52) is equal to # zeros of f'. Think of f' as a meromorphic section of κ.)

11. Give another derivation of (9.52) by triangulating S^2 so that the points $q_j = f(p_j)$ are among the vertices, pulling this triangulation back to M, and comparing the numbers of vertices, edges, and faces.

The formula (9.52) is called Hurwitz' formula.

12. Let X be a "real" vector field on a compact Riemann surface M. Assume M is given a Riemannian metric compatible with its complex structure, so that $J : T_xM \to T_xM$ is an isometry. Picture X as a section of the complex line bundle $\mathcal{T} = \kappa^{-1}$. Show that X generates a group of conformal diffeomorphisms of M if and only if it is a *holomorphic* section of κ^{-1}. If g is the genus of M, show that

$$\begin{aligned} g \ge 2 &\Longrightarrow \mathcal{O}(\kappa^{-1}) = 0, \\ g = 1 &\Longrightarrow \dim_{\mathbb{C}} \mathcal{O}(\kappa^{-1}) = 1, \\ g = 0 &\Longrightarrow \dim_{\mathbb{C}} \mathcal{O}(\kappa^{-1}) = 3. \end{aligned}$$

Deduce the dimension of Lie groups of conformal diffeomorphisms in these cases. Compare the conclusion in case $g \ge 2$ with that of Exercise 5 of §4, given (see Chap. 14, §2) that one could choose a Riemannian metric of curvature -1. Compare the $g = 1$ case with Exercise 6 of §4.

13. Considering $\mathcal{M}(\kappa, p) = \{u \in \mathcal{O}(\kappa) : u(p) = 0\} \approx \mathcal{O}(\kappa \otimes E_p)$, show that

$$g \ge 1 \Longrightarrow \dim \mathcal{M}(\kappa, p) = g - 1.$$

Deduce that, for each $p \in M$, there exists $u \in \mathcal{O}(\kappa)$ such that $u(p) \neq 0$, provided $g \geq 1$. *Hint.* Use (9.17) to get $\dim \mathcal{O}(\kappa \otimes E_p) - \dim \mathcal{O}(E_p^{-1}) = g - 2$. Then show that $\dim \mathcal{O}(E_p^{-1}) = \dim \mathcal{M}(-p) = 1$ if $g \geq 1$. (Cf. Proposition 9.9)

14. Consider $\overline{\partial}_\kappa : H^s(M, \kappa) \to H^{s-1}(M, \kappa \otimes \overline{\kappa}) \approx H^{s-1}(M)$. Show that the range of $\overline{\partial}_\kappa$ has codimension 1. *Hint.* As in (9.8), the adjoint is $-\overline{\partial} : H^{1-s}(M) \to H^{-s}(M, \overline{\kappa})$.
15. Let u_j be meromorphic 1-forms on neighborhoods $\mathcal{O}_j$ of p_j $(1 \leq k \leq K)$, with poles at p_j. Use Exercise 14 to show there exists $u \in \mathcal{M}(\kappa)$ such that $u - u_j|_{\mathcal{O}_j}$ is pole free for each j, if and only if $\sum_{j=1}^K \operatorname{Res}_{p_j} u_j = 0$.
16. Let $E \to M$ be a holomorphic vector bundle over a compact Riemann surface, of rank k. That is, each fiber E_p has complex dimension k. Modify the proof of Theorem 9.1 to show that

$$\dim \mathcal{O}(E) - \dim \mathcal{O}(E' \otimes \kappa) = c_1(E)[M] - \frac{k}{2} c_1(\kappa)[M].$$

Here E' is the dual bundle of E. (*Hint.* Obtain an analogue of (9.15) and use $\operatorname{Ch}(E) = \operatorname{Tr} e^{-\Phi/2\pi i}$, as in (6.36), where Φ is the $\operatorname{End}(E)$-valued curvature form of a connection on E, to get

$$e^{-c_1(\kappa)/2} \operatorname{Ch}(E) = c_1(e) - \frac{k}{2} c_1(\kappa).)$$

10. Direct attack in 2-D

Here we produce a direct analysis of the index formula for a first-order, elliptic operator

$$D : C^\infty(M, E_0) \longrightarrow C^\infty(M, E_1) \tag{10.1}$$

of Dirac type when $\dim M = 2$. In view of (5.11), if $k_j(t, x, y)$ are the integral kernels of e^{-tD^*D} and e^{-tDD^*}, $j = 0, 1$, then

$$k_j(t, x, x) \sim a_{j0}(x) t^{-1} + a_{j1}(x) + a_{j2}(x) t + \cdots, \tag{10.2}$$

as $t \searrow 0$, and

$$\operatorname{Index} D = \int_M \big[a_{01}(x) - a_{11}(x) \big] \, dV(x). \tag{10.3}$$

As shown in Chap. 7, §14, we can produce explicit formulas for $a_{j1}(x)$ via calculations using the Weyl calculus.

Thus, pick local frame fields for E_0 and E_1 so that, in a local coordinate chart, $D = A(X, D)$, with

$$A(x, \xi) = \sum A_j(x) \xi_j + C(x), \tag{10.4}$$

a $K \times K$ matrix-valued symbol. Assume that

$$\begin{aligned} D^*D &= g(X,D) + \ell_0(X,D) + B_0(x), \\ DD^* &= g(X,D) + \ell_1(X,D) + B_1(x), \end{aligned} \tag{10.5}$$

where $g(x,\xi)$ defines a metric tensor, while $\ell_j(x,\xi)$ and $B_j(x)$ are $K \times K$ matrix-valued, and

$$\ell_\nu(x,\xi) = \sum_j \ell_j^{(\nu)}(x)\xi_j. \tag{10.6}$$

By (14.86) of Chap. 7, we have the following:

Proposition 10.1. *If D is an operator of Dirac type satisfying the hypotheses above and $\dim M = 2$, then Index D is equal to*

$$\frac{1}{4\pi}\int_M \left\{ Tr\, \sum_j \left[\ell_j^{(0)}(x)^2 - \ell_j^{(1)}(x)^2\right] + \ Tr\left[B_1(x) - B_0(x)\right]\right\} dV. \tag{10.7}$$

Of course, the individual terms in the integrand in (10.7) are not generally globally well defined on M; only the total is. We want to express these terms directly in terms of the symbol of D. Assuming the adjoint is computed using $L^2(U,dx)$, we have $D^*D = L_0(X,D)$ and $DD^* = L_1(X,D)$, with

$$\begin{aligned} L_0(x,\xi) &= A(x,\xi)^*A(x,\xi) + \frac{i}{2}\{A^*,A\}, \\ L_1(x,\xi) &= A(x,\xi)A(x,\xi)^* + \frac{i}{2}\{A,A^*\}. \end{aligned} \tag{10.8}$$

Hence

$$\begin{aligned} \ell_0(x,\xi) &= A_1(x,\xi)^*C(x) + C(x)^*A_1(x,\xi) + \frac{i}{2}\{A_1^*,A_1\}, \\ \ell_1(x,\xi) &= A_1(x,\xi)C(x)^* + C(x)A_1(x,\xi)^* + \frac{i}{2}\{A_1,A_1^*\}, \end{aligned} \tag{10.9}$$

where $A_1(x,\xi) = \sum A_j(x)\xi_j$, and

$$\begin{aligned} B_0(x) &= C(x)^*C(x) + \frac{i}{2}\{C^*,A_1\} + \frac{i}{2}\{A_1^*,C\}, \\ B_1(x) &= C(x)C(x)^* + \frac{i}{2}\{C,A_1^*\} + \frac{i}{2}\{A_1,C^*\}. \end{aligned} \tag{10.10}$$

Suppose that, for a given point $x_0 \in M$, we arrange $C(x_0) = 0$. Then

$$
\begin{aligned}
\ell_0(x_0,\xi) &= \frac{i}{2}\{A_1^*, A_1\} = \frac{i}{2}\sum_j \Big(\frac{\partial A_1^*}{\partial \xi_j}\frac{\partial A_1}{\partial x_j} - \frac{\partial A_1^*}{\partial x_j}\frac{\partial A_1}{\partial \xi_j}\Big),\\
\ell_1(x_0,\xi) &= \frac{i}{2}\{A_1, A_1^*\} = \frac{i}{2}\sum_j \Big(\frac{\partial A_1}{\partial \xi_j}\frac{\partial A_1^*}{\partial x_j} - \frac{\partial A_1}{\partial x_j}\frac{\partial A_1^*}{\partial \xi_j}\Big),
\end{aligned}
\tag{10.11}
$$

and

$$
\begin{aligned}
B_0(x_0) &= \frac{i}{2}\{C^*, A_1\} + \frac{i}{2}\{A_1^*, C\}\\
&= \frac{i}{2}\sum_j \Big(-\frac{\partial C^*}{\partial x_j}\frac{\partial A_1}{\partial \xi_j} + \frac{\partial A_1^*}{\partial \xi_j}\frac{\partial C}{\partial x_j}\Big),\\
B_1(x_0) &= \frac{i}{2}\{C, A_1^*\} + \frac{i}{2}\{A_1, C^*\}\\
&= \frac{i}{2}\sum_j \Big(-\frac{\partial C}{\partial x_j}\frac{\partial A_1^*}{\partial \xi_j} + \frac{\partial A_1}{\partial \xi_j}\frac{\partial C^*}{\partial x_j}\Big).
\end{aligned}
\tag{10.12}
$$

Note that if $A_1(x,\xi)$ is scalar, then $\ell_0(x_0,\xi) = -\ell_1(x_0,\xi)$ (granted that $C(x_0) = 0$). Hence their contributions to the integrand in (10.7) cancel. Also, if $A_1(x,\xi)$ is scalar, then $B_1(x_0) = -B_0(x_0)$. Thus, at x_0, the integrand in (10.7) is equal to

$$
2\ \mathrm{Tr}\ B_1(x_0) = -\ \mathrm{Tr}\ \sum_j \Big(\overline{A}_j \frac{\partial C}{\partial x_j} - A_j \frac{\partial C^*}{\partial x_j}\Big)
\tag{10.13}
$$

in this case. This situation arises for elliptic differential operators on sections of complex line bundles. In such a case, $C(x)$ is also scalar, and we can rewrite (10.13) as

$$
-2\ \mathrm{Im}\ \sum_j \overline{A}_j \frac{\partial C}{\partial x_j}.
\tag{10.14}
$$

Let's take a look at the operator $D_L : C^\infty(M, L) \to C^\infty(M, L \otimes \overline{\kappa})$, where M is a Riemann surface, $L \to M$ is a complex line bundle, with a Hermitian metric and a metric connection ∇, and, for a vector field X,

$$
\langle D_L u, X\rangle = \nabla_X u + i\nabla_{JX} u.
\tag{10.15}
$$

This is the same as (9.11)–(9.12), up to a factor of 2. Here J is the complex structure on TM. We can assume M has a Riemannian metric with respect to which J is rotation by 90°. Pick $x_0 \in M$. Use a geodesic normal coordinate system centered at x_0, so the metric tensor g_{jk} satisfies

$$
\nabla g_{jk}(x_0) = 0.
\tag{10.16}
$$

Let $X(x_0) = \partial/\partial x_1$, and define X by parallel transport radially from x_0 (along geodesics). Then

$$X(x) = a_1^1(x)\frac{\partial}{\partial x_1} + a_1^2(x)\frac{\partial}{\partial x_2}, \tag{10.17}$$

with

$$a_1^1(x_0) = 1, \quad a_1^2(x_0) = 0, \quad \nabla a_1^j(x_0) = 0. \tag{10.18}$$

Furthermore,

$$JX(x) = a_2^1(x)\frac{\partial}{\partial x_1} + a_2^2(x)\frac{\partial}{\partial x_2}, \tag{10.19}$$

with

$$a_2^1(x_0) = 0, \quad a_2^2(x_0) = 1, \quad \nabla a_2^j(x_0) = 0. \tag{10.20}$$

Next, let φ be a local section of L such that $\varphi(x_0)$ has norm 1, and $\varphi(x)$ is obtained from $\varphi(x_0)$ by radial parallel translation. Thus

$$u = v\varphi \Longrightarrow \nabla_{\partial_j} u = (\partial_j v + i\theta_j v)\varphi, \tag{10.21}$$

where the connection coefficients satisfy

$$\theta_j(x_0) = 0. \tag{10.22}$$

In such a coordinate system, and with respect to such choices, the operator D_L takes the form

$$D_L(v\varphi) = \frac{1}{i}\sum\Big[A_j\frac{\partial v}{\partial x_j} - A_j\theta_j v\Big]\varphi\otimes\vartheta, \tag{10.23}$$

where

$$A_j = i\big(a_1^j + ia_2^j\big) \tag{10.24}$$

and where $\vartheta \in C^\infty(U, \overline{\kappa})$ satisfies

$$\langle X, \vartheta\rangle = 1, \quad \langle JX, \vartheta\rangle = i.$$

Then $D_L^* : C^\infty(M, L\otimes\overline{\kappa}) \to C^\infty(M, L)$ is given by

$$D_L^*(w\,\varphi\otimes\vartheta) = \frac{1}{i}\sum g^{-1/2}\Big[\overline{A}_j\frac{\partial}{\partial x_j} + (\partial_j\overline{A}_j + \overline{A}_j\overline{\theta}_j)\Big]\big(g^{1/2}w\big)\varphi. \tag{10.25}$$

Now we want to take adjoints using $L^2(U, dx)$ rather than $L^2(U, \sqrt{g}dx)$, so we conjugate by $g^{1/4}$, and replace D_L by

$$\tilde{D}_L = \frac{1}{i}\sum\Big[g^{1/4}A_j\frac{\partial}{\partial x_j}\big(g^{-1/4}v\big) - A_j\theta_j v\Big]. \tag{10.26}$$

Thus we are in the situation of considering an operator of the form (10.4), with A_j given by (10.24) and

$$C(x) = \sum\left[\frac{i}{2}\frac{\partial A_j}{\partial x_j} - A_j\theta_j - \frac{1}{4}g^{-1}\frac{\partial g}{\partial x_j}A_j\right]. \tag{10.27}$$

Thus $C(x_0) = 0$, by (10.18)–(10.22), while

$$\partial_k C(x_0) = \sum_j\left[-A_j(\partial_k\theta_j) + \frac{i}{2}\partial_k\partial_j A_j - \frac{1}{4}A_j(\partial_k\partial_j g)\right]. \tag{10.28}$$

Now $\partial_k\theta_j(x_0)$ is given by the curvature of ∇ on L:

$$\frac{\partial\theta_j}{\partial x_k}(x_0) = \frac{1}{2}F_{jk}(x_0). \tag{10.29}$$

Meanwhile, as shown in §3 of Appendix C, $\partial_k\partial_j A_j$ can be expressed in terms of the Riemannian curvature:

$$\partial_j\partial_k a^\ell_m(x_0) = -\frac{1}{6}R_{\ell jmk} - \frac{1}{6}R_{\ell kmj}, \tag{10.30}$$

and of course so can $\partial_k\partial_j g(x_0)$. Consequently, at x_0, the formula (10.14) for the integrand in (10.7) becomes

$$-\frac{2}{i}F_{12} + \frac{1}{2}S(x_0). \tag{10.31}$$

Note that $S/2 = K$, the Gauss curvature. Thus the formula (10.7) becomes

$$\begin{aligned} \text{Index } D_L &= \frac{1}{4\pi}\int_M \Big(-\frac{2}{i}F_{12} + K\Big)\, dV \\ &= -\frac{1}{2\pi i}\int_M \omega_L + \frac{1}{4\pi}\int_M K\, dV, \end{aligned} \tag{10.32}$$

where ω_L is the curvature form of L. We have the identities

$$-\frac{1}{2\pi i}\int_M \omega_L = c_1(L)[M], \qquad \frac{1}{4\pi}\int_M K\, dV = \frac{1}{2}\chi(M), \tag{10.33}$$

the latter being the Gauss–Bonnet theorem.

Now, if $L \to M$ is a holomorphic line bundle, then $(1/2)D_L$ has the same principal symbol, hence the same index, as

(10.34) $$\overline{\partial}_L : C^\infty(M, L) \longrightarrow C^\infty(M, L \otimes \overline{\kappa}).$$

Hence we obtain the Riemann–Roch formula:

(10.35) $$\text{Index}\,\overline{\partial}_L = c_1(L)[M] + \frac{1}{2}\chi(M),$$

in agreement with (9.17).

We finish with a further comment on the Gauss–Bonnet formula; $\chi(M)$ is the index of

(10.36) $$d + \delta : \Lambda^0 M \oplus \Lambda^2 M \longrightarrow \Lambda^1 M$$

if $\dim M = 2$. If M is oriented, both $\Lambda^1 M$ and $(\Lambda^0 \oplus \Lambda^2)M$ get structures of complex line bundles via the Hodge $*$ operator; use

(10.37) $$J = * \text{ on } \Lambda^1, \quad J = -* : \Lambda^0 \to \Lambda^2, \quad J = * : \Lambda^2 \to \Lambda^0.$$

It follows easily that $(d + \delta)J = J(d + \delta)$, so we get a $\mathbb{C}$-linear differential operator

(10.38) $$\vartheta : \Lambda_e M \longrightarrow \Lambda_o M,$$

where $\Lambda_e = \Lambda^0 \oplus \Lambda^2,\ \Lambda_o = \Lambda^1$, regarded as complex *line* bundles, so

$$\text{Index}\,\vartheta = \frac{1}{2}\,\text{Index}(d + \delta).$$

Ker ϑ is a one-dimensional complex vector space:

$$\text{Ker}\,\vartheta = \text{span}(1) = \text{span}(*1).$$

The cokernel of $d + \delta$ in (10.36) consists of the space $\mathcal{H}_1$ of (real) harmonic 1-forms on M. This is invariant under $*$, so it becomes a complex vector space:

(10.39) $$\dim_{\mathbb{C}} \mathcal{H}_1 = \frac{1}{2}\dim_{\mathbb{R}} \mathcal{H}_1 = g.$$

Thus

(10.40) $$\text{Index}\,\vartheta = \frac{1}{2}(2 - 2g) = 1 - g.$$

When one applies an analysis parallel to that above, leading to (10.32), one gets

$$\text{Index}\, \vartheta = \frac{1}{4\pi} \int_M K\, dV. \tag{10.41}$$

Putting together (10.40) and (10.41), we again obtain the Gauss–Bonnet formula, for a compact, oriented surface.

Exercises

1. Use (10.36)–(10.39) to give another proof of (9.24), that is,

$$\dim \mathcal{O}(\kappa) = \frac{1}{2} \dim \mathcal{H}^1(M, \mathbb{C}) = g.$$

In Exercises 2–4, suppose $E_j \to M$ are complex line bundles over M, a compact manifold of dimension 2, and suppose

$$D : C^\infty(M, E_0) \longrightarrow C^\infty(M, E_1)$$

is a first-order, elliptic differential operator.

2. Show that the symbol of D induces an $\mathbb{R}$-linear isomorphism

$$\sigma_D(x) : T_x^* \longrightarrow \mathcal{L}(E_{0x}, E_{1x}). \tag{10.42}$$

Hence M has a complex structure, making this $\mathbb{C}$-linear. This gives M an orientation; reversing the orientation makes (10.42) conjugate linear.

3. If M is oriented so that (10.42) is conjugate linear, show that D has a principal symbol homotopic to that of D_L, given by (10.15), with $L = E_0,\ L \otimes \overline{\kappa} \approx E_1$. Deduce that

$$\text{Index}\, D = \frac{1}{2} c_1(E_0)[M] + \frac{1}{2} c_1(E_1)[M]. \tag{10.43}$$

4. What happens to the formula for Index D^*?

In Exercises 5–8, $S_0^2 T^*$ denotes the bundle of symmetric second-order tensors with trace zero on a Riemannian manifold M, and $S_0^{1,1}$ denotes the bundle of symmetric tensors of type $(1,1)$ with trace 0. The metric tensor provides an isomorphism of these two bundles.

5. If M is a compact, oriented 2-fold, with associated complex structure $J : T_x M \to T_x M$, show that a complex structure is defined on $S_{0x}^{1,1} \subset \text{Hom}\, T_x$ by

$$\mathfrak{J}(A) = \frac{1}{2}[J, A] = JA. \tag{10.44}$$

Thus $S_0^{1,1}$ and $S_0^2 T^*$ become complex line bundles.

6. Recall the first-order operator considered in (4.29)–(4.31):

$$D_{TF} : C^\infty(M, T) \longrightarrow C^\infty(M, S_0^2 T^*), \quad D_{TF} X = \text{Def}\, X - \frac{1}{2}(\text{div}\, X) g, \tag{10.45}$$

in case $n = \dim M = 2$. If T and $S_0^2 T^*$ are regarded as complex line bundles, show that D_{TF} is $\mathbb{C}$-linear.

7. Recall that $\ker D_{TF}$ consists of vector fields that generate conformal diffeomorphisms of M, hence of holomorphic sections of $\mathcal{T} = \kappa^{-1}$. Show that there is an isomorphism $S_0^2 T^* \approx \kappa^{-1} \otimes \overline{\kappa}$ transforming (10.45) to

$$\overline{\partial} : C^\infty(M, \kappa^{-1}) \longrightarrow C^\infty(M, \kappa^{-1} \otimes \overline{\kappa}). \tag{10.46}$$

Note that Index $\overline{\partial} = -(3g - 3)$ in this case, if g is the genus of M.

8. In view of (4.30), the orthogonal complement of the range of D_{TF} is the finite dimensional space

$$\mathcal{V} = \{u \in C^\infty(M, S_0^2 T^*) : \operatorname{div} u = 0\}. \tag{10.47}$$

Comparing (10.45) and (10.46), show that $\mathcal{V} \approx \mathcal{O}(\kappa \otimes \kappa)$. If M has genus $g \geq 2$, $\overline{\partial}$ in (10.46) is injective (by Exercise 12, §9). Deduce that

$$\dim_{\mathbb{R}} \mathcal{V} = 6g - 6, \quad \text{if } g \geq 2. \tag{10.48}$$

Compare Exercise 5 of §9. For $g = 0$, compare Exercise 7 of §4.
For connections with the dimension of Teichmuller space, see [Tro].

11. Index of operators of harmonic oscillator type

In this section we study elliptic operators of harmonic oscillator type, introduced in §15 of Chap. 7. We recall that a symbol $p(x, \xi)$ belongs to $\mathcal{S}_1^m(\mathbb{R}^n)$ if it is smooth in $(x, \xi) \in \mathbb{R}^n \times \mathbb{R}^n$ and satisfies estimates

$$|D_x^\beta D_\xi^\alpha p(x, \xi)| \leq C_{\alpha\beta}(1 + |x| + |\xi|)^{m - |\alpha| - |\beta|}. \tag{11.1}$$

The associated operator $P = p(X, D) \in OP\mathcal{S}_1^m(\mathbb{R}^n)$ is defined using the Weyl calculus. The operator is elliptic provided that, for $|x|^2 + |\xi|^2$ large enough,

$$|p(x, \xi)^{-1}| \leq C(1 + |x| + |\xi|)^{-m}. \tag{11.2}$$

In such a case, P has a parametrix $Q \in OP\mathcal{S}_1^{-m}(\mathbb{R}^n)$, such that $PQ - I$ and $QP - I$ belong to $OP\mathcal{S}_1^{-\infty}(\mathbb{R}^n)$. The class $\mathcal{S}^m(\mathbb{R}^n)$ of classical symbols is defined to consist of elements $p(x, \xi) \in \mathcal{S}_1^m(\mathbb{R}^n)$ such that

$$p(x, \xi) \sim \sum_{j \geq 0} p_j(x, \xi), \tag{11.3}$$

where $p_j(x, \xi) \in \mathcal{S}_1^{m-2j}(\mathbb{R}^n)$ is homogeneous of degree $m - 2j$ in (x, ξ) for $|x|^2 + |\xi|^2 \geq 1$. If such a symbol satisfies the ellipticity condition (11.2), then $P = p(X, D)$ has parametrix $Q \in OP\mathcal{S}^{-m}(\mathbb{R}^n)$. A paradigm example of such an operator is the harmonic oscillator

$$H = -\Delta + |x|^2, \tag{11.4}$$

which is elliptic in $OP\mathcal{S}^2(\mathbb{R}^n)$, with symbol $|x|^2 + |\xi|^2$. It is a positive definite operator, and, as shown in Chap. 7,

$$H^s \in OP\mathcal{S}^{2s}(\mathbb{R}^n), \quad \forall\, s \in \mathbb{R}. \tag{11.5}$$

There are Sobolev-type spaces $\mathcal{H}^s(\mathbb{R}^n),\ s \in \mathbb{R}$, such that, for $s = k \in \mathbb{Z}^+$,

$$\mathcal{H}^k(\mathbb{R}^n) = \{u \in L^2(\mathbb{R}^n) : x^\alpha D_x^\beta u \in L^2(\mathbb{R}^n),\ \forall\, |\alpha| + |\beta| \le k\}. \tag{11.6}$$

As shown in Chap. 7, if $P \in OP\mathcal{S}^m(\mathbb{R}^n)$, then, for all $s \in \mathbb{R}$,

$$P : \mathcal{H}^s(\mathbb{R}^n) \longrightarrow \mathcal{H}^{s-m}(\mathbb{R}^n), \tag{11.7}$$

and if P is elliptic, this map is Fredholm. We want to study its index. For simplicity, we stick to operators with symbols of classical type.

If $P = p(X, D)$ is an elliptic operator ($k \times k$ matrix valued), with symbol expansion of the form, we call $p_0(x, \xi)$ the principal symbol. Recall we assume $p_0(x, \xi)$ is homogeneous of order m for $|x|^2 + |\xi|^2 \ge 1$. We then have the symbol map

$$\begin{aligned} &\sigma_P : S^{2n-1} \longrightarrow G\ell(k, \mathbb{C}), \\ &\sigma_P(x, \xi) = p_0(x, \xi), \quad |x|^2 + |\xi|^2 = 1. \end{aligned} \tag{11.8}$$

Note that $P \in OP\mathcal{S}^m(\mathbb{R}^n)$ and $PH^\mu \in OP\mathcal{S}^{m+2\mu}(\mathbb{R}^n)$ have the same symbol map, and they have the same index, one on $\mathcal{H}^s(\mathbb{R}^n) \to \mathcal{H}^{s-m}(\mathbb{R}^n)$ and the other on $\mathcal{H}^s(\mathbb{R}^n) \to \mathcal{H}^{s-m-2\mu}(\mathbb{R}^n)$. Basic Fredholm theory gives the following.

Proposition 11.1. *Given elliptic $k \times k$ systems $P_j \in OP\mathcal{S}^{m_j}(\mathbb{R}^n)$, if σ_{P_1} and σ_{P_2} are homotopic maps from S^{2n-1} to $G\ell(k, \mathbb{C})$, then* Index $P_1 =$ Index P_2.

Let us take $n = 1$ and $k = 1$ and look for specific index formulas. In this case, given elliptic scalar $P \in OP\mathcal{S}^m(\mathbb{R})$, we have

$$\sigma_P : S^1 \longrightarrow G\ell(1, \mathbb{C}) = \mathbb{C} \setminus 0. \tag{11.9}$$

Such a map is specified up to homotopy by the winding number

$$\operatorname{ind} \sigma_P = \frac{1}{2\pi i} \int_{S^1} \frac{\sigma_P'(\zeta)}{\sigma_P(\zeta)}\, d\zeta, \tag{11.10}$$

where $\zeta = x + i\xi$. If P_1 and P_2 are two such elliptic operators, we have

$$\text{Index } P_1 P_2 = \text{Index } P_1 + \text{Index } P_2, \tag{11.11}$$

and

$$\operatorname{ind} \sigma_{P_1 P_2} = \operatorname{ind} \sigma_{P_1} + \operatorname{ind} \sigma_{P_2}. \tag{11.12}$$

Let us consider the operator

$$D_1 = \frac{\partial}{\partial x_1} + x_1, \tag{11.13}$$

acting on functions of $x_1 \in \mathbb{R}$. Its symbol is $x_1 + i\xi_1$, so

$$\operatorname{ind} \sigma_{D_1} = 1. \tag{11.14}$$

Note that $D_1^* = -\partial_1 + x_1$, and

$$D_!^* D_1 = -\partial_1^2 + x_1^2 - 1, \quad D_1 D_1^* = -\partial_1^2 + x_1^2 + 1. \tag{11.15}$$

We have

$$\begin{gathered} \operatorname{Ker} D_1 = \operatorname{Span}\{e^{-x_1^2/2}\}, \ \text{ and} \\ D_1 D_1^* \geq 2I \Longrightarrow \operatorname{Ker} D_1^* = 0, \end{gathered} \tag{11.16}$$

hence

$$\operatorname{Index} D_1 = 1. \tag{11.17}$$

Putting together (11.9)–(11.17) and Proposition 11.1, we have the following.

Proposition 11.2. *If $P \in OPS^m(\mathbb{R})$ is a scalar elliptic operator, then*

$$\operatorname{Index} P = \operatorname{ind} \sigma_P. \tag{11.18}$$

We next keep $n = 1$ and let $P \in OPS^m(\mathbb{R})$ be an elliptic $k \times k$ system, so

$$\sigma_P : S^1 \longrightarrow G\ell(k, \mathbb{C}). \tag{11.19}$$

We want to classify these maps, up to homotopy. To do this, we bring in the following topological fact about

$$S\ell(k, \mathbb{C}) = \{A \in G\ell(k, \mathbb{C}) : \det A = 1\}, \tag{11.20}$$

namely

$$S\ell(k, \mathbb{C}) \ \text{ is simply connected.} \tag{11.21}$$

Using this fact, we prove the following.

Proposition 11.3. *Given a symbol map (11.19), define* $\tilde{\sigma}_P : S^1 \to G\ell(k, \mathbb{C})$ *by*

$$\tilde{\sigma}_P(x,\xi) = \begin{pmatrix} \det \sigma_P(x,\xi) & 0 \\ 0 & I \end{pmatrix}, \tag{11.22}$$

where I *denotes the* $(k-1)\times(k-1)$ *identity matrix. Then* σ_P *and* $\tilde{\sigma}_P$ *are homotopic.*

Proof. Given (11.19) and (11.22), we set

$$\gamma_1 = \tilde{\sigma}_P\,\sigma_P^{-1} : S^1 \longrightarrow S\ell(k,\mathbb{C}). \tag{11.23}$$

Using (11.21), we can deform γ_1 to $\gamma_0 \equiv I$, through $\gamma_\tau : S^1 \to S\ell(k,\mathbb{C})$, $0 \le \tau \le 1$. A homotopy from σ_P to $\tilde{\sigma}_P$ is then given by $\sigma_P^\tau(x,\xi) = \gamma_\tau(x,\xi)\sigma_P(x,\xi)$, $0 \le \tau \le 1$.

We have a scalar operator $\widetilde{P} \in OP\mathcal{S}^{km}(\mathbb{R})$, defined uniquely mod $OP\mathcal{S}^{km-2}(\mathbb{R})$ by the condition

$$\sigma_{\widetilde{P}} = \det \sigma_P. \tag{11.24}$$

Then

$$\text{Index } P = \text{Index} \begin{pmatrix} \widetilde{P} & \\ & I \end{pmatrix}, \tag{11.25}$$

which by Proposition 11.2 is given by ind det σ_P. We have proved the following.

Proposition 11.4. *If* $P \in OP\mathcal{S}^m(\mathbb{R})$ *is an elliptic* $k \times k$ *system,*

$$\text{Index } P = \text{ind} \det \sigma_P. \tag{11.26}$$

Returning to (11.21), we note that it is equivalent to the result

$$SU(k) \ \text{ is simply connected.} \tag{11.27}$$

To see this, we use the polar decomposition

$$A \in G\ell(k,\mathbb{C}) \Longrightarrow A = U(A)\Pi(A), \tag{11.28}$$

where

$$\Pi(A) = (A^*A)^{1/2} \ \text{ is positive definite,}$$
$$U(A) = A(A^*A)^{-1/2} \in U(k).$$

With this, we can define a 1-parameter family of maps

$$\vartheta_\tau : G\ell(k,\mathbb{C}) \longrightarrow G\ell(k,\mathbb{C}), \quad \tau \in [0,1], \tag{11.29}$$

by

$$\vartheta_\tau(A) = U(A)\Pi(A)^\tau. \tag{11.30}$$

We have

$$\vartheta_0(A) = U(A), \quad \vartheta_1(A) = A. \tag{11.31}$$

This makes $U(k)$ a deformation retract of $G\ell(k,\mathbb{C})$. As a consequence, each continuous map $\sigma : S^{2n-1} \to G\ell(k,\mathbb{C})$ is homotopic to the map $\vartheta_0 \circ \sigma : S^{2n-1} \to U(k)$. Note that

$$\det \Pi(A) = |\det A|, \quad \det U(A) = \frac{\det A}{|\det A|}, \tag{11.32}$$

so

$$\vartheta_0 : S\ell(k,\mathbb{C}) \longrightarrow SU(k), \tag{11.33}$$

and ϑ_τ makes $SU(k)$ a deformation retract of $S\ell(k,\mathbb{C})$. This establishes the equivalence of (11.21) and (11.27). In case $k = 2$, we have

$$SU(2) = \left\{ \begin{pmatrix} a & -\overline{b} \\ b & \overline{a} \end{pmatrix} : a, b \in \mathbb{C},\ |a|^2 + |b|^2 = 1 \right\} \approx S^3, \tag{11.34}$$

which is clearly simply connected. For $k > 2$, (11.27) is a special case of (11.56) below.

Let us now take $n \geq 2$ and consider a $k \times k$ elliptic system $P \in OP\mathcal{S}^m(\mathbb{R}^n)$, giving a symbol map (11.8). Making use of the deformation retract (11.29)–(11.31), we see that σ_P is homotopic to a symbol map

$$\sigma_{P^\#} : S^{2n-1} \longrightarrow U(k), \tag{11.35}$$

for an operator $P^\# \in OP\mathcal{S}^m(\mathbb{R}^n)$, uniquely defined mod $OP\mathcal{S}^{m-2}(\mathbb{R}^n)$, and Index P = Index $P^\#$. For $k = 1$, we have the following topological result.

Lemma 11.5. *If $n \geq 2$, every continuous map $\sigma : S^{2n-1} \to U(1) = S^1$ is homotopic to a constant map.*

Proof. Indeed, since S^{2n-1} is simply connected for $n \geq 2$, σ lifts to a continuous map $\tilde{\sigma} : S^{2n-1} \to \mathbb{R}$, which is clearly homotopic to a constant map.

In light of this, if we have (11.35) and set (as in (11.22))

$$\tilde{\sigma}(x,\xi) = \begin{pmatrix} \det \sigma_{P^\#}(x,\xi) & 0 \\ 0 & I \end{pmatrix}, \quad \tilde{\sigma} : S^{2n-1} \to U(k), \tag{11.36}$$

then, for $n \geq 2$, $\tilde{\sigma}$ is homotopic to a constant. Hence $\sigma_{P^\#}$ and

$$\sigma^b : S^{2n-1} \to SU(k), \quad \sigma^b(x,\xi) = \tilde{\sigma}(x,\xi)^{-1}\sigma_{P^\#}(x,\xi), \tag{11.37}$$

are homotopic. Given $\mu \in \mathbb{R}$, this is the symbol map of an operator $\widehat{P} \in OP\mathcal{S}^\mu(\mathbb{R}^n)$, uniquely determined up to a lower order operator. We have the following result.

Proposition 11.6. *For $n \geq 2$, if $P \in OP\mathcal{S}^m(\mathbb{R}^n)$ is an elliptic $k \times k$ system, there exists for each $\mu \in \mathbb{R}$ an elliptic $k \times k$ system $\widehat{P} \in OP\mathcal{S}^\mu(\mathbb{R}^n)$ whose symbol map*

$$\sigma_{\widehat{P}} : S^{2n-1} \longrightarrow SU(k) \tag{11.38}$$

is homotopic to σ_P, as maps $S^{2n-1} \to G\ell(k,\mathbb{C})$. Hence

$$\text{Index}\, P = \text{Index}\, \widehat{P}. \tag{11.39}$$

Let us now specialize to $n = 2$. By Lemma 11.5, every scalar elliptic $P \in OP\mathcal{S}^m(\mathbb{R}^2)$ must have index 0. We construct an elliptic 2×2 system with nonzero index as follows. With D_1 as in (11.13), set

$$\begin{aligned} D_2 &= \begin{pmatrix} \partial_1 + x_1 & \partial_2 - x_2 \\ \partial_2 + x_2 & -\partial_1 + x_1 \end{pmatrix} \\ &= \begin{pmatrix} D_1 & -L_2^* \\ L_2 & D_1^* \end{pmatrix}, \end{aligned} \tag{11.40}$$

where

$$L_2 = \partial_2 + x_2, \quad L_2^* = -\partial_2 + x_2. \tag{11.41}$$

Note that

$$\sigma_{D_2} = \begin{pmatrix} x_1 + i\xi_1 & -x_2 + i\xi_2 \\ x_2 + i\xi_2 & x_1 - i\xi_2 \end{pmatrix}, \quad \text{so } \sigma_{D_2} : S^3 \to SU(2) \approx S^3 \tag{11.42}$$

is essentially the identity map. A computation gives

$$D_2^* D_2 = \begin{pmatrix} D_1^* D_1 + L_2^* L_2 & \\ & D_1 D_1^* + L_2 L_2^* \end{pmatrix}, \tag{11.43}$$

and

$$D_2 D_2^* = \begin{pmatrix} D_1 D_1^* + L_2^* L_2 & \\ & D_1^* D_1 + L_2 L_2^* \end{pmatrix}. \tag{11.44}$$

We recall the formulas for $D_1^* D_1$ and $D_1 D_1^*$ in (11.15). Similarly,

$$L_2^* L_2 = -\partial_2^2 + x_2^2 - 1, \quad L_2 L_2^* = -\partial_2^2 + x_2^2 + 1. \tag{11.45}$$

Hence $\operatorname{Ker} L_2 = \operatorname{Span}\{e^{-x_2^2/2}\}$ and $L_2 L_2^* \geq 2I$, and we have for the four diagonal elements of (11.43)–(11.44) that

$$\begin{aligned} \dim \operatorname{Ker}(D_1^* D_1 + L_2^* L_2) &= 1, \\ D_1 D_1^* + L_2 L_2^* &\geq 4I, \\ D_1 D_1^* + L_2^* L_2 &\geq 2I, \\ D_1^* D_1 + L_2 L_2^* &\geq 2I. \end{aligned} \tag{11.46}$$

Hence

$$\dim \operatorname{Ker} D_2 = 1, \quad \dim \operatorname{Ker} D_2^* = 0, \tag{11.47}$$

so

$$\operatorname{Index} D_2 = 1. \tag{11.48}$$

Now consider an arbitrary 2×2 elliptic system $P \in OPS^m(\mathbb{R}^2)$. As in (11.38), we have an adjusted operator $\widehat{P}$, with the same index as P, and

$$\sigma_{\widehat{P}} : S^3 \longrightarrow SU(2) \approx S^3. \tag{11.49}$$

The homotopy class of this map is an element of $\pi_3(S^3)$. Results on this homotopy group, which we will discuss in more detail below, imply the following.

Proposition 11.7. *Let $P \in OPS^m(\mathbb{R}^2)$ be a 2×2 elliptic system. For $\sigma_{\widehat{P}}$ as in (11.49), there is a unique integer ℓ such that either*

$$\begin{aligned} &\ell > 0 \ \textit{and} \ \sigma_{\widehat{P}} \ \textit{is homotopic to} \ \sigma_{D_2^\ell}, \\ &\ell = 0 \ \textit{and} \ \sigma_{\widehat{P}} \ \textit{is homotopic to a constant map}, \\ &\ell < 0 \ \textit{and} \ \sigma_{\widehat{P}} \ \textit{is homotopic to} \ \sigma_{(D_2^*)^{|\ell|}}. \end{aligned} \tag{11.50}$$

We denote this ℓ by

$$\ell = \operatorname{ind} \sigma_{\widehat{P}}. \tag{11.51}$$

Then

$$\text{Index } P = \text{Index } \widehat{P} = \text{ind}\, \sigma_{\widehat{P}}. \tag{11.52}$$

To see that (11.52) follows from (11.50), note that in the first case Index $P =$ Index $D_2^\ell = \ell$ and in the third case Index $P = \text{Index}(D_2^*)^{|\ell|} = -|\ell| = \ell$.

We now discuss some homotopy theory behind (11.50). It is convenient to place this in a more general setting. If M is a smooth, connected manifold and $j \in \mathbb{N}$, $\pi_j(M)$ denotes the set of homotopy classes of continuous maps $\varphi : S^j \to M$ (which is equivalent to the set of homotopy classes of smooth maps). This can be given a group structure as follows. Fix $p_0 \in S^j$, $q_0 \in M$. Given maps $\varphi, \psi : S^j \to M$, one can produce maps homotopic to these that take p_0 to q_0, so assume φ and ψ have this property. Now take S^j and collapse its "equator," which is homeomorphic to S^{j-1}, to a point. You obtain two copies of S^j, joined at a point, which we identify with p_0. Then map the top sphere to M by ψ and the bottom sphere to M by φ, and compose with the collapse map, to get a map $\sigma : S^j \to M$, whose homotopy class $[\sigma] = [\varphi] \cdot [\psi]$.

In case G is a connected Lie group, there is another way to define a product on $\pi_j(G)$. Namely, if $\varphi, \psi : S^j \to G$, consider the map $\varphi \cdot \psi : S^j \to G$ given by $(\varphi \cdot \psi)(x) = \varphi(x)\psi(x)$, using the product on G. If φ and $\tilde{\varphi}$ are homotopic (write $\varphi \sim \tilde{\varphi}$) and also $\psi \sim \tilde{\psi}$, we have $\varphi \cdot \psi \sim \tilde{\varphi} \cdot \tilde{\psi}$, so this gives a product on $\pi_j(G)$. It is a basic fact that this product on $\pi_j(G)$ agrees with the previously defined one; cf. [Spa], Chap. 1.

What makes (11.50) work is the $j = 3$ case of the following fundamental result of H. Hopf.

Proposition 11.8. *For each $j \in \mathbb{N}$,*

$$\pi_j(S^j) \approx \mathbb{Z}, \tag{11.53}$$

and (the homotopy class of) the identity map Id : $S^j \to S^j$ is a generator.

In fact, if $\varphi, \psi : S^j \to S^j$ are smooth, they have degrees, defined in Chap. 1, §19, and the Hopf theorem says they are homotopic if and only if they have the same degree. Cf. [Spa], p. 398.

Under the identification (11.34) of $SU(2)$ with $(a, b) \in S^3$, $\sigma_{D_2} : S^3 \to S^3$ is the identity map, and $\sigma_{D_2^\ell} \in \pi_3(S^3)$ is an ℓ-fold product, hence corresponds to $\ell \in \mathbb{Z}$ under this isomorphism, while $\sigma_{D_2^*} = -1 \in \pi_3(S^3)$, and $\sigma_{(D_2^*)^{|\ell|}} = -|\ell| \in \pi_3(S^3)$.

Let us next consider a $k \times k$ elliptic system $P \in OPS^m(\mathbb{R}^2)$, giving rise to $P^\#$ as in (11.35) and $\widehat{P}$ as in (11.38), all having the same index. The following result is useful.

Proposition 11.9. *For each $k \in \mathbb{N}$, the natural inclusion $SU(k) \hookrightarrow U(k)$ induces an isomorphism*

$$\pi_j(SU(k)) \xrightarrow{\approx} \pi_j(U((k)), \quad \textit{if } \ j > 1. \tag{11.54}$$

Furthermore, the inclusions $U(k) \hookrightarrow U(k+\ell)$ *and* $SU(k) \hookrightarrow SU(k+\ell)$*, given by*

$$A \mapsto \begin{pmatrix} A & \\ & I \end{pmatrix}, \tag{11.55}$$

where I *denotes the* $\ell \times \ell$ *identity matrix, induce isomorphisms*

(11.56)
$$\pi_j(U(k)) \xrightarrow{\approx} \pi_j(U(k+\ell)), \quad \pi_j(SU(k)) \xrightarrow{\approx} \pi_j(SU(k+\ell)), \quad \textit{if } \ j \le 2k-1.$$

We mention that a proof of (11.54) requires just a few arguments beyond the proof of Proposition 11.6. The proof of (11.56), with $\ell = 1$, which then proceeds inductively, follows by applying the "homotopy exact sequence" to

$$U(k+1)/U(k) \approx S^{2k+1}, \quad SU(k+1)/SU(k) \approx S^{2k+1}. \tag{11.57}$$

See (11.82) below. According to Proposition 11.9, when $j = 3$, (11.56) holds for $k \ge 2$. Taking (11.53) into account, we have

$$\pi_3(SU(k)) \approx \pi_3(U(k)) \approx \mathbb{Z}, \quad \forall\, k \ge 2. \tag{11.58}$$

We can now augment Proposition 11.7 as follows.

Proposition 11.10. *Let* $P \in OP\mathcal{S}^m(\mathbb{R}^2)$ *be a* $k \times k$ *elliptic system,* $k > 2$*. For* $\sigma_{\widehat{P}} : S^3 \to SU(k)$ *as in (11.49), there is a unique integer* ℓ *such that, with* I *denoting the* $(\ell-2) \times (\ell-2)$ *identity matrix, either*

$$\begin{aligned} &\ell > 0 \ \textit{ and } \ \sigma_{\widehat{P}} \ \textit{ is homotopic to } \begin{pmatrix} \sigma_{D_2^\ell} & \\ & I \end{pmatrix}, \\ &\ell = 0 \ \textit{ and } \ \sigma_{\widehat{P}} \ \textit{ is homotopic to a constant map}, \\ &\ell < 0 \ \textit{ and } \ \sigma_{\widehat{P}} \ \textit{ is homotopic to } \begin{pmatrix} \sigma_{(D_2^*)^{|\ell|}} & \\ & I \end{pmatrix}. \end{aligned} \tag{11.59}$$

We denote this ℓ *by*

$$\ell = \operatorname{ind} \sigma_{\widehat{P}}. \tag{11.60}$$

Then

$$\operatorname{Index} P = \operatorname{Index} \widehat{P} = \operatorname{ind} \sigma_{\widehat{P}}. \tag{11.61}$$

We now turn to higher dimensions. Our next task is to construct, for each j an elliptic system $D_j \in OP\mathcal{S}^1(\mathbb{R}^j)$ (actually a system of differential operators) of index 1. The construction is inductive. Assume we have such an elliptic system D_{n-1}, with the properties

(11.62) $$\dim \operatorname{Ker} D_{n-1} = 1,$$

and

(11.63) $$D_{n-1}D_{n-1}^* \geq 2I.$$

By (11.15)–(11.16) we have this for $n-1=1$, and by (11.43)–(11.47) we have this for $n-1=2$. We then set

(11.64) $$D_n = \begin{pmatrix} D_{n-1} & \partial_n - x_n \\ \partial_n + x_n & D_{n-1}^* \end{pmatrix} = \begin{pmatrix} D_{n-1} & -L_n^* \\ L_n & D_{n-1}^* \end{pmatrix},$$

where

(11.65) $$L_n = \partial_n + x_n, \quad L_n^* = -\partial_n + x_n.$$

Parallel to (11.43)–(11.44), a computation gives

(11.66) $$D_n^* D_n = \begin{pmatrix} D_{n-1}^* D_{n-1} + L_n^* L_n & \\ & D_{n-1}D_{n-1}^* + L_n L_n^* \end{pmatrix},$$

and

(11.67) $$D_n D_n^* = \begin{pmatrix} D_{n-1} D_{n-1}^* + L_n^* L_n & \\ & D_{n-1}^* D_{n-1} + L_n L_n^* \end{pmatrix}.$$

Parallel to (11.45), we have

(11.68) $$L_n^* L_n = -\partial_n^2 + x_n^2 - 1, \quad L_n L_n^* = -\partial_n^2 + x_n^2 + 1.$$

We see that L_n annihlates $e^{-x_n^2/2}$ and $L_n L_n^* \geq 2I$. Hence, parallel to (11.46), we have

(11.69) $$\begin{aligned} \dim \operatorname{Ker}(D_{n-1}^* D_{n-1} + L_n^* L_n) &= 1, \\ D_{n-1}D_{n-1}^* + L_n L_n^* &\geq 4I, \\ D_{n-1}D_{n-1}^* + L_n^* L_n &\geq 2I, \\ D_{n-1}^* D_{n-1} + L_n L_n^* &\geq 2I. \end{aligned}$$

Consequently, we have

(11.70) $$\dim \operatorname{Ker} D_n = 1,$$

and

$$D_n D_n^* \geq 2I, \tag{11.71}$$

hence

$$\operatorname{Index} D_n = 1. \tag{11.72}$$

This completes the inductive construction. Note that the matrix doubles in size at each iteration, so D_n is a $2^{n-1} \times 2^{n-1}$ matrix of differential operators.

We can extend Proposition 11.10, using the following fundamental result of R. Bott. Cf. [Mil], §23.

Proposition 11.11. *For $n \in \mathbb{N}$,*

$$\pi_{2n-1}(U(k)) \approx \pi_{2n-1}(SU(k)) \approx \mathbb{Z}, \quad \text{if } k \geq n. \tag{11.73}$$

Note that (11.58) is the case $n = 2$ of this result. Given this proposition, the calculation (11.72) implies the following.

Proposition 11.12. *For $n \in \mathbb{N}$, the map*

$$\sigma_{D_n} : S^{2n-1} \longrightarrow U(2^{n-1}) \tag{11.74}$$

defines a generator of $\pi_{2n-1}(U(2^{n-1}))$.

Note: The calculation (11.66) implies $\sigma_{D_n^* D_n}(x, \xi) = \sigma_{D_n}(x, \xi)^* \sigma_{D_n}(x, \xi) = I$, for $|x|^2 + |\xi|^2 = 1$.

From here, we have the following extension of Proposition 11.10.

Proposition 11.13. *Let $P \in OPS^m(\mathbb{R}^n)$ be a $k \times k$ elliptic system, with associated symbol map $\sigma_{\widehat{P}} : S^{2n-1} \to SU(k)$. If $k = 2^{n-1}$, there is a unique integer ℓ such that either*

$$\begin{aligned} &\ell > 0 \ \text{ and } \ \sigma_{\widehat{P}} \ \text{ is homotopic to } \sigma_{D_n^\ell}, \\ &\ell = 0 \ \text{ and } \ \sigma_{\widehat{P}} \ \text{ is homotopic to a constant map}, \\ &\ell < 0 \ \text{ and } \ \sigma_{\widehat{P}} \ \text{ is homotopic to } \sigma_{(D_n^*)^{|\ell|}}. \end{aligned} \tag{11.75}$$

we denote this ℓ by

$$\ell = \operatorname{ind} \sigma_{\widehat{P}}. \tag{11.76}$$

Then

$$\text{Index } P = \text{Index } \widehat{P} = \text{ind } \sigma_{\widehat{P}}. \tag{11.77}$$

If $k < 2^{n-1}$, *then* Index P = Index $\widehat{P}$ = Index $\widetilde{P}$, *where*

$$\sigma_{\widetilde{P}} = \begin{pmatrix} \sigma_{\widehat{P}} & \\ & I \end{pmatrix}, \tag{11.78}$$

I *being the* $(2^{n-1} - k) \times (2^{n-1} - k)$ *identity matrix, and the considerations above apply to give* Index $\widetilde{P}$, *hence* Index P.

If $k > 2^{n-1}$, *then there is a unique integer* ℓ *such that either*

$$\begin{aligned} &\ell > 0 \ \text{ and } \ \sigma_{\widehat{P}} \ \text{ is homotopic to } \begin{pmatrix} \sigma_{D_n^\ell} & \\ & I \end{pmatrix}, \\ &\ell = 0 \ \text{ and } \ \sigma_{\widehat{P}} \ \text{ is homotopic to a constant map,} \\ &\ell < 0 \ \text{ and } \ \sigma_{\widehat{P}} \ \text{ is homotopic to } \begin{pmatrix} \sigma_{(D_n^*)^{|\ell|}} & \\ & I \end{pmatrix}, \end{aligned} \tag{11.79}$$

I *being the* $(k - 2^{n-1}) \times (k - 2^{n-1})$ *identity matrix, and analogues of (11.76)–(11.77) hold.*

Remark: An integral formula for Index P is given in [Fed]; see also [Ho].

Also of use in index theory is the following complement to Proposition 11.11.

Proposition 11.14. *Given* $k \geq 1$,

$$j \notin \{1, 3, \dots, 2k-1\} \Longrightarrow \pi_j(U(k)) \ \text{ is finite.} \tag{11.80}$$

Thanks to Shrawan Kumar for mentioning this and for explaining the proof, which we now sketch. One ingredient is the result that

$$\pi_j(S^{2k-1}) \ \text{ is finite for all } \ j \neq 2k-1. \tag{11.81}$$

See [Spa], p. 515. The proof of (11.80) goes by induction on k. The case $k = 1$ is clear. The case $k = 2$ follows from (11.54), which reduces (11.80) with $k = 2$ to the assertion that $\pi_j(SU(2)) = \pi_j(S^3)$ is finite for $j \neq 3$. To do the inductive step, we assume that

$$j \neq \{1, 3, \dots, 2k-3\} \Longrightarrow \pi_j(U(k-1)) \ \text{ is finite,} \tag{11.82}$$

and aim to deduce (11.80). Another ingredient for this is the homotopy exact sequence for $U(k)/U(k-1) = S^{2k-1}$, which includes the segment

$$\pi_{j+1}(S^{2k-1}) \to \pi_j(U(k-1)) \to \pi_j(U(k)) \to \pi_j(S^{2k-1}), \tag{11.83}$$

cf. [Mil], p. 128. We tensor with $\mathbb{Q}$, denoting $\pi_j(X) \otimes \mathbb{Q}$ by $\pi_j^{\mathbb{Q}}(X)$.

$$\pi_{j+1}^{\mathbb{Q}}(S^{2k-1}) \to \pi_j^{\mathbb{Q}}(U(k-1)) \to \pi_j^{\mathbb{Q}}(U(k)) \to \pi_j^{\mathbb{Q}}(S^{2k-1}). \tag{11.84}$$

By (11.81),

$$\pi_j^{\mathbb{Q}}(S^{2k-1}) = 0 \ \text{ if } \ j \neq 2k-1. \tag{11.85}$$

Thus

$$j \notin \{2k-2, 2k-1\} \Longrightarrow \pi_j^{\mathbb{Q}}(U(k)) \approx \pi_j^{\mathbb{Q}}(U(k-1)). \tag{11.86}$$

With this, (11.82) leads to

$$\pi_j^{\mathbb{Q}}(U(k)) = 0 \ \text{ if } \ j \notin \{1, 3, \dots, 2k-3\} \ \text{ and } \ j \notin \{2k-2, 2k-1\}. \tag{11.87}$$

On the other hand, setting $j = 2k-2$ in (11.84) gives

$$\mathbb{Q} \to \pi_{2k-2}^{\mathbb{Q}}(U(k-1)) \to \pi_{2k-2}^{\mathbb{Q}}(U(k)) \to 0, \tag{11.88}$$

so

$$\pi_{2k-2}^{\mathbb{Q}}(U(k-1)) = 0 \Longrightarrow \pi_{2k-2}^{\mathbb{Q}}(U(k)) = 0, \tag{11.89}$$

giving (11.80).

See the exercises for an application of Proposition 11.14.

Remark: S. Kumar has also shown the author how further arguments yield, for $k \geq 2$,

$$\begin{aligned} \pi_{2k+1}(U(k)) = 0 \qquad & \text{if } k \text{ is odd}, \\ \mathbb{Z}/(2) \quad & \text{if } k \text{ is even}. \end{aligned} \tag{11.90}$$

In case $k = 2$, one has

$$\pi_5(U(2)) = \pi_5(SU(2)) = \pi_5(S^3) = \mathbb{Z}/(2). \tag{11.91}$$

See [Spa], p. 520.

Exercises

1. Give a Clifford algebra description of the operator D_n in (11.64).
2. Show that if $k \geq n$, there exists for each $\ell \in \mathbb{Z}$ a $k \times k$ elliptic system

$$P \in OP\mathcal{S}^m(\mathbb{R}^n) \text{ such that } \text{Index } P = \ell.$$

3. Suppose you know that

$$\pi_{2n-1}(U(k)) \text{ is a finite group.}$$

(By (11.73) this would require $k < n$.) Show that if $P \in OP\mathcal{S}^m(\mathbb{R}^n)$ is a $k \times k$ elliptic system,

$$\text{Index } P = 0.$$

(*Hint*. Index $P^j = j$ Index P.)

4. Using Exercise 3 and Proposition 11.14, show that if $P \in OP\mathcal{S}^m(\mathbb{R}^n)$ is a $k \times k$ elliptic system,

$$k < n \Longrightarrow \text{Index } P = 0.$$

References

[An] M. Anderson, Ricci curvature bounds and Einstein metrics on compact manifolds, *J. AMS* 2(1989), 455–490.

[ABP] M. Atiyah, R. Bott, and V. Patodi, On the heat equation and the index theorem, *Inventiones Math*. 19(1973), 279–330.

[ABS] M. Atiyah, R. Bott, and A. Shapiro, Clifford modules, *Topology* 3(1964), 3–38.

[AHS] M. Atiyah, N. Hitchen, and I. Singer, Self-duality in four-dimensional Riemannian geometry, *Proc. Royal Soc. London A* 362(1978), 425–461.

[AS] M. Atiyah and I. Singer, The index of elliptic operators I, *Ann. Math*. 87(1968), 484–530; III, *Ann. Math*. 87(1968), 546–664; IV, *Ann. Math*. 93(1971), 119–138; V, *Ann. Math*. 93(1971), 139–149.

[BDT] P. Baum, R. Douglas, and M. Taylor, Cycles and relative cycles in analytic K-homology, *J. Diff. Geom*. 30(1989), 761–804.

[Ber] F. Berezin, The Method of Second Quantization, *Academic*, New York, 1966.

[BGM] M. Berger, P. Gauduchon, and E. Mazet, *Le Spectre d'une Variété Riemannienne*, LNM #194, Springer, New York, 1971.

[BGV] N. Berline, E. Getzler, and M. Vergne, *Heat Kernels and Dirac Operators*, Springer, New York, 1992.

[BV] N. Berline and M. Vergne, A computation of the equivariant index of the Dirac operator, *Bull. Soc. Math. France* 113(1985), 305–345.

[BV2] N. Berline and M. Vergne, A proof of Bismut local index theorem for a family of Dirac operators, *Topology* 26(1987), 438–464.

[Bes] A. Besse, *Einstein Manifolds*, Springer, New York, 1987.

[Bi] J. Bismut, The Atiyah–Singer theorems for classical elliptic operators, a probabilistic approach, *J. Func. Anal*. 57(1984), 56–99.

[Bi2] J. Bismut, The Atiyah–Singer index theorem for families of Dirac operators: two heat equation proofs, *Invent. Math*. 83(1986), 91–151.

[BiC] J. Bismut and J. Cheeger, Families index for manifolds with boundary, superconnections, and cones, I, *J. Func. Anal.* 89(1990), 313–363; II, *J. Func. Anal.* 90(1990), 306–354.

[Bl] B. Blackadar, *K-theory for Operator Algebras*, Springer, New York, 1986.

[BTu] R. Bott and L. Tu, *Differential Forms in Algebraic Topology,* Springer, New York, 1982.

[Ch] J. Cheeger, Analytic torsion and the heat equation, *Ann. Math.* 109(1979), 259–322.

[Cher] S. S. Chern, A simple intrinsic proof of the Gauss–Bonnet formula for closed Riemannian manifolds., *Ann. Math.* 45(1944), 747–752.

[Chv] C. Chevalley, *Theory of Lie Groups*, Princeton University Press, Princeton, NJ, 1946.

[Con] A. Connes, *Noncommutative Geometry*, Academic, New York, 1994.

[CS] H. Cycon, R. Froese, W. Kirsch, and B. Simon, *Schrödinger Operators*, Springer, New York, 1987.

[D] S. Donaldson, The Seiberg–Witten equations and 4-manifold topology, *Bull. AMS* 33(1996), 45–70.

[Don] H. Donnelly, Local index theorems for families, *Mich. Math. J.* 35(1988), 11–20.

[EGH] T. Eguchi, P. Gilkey, and A. Hanson, Gravitation, Gauge Theories, and Differential Geometry, *Physics Reports*, Vol. 66, no. 6(1980).

[Fed] B. Fedosov, Direct proof of the formula for the index of an elliptic system in Euclidean space, *Funct. Anal. Appl.* 4(1970), 339–341.

[FU] D. Freed and K. Uhlenbeck, *Instantons and Four-Manifolds*, Springer, New York, 1984.

[Gt1] E. Getzler, Pseudodifferential operators on supermanifolds and the Atiyah–Singer index theorem, *Comm. Math. Phys*. 92(1983), 163–178.

[Gt2] E. Getzler, A short proof of the local Atiyah–Singer index theorem, *Topology* 25(1986), 111–117.

[Gil] P. Gilkey, *Invariance Theory, the Heat Equation, and the Atiyah–Singer Index Theorem*, CRC, Boca Raton, FL, 1995.

[Gu] R. Gunning, *Lectures on Riemann Surfaces*, Princeton University Press, Princeton, NJ, 1967.

[Har] R. Hartshorne, *Algebraic Geometry*, Springer, New York, 1977.

[Hi] N. Hicks, *Notes on Differential Geometry*, Van Nostrand, New York, 1965.

[Ho] L. Hörmander, The Weyl calculus of pseudo-differential operators, *Comm. Pure Appl. Math.* 32(1979), 359–443.

[Hir] F. Hirzebruch, *Topological Methods in Algebraic Geometry*, Springer, New York, 1966.

[KN] S. Kobayashi and N. Nomizu, *Foundations of Differential Geometry*, Interscience, New York, Vol.1, 1963; Vol. 2, 1969.

[Ko1] K. Kodaira, The theorem of Riemann–Roch on compact analytic surfaces, *Am. J. Math*. 73(1951), 813–875.

[Ko2] K. Kodaira, The theorem of Riemann–Roch for adjoint systems on 3-dimensional algebraic varieties, *Ann. Math.* 56(1952), 298–342.

[Kot] T. Kotake, An analytical proof of the classical Riemann–Roch theorem, *Proc. Symp. Pure Math.* 16(1970), 137–146.

[LM] H. B. Lawson and M. L. Michelson, *Spin Geometry*, Princeton University Press, Princeton, NJ, 1989.

[Lic] A. Lichnerowicz, Spineurs harmoniques, *C.R. Acad. Sci. Paris, Ser. A* 257(1963), 7–9.

[MS] H. McKean and I. Singer, Curvature and the eigenvalues of the Laplacian, *J. Diff. Geom*. 1(1967), 43–69.

[Mel] R. Melrose, *Index Theory on Manifolds with Corners*, A. K. Peters, Boston, 1994.

[Mil] J. Milnor, *Morse Theory*, Princeton University Press, Princeton, NJ, 1963.

[MiS] J. Milnor and J. Stasheff, *Characteristic Classes*, Princeton University Press, Princeton, NJ, 1974.

[Mor] J. Morgan, *The Seiberg–Witten Equations and Applications to the Topology of Smooth Four-Manifolds,* Princeton University Press, Princeton, NJ, 1996.

[Pal] R. Palais, ed., *Seminar on the Atiyah–Singer Index Theorem*, Princeton University Press, Princeton, NJ, 1963.

[Pt1] V. Patodi, Curvature and the eigenforms of the Laplace operator, *J. Diff. Geom*. 5(1971), 233–249.

[Pt2] V. Patodi, An analytic proof of the Riemann–Roch–Hirzebruch theorem for Kahler manifolds, *J. Diff. Geom*. 5(1971), 251–283.

[Poo] W. Poor, *Differential Geometric Structures*, McGraw-Hill, New York, 1981.

[Roe] J. Roe, *Elliptic Operators, Topology, and Asymptotic Methods*, Longman, New York, 1988.

[Rog] A. Rogers, A superspace path integral proof of the Gauss–Bonnet–Chern theorem, *J. Geom. Phys*. 4(1987), 417–437.

[Spa] E. Spanier, *Algebraic Topology*, McGraw-Hill, New York, 1966.

[Spi] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, Vol. 1–5, Publish or Perish, Berkeley, 1979.

[Stb] S. Sternberg, *Lectures on Differential Geometry*, Prentice-Hall, Englewood Cliffs, NJ, 1964.

[Tro] A. Tromba, A classical variational approach to Teichmuller theory, pp.155–185 in LNM #1365, Springer, New York, 1989.

[Wit] E. Witten, Constraints on supersymmetry breaking, *Nucl. Phys.* B202(1982), 253.

11

Brownian Motion and Potential Theory

Introduction

Diffusion can be understood on several levels. The study of diffusion on a macroscopic level, of a substance such as heat, involves the notion of the flux of the quantity. If $u(t,x)$ measures the intensity of the quantity that is diffusing, the flux J across the boundary of a region $\mathcal{O}$ in x-space satisfies the identity

$$\frac{\partial}{\partial t}\int_{\mathcal{O}} u(t,x)\,dV(x) = -\int_{\partial\mathcal{O}} \nu\cdot J\;dS(x), \tag{0.1}$$

as long as the substance is being neither created nor destroyed. By the divergence theorem, this implies

$$\frac{\partial u}{\partial t} = -\operatorname{div} J. \tag{0.2}$$

The mechanism of diffusion creates a flux in the direction from greater concentration to lesser concentration. In the simplest model, the quantitative relation specified is that the flux is proportional to the x-gradient of u:

$$J = -D \operatorname{grad} u, \tag{0.3}$$

with $D > 0$. Applying (0.2), we obtain for u the PDE

$$\frac{\partial u}{\partial t} = D\,\Delta u, \tag{0.4}$$

in case D is constant. In such a case we can make $D = 1$, by rescaling, and this PDE is the one usually called “the heat equation.”

Many real diffusions result from jitterings of microscopic or submicroscopic particles, in a fashion that appears random. This motivates a probabilistic attack on diffusion, including creating probabilistic tools to analyze the heat equation. This is the topic of the present chapter.

M.E. Taylor, *Partial Differential Equations II: Qualitative Studies of Linear Equations*, Applied Mathematical Sciences 116, DOI 10.1007/978-1-4419-7052-7_5,

In §1 we give a construction of Wiener measure on the space of paths in $\mathbb{R}^n$, governed by the hypothesis that a particle located at $x \in \mathbb{R}^n$ at time t_1 will have the probability $P(t, x, U)$ of being in an open set $U \subset \mathbb{R}^n$ at time $t_1 + t$, where

$$P(t, x, U) = \int_U p(t, x, y)\, dy, \tag{0.5}$$

and $p(t, x, y)$ is the fundamental solution to the heat equation. We prove that, with respect to Wiener measure, almost every path is continuous, and we establish a modulus of continuity. Our choice of $e^{t\Delta}$ rather than $e^{t\Delta/2}$ to define such probabilities differs from the most popular convention and leads to minor differences in various formulas. Of course, translation between the two conventions is quite easy.

In §2 we establish the Feynman–Kac formula, for the solution to

$$\frac{\partial u}{\partial t} = \Delta u + V(x)u, \tag{0.6}$$

in terms of an integral over path space. A limiting argument made in §3 gives us formulas for the solution to (0.4) on a bounded domain Ω, with Dirichlet boundary conditions. This also leads to formulas for solutions to

$$\Delta u = f \text{ on } \Omega, \quad u = 0 \text{ on } \partial\Omega, \tag{0.7}$$

and

$$\Delta u = 0 \text{ on } \Omega, \quad u = g \text{ on } \partial\Omega. \tag{0.8}$$

A different, and more natural, formula for the solution to (0.8) is derived in §5, after the development in §4 of a tool known as the "strong Markov property." In §6 we present a study of the Newtonian capacity of a compact set $K \subset \mathbb{R}^n$, in the case $n \geq 3$, which is related to the probability that a Brownian path starting outside K will hit K. We give Wiener's criterion for a point y in $\partial\Omega$ to be regular for the Dirichlet problem (0.8), in terms of the capacity of $K_r = \{z \in \partial\Omega : |z - y| \leq r\}$, as $r \to 0$, which has a natural probabilistic proof.

In §7 we introduce the notion of the stochastic integral, such as

$$\int_0^t f\big(s, \omega(s)\big)\, d\omega(s), \tag{0.9}$$

which is not straightforward since almost all Brownian paths fail to have locally bounded variation. We show how the solution to

$$\frac{\partial u}{\partial t} = \Delta u + Xu \tag{0.10}$$

can be given in terms of an integral over path space, whose integrand involves a stochastic integral, in case X is a first-order differential operator. The derivation of this formula, like the derivation of the Feynman–Kac formula in §2, uses a tool from functional analysis known as the Trotter product formula, which we establish in Appendix A at the end of this chapter.

In §8 we consider a more general sort of stochastic integral, needed to solve stochastic differential equations:

$$d\mathfrak{X} = b(t, \mathfrak{X})\, dt + \sigma(t, \mathfrak{X})\, d\omega, \tag{0.11}$$

which we study in §9. Via Ito's formulas, stochastic differential equations can be used to treat diffusion equations of the form

$$\frac{\partial u}{\partial t} = \sum A_{jk}(x)\, \partial_j \partial_k u + \sum b_j(x)\, \partial_j u + V(x)u, \tag{0.12}$$

in terms of path space integrals. We look at this in §10. Results there, specialized to (0.10), yield a formula with a different appearance than that derived in §7. The identity of these two formulas leads to a formula of Cameron-Martin-Girsanov, representing the "Jacobian determinant" of a certain nonlinear transformation of path space.

An important topic that we do not treat here is Malliavin's stochastic calculus of variations, introduced in [Mal], which has had numerous interesting applications to PDE. We refer the reader to [Stk2] and [B] for material on this, and further references.

1. Brownian motion and Wiener measure

One way to state the probabilistic connection with the heat equation

$$\frac{\partial u}{\partial t} = \Delta u \tag{1.1}$$

is in terms of the heat kernel, $p(t, x, y)$, satisfying

$$e^{t\Delta} f(x) = \int p(t, x, y) f(y)\, dV(y). \tag{1.2}$$

If Δ in (1.1) is the Friedrichs extension of the Laplacian on any Riemannian manifold M, the maximum principle implies

$$p(t, x, y) \geq 0. \tag{1.3}$$

In many cases, including all compact M and $M = \mathbb{R}^n$, we also have

$$\int p(t,x,y)\, dV(y) = 1. \tag{1.4}$$

Consequently, for each $x \in M$, $p(t,x,y)\, dV(y)$ defines a probability distribution, which we can interpret as giving the probability that a particle starting at the point x at time 0 will be in a given region in M at time t.

Restricting our attention to the case $M = \mathbb{R}^n$, we proceed to construct a probability measure, known as "Wiener measure," on the set of paths $\omega : [0,\infty) \to \mathbb{R}^n$, undergoing a random motion, sometimes called Brownian motion, described as follows. Given $t_1 < t_2$ and that $\omega(t_1) = x_1$, the probability density for the location of $\omega(t_2)$ is

$$e^{t\Delta}\delta_{x_1}(x) = p(t, x - x_1) = (4\pi t)^{-n/2} e^{-|x-x_1|^2/4t}, \quad t = t_2 - t_1. \tag{1.5}$$

The motion of a random path for $t_1 \le t \le t_2$ is supposed to be independent of its past history. Thus, given $0 < t_1 < t_2 < \cdots < t_k$, and given Borel sets $E_j \subset \mathbb{R}^n$, the probability that a path, starting at $x = 0$ at $t = 0$, lies in E_j at time t_j for each $j \in [1,k]$ is

$$\int_{E_1} \cdots \int_{E_k} p(t_k - t_{k-1}, x_k - x_{k-1}) \cdots p(t_1, x_1)\, dx_k \cdots dx_1. \tag{1.6}$$

It is not obvious that there is a countably additive measure characterized by these properties, and Wiener's result was a great achievement. The construction we give here is a slight modification of one in Appendix A of [Nel2].

Anticipating that Wiener measure is supported on the set of continuous paths, we will take a path to be characterized by its locations at all positive *rational* t. Thus, we consider the set of "paths"

$$\mathfrak{P} = \prod_{t \in \mathbb{Q}^+} \dot{\mathbb{R}}^n. \tag{1.7}$$

Here, $\dot{\mathbb{R}}^n$ is the one-point compactification of $\mathbb{R}^n$ (i.e., $\dot{\mathbb{R}}^n = \mathbb{R}^n \cup \{\infty\}$). Thus $\mathfrak{P}$ is a compact, metrizable space. We construct Wiener measure W as a positive Borel measure on $\mathfrak{P}$.

By the Riesz theorem, it suffices to construct a positive linear functional $E : C(\mathfrak{P}) \to \mathbb{R}$, on the space $C(\mathfrak{P})$ of real-valued, continuous functions on $\mathfrak{P}$, satisfying $E(1) = 1$. We first define E on the subspace $\mathcal{C}^\#$, consisting of continuous functions that depend on only finitely many of the factors in (1.7); that is, functions on $\mathfrak{P}$ of the form

$$\varphi(\omega) = F\big(\omega(t_1), \dots, \omega(t_k)\big), \quad t_1 < \cdots < t_k, \tag{1.8}$$

where F is continuous on $\prod_1^k \dot{\mathbb{R}}^n$, and $t_j \in \mathbb{Q}^+$. To be consistent with (1.6), we take

$$E(\varphi) = \int \cdots \int p(t_1, x_1) p(t_2 - t_1, x_2 - x_1) \cdots p(t_k - t_{k-1}, x_k - x_{k-1}) F(x_1, \dots, x_k)\, dx_k \cdots dx_1. \tag{1.9}$$

If $\varphi(\omega)$ in (1.8) actually depends only on $\omega(t_\nu)$ for some proper subset $\{t_\nu\}$ of $\{t_1, \dots, t_k\}$, there arises a formula for $E(\varphi)$ with a different appearance from (1.9). The fact that these two expressions are equal follows from the semigroup property of $e^{t\Delta}$. From this it follows that $E : \mathcal{C}^\# \to \mathbb{R}$ is well defined. It is also a positive linear functional, satisfying $E(1) = 1$.

Now, by the Stone-Weierstrass theorem, $\mathcal{C}^\#$ is dense in $C(\mathfrak{P})$. Since $E : \mathcal{C}^\# \to \mathbb{R}$ is a positive linear functional and $E(1) = 1$, it follows that E has a unique continuous extension to $C(\mathfrak{P})$, possessing these properties. Thus there is a unique probability measure W on $\mathfrak{P}$ such that

$$E(\varphi) = \int_{\mathfrak{P}} \varphi(\omega)\, dW(\omega). \tag{1.10}$$

This is the Wiener measure.

Proposition 1.1. *The set $\mathfrak{P}_0$ of paths from $\mathbb{Q}^+$ to $\mathbb{R}^n$, which are uniformly continuous on bounded subsets of $\mathbb{Q}^+$ (and which thus extend uniquely to continuous paths from $[0, \infty)$ to $\mathbb{R}^n$), is a Borel subset of $\mathfrak{P}$ with Wiener measure* 1.

For a set S, let $\operatorname{osc}_S(\omega)$ denote $\sup_{s,t \in S} |\omega(s) - \omega(t)|$. Set

$$E(a, b, \varepsilon) = \big\{\omega \in \mathfrak{P} : \operatorname{osc}_{[a,b]}(\omega) > 2\varepsilon\big\}; \tag{1.11}$$

here $[a, b]$ denotes $\{s \in \mathbb{Q}^+ : a \le s \le b\}$. Its complement is

$$E^c(a, b, \varepsilon) = \bigcap_{t,s \in [a,b]} \big\{\omega \in \mathfrak{P} : |\omega(s) - \omega(t)| \le 2\varepsilon\big\}, \tag{1.12}$$

which is closed in $\mathfrak{P}$. Below we will demonstrate the following estimate on the Wiener measure of $E(a, b, \varepsilon)$:

$$W\big(E(a, b, \varepsilon)\big) \le 2\rho\Big(\frac{\varepsilon}{2}, |b - a|\Big), \tag{1.13}$$

where

$$\rho(\varepsilon, \delta) = \sup_{t \le \delta} \int_{|x| > \varepsilon} p(t, x)\, dx, \tag{1.14}$$

with $p(t,x) = e^{t\Delta}\delta(x)$, as in (1.5). In fact, the sup is assumed at $t = \delta$, so

$$\rho(\varepsilon,\delta) = \int\limits_{|y|>\varepsilon/\sqrt{\delta}} p(1,y)\,dy = \psi_n\Big(\frac{\varepsilon}{\sqrt{\delta}}\Big), \tag{1.15}$$

where

$$\psi_n(r) = (4\pi)^{-n/2} \int\limits_{|y|>r} e^{-|y|^2/4}\,dy \le \alpha_n r^{n-1} e^{-r^2/4}, \tag{1.16}$$

as $r \to \infty$.

The relevance of the analysis of $E(a,b,\varepsilon)$ is that if we set

$$F(k,\varepsilon,\delta) = \big\{\omega \in \mathfrak{P} : \exists\, J \subset [0,k] \cap \mathbb{Q}^+, \ell(J) \le \delta,\ \operatorname{osc}_J(\omega) > 4\varepsilon\big\}, \tag{1.17}$$

where $\ell(J)$ is the length of the interval J, then

$$F(k,\varepsilon,\delta) = \bigcup \big\{E(a,b,2\varepsilon) : [a,b] \subset [0,k], |b-a| \le \delta\big\} \tag{1.18}$$

is an open set, and, via (1.13), we have

$$W\big(F(k,\varepsilon,\delta)\big) \le 2k\,\frac{\rho(\varepsilon,\delta)}{\delta}. \tag{1.19}$$

Furthermore, with $F^c(k,\varepsilon,\delta) = \mathfrak{P} \setminus F(k,\varepsilon,\delta)$,

$$\begin{aligned} \mathfrak{P}_0 &= \big\{\omega : \forall k < \infty, \forall \varepsilon > 0, \exists \delta > 0 \text{ such that } \omega \in F^c(k,\varepsilon,\delta)\big\} \\ &= \bigcap_k \bigcap_{\varepsilon=1/\nu} \bigcup_{\delta=1/\mu} F^c(k,\varepsilon,\delta) \end{aligned} \tag{1.20}$$

is a Borel set (in fact, an $\mathcal{F}_{\sigma\delta}$ set), and we can conclude that $W(\mathfrak{P}_0) = 1$ from (1.19), given the observation that, for any $\varepsilon > 0$,

$$\frac{\rho(\varepsilon,\delta)}{\delta} \longrightarrow 0, \quad \text{as } \delta \to 0, \tag{1.21}$$

which follows immediately from (1.15) and (1.16). Thus, to complete the proof of Proposition 1.1, it remains to establish the estimate (1.13).

Lemma 1.2. *Given $\varepsilon, \delta > 0$, take ν numbers $t_j \in \mathbb{Q}^+$, $0 \le t_1 < \cdots < t_\nu$, such that $t_\nu - t_1 \le \delta$. Let*

$$A = \big\{\omega \in \mathfrak{P} : |\omega(t_1) - \omega(t_j)| > \varepsilon, \textit{ for some } j = 1,\dots,\nu\big\}. \tag{1.22}$$

Then

$$W(A) \le 2\rho\big(\frac{\varepsilon}{2}, \delta\big). \tag{1.23}$$

Proof. Let

$$\begin{aligned} B &= \big\{\omega : |\omega(t_1) - \omega(t_\nu)| > \frac{\varepsilon}{2}\big\}, \\ C_j &= \big\{\omega : |\omega(t_j) - \omega(t_\nu)| > \frac{\varepsilon}{2}\big\}, \\ D_j &= \{\omega : |\omega(t_1) - \omega(t_j)| > \varepsilon \text{ and} \\ &\qquad\qquad |\omega(t_1) - \omega(t_k)| \le \varepsilon, \forall\, k \le j-1\}. \end{aligned} \tag{1.24}$$

Then $A \subset B \cup \bigcup_{j=1}^{\nu} \big(C_j \cap D_j\big)$, so

$$W(A) \le W(B) + \sum_{j=1}^{\nu} W\big(C_j \cap D_j\big). \tag{1.25}$$

Clearly, $W(B) \le \rho(\varepsilon/2, \delta)$. Furthermore, via (1.8)–(1.9), if we set

$$\begin{aligned} D\big(\omega(t_1), \dots, \omega(t_j)\big) &= 1, \ \text{ if } \omega \in D_j, \ 0 \text{ otherwise}, \\ C\big(\omega(t_j), \omega(t_\nu)\big) &= 1, \ \text{ if } \omega \in C_j, \ 0 \text{ otherwise}, \end{aligned}$$

we have $C(x_j, x_\nu) = C_1(x_j - x_\nu)$ and

$$\begin{aligned} &W(C_j \cap D_j) \\ &= \int \cdots \int D(x_1, \dots, x_j) C(x_j, x_\nu) p(t_1, x_1) p(t_2 - t_1, x_2 - x_1) \cdots \\ &\qquad\qquad p(t_j - t_{j-1}, x_j - x_{j-1}) p(t_\nu - t_j, x_\nu - x_j)\, dx_\nu dx_j \cdots dx_1 \\ &\le \rho\big(\frac{\varepsilon}{2}, \delta\big) \int \cdots \int D(x_1, \dots, x_j) p(t_1, x_1) \cdots p(t_j - t_{j-1}, x_j - x_{j-1}) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \cdot\, dx_j \cdots dx_1 \\ &\le \rho\big(\frac{\varepsilon}{2}, \delta\big) W(D_j), \end{aligned} \tag{1.26}$$

so

$$\sum_j W\big(C_j \cap D_j\big) \le \rho\big(\frac{\varepsilon}{2}, \delta\big), \tag{1.27}$$

since the D_j are mutually disjoint. This proves (1.23).

Let us note an intuitive approach to (1.26). Since D_j describes properties of $\omega(t)$ for $t \in [t_1, t_j]$ and C_j describes a property of $\omega(t_\nu) - \omega(t_j)$, these sets describe *independent* events, so $W(C_j \cap D_j) = W(C_j)W(D_j)$; meanwhile $W(C_j) \le \rho(\varepsilon/2, \delta)$.

We continue the demonstration of (1.13). Now, given such t_j as in the statement of Lemma 1.2, if we set

$$E = \{\omega : |\omega(t_j) - \omega(t_k)| > 2\varepsilon, \text{ for some } j,k \in [1,\nu]\}, \tag{1.28}$$

it follows that

$$W(E) \le 2\rho\Big(\frac{\varepsilon}{2}, \delta\Big), \tag{1.29}$$

since E is a subset of A, given by (1.22). Now, $E(a,b,\varepsilon)$, given by (1.11), is a countable increasing union of sets of the form (1.28), obtained, say, by letting $\{t_1, \dots, t_\nu\}$ consist of all $t \in [a,b]$ that are rational with denominator $\le K$, and taking $K \nearrow +\infty$. Thus we have (1.13), and the proof of Proposition 1.1 is complete.

We make the natural identification of paths $\omega \in \mathfrak{P}_0$ with continuous paths $\omega : [0,\infty) \to \mathbb{R}^n$. Note that a function φ on $\mathfrak{P}_0$ of the form (1.8), with $t_j \in \mathbb{R}^+$, not necessarily rational, is a pointwise limit on $\mathfrak{P}_0$ of functions in $C^\#$, as long as F is continuous on $\prod_1^k \dot{\mathbb{R}}^n$, and consequently such φ is measurable. Furthermore, (1.9) continues to hold, by the dominated convergence theorem.

An alternative approach to the construction of W would be to replace (1.7) by $\widetilde{\mathfrak{P}} = \prod\{\dot{\mathbb{R}}^n : t \in \mathbb{R}^+\}$. With the product topology, this is compact but not metrizable. The set of continuous paths is a Borel subset of $\widetilde{\mathfrak{P}}$, but not a Baire set, so some extra measure-theoretic considerations arise if one takes this route.

Looking more closely at the estimate (1.19) of the measure of the set $F(k,\varepsilon,\delta)$, defined by (1.17), we note that you can take $\varepsilon = K\sqrt{\delta \log 1/\delta}$, in which case

$$\rho(\varepsilon,\delta) = \psi_n\Big(K\sqrt{\log\frac{1}{\delta}}\Big) \le C_n\Big(\log\frac{1}{\delta}\Big)^{n/2-1}\delta^{K^2/4}. \tag{1.30}$$

Then we obtain the following refinement of Proposition 1.1.

Proposition 1.3. *For almost all $\omega \in \mathfrak{P}$, we have the modulus of continuity $8\sqrt{\delta \log 1/\delta}$, that is, given $0 \le s,t \le k < \infty$,*

$$\limsup_{|s-t|=\delta\to 0}\Big(\big|\omega(s)-\omega(t)\big| - 8\sqrt{\delta\log\frac{1}{\delta}}\Big) \le 0. \tag{1.31}$$

In fact, (1.30) gives $W(S_k) = 1$, where S_k is the set of paths satisfying (1.31), with 8 replaced by $8 + 1/k$, and then $\bigcap_k S_k$ is precisely the set of paths satisfying (1.31).

This result is not quite sharp; P. Levy showed that, for almost all $\omega \in \mathfrak{P}$, with $\mu(\delta) = 2\sqrt{\delta \log 1/\delta}$, $0 \le s, t \le k < \infty$,

$$\limsup_{|s-t|\to 0} \frac{|\omega(s) - \omega(t)|}{\mu(|s-t|)} = 1. \tag{1.32}$$

See [McK] for a proof. We also refer to [McK] for a proof of the result, due to Wiener, that almost all paths ω are nowhere differentiable.

By comparison with (1.31), note that if we define functions X_t on $\mathfrak{P}$, taking values in $\mathbb{R}^n$, by

$$X_t(\omega) = \omega(t), \tag{1.33}$$

then a simple application of (1.8)–(1.10) yields

$$\|X_t\|^2_{L^2(\mathfrak{P})} = \int |x|^2 p(t,x)\, dx = 2nt, \tag{1.34}$$

and more generally

$$\|X_t - X_s\|_{L^2(\mathfrak{P})} = \sqrt{2n}\, |s-t|^{1/2}. \tag{1.35}$$

Note that (1.35) depends on n, while (1.32) does not.

Via a simple translation of coordinates, we have a similar construction for the set of Brownian paths ω starting at a general point $x \in \mathbb{R}^\ell$, yielding the positive functional $E_x : C(\mathfrak{P}) \to \mathbb{R}$, and Wiener measure W_x, such that

$$E_x(\varphi) = \int_{\mathfrak{P}} \varphi(\omega)\, dW_x(\omega). \tag{1.36}$$

When $\varphi(\omega)$ is given by (1.8), $E_x(\varphi)$ has the form (1.9), with the function $p(t_1, x_1)$ replaced by $p(t_1, x_1 - x)$. To put it another way, $E_x(\varphi)$ has the form (1.9) with $F(x_1, \dots, x_k)$ replaced by $F(x_1 + x, \dots, x_k + x)$.

We will often use such notation as

$$E_x\big(f(\omega(t))\big)$$

instead of $\int_{\mathfrak{P}} f\big(X_t(\omega)\big) dW_x(\omega)$ or $E_x\big(f(X_t(\omega))\big)$.

The following simple observation is useful.

Proposition 1.4. *If $\varphi \in C(\mathfrak{P})$, then $E_x(\varphi)$ is continuous in x.*

Proof. Continuity for $\varphi \in \mathcal{C}^\#$, the set of functions of the form (1.8), is clear from (1.9) and its extension to $x \ne 0$ discussed above. Since $\mathcal{C}^\#$ is dense in $C(\mathfrak{P})$, the result follows easily.

Exercises

1. Given $a > 0$, define a transformation $D_a : \mathfrak{P}_0 \to \mathfrak{P}_0$ by
$$(D_a\omega)(t) = a\omega(a^{-2}t).$$
Show that D_a preserves the Wiener measure W. This transformation is called Brownian scaling.
2. Let $\widetilde{\mathfrak{P}}_0 = \{\omega \in \mathfrak{P}_0 : \lim_{s\to\infty} s^{-1}\omega(s) = 0\}$. Show that $W(\widetilde{\mathfrak{P}}_0) = 1$. Define a transformation $\rho : \widetilde{\mathfrak{P}}_0 \to \mathfrak{P}_0$ by
$$(\rho\omega)(t) = t\omega(t^{-1}),$$
for $t > 0$. Show that ρ preserves the Wiener measure W.
3. Given $a > 0$, define a transformation $R_a : \mathfrak{P}_0 \to \mathfrak{P}_0$ by
$$\begin{aligned}(R_a\omega)(t) = \ & \omega(t), && \text{for } 0 \le t \le a,\\ & 2\omega(a) - \omega(t), && \text{for } t \ge a.\end{aligned}$$
Show that R_a preserves the Wiener measure W.
4. Show that $L^p(\mathfrak{P}_0, dW_0)$ is separable, for $1 \le p < \infty$. (*Hint*: $\mathfrak{P}$ is a compact metric space. Show that $C(\mathfrak{P})$ is separable.)
5. If $0 \le a_1 < b_1 \le a_2 < b_2$, show that $X_{b_1} - X_{a_1}$ is orthogonal to $X_{b_2} - X_{a_2}$ in $L^2(\mathfrak{P}, dW_x, \mathbb{R}^n)$, where $X_t(\omega) = \omega(t)$, as in (1.33).
6. Verify the following identities (when $n = 1$):

$$E_x\Big(e^{\lambda(\omega(t)-\omega(s))}\Big) = e^{|t-s|\lambda^2}, \tag{1.37}$$

$$E_x\Big(\big[\omega(t) - \omega(s)\big]^{2k}\Big) = \frac{(2k)!}{k!}|t-s|^k, \tag{1.38}$$

$$E\big(\omega(s)\omega(t)\big) = 2\min(s,t). \tag{1.39}$$

7. Show that $e^{\lambda|\omega(t)|^2} \in L^2(\mathfrak{P}_0, dW_0)$ if and only if $\lambda < 1/8t$.

2. The Feynman–Kac formula

To illustrate the application of Wiener measure to PDE, we now derive a formula, known as the Feynman–Kac formula, for the solution operator $e^{t(\Delta-V)}$ to

$$\frac{\partial u}{\partial t} = \Delta u - Vu, \quad u(0) = f, \tag{2.1}$$

given f in an appropriate Banach space, such as $L^p(\mathbb{R}^n)$, $1 \le p < \infty$, or $f \in C_o(\mathbb{R}^n)$, the space of continuous functions on $\mathbb{R}^n$ vanishing at infinity. To start, we will assume V is bounded and continuous on $\mathbb{R}^n$. Following [Nel2], we will use the Trotter product formula

$$e^{t(\Delta-V)}f = \lim_{k\to\infty}\Big(e^{(t/k)\Delta}e^{-(t/k)V}\Big)^k f. \tag{2.2}$$

For any k, $\left(e^{(t/k)\Delta}e^{-(t/k)V}\right)^k f$ is expressed as a k-fold integral:

$$\begin{aligned}(2.3)\qquad &\left(e^{(t/k)\Delta}e^{-(t/k)V}\right)^k f(x)\\ &= \int\cdots\int f(x_k)e^{-(t/k)V(x_k)}\,p\Big(\frac{t}{k},x_k-x_{k-1}\Big)\,e^{(t/k)V(x_{k-1})}\cdots\\ &\qquad\cdot\, e^{-(t/k)V(x_1)}\,p\Big(\frac{t}{k},x-x_1\Big)\,dx_1\cdots dx_k.\end{aligned}$$

Comparison with (1.36) gives

$$(2.4)\qquad \left(e^{(t/k)\Delta}e^{-(t/k)V}\right)^k f(x) = E_x(\varphi_k),$$

where

$$(2.5)\qquad \varphi_k(\omega) = f\big(\omega(t)\big)\,e^{-S_k(\omega)},\quad S_k(\omega)=\frac{t}{k}\sum_{j=1}^{k} V\Big(\omega\Big(\frac{jt}{k}\Big)\Big).$$

We are ready to prove the Feynman–Kac formula.

Proposition 2.1. *If V is bounded and continuous on $\mathbb{R}^n$, and $f\in C(\mathbb{R}^n)$ vanishes at infinity, then, for all $x\in\mathbb{R}^n$,*

$$(2.6)\qquad e^{t(\Delta-V)}f(x) = E_x\left(f\big(\omega(t)\big)e^{-\int_0^t V(\omega(\tau))\,d\tau}\right).$$

Proof. We know that $e^{t(\Delta-V)}f$ is equal to the limit of (2.4) as $k\to\infty$, in the sup norm. Meanwhile, since almost all $\omega\in\mathfrak{P}$ are continuous paths, $S_k(\omega)\to\int_0^t V(\omega(\tau))d\tau$ boundedly and a.e. on $\mathfrak{P}$. Hence, for each $x\in\mathbb{R}^n$, the right side of (2.4) converges to the right side of (2.6). This finishes the proof.

Note that if V is real-valued and in $L^\infty(\mathbb{R}^n)$, then $e^{t(\Delta-V)}$ is defined on $L^\infty(\mathbb{R}^n)$, by duality from its action on $L^1(\mathbb{R}^n)$, and

$$(2.7)\qquad f_\nu\in C_0^\infty(\mathbb{R}^n),\ f_\nu\nearrow 1 \Longrightarrow e^{t(\Delta-V)}f_\nu\nearrow e^{t(\Delta-V)}1.$$

Thus, if V is real-valued, bounded, and continuous, then, for all $x\in\mathbb{R}^n$,

$$(2.8)\qquad e^{t(\Delta-V)}1(x) = E_x\left(e^{-\int_0^t V(\omega(\tau))d\tau}\right).$$

We can extend these identities to some larger classes of V. First we consider the nature of the right side of (2.6) for more general V.

Lemma 2.2. *Fix $t \in [0,\infty)$. If $V \in L^\infty(\mathbb{R}^n)$, then*

$$I_V(\omega) = \int_0^t V(\omega(\tau))\, d\tau \tag{2.9}$$

is well defined in $L^\infty(\mathfrak{P})$. If V_ν is a bounded sequence in $L^\infty(\mathbb{R}^n)$ and $V_\nu \to V$ in measure, then $I_{V_\nu} \to I_V$ boundedly and in measure on $\mathfrak{P}$. This is true for each measure $W_x,\ x \in \mathbb{R}^n$.

Proof. Here, L^∞ is the set of equivalence classes (mod a.e. equality) of bounded measurable functions, that is, elements of $\mathcal{L}^\infty(\mathbb{R}^n)$. Suppose $W \in \mathcal{L}^\infty(\mathbb{R}^n)$ is a pre-image of V. Then $\int_0^t W(\omega(\tau))\, d\tau = \iota_W(\omega)$ is defined and measurable, and $\|\iota_W\|_{\mathcal{L}^\infty(\mathfrak{P})} \le \|W\|_{\mathcal{L}^\infty(\mathbb{R}^n)} t$. If $W^\#$ is also a pre-image of V, then $W = W^\#$ almost everywhere on $\mathbb{R}^n$. Look at U, defined on $\mathfrak{P} \times \mathbb{R}^+$ by

$$U(\omega, s) = W(\omega(s)) - W^\#(\omega(s)).$$

This is measurable. Let $K \subset \mathbb{R}^n$ be the set where $W(x) \ne W^\#(x)$; this has measure 0. Now, for fixed s, the set of $\omega \in \mathfrak{P}$ such that $\omega(s) \in K$ has Wiener measure 0. By Fubini's theorem it follows that $U = 0$ a.e. on $\mathfrak{P} \times \mathbb{R}^+$, and hence, for almost all $\omega \in \mathfrak{P}$, $U(\omega, \cdot) = 0$ a.e. on $\mathbb{R}^+$. Thus $\int_0^t W^\#(\omega(\tau))\, d\tau = \int_0^t W(\omega(\tau))\, d\tau$ for a.e. $\omega \in \mathfrak{P}$, so I_V is well defined in $L^\infty(\mathfrak{P})$ for each $V \in L^\infty(\mathbb{R}^n)$. Clearly, $\|I_V\|_{L^\infty} \le \|V\|_{L^\infty} t$.

If $V_\nu \to V$ boundedly and in measure, in view of the previous argument we can assume without loss of generality that, upon passing to a subsequence, $V_\nu(x) \to V(x)$ for all x. Consider

$$U_\nu(\omega, s) = V(\omega(s)) - V_\nu(\omega(s)),$$

which is bounded in $L^\infty(\mathfrak{P}\times\mathbb{R}^+)$. This converges to 0 for each $(\omega, s) \in \mathfrak{P}\times\mathbb{R}^+$, so by Fubini's theorem again, $\int_0^t U_\nu(\omega, s)\, ds \to 0$ for a.e. ω. This completes the proof.

A similar argument yields the following.

Lemma 2.3. *If $V \in L^1_{loc}(\mathbb{R}^n)$ is bounded from below, then*

$$e_V(\omega) = e^{-\int_0^t V(\omega(\tau))\, d\tau} \tag{2.10}$$

is well defined in $L^\infty(\mathfrak{P})$. If $V_\nu \in L^1_{loc}(\mathbb{R}^n)$ are uniformly bounded below and $V_\nu \to V$ in L^1_{loc}, then $e_{V_\nu} \to e_V$ boundedly and in measure on $\mathfrak{P}$.

Thus, if $V \in L^1_{\text{loc}}(\mathbb{R}^n),\ V \ge -K > -\infty$, take bounded, continuous V_ν such that $V_\nu \ge -K$ and $V_\nu \to V$ in L^1_{loc}. We have $\|e^{t(\Delta - V_\nu)}\| \le e^{Kt}$ for all ν, where $\|\cdot\|$ can be the operator norm on $L^p(\mathbb{R}^n)$ or on $C_o(\mathbb{R}^n)$. Now, if we replace V

by V_ν in (2.6), then Lemma 2.3 implies that, for any $f \in C_0^\infty(\mathbb{R}^n)$, the right side converges, for each x, namely,

$$(2.11)\qquad E_x\Big(f\big(\omega(t)\big)e^{-\int_0^t V_\nu(\omega(\tau))\,d\tau}\Big) \longrightarrow P(t)f(x), \quad \text{as } \nu \to \infty.$$

Clearly $|P(t)f(x)| \le e^{Kt}E_x(|f|) \le e^{Kt}\|f\|_{L^\infty}$. Consequently, for each $x \in \mathbb{R}^n$, if $f \in C_0^\infty(\mathbb{R}^n)$,

$$(2.12)\qquad e^{t(\Delta - V_\nu)}f(x) \longrightarrow P(t)f(x) = E_x\Big(f\big(\omega(t)\big)e^{-\int_0^t V(\omega(\tau))\,d\tau}\Big).$$

It follows that $P(t) : C_0^\infty(\mathbb{R}^n) \to L^\infty(\mathbb{R}^n)$. Since

$$(2.13)\qquad \big|e^{t(\Delta - V_\nu)}f(x)\big| \le e^{Kt}e^{t\Delta}|f|(x),$$

we also have $P(t) : C_0^\infty(\mathbb{R}^n) \to L^1(\mathbb{R}^n)$. Furthermore, we can pass to the limit in the PDE $\partial u_\nu/\partial t = \Delta u_\nu - V_\nu u_\nu$ for $u_\nu = e^{t(\Delta - V_\nu)}f$, to obtain for $u(t) = P(t)f$ the PDE

$$(2.14)\qquad \frac{\partial u}{\partial t} = \Delta u - Vu, \quad u(0) = f.$$

If $\Delta - V$, with domain $\mathcal{D} = \mathcal{D}(\Delta) \cap \mathcal{D}(V)$, is self-adjoint, or has self-adjoint closure A, the uniqueness result of Proposition 9.11 in Appendix A, Functional Analysis, guarantees that $P(t)f = e^{tA}f$. For examples of such self-adjointness results on $\Delta - V$, see Chap. 8, §2, and the exercises following that section. Thus the identity (2.6) extends to such V, for example, to $V \in L^\infty(\mathbb{R}^n)$; so does the identity (2.8).

We can derive a similar formula for the solution operator $S(t,0)$ to

$$(2.15)\qquad \frac{\partial u}{\partial t} = \Delta u - V(t,x)u, \quad u(0) = f,$$

using the time-dependent Trotter product formula, Proposition A.5, and its consequence, Proposition A.6. Thus, we obtain

$$(2.16)\qquad S(t,0)f(x) = E_x\Big(f\big(\omega(t)\big)e^{-\int_0^t V(\tau,\omega(\tau))\,d\tau}\Big)$$

when $V(t) \in C\big([0,\infty), BC(\mathbb{R}^n)\big)$, $BC(\mathbb{R}^n)$ denoting the space of bounded continuous functions on $\mathbb{R}^n$. By arguments such as those used above, we can extend this identity to larger classes of functions $V(t)$.

Exercises

1. Given $\varepsilon > 0, \lambda \in \mathbb{R}$, compute the integral operator giving

$$(2.17)\qquad e^{t(\partial_x^2 - \varepsilon x^2 - \lambda x)}f(x).$$

(*Hint*: Use $\varepsilon x^2 + \lambda x = \varepsilon(x + \lambda/2\varepsilon)^2 - \lambda^2/4\varepsilon$ to reduce this to the problem of computing the integral operator giving

$$e^{t(\partial_x^2 - \varepsilon x^2)} g(x). \tag{2.18}$$

For this, see the material on the harmonic oscillator in §6 of Chap. 8, in particular, Mehler's formula.)

2. Obtain a formula for

$$E_x\Big(e^{-\varepsilon \int_0^t \omega(s)^2\, ds - \lambda \int_0^t \omega(s)\, ds}\Big) = e^{t(\partial_x^2 - \varepsilon x^2 - \lambda x)} 1(x), \tag{2.19}$$

in the case of one-dimensional Brownian motion. (*Hint*: Use the formula

$$\begin{aligned} e^{t(\partial_x^2 - \varepsilon x^2)} 1(x) &= a(t) e^{-b(t)x^2}, \\ a(t) = \big(\cosh 2\sqrt{\varepsilon} t\big)^{-1/2}, &\quad b(t) = \frac{1}{2}\sqrt{\varepsilon} \tanh 2\sqrt{\varepsilon} t, \end{aligned} \tag{2.20}$$

which follows from the formula for (2.18). Alternatively, verify (2.20) directly, examining the system of ODE

$$a'(t) = -2a(t)b(t), \quad b'(t) = \varepsilon - 4b(t)^2.)$$

3. Pass to the limit $\varepsilon \searrow 0$ in (2.19), to evaluate

$$E_x\Big(e^{-\lambda \int_0^t \omega(s) ds}\Big). \tag{2.21}$$

Note that the monotone convergence theorem applies.

Exercises 4 and 5 will investigate

$$\psi(\varepsilon) = W_0\left(\left\{\omega \in \mathfrak{P} : \int_0^a \omega(s)^2\, ds < \varepsilon\right\}\right) = P\left(\int_0^a \omega(s)^2\, ds < \varepsilon\right). \tag{2.22}$$

4. Using Exercise 2, show that, for all $\lambda > 0$,

$$\begin{aligned} \int_0^\infty \psi'(s) e^{-\lambda s}\, ds &= E_0\Big(e^{-\lambda \int_0^a \omega(s)^2\, ds}\Big) \\ &= \big(\cosh 2a\sqrt{\lambda}\big)^{-1/2} = \sqrt{2} e^{-a\sqrt{\lambda}} \big(1 + e^{-4a\sqrt{\lambda}}\big)^{-1/2}. \end{aligned} \tag{2.23}$$

Other derivations of (2.23) can be found in [CM] and [Lev].

5. The subordination identity, given as (5.22) in Chap. 3, implies

$$\int_0^\infty \varphi_a(s) e^{-\lambda s}\, ds = \sqrt{2} e^{-a\sqrt{\lambda}} \quad \text{if} \quad \varphi_a(s) = \frac{a}{\sqrt{2\pi}} s^{-3/2} e^{-a^2/4s}.$$

Deduce that

$$\psi'(s) = \varphi_a(s) - \frac{1}{2}\varphi_{5a}(s) + \frac{3}{8}\varphi_{9a}(s) - \cdots,$$

hence that

$$\begin{aligned} &\frac{d}{d\varepsilon} P\left(\int_0^a \omega(s)^2\, ds < \varepsilon\right) \\ &= \frac{a}{\sqrt{2\pi}} \varepsilon^{-3/2} \Big[e^{-a^2/4\varepsilon} - \frac{1}{2}\cdot 5 e^{-25a^2/4\varepsilon} + \frac{3}{8}\cdot 9 e^{-81a^2/4\varepsilon} - \cdots\Big]. \end{aligned} \tag{2.24}$$

Show that the terms in this alternating series have progressively decreasing magnitude provided $\varepsilon/a^2 \le 1/2$. (*Hint*: Use the power series

$$(1+y)^{-1/2} = 1 - \frac{1}{2}y + \frac{3}{8}y^2 - \cdots$$

with $y = e^{-4a\sqrt{\lambda}}$.)

6. Suppose now that $\omega(t)$ is Brownian motion in $\mathbb{R}^n$. Show that

$$E_0\left(e^{-\lambda \int_0^a |\omega(s)|^2\, ds}\right) = \left(\cosh 2a\sqrt{\lambda}\right)^{-n/2}.$$

Deduce that in the case $n = 2$,

$$\frac{d}{d\varepsilon} P\left(\int_0^a |\omega(s)|^2\, ds < \varepsilon\right) = \frac{2a}{\sqrt{\pi}} \varepsilon^{-3/2}\left[e^{-a^2/\varepsilon} - 3e^{-9a^2/\varepsilon} + 5e^{-25a^2/\varepsilon} - \cdots\right].$$

Show that the terms in this alternating series have progressively decreasing magnitude provided $\varepsilon \le 2a^2$.

3. The Dirichlet problem and diffusion on domains with boundary

We can use results of §2 to provide connections between Brownian motion and the Dirichlet boundary problem for the Laplace operator. We begin by extending Lemma 2.3 to situations where $V_\nu \nearrow V$, with $V(x)$ possibly equal to $+\infty$ on a big set. We have the following analogue of Lemma 2.3.

Lemma 3.1. *Let $V_\nu \in L^1_{loc}(\mathbb{R}^n)$, $-K \le V_\nu \nearrow V$, with possibly $V(x) = +\infty$ on a set of positive measure. Then $e_V(\omega)$, given by (2.10), is well defined in $L^\infty(\mathfrak{P})$, provided we set $e^{-\infty} = 0$, and $e_{V_\nu} \to e_V$ boundedly and in measure on Ω, for each t.*

Proof. This follows from the monotone convergence theorem.

Thus we again have convergence with bounds in (2.11)–(2.13). We will look at a special class of such sequences. Let $\Omega \subset \mathbb{R}^n$ be open, with smooth boundary (in fact, Lipschitz boundary will more than suffice), and set $E = \mathbb{R}^n \setminus \Omega$. Let $V_\nu \ge 0$ be continuous and bounded on $\mathbb{R}^n$ and satisfy

$$V_\nu = 0 \text{ on } \overline{\Omega}, \quad V_\nu \ge \nu \text{ on } E_\nu, \quad V_\nu \nearrow, \tag{3.1}$$

where E_ν is the set of points of distance $\ge 1/\nu$ from $\overline{\Omega}$. Given $f \in L^2(\mathbb{R}^n)$, $g \in L^2(\Omega)$, set $P_\Omega f = f|_\Omega \in L^2(\Omega)$, and define $E_\Omega g \in L^2(\mathbb{R}^n)$ to be $g(x)$ for $x \in \Omega$, 0 for $x \in E = \mathbb{R}^n \setminus \Omega$.

Proposition 3.2. *Under the hypotheses above, if $f \in L^2(\mathbb{R}^n)$, then*

$$e^{t(\Delta - V_\nu)} f \longrightarrow E_\Omega e^{t\Delta_\Omega}\left(P_\Omega f\right), \tag{3.2}$$

as $\nu \to \infty$, where Δ_Ω is the Laplace operator with Dirichlet boundary condition on Ω.

Proof. We will first show that, for any $\lambda > 0$,

$$(\lambda - \Delta + V_\nu)^{-1} f \to E_\Omega (\lambda - \Delta_\Omega)^{-1} P_\Omega f. \tag{3.3}$$

Indeed, denote the left side of (3.3) by u_ν, so $(\lambda - \Delta + V_\nu)u_\nu = f$. Taking the inner product with u_ν, we have

$$\lambda \|u_\nu\|^2_{L^2} + \|\nabla u_\nu\|^2_{L^2} + \int V_\nu |u_\nu|^2\, dx = (f, u_\nu) \le \frac{\lambda}{2}\|u_\nu\|^2_{L^2} + \frac{1}{2\lambda}\|f\|^2_{L^2}, \tag{3.4}$$

so

$$\frac{\lambda}{2}\|u_\nu\|^2_{L^2} + \|\nabla u_\nu\|^2_{L^2} + \int V_\nu |u_\nu|^2\, dx \le \frac{1}{2\lambda}\|f\|^2_{L^2}. \tag{3.5}$$

Thus, for fixed $\lambda > 0$, $\{u_\nu : \nu \in \mathbb{Z}^+\}$ is bounded in $H^1(\mathbb{R}^n)$, while $\int_{E_\nu} |u_\nu|^2\, dx \le C/\nu$. Thus $\{u_\nu\}$ has a weak limit point $u \in H^1(\mathbb{R}^n)$, and $u = 0$ on $\cup E_\nu$. The regularity hypothesized for $\partial\Omega$ implies $u \in H^1_0(\Omega)$. Clearly, $(\lambda - \Delta)u = f$ on Ω, so (3.3) follows, with weak convergence in $H^1(\mathbb{R}^n)$. But note that, parallel to (3.4),

$$\lambda \|u\|^2_{L^2} + \|\nabla u\|^2_{L^2} = (f, u) = \lim_{\nu\to\infty} (f, u_\nu),$$

so

$$\lambda \|u\|^2_{L^2} + \|\nabla u\|^2_{L^2} \ge \limsup_{\nu\to\infty} \lambda \|u_\nu\|^2_{L^2} + \|\nabla u_\nu\|^2_{L^2}. \tag{3.6}$$

Hence, in fact, we have H^1-norm convergence in (3.3), and a fortiori L^2-norm convergence.

Now consider the set $\mathcal{F}$ of real-valued $\varphi \in C_o([0,\infty))$ such that, for all $f \in L^2(\mathbb{R}^n)$,

$$\varphi(-\Delta + V_\nu) f \longrightarrow E_\Omega \varphi(-\Delta_\Omega) P_\Omega f, \ \text{ in } L^2(\mathbb{R}^n)\text{-norm}, \tag{3.7}$$

where $\varphi(H)$ is defined via the spectral theorem for a self-adjoint operator H. (Material on this functional calculus can be found in §1 of Chap. 8.) The analysis above shows that, for each $\lambda > 0$, $r_\lambda(s) = (\lambda + s)^{-1}$ belongs to $\mathcal{F}$. Since $P_\Omega E_\Omega$ is the identity on $L^2(\Omega)$, it is clear that $\mathcal{F}$ is an algebra; it is also easily seen to be a closed subset of $C_o([0,\infty))$. Since it contains r_λ for $\lambda > 0$, it separates points, so by the Stone-Weierstrass theorem all real-valued $\varphi \in C_o([0,\infty))$ belong to $\mathcal{F}$. This proves (3.2).

The version of (2.12) we have this time is the following.

Proposition 3.3. *Let $\Omega \subset \mathbb{R}^n$ be open, with smooth boundary, or more generally with the property that*

$$\{u \in H^1(\mathbb{R}^n) : supp\, u \subset \overline{\Omega}\} = H^1_0(\Omega).$$

Let $F \in C_0^\infty(\mathbb{R}^n),\ f = F|_\Omega$. Then, for all $x \in \overline{\Omega},\ t \geq 0$,

$$e^{t\Delta} f(x) = E_x\Big(f\big(\omega(t)\big)e^{-\int_0^t \ell_\Omega(\omega(\tau))\,d\tau}\Big). \tag{3.8}$$

On the left, $e^{t\Delta}$ is the solution operator to the heat equation on $\mathbb{R}^+ \times \Omega$ with Dirichlet boundary condition on $\partial\Omega$, and in the expression on the right

$$\ell_\Omega(x) = 0 \ \text{ on } \overline{\Omega}, \quad +\infty \ \text{ on } \mathbb{R}^n \setminus \overline{\Omega} = \overset{\circ}{E}. \tag{3.9}$$

Note that, for ω continuous,

$$e^{-\int_0^t \ell_\Omega(\omega(\tau))\,d\tau} = \psi_{\overline{\Omega}}(\omega, t) = \begin{matrix} 1 & \text{if } \omega([0,t]) \subset \overline{\Omega}, \\ 0 & \text{otherwise.} \end{matrix} \tag{3.10}$$

The second identity defines $\psi_{\overline{\Omega}}(\omega, t)$. Of course, for ω continuous, $\omega([0,t]) \subset \overline{\Omega}$ if and only if $\omega([0,t] \cap \mathbb{Q}) \subset \overline{\Omega}$.

We now extend Proposition 3.3 to the case where $\Omega \subset \mathbb{R}^n$ is open, with no regularity hypothesis on $\partial\Omega$. Choose a sequence Ω_j of open regions with smooth boundary, such that $\Omega_j \subset\subset \Omega_{j+1} \subset\subset \cdots,\ \bigcup_j \Omega_j = \Omega$. Let Δ_j denote the Laplace operator on Ω_j, with Dirichlet boundary condition, and let Δ denote that of Ω, also with Dirichlet boundary condition.

Lemma 3.4. *Given $f \in L^2(\Omega),\ t \geq 0$,*

$$e^{t\Delta} f = \lim_{j\to\infty} E_j e^{t\Delta_j} P_j f, \tag{3.11}$$

where $P_j f = f|_{\Omega_j}$ and, for $g \in L^2(\Omega_j),\ E_j g(x) = g(x)$ for $x \in \Omega_j,\ 0$ for $x \in \Omega \setminus \Omega_j$.

Proof. Methods of Chap. 5, §5, show that, for $\lambda > 0$,

$$E_j(\lambda - \Delta_j)^{-1} P_j f \to (\lambda - \Delta)^{-1} f \tag{3.12}$$

in L^2-norm, and then (3.11) follows from this, by reasoning used in the proof of Proposition 3.2.

Suppose $f \in C_0^\infty(\Omega_L)$. Then, for $j \geq L,\ E_j e^{t\Delta_j} f \to e^{t\Delta} f$ in L^2-norm, as we have just seen. Furthermore, local regularity implies

$$E_j e^{t\Delta_j} f \longrightarrow e^{t\Delta} f \quad \text{locally uniformly on } \Omega. \tag{3.13}$$

Thus, given such f, and any $x \in \Omega$ (hence $x \in \Omega_j$ for j large),

$$e^{t\Delta} f(x) = \lim_{j\to\infty} E_x \Big(f\big(\omega(t)\big) \psi_{\overline{\Omega}_j}(\omega, t) \Big). \tag{3.14}$$

Now, as $j \to \infty$,

$$\psi_{\overline{\Omega}_j}(\omega, t) \nearrow \psi_{\Omega}(\omega, t), \tag{3.15}$$

where we define

$$\begin{aligned} \psi_\Omega(\omega, t) = 1 \quad & \text{if } \omega([0,t]) \subset \Omega, \\ 0 \quad & \text{otherwise.} \end{aligned} \tag{3.16}$$

This yields the following:

Proposition 3.5. *For any open $\Omega \subset \mathbb{R}^n$, given $f \in C_0^\infty(\Omega),\ x \in \Omega$,*

$$e^{t\Delta} f(x) = E_x \big(f\big(\omega(t)\big) \psi_\Omega(\omega, t) \big). \tag{3.17}$$

In particular, if Ω has smooth boundary, one can use either $\psi_\Omega(\omega, t)$ or $\psi_{\overline{\Omega}}(\omega, t)$ in the formula for $e^{t\Delta} f(x)$. However, if $\partial\Omega$ is not smooth, it is $\psi_\Omega(\omega, t)$ that one must use.

It is useful to extend this result to more general f. Suppose $f_j \in C_0^\infty(\Omega),\ f \in L^2(\Omega)$, and $f_j(x) \searrow f(x)$ for each $x \in \Omega$. Then, for any $t > 0,\ e^{t\Delta} f_j \to e^{t\Delta} f$ in $L^2(\Omega) \cap C^\infty(\Omega)$, while, for each $x \in \Omega,\ E_x\big(f_j(\omega(t))\psi_\Omega(\omega, t)\big)$ converges $\searrow$ to the right side of (3.17), by the monotone convergence theorem. Hence (3.17) holds for all such f; denote this class by $\mathcal{L}(\Omega)$. Clearly, the characteristic function $\chi_K \in \mathcal{L}(\Omega)$ for each compact $K \subset \Omega$.

By the same reasoning, the class of functions in $L^2(\Omega)$ for which (3.17) holds is closed under forming monotone limits, either $f_j \nearrow f$ or $f_j \searrow f$, of sequences bounded in $L^2(\Omega)$. An argument used in Lemma 2.2 shows that modifying $f \in L^2(\Omega)$ on a set of measure zero does not change the right side of (3.17). If $S \subset \Omega$ is measurable, then

$$\chi_S(x) = \lim_{j\to\infty} \chi_{K_j}(x), \text{ a.e.,}$$

for an increasing sequence of compact sets $K_j \subset S$, so (3.17) holds for $f = \chi_S$. Thus it holds for finite linear combinations of such characteristic functions, and an easy limiting argument gives the following:

Proposition 3.6. *The identity* (3.17) *holds for all* $f \in L^2(\Omega)$ *when* $t > 0,\ x \in \Omega$.

Suppose now that Ω is bounded. Then, for $f \in L^p(\Omega),\ 1 \le p \le \infty$,

$$-\Delta^{-1} f = \int_0^\infty e^{t\Delta} f\, dt, \tag{3.18}$$

the integral being absolutely convergent in L^p-norm. If $f \in C_0^\infty(\Omega)$, we hence have, for each $x \in \Omega$,

$$-\Delta^{-1} f(x) = E_x\left(\int_0^\infty f\big(\omega(t)\big)\psi_\Omega(\omega, t)\, dt\right). \tag{3.19}$$

Furthermore, by an argument such as used to prove Proposition 3.6, this identity holds for almost every $x \in \Omega$, given $f \in L^2(\Omega)$, and for *every* x if $f_j \in C_0^\infty(\Omega)$ and $f_j(x) \nearrow f(x)$ for all x. In particular, for Ω bounded,

$$-\Delta^{-1} 1(x) = E_x\big(\vartheta_\Omega(\omega)\big), \quad x \in \Omega, \tag{3.20}$$

where, if ω is a continuous path starting inside Ω, we define

$$\begin{aligned} \vartheta_\Omega(\omega) = \int_0^\infty \psi_\Omega(\omega, t) dt &= \sup\{t : \omega([0, t]) \subset \Omega\} \\ &= \min\{t : \omega(t) \in \partial\Omega\}. \end{aligned} \tag{3.21}$$

In other words, $\vartheta_\Omega(\omega)$ is the first time $\omega(t)$ hits $\partial\Omega$; it is called the "first exit time." Since $\Delta^{-1} 1 \in C^\infty(\Omega)$, it is clear that the first exit time for a path starting at any $x \in \Omega$ is finite for W_x-almost every ω when Ω is bounded. (If ω starts at a point in $\partial\Omega$ or in $\mathbb{R}^n \setminus \overline{\Omega}$, set $\vartheta_\Omega(\omega) = 0$.) Note that we can write

$$-\Delta^{-1} f(x) = E_x\left(\int_0^{\vartheta_\Omega(\omega)} f\big(\omega(t)\big)\, dt\right). \tag{3.22}$$

If $\partial\Omega$ is smooth enough for Proposition 3.3 to hold, we have the formula (3.19), with $\psi_\Omega(\omega, t)$ replaced by $\psi_{\overline{\Omega}}(\omega, t)$, valid for all $x \in \overline{\Omega}$. In particular, for smooth bounded Ω,

$$-\Delta^{-1} 1(x) = E_x\big(\vartheta_{\overline{\Omega}}(\omega)\big), \quad x \in \overline{\Omega}, \tag{3.23}$$

where we define

$$\vartheta_{\overline{\Omega}}(\omega) = \inf\big\{t : \omega(t) \in \mathbb{R}^n \setminus \overline{\Omega}\big\} = \max\big\{t : \omega\big([0, t]\big) \subset \overline{\Omega}\big\}. \tag{3.24}$$

(If $\omega(0) \in \mathbb{R}^n \setminus \overline{\Omega}$, set $\vartheta_{\overline{\Omega}}(\omega) = 0$.) Comparing this with (3.20), noting that $\vartheta_{\overline{\Omega}}(\omega) \ge \vartheta_\Omega(\omega)$, we have the next result.

Proposition 3.7. *If Ω is bounded and $\partial\Omega$ is smooth enough for Proposition 3.3 to hold, then*

$$x \in \Omega \Longrightarrow \vartheta_\Omega(\omega) = \vartheta_{\overline{\Omega}}(\omega), \ \textit{for } W_x\textit{- almost every } \omega, \tag{3.25}$$

and

$$x \in \partial\Omega \Longrightarrow \vartheta_{\overline{\Omega}}(\omega) = 0, \ \textit{for } W_x\textit{- almost every } \omega. \tag{3.26}$$

The probabilistic interpretation of this result is that, for any $x \in \overline{\Omega}$, once a Brownian path ω starting at x hits $\partial\Omega$, it penetrates into the interior of $\mathbb{R}^n \setminus \overline{\Omega}$ within an arbitrarily short time, for W_x-almost all ω. From here one can show that, given $x \in \partial\Omega$, W_x-a.e. path ω spends a positive amount of time in both Ω and $\mathbb{R}^n \setminus \overline{\Omega}$, on any time interval $[0, s_0]$, for any $s_0 > 0$, however small. This is one manifestation of how wiggly Brownian paths are.

Note that taking $f = 1$ in (3.17) gives, for all $x \in \Omega$, any open set in $\mathbb{R}^n$,

$$e^{t\Delta}1(x) = W_x\big(\{\omega : \vartheta_\Omega(\omega) > t\}\big), \quad x \in \Omega, \tag{3.27}$$

the right side being the probability that a path starting in Ω at x has first exit time $> t$. Meanwhile, if $\partial\Omega$ is regular enough for Proposition 3.3 to hold, then

$$e^{t\Delta}1(x) = W_x\left(\{\omega : \vartheta_{\overline{\Omega}}(\omega) > t\}\right). \tag{3.28}$$

Comparing these identities extends Proposition 3.7 to unbounded Ω.

The following is an interesting consequence of (3.28).

Proposition 3.8. *For one-dimensional Brownian motion, starting at the origin, given $t > 0,\ \lambda > 0$,*

$$W\big(\{\omega : \sup_{0\le s\le t} \omega(s) \ge \lambda\}\big) = 2W\big(\{\omega : \omega(t) \ge \lambda\}\big). \tag{3.29}$$

Proof. The right side is $\int_\lambda^\infty p(t,x)\,dx$, with $p(t,x) = e^{td^2/dx^2}\delta(x) = (4\pi t)^{-1/2}e^{-x^2/4t}$, the $n = 1$ case of (1.5). The left side of (3.29) is the same as $W\big(\{\omega : \vartheta_{(-\infty,\lambda)}(\omega) < t\}\big)$, which by (3.28) is equal to $1 - e^{tL}1(0)$ if $L = d^2/dx^2$ on $(-\infty, \lambda)$, with Dirichlet boundary condition at $x = \lambda$. By the method of images we have, for $x < \lambda$,

$$e^{tL}1(x) = \int p(t,y)H(\lambda - x + y)\,dy,$$

where $H(s) = 1$ for $s > 0,\ -1$ for $s < 0$. From this, the identity (3.29) readily follows.

We next derive an expression for the Poisson integral formula, for the solution $\mathrm{PI}\, f = u$ to

$$\Delta u = 0 \text{ on } \Omega, \quad u|_{\partial\Omega} = f. \tag{3.30}$$

This can be expressed in terms of the integral kernel $G(x, y)$ of Δ^{-1} if $\partial\Omega$ is smooth. In fact, an application of Green's formula gives

$$\mathrm{PI}\, f(x) = \int_{\partial\Omega} f(y) \frac{\partial}{\partial \nu_y} G(x, y)\, dS(y), \tag{3.31}$$

where ν_y is the outward normal to $\partial\Omega$ at y. A closely related result is the following. Let f be defined and continuous on a neighborhood of $\partial\Omega$. Given small $\delta > 0$, set

$$S_\delta = \{x \in \Omega : \mathrm{dist}(x, \partial\Omega) < \delta\}, \tag{3.32}$$

and define u_δ by

$$\begin{gathered} \Delta u_\delta = \delta^{-2} f_\delta \text{ on } \Omega, \quad u_\delta = 0 \text{ on } \partial\Omega, \\ f_\delta = f \text{ on } S_\delta,\ 0 \text{ on } \Omega \setminus S_\delta. \end{gathered} \tag{3.33}$$

Lemma 3.9. *If $\partial\Omega$ is smooth, then, locally uniformly on Ω,*

$$\lim_{\delta\to 0} u_\delta = -\frac{1}{2} PI\, f. \tag{3.34}$$

Proof. If ν is the outward normal, we have

$$\begin{aligned} u_\delta(x) &= \delta^{-2} \int_{\partial\Omega} \int_0^\delta G(x, y - s\nu) f(y)\, ds\, dS(y) + o(1) \\ &= -\delta^{-2} \int_{\partial\Omega} f(y) \frac{\partial}{\partial \nu_y} G(x, y) \Big(\int_0^\delta s\, ds \Big)\, dS(y) + o(1) \\ &= -\frac{1}{2} \int_{\partial\Omega} f(y) \frac{\partial}{\partial \nu_y} G(x, y)\, dS(y) + o(1), \end{aligned} \tag{3.35}$$

so the result follows from (3.31).

Comparing this with (3.22), we conclude that when $\partial\Omega$ is smooth,

$$\mathrm{PI}\, f(x) = \lim_{\delta\searrow 0} \frac{2}{\delta^2} E_x \left(\int_0^{\vartheta_\Omega(\omega)} f\big(\omega(t)\big) \iota_{S_\delta}(\omega, t)\, dt \right), \tag{3.36}$$

where S_δ is as in (3.32), and, for $S \subset \Omega$,

$$\iota_S(\omega,t) = 1 \quad \text{if } \omega(t) \in S, \qquad 0 \quad \text{otherwise.} \tag{3.37}$$

We will discuss further formulas for PI f in §5.

Exercises

1. Looking at the definitions, check that $\psi_{\overline{\Omega}}(\omega,t)$ and $\vartheta_{\overline{\Omega}}(\omega)$ are measurable when $\Omega \subset \mathbb{R}^n$ is open with smooth boundary and that $\psi_\Omega(\omega,t)$ and $\vartheta_\Omega(\omega)$ are measurable, for general open $\Omega \subset \mathbb{R}^n$.
2. Show that if $x \in \mathcal{O}$, then

$$\{\omega \in \mathfrak{P}_0 : \vartheta_{\overline{\mathcal{O}}}(\omega) < t_0\} = \bigcup_{s\in[0,t_0)\cap\mathbb{Q}} \{\omega \in \mathfrak{P}_0 : \omega(s) \in \mathbb{R}^n \setminus \overline{\mathcal{O}}\}. \tag{3.38}$$

3. For any finite set $\mathcal{S} = \{s_1,\dots,s_K\} \subset \mathbb{Q}^+,\ N \in \mathbb{Z}^+$, set

$$F_{N,\mathcal{S}}(\omega) = \Phi_{N,\mathcal{S}}\big(\omega(s_1)\dots\omega(s_K)\big),$$
$$\Phi_{N,\mathcal{S}}(x_1,\dots,x_K) = \min\Big(N, \min\{s_\nu : x_\nu \in \mathbb{R}^n \setminus \overline{\mathcal{O}}\}\Big).$$

Show that, for any continuous path ω,

$$\vartheta_{\overline{\mathcal{O}}}(\omega) = \sup_N \inf_{\mathcal{S}} F_{N,\mathcal{S}}(\omega). \tag{3.39}$$

Note that the collection of such sets $\mathcal{S}$ is countable.
4. If $\mathfrak{P}_{\overline{\mathcal{O}},N} = \{\omega \in \mathfrak{P}_0 : \vartheta_{\overline{\mathcal{O}}}(\omega) \le N\}$ and $\mathcal{O}$ is bounded, show that

$$W_x\big(\mathfrak{P}_0 \setminus \mathfrak{P}_{\overline{\mathcal{O}},N}\big) \le CN^{-1}. \tag{3.40}$$

(*Hint*: Use (3.23).)
5. If $\omega \in \mathfrak{P}_{\overline{\mathcal{O}},N}$, show that

$$\vartheta_{\overline{\mathcal{O}}}(\omega) = \lim_{\nu\to\infty} \vartheta_{\nu,N}(\omega), \tag{3.41}$$

where

$$\vartheta_{\nu,N}(\omega) = \min\Big(N, \inf\big\{s \in 2^{-\nu}\mathbb{Z}^+ : \omega(s) \notin \overline{\mathcal{O}}\big\}\Big).$$

Write $\vartheta_{\nu,N}(\omega) = \Phi_{\nu,N}\big(\omega(s_1),\dots,\omega(s_L)\big)$, where $\Phi_{\nu,N}$ has a form similar to $\Phi_{N,\mathcal{S}}$ in Exercise 3.
6. For one-dimensional Brownian motion, establish the following, known as Kolmogorov's inequality:

$$W\big(\{\omega : \sup_{0\le s\le t} |\omega(s)| \ge \varepsilon\}\big) \le \frac{2t}{\varepsilon^2}, \quad \varepsilon > 0. \tag{3.42}$$

(*Hint*: Write the left side of (3.42) as $W\big(\{\omega : \vartheta_{(-\varepsilon,\varepsilon)}(\omega) < t\}\big)$, and relate this to the heat equation on $\Omega = [-\varepsilon,\varepsilon]$, with Dirichlet boundary condition, in a fashion parallel to the proof of Proposition 3.8.)

Note that this estimate is nontrivial only for $t < \varepsilon^2/2$. By Brownian scaling, it suffices to consider the case $\varepsilon = 1$. Compare the estimate

$$W\left(\left\{\omega : \sup_{0\le s\le t} |\omega(s)| \ge \varepsilon\right\}\right) \le 4\int_{\varepsilon}^{\infty} p(t,x)\, dx,$$

which follows from (3.29).

7. Given $\Omega \subset \mathbb{R}^n$ open, with complement K, and Δ with Dirichlet boundary condition on $\partial\Omega$, show that, for $x \in \Omega$,

$$W_x\big(\{\omega : \vartheta_\Omega(\omega) = \infty\}\big) = H_K(x), \tag{3.43}$$

where

$$H_K(t,x) = e^{t\Delta}1(x) \searrow H_K(x), \quad \text{as } t \nearrow \infty. \tag{3.44}$$

8. Suppose that $K = \mathbb{R}^n \setminus \Omega$ is compact, and suppose there exists $\widetilde{H}_K(x) \in C(\overline{\Omega})$, harmonic on Ω, such that $\widetilde{H}_K = 0$ on ∂K and $\widetilde{H}_K(x) \to 1$, as $|x| \to \infty$. Show that

$$H_K(t,x) \ge \widetilde{H}_K(x), \quad \text{for all } t < \infty.$$

(*Hint*: Show that $\Delta H_K(t,x) \le 0$ and that $H_K(t,x) \to 1$ as $|x| \to \infty$, and use the maximum principle.)
Deduce that if such $\widetilde{H}_K(x)$ exists, then $W_x\big(\{\omega : \vartheta_\Omega(\omega) = \infty\}\big) > 0$.

9. In the context of Exercise 8, show that if such $\widetilde{H}_K$ exists, then in fact

$$H_K(x) = \widetilde{H}_K(x), \quad \text{for all } x \in \overline{\Omega}. \tag{3.45}$$

(*Hint*: Show that H_K must be harmonic in Ω and that $\limsup_{|x|\to\infty} H_K(x) \le 1$.)
By explicit construction, produce such a function on $\mathbb{R}^n \setminus B$ when B is a ball of radius $a > 0$, provided $n \ge 3$.

10. Using Exercises 7–9, show that when $n \ge 3$,

$$W_x\big(\{\omega : |\omega(t)| \to \infty \text{ as } t \to \infty\}\big) = 1. \tag{3.46}$$

(*Hint*: Given $R > 0$, the probability that $|\omega(t)| \ge R$ for some t is 1. If $R >> a$, and $|\omega(t_0)| \ge R$, show that the probability that $|\omega(t_0 + s)| \le a$ for some $s > 0$ is small, using (3.43) for $K = B_a = \{x : |x| \le a\}$.) To restate (3.46), one says that Brownian motion in $\mathbb{R}^n$ is "non-recurrent," for $n \ge 3$.

11. If $n \le 2$ and $K = B_a$, show that $H_K(t,x) = 0$ in (3.44), and hence the probability defined in (3.46) is zero. Deduce that if $n \le 2$ and $U \subset \mathbb{R}^n$ is a nonempty open set, almost every Brownian path ω visits U at an infinite sequence of times $t_\nu \to \infty$. One says that Brownian motion in $\mathbb{R}^n$ is "recurrent," for $n \le 2$.

12. Relate the formula (3.34) for PI f to representations of PI f by double-layer potentials, discussed in §11 of Chap. 7. Where is the second layer coming from?

13. If Ω is a bounded domain with smooth boundary, show that (3.36) remains true with S_δ replaced by

$$\widetilde{S}_\delta = \{x \in \mathbb{R}^n \setminus \overline{\Omega} : \text{dist}(x, \partial\Omega) < \delta\}$$

and with $\vartheta_\Omega(\omega)$ replaced by $\vartheta_{\Omega_\delta}(\omega)$, where $\Omega_\delta = \overline{\Omega} \cup \widetilde{S}_\delta$. (*Hint*: Start by showing that $\widetilde{u}_\delta(x) \to -(1/2)\mathrm{PI}\, f(x)$, for $x \in \Omega$, where, in place of (3.33),

$$\Delta \widetilde{u}_\delta = \delta^{-2} f_\delta \text{ on } \Omega_\delta, \quad \widetilde{u}_\delta = 0 \text{ on } \partial\Omega_\delta,$$

with $f_\delta = f$ on $\widetilde{S}_\delta$, 0 on $\overline{\Omega}$.

4. Martingales, stopping times, and the strong Markov property

Given $t \in [0, \infty)$, let $\mathfrak{B}_t$ be the σ-field of subsets of $\mathfrak{P}_0$ generated by sets of the form

$$\{\omega \in \mathfrak{P}_0 : \omega(s) \in E\}, \tag{4.1}$$

where $s \in [0, t]$ and E is a Borel subset of $\mathbb{R}^n$. One easily sees that each element of $\mathfrak{B}_t$ is a Borel set in $\mathfrak{P}$. As t increases, $\mathfrak{B}_t$ is an increasing family of σ-fields, each consisting of sets which are W_x-measurable, for all $x \in \mathbb{R}^n$. Set $\mathfrak{B}_\infty = \sigma\big(\bigcup_{t<\infty} \mathfrak{B}_t\big)$.

Given $f \in L^1(\mathfrak{P}_0, \mathfrak{B}_\infty, dW_x)$, we can define the conditional expectation

$$E_x\big(f\,\big|\mathfrak{B}_t\big), \tag{4.2}$$

a function measurable with respect to $\mathfrak{B}_t$, as follows. Denote by $W_{x,t}$ the restriction of the Wiener measure W_x to the σ-field $\mathfrak{B}_t$. Then

$$\lambda(S) = \int_S f(\omega)\, dW_x(\omega) = E_x(f\chi_S) \tag{4.3}$$

defines a countably additive set function on $\mathfrak{B}_t$, which is absolutely continuous with respect to $W_{x,t}$, so by the Radon-Nikodym theorem there exists a $\mathfrak{B}_t$-measurable function Φ_t, uniquely defined $W_{x,t}$-almost everywhere, such that (4.3) is equal to $\int_S \Phi_t(\omega)\, dW_{x,t}(\omega)$, for all $S \in \mathfrak{B}_t$. This function is $E_x(f|\mathfrak{B}_t)$. Clearly,

$$f \in L^1(\mathfrak{P}_0, \mathfrak{B}_\infty, dW_x) \Longrightarrow E_x\big(f\,\big|\mathfrak{B}_t\big) \in L^1(\mathfrak{P}_0, \mathfrak{B}_t, dW_{x,t}). \tag{4.4}$$

This construction of conditional expectation generalizes in the obvious way to any situation where f is measurable with respect to some σ-field $\mathfrak{F}$, and is L^1 with respect to a given probability measure on $\mathfrak{F}$, and one wants to define the conditional expectation $E(f|\mathfrak{F}_0)$ with respect to some sub-σ-field $\mathfrak{F}_0$ of $\mathfrak{F}$.

Note that we can regard $L^1(\mathfrak{P}_0, \mathfrak{B}_t, dW_{x,t})$ naturally as a closed linear subspace of $L^1(\mathfrak{P}_0, \mathfrak{B}_\infty, dW_x)$. Then the map $f \mapsto E_x(f|\mathfrak{B}_t)$ is a projection. Similarly, we have

$$f \in L^2(\mathfrak{P}_0, \mathfrak{B}_\infty, dW_x) \Longrightarrow E_x\big(f\,\big|\mathfrak{B}_t\big) \in L^2(\mathfrak{P}_0, \mathfrak{B}_t, dW_{x,t}),$$

and in this case $E_x(f|\mathfrak{B}_t)$ is simply the orthogonal projection of f onto $L^2(\mathfrak{P}_0, \mathfrak{B}_t, dW_{x,t})$, regarded as a linear subspace of $L^2(\mathfrak{P}_0, \mathfrak{B}_\infty, dW_x)$. The reader might think of this in light of von Neumann's proof of the Radon-Nikodym theorem, which is sketched in the exercises for §2 of Appendix A.

The following is a statement that Brownian motion possesses the *Markov property*.

Proposition 4.1. *Given* $s, t > 0,\ f \in C(\dot{\mathbb{R}}^n)$,

$$E_x\big(f(\omega(t+s))\big|\mathfrak{B}_s\big) = E_{\omega(s)}\big(f(\omega(t))\big), \quad \textit{for } W_x\textit{-almost all } \omega. \tag{4.5}$$

Proof. The right side of (4.5) is $\mathfrak{B}_s$-measurable, so the identity is equivalent to the statement that

$$\int_S f(\omega(t+s))\, dW_x(\omega) = \int_S \Big(\int f(\widetilde{\omega}(t))\, dW_{\omega(s)}(\widetilde{\omega})\Big)\, dW_x(\omega), \tag{4.6}$$

for all $S \in \mathfrak{B}_s$. It suffices to verify (4.6) for all S of the form

$$S = \{\omega \in \mathfrak{P}_0 : \omega(t_1) \in E_1, \dots, \omega(t_K) \in E_K\},$$

given $t_j \in [0, s],\ E_j$ Borel sets in $\mathbb{R}^n$. For such S, (4.6) follows directly from the characterization of the Wiener integral given in §1, that is, from (1.6)–(1.9) in the case $x = 0$, together with the identity

$$\int f(\widetilde{\omega}(t))\, dW_y(\widetilde{\omega}) = E\big(f(y + \omega(t))\big) \tag{4.7}$$

used to define (1.36).

We can easily extend (4.5) to

$$E_x\big(F(\omega(s+t_1), \dots, \omega(s+t_k))\big|\mathfrak{B}_s\big) = E_{\omega(s)}\big(F(\omega(t_1), \dots, \omega(t_k))\big), \tag{4.8}$$

for W_x-almost all ω, given $t_1, \dots, t_k > 0$, and F continuous on $\prod_1^k \dot{\mathbb{R}}^n$, as in (1.8). Also, standard limiting arguments allow us to enlarge the class of functions F for which this works. We then get the following more definitive statement of the Markov property.

Proposition 4.2. *For* $s > 0$, *define the map*

$$\sigma_s : \mathfrak{P}_0 \longrightarrow \mathfrak{P}_0, \quad (\sigma_s\omega)(t) = \omega(t+s). \tag{4.9}$$

Then, given φ bounded and $\mathfrak{B}_\infty$-measurable, we have

$$E_x\big(\varphi \circ \sigma_s \big| \mathfrak{B}_s\big) = E_{\omega(s)}(\varphi), \ \textit{for } W_x\textit{-almost all } \omega. \tag{4.10}$$

The following is a useful restatement of Proposition 4.2.

Corollary 4.3. *For $s > 0$, define the map*

$$\vartheta_s : \mathfrak{P}_0 \to \mathfrak{P}_0, \quad (\vartheta_s \omega)(t) = \omega(t+s) - \omega(s). \tag{4.11}$$

Then, given $\varphi \in L^1(\mathfrak{P}_0, dW_0)$, we have

$$E_x\big(\varphi \circ \vartheta_s \big| \mathfrak{B}_s\big) = E_0(\varphi). \tag{4.12}$$

In particular,

$$E_x\big(f\big(\vartheta_s \omega(t)\big) \big| \mathfrak{B}_s\big) = E_0\big(f(\omega(t))\big). \tag{4.13}$$

Note that (4.12) implies ϑ_s is measure preserving, in the sense that

$$W_x\big(\vartheta_s^{-1}(S)\big) = W_0(S), \tag{4.14}$$

for W_0-measurable sets S. The map ϑ_s is not one-to-one, of course, but it is *onto* the set of paths in $\mathfrak{P}_0$ satisfying $\omega(0) = 0$.

The Markov property also implies certain independence properties. A function $\varphi \in L^1(\mathfrak{P}_0, dW_x)$ is said to be independent of the σ-algebra $\mathfrak{B}_t$ provided that, for all continuous F,

$$\int_S F\big(\varphi(\omega)\big)\, dW_x(\omega) = W_x(S) E_x(F \circ \varphi), \quad \forall\, S \in \mathfrak{B}_t. \tag{4.15}$$

An equivalent condition is

$$E_x\big(F(\varphi)\psi\big) = E_x\big(F(\varphi)\big) E_x(\psi), \quad \forall\, \psi \in L^1(\mathfrak{P}_0, \mathfrak{B}_t, dW_x), \tag{4.16}$$

given $F(\varphi)\psi \in L^1(\mathfrak{P}_0, dW_x)$, and another equivalent condition is

$$E_x\big(F(\varphi)|\mathfrak{B}_t\big) = E_x\big(F(\varphi)\big). \tag{4.17}$$

In turn, this identity holds whenever the left side is constant. From Corollary 4.3 we deduce:

Corollary 4.4. *For $s \ge 0$, $\vartheta_s \omega(t) = \omega(t+s) - \omega(s)$ is independent of $\mathfrak{B}_s$.*

Proof. By (4.13),

$$E_x\big(F(\omega(t+s) - \omega(s)) \big| \mathfrak{B}_s\big) = E_0\big(F(\omega(t))\big), \tag{4.18}$$

which is constant.

The Markov property gives rise to martingales. By definition (valid in general for an increasing family $\mathfrak{B}_t$ of σ-fields), a *martingale* is a family $F_t \in L^1(\mathfrak{P}_0, \mathfrak{B}_t, dW_{x,t})$ such that

$$E_x\big(F_t\big|\mathfrak{B}_s\big) = F_s \ \text{ when } s < t. \tag{4.19}$$

If $E_x(F_t|\mathfrak{B}_s) \ge F_s$ for $s < t$, $\{F_t\}$ is called a *submartingale* over $\mathfrak{B}_t$. The following is a very useful class of martingales.

Proposition 4.5. *Let $h(t,x)$ be smooth in $t \ge 0, x \in \mathbb{R}^n$, and satisfy $|h(t,x)| \le C_\varepsilon e^{\varepsilon|x|^2}$ for all $\varepsilon > 0$, and the backward heat equation*

$$\frac{\partial h}{\partial t} = -\Delta h. \tag{4.20}$$

Then $\mathfrak{h}_t(\omega) = h(t,\omega(t))$ is a martingale over $\mathfrak{B}_t$.

Proof. The hypothesis on $h(t,x)$ implies that, for $t, s > 0$,

$$h(s,x) = \int p(t,y)h(t+s,x-y)\,dy, \tag{4.21}$$

where $p(t,x) = e^{t\Delta}\delta(x)$ is given by (1.5). Now

$$\begin{aligned} E_x\big(\mathfrak{h}_{t+s}\big|\mathfrak{B}_s\big) &= E_x\big(h(t+s,\omega(t+s))\big|\mathfrak{B}_s\big) \\ &= E_{\omega(s)}\big(h(t+s,\omega(t))\big), \end{aligned} \tag{4.22}$$

for W_x-almost all ω, by (4.5). This is equal to

$$\int p(t,y-\omega(s))\,h(t+s,y)\,dy, \tag{4.23}$$

by the characterization (1.9) of expectation, adjusted as in (1.36), and by (4.21) this is equal to $h(s,\omega(s)) = \mathfrak{h}_s(\omega)$.

Corollary 4.6. *For one-dimensional Brownian motion, the following are martingales over $\mathfrak{B}_t$:*

$$\mathfrak{x}_t(\omega) = \omega(t), \quad \mathfrak{q}_t(\omega) = \omega(t)^2 - 2t, \quad \mathfrak{z}_t(\omega) = e^{a\omega(t)-a^2t}, \tag{4.24}$$

given $a > 0$.

One important property of martingales is the following martingale maximal inequality.

Proposition 4.7. *If F_t is a martingale over $\mathfrak{B}_t$, then, given any countable set $\{t_j\} \subset \mathbb{R}^+$, the "maximal function"*

$$F^*(\omega) = \sup_j\ F_{t_j}(\omega) \tag{4.25}$$

satisfies, for all $\lambda > 0$,

$$W_x\big(\{\omega : F^*(\omega) > \lambda\}\big) \le \frac{1}{\lambda} \| F_t \|_{L^1(\mathfrak{P}_0, dW_x)}. \tag{4.26}$$

Of course, the assumption that F_t is a martingale implies that $\|F_t\|_{L^1}$ is independent of t.

Proof. It suffices to demonstrate this for an arbitrary finite subset $\{t_j\}$ of $\mathbb{R}^+$. Thus we can work with $f_j(\omega) = F_{t_j}(\omega), \mathfrak{B}_j = \mathfrak{B}_{t_j}, 1 \le j \le N$, and take $t_1 < t_2 < \cdots < t_N$, and the martingale hypothesis is that $E_x(f_k \big| \mathfrak{B}_j) = f_j$ when $j < k$. There is no loss in assuming $f_N(\omega) \ge 0$, so all $f_j(\omega) \ge 0$. Now consider

$$S_\lambda = \{\omega : f^*(\omega) > \lambda\} = \{\omega : \text{some } f_j(\omega) > \lambda\}. \tag{4.27}$$

There is a pairwise-disjoint decomposition

$$S_\lambda = \bigcup_{j=1}^{N} S_{\lambda j}, \quad S_{\lambda j} = \{\omega : f_j(\omega) > \lambda \text{ but } f_\ell(\omega) \le \lambda \text{ for } \ell < j\}. \tag{4.28}$$

Note that $S_{\lambda j}$ is $\mathfrak{B}_j$-measurable. Consequently, we have

$$\begin{aligned} &\int_{S_\lambda} f_N(\omega)\, dW_x(\omega) \\ &= \sum_{j=1}^{N} \int_{S_{\lambda j}} f_N(\omega)\, dW_x(\omega) = \sum_{j=1}^{N} \int_{S_{\lambda j}} f_j(\omega)\, dW_x(\omega) \\ &\ge \sum_{j=1}^{N} \lambda\, W_x\big(S_{\lambda j}\big) = \lambda\, W_x(S_\lambda). \end{aligned} \tag{4.29}$$

This yields (4.26), in this special case, and the proposition is hence proved.

Applying the martingale maximal inequality to $\mathfrak{z}_t(\omega) = e^{a\omega(t) - a^2 t}$, we obtain the following.

Corollary 4.8. *For one-dimensional Brownian motion, given* $t > 0$,

$$W_0\big(\{\omega \in \mathfrak{P}_0 : \sup_{0 \le s \le t} \omega(s) - as > \lambda\}\big) \le e^{-a\lambda}. \tag{4.30}$$

Proof. The set whose measure is estimated in (4.30) is

$$\{\omega \in \mathfrak{P}_0 : \sup_{0 \le s \le t} e^{a\omega(s) - a^2 s} > e^{a\lambda}\}.$$

Since paths in $\mathfrak{P}_0$ are continuous, one can take the sup over $[0, t] \cap \mathbb{Q}$, which is countable, so (4.26) applies. Note that $E_0(\mathfrak{z}_t) = 1$.

We turn to a discussion of the *strong* Markov property of Brownian motion. For this, we need the notion of a stopping time. A function τ on $\mathfrak{P}_0$ with values in $[0, +\infty]$ is called a *stopping time* provided that, for each $t \geq 0$, $\{\omega \in \mathfrak{P}_0 : \tau(\omega) < t\}$ belongs to the σ-field $\mathfrak{B}_t$. It follows from (3.39) that $\vartheta_{\overline{\mathcal{O}}}$ is a stopping time. So is $\vartheta_{\mathcal{O}}$.

Given a stopping time τ, define $\mathfrak{B}_{\tau+}$ to be the σ-algebra of sets $S \in \mathfrak{B}_\infty$ such that $S \cap \{\omega : \tau(\omega) < t\}$ belongs to $\mathfrak{B}_t$ for each $t \geq 0$. Note that τ is measurable with respect to $\mathfrak{B}_{\tau+}$. The hypothesis that τ is a stopping time means precisely that the whole set $\mathfrak{P}_0$ satisfies the criteria for membership in $\mathfrak{B}_{\tau+}$. We note that any $t \in [0, \infty)$, regarded as a constant function on $\mathfrak{P}_0$, is a stopping time and that, in this case, $\mathfrak{B}_{t+} = \bigcap_{s>t} \mathfrak{B}_s$.

The following analogue of Propositions 4.1 and 4.2 is one statement of the strong Markov property.

Proposition 4.9. *If τ is a stopping time such that $\tau(\omega) < \infty$ for W_x-almost all ω, and if $t > 0$, then*

$$E_x\Big(f\big(\omega(\tau + t)\big)\big|\mathfrak{B}_{\tau+}\Big) = E_{\omega(\tau)}\big(f(\omega(t))\big), \tag{4.31}$$

for W_x-almost all ω. More generally, with

$$(\sigma_\tau \omega)(t) = \omega(t + \tau),$$

and φ bounded and $\mathfrak{B}_\infty$-measurable, we have

$$E_x\big(\varphi \circ \sigma_\tau \big| \mathfrak{B}_{\tau+}\big) = E_{\omega(\tau)}(\varphi), \tag{4.32}$$

for W_x-almost all ω.

As in (4.6), the content of (4.31) is that

$$\int_S f\big(\omega(\tau + t)\big)\, dW_x(\omega) = \int_S \Big(\int f\big(\omega^\#(t)\big)\, dW_{\omega(\tau)}(\omega^\#)\Big)\, dW_x(\omega), \tag{4.33}$$

given $S \in \mathfrak{B}_{\tau+}$. In other words, given that $S \cap \{\omega : \tau(\omega) < t'\} \in \mathfrak{B}_{t'}$, for each $t' \geq 0$. There is no loss in taking $x = 0$, and we can rewrite (4.33) as

$$\int_S f\big(\omega(\tau + t)\big)\, dW(\omega) = \int_S \int f\big(\omega^\#(t) + \omega(\tau)\big)\, dW(\omega^\#)\, dW(\omega). \tag{4.34}$$

It is useful to approximate τ by discretization:

$$\tau_\nu(\omega) = 2^{-\nu}k, \ \text{if } 2^{-\nu}(k-1) \leq \tau(\omega) < 2^{-\nu}k. \tag{4.35}$$

Thus

$$\{\omega : \tau_\nu(\omega) < t\} = \{\omega : \tau(\omega) < 2^{-\nu}k\} \in \mathfrak{B}_t, \tag{4.36}$$

so each τ_ν is a stopping time. Note that

$$\begin{aligned} A_{\nu k} &= \{\omega : \tau_\nu(\omega) = 2^{-\nu}k\} \\ &= \{\omega : \tau(\omega) < 2^{-\nu}k\} \setminus \{\omega : \tau(\omega) < 2^{-\nu}(k-1)\} \end{aligned} \tag{4.37}$$

belongs to $\mathfrak{B}_{2^{-\nu}k}$.

If τ is replaced by τ_ν, the left side of (4.34) becomes

$$\sum_{\nu,k} \int_{S\cap A_{\nu k}} f\big(\omega(t+2^{-\nu}k)\big)\, dW(\omega), \tag{4.38}$$

and the right side of (4.34) becomes

$$\sum_{\nu,k} \int_{S\cap A_{\nu k}} \int f\big(\omega^\#(t) + \omega(2^{-\nu}k)\big)\, dW(\omega^\#)\, dW(\omega). \tag{4.39}$$

Note that if $S \in \mathfrak{B}_{\tau+}$, then $S \cap A_{\nu k} \in \mathfrak{B}_{2^{-\nu}k}$. Thus, the fact that each term in the sum (4.38) is equal to the corresponding term in (4.39) follows from (4.6). Consequently, we have

$$\int_S f\big(\omega(\tau_\nu + t)\big)\, dW(\omega) = \int_S \int f\big(\omega^\#(t) + \omega(\tau_\nu)\big)\, dW(\omega^\#)\, dW(\omega), \tag{4.40}$$

for all ν, if $S \in \mathfrak{B}_{\tau+}$. The desired identity (4.34) follows by taking $\nu \to \infty$, if $f \in C(\dot{\mathbb{R}}^n)$. Passing from this to (4.32) is then done as in the proof of Proposition 4.2.

In particular, the extension of (4.31) analogous to (4.8), in the special case $F(x_1, x_2) = f(x_2 - x_1)$, yields the identity

$$\begin{aligned} \int_S f\big(\omega(\tau+t) - \omega(\tau)\big)\, dW(\omega) &= \int_S \int f\big(\omega^\#(t)\big)\, dW(\omega^\#)\, dW(\omega) \\ &= E\big(f(\omega(t))\big) \cdot W(S), \end{aligned} \tag{4.41}$$

given $S \in \mathfrak{B}_{\tau+}$. This, together with the extension to $F(x_1, \dots x_K)$, says that $\omega(\tau + t) - \omega(\tau) = \beta(t)$ has the probability distribution of a Brownian motion, independent of $\mathfrak{B}_{\tau+}$. This is a common form in which the strong Markov property is stated.

It is sometimes useful to consider stopping times for which $\{\omega : \tau(\omega) = \infty\}$ has positive measure. In such a case, the extension of Proposition 4.9 is that (4.32) holds for W_x-almost ω in the set $\{\omega : \tau(\omega) < \infty\}$. Thus, for example, (4.33) and (4.34) hold, given $S \in \mathfrak{B}_{\tau+}$ and $S \subset \{\omega : \tau(\omega) < \infty\}$.

We next look at some operator-theoretic properties of

$$(4.42)\qquad \begin{aligned} Q_t &: L^2(\mathfrak{P}_0, dW_0) \to L^2(\mathfrak{P}_0, dW_0), \quad Q_t\varphi = E_0(\varphi|\mathfrak{B}_t),\\ \Theta_t &: L^2(\mathfrak{P}_0, dW_0) \to L^2(\mathfrak{P}_0, dW_0), \quad \Theta_t\varphi(\omega) = \varphi(\vartheta_t\omega), \end{aligned}$$

where ϑ_t is given by (4.11). For each $t \ge 0$, Q_t is an orthogonal projection, and $Q_sQ_t = Q_tQ_s = Q_s$, for $s \le t$. Note that (4.13) implies

$$(4.43)\qquad Q_t\Theta_t = Q_0,$$

since Q_0 is the orthogonal projection of $L^2(\mathfrak{P}_0, dW_0)$ onto

$$(4.44)\qquad \mathcal{R}(Q_0) = \text{ set of constant functions.}$$

Proposition 4.10. *The family Θ_t, $t \in [0,\infty)$, is a strongly continuous semigroup of isometries of $L^2(\mathfrak{P}_0, dW_0)$, with*

$$(4.45)\qquad \mathcal{R}(\Theta_t) \subset \mathit{Ker}(Q_t - Q_0) = \{\varphi : E_0(\varphi|\mathfrak{B}_t) = \mathit{const.}\}.$$

Proof. That Θ_t is an isometry follows from the measure-preserving property (4.14). If we apply Q_0 to (4.43), we get $Q_0\Theta_t = Q_0$; hence $(Q_t - Q_0)\Theta_t = 0$, which yields (4.45).

The semigroup property follows from a straightforward calculation:

$$(4.46)\qquad \vartheta_\sigma\vartheta_s\omega = \vartheta_{\sigma+s}\omega \Longrightarrow \Theta_{s+\sigma} = \Theta_s\Theta_\sigma.$$

The convergence

$$(4.47)\qquad \Theta_s\varphi \to \Theta_t\varphi \quad \text{in } L^2(\mathfrak{P}_0, dW_0), \text{ as } s \to t,$$

is easy to demonstrate for $\varphi(\omega)$ of the form (1.8), that is,

$$(4.48)\qquad \varphi(\omega) = f\big(\omega(t_1), \dots, \omega(t_k)\big),$$

with f continuous on $\dot{\mathbb{R}}^n \times \cdots \times \dot{\mathbb{R}}^n$ (k factors). In fact, $\varphi(\vartheta_s(\omega)) = \varphi_s(\omega) \to \varphi_t(\omega)$ boundedly and pointwise on $\mathfrak{P}_0$ for such φ. Since the set of φ of the form (4.48) is dense in $L^2(\mathfrak{P}_0, dW_0)$, (4.47) follows.

Proposition 4.11. *The family of orthogonal projections Q_t is strongly continuous in $t \in [0,\infty)$.*

Proof. It is easy to verify that, for any $\varphi \in L^2(\mathfrak{P}_0, dW_0)$,

$$(4.49)\qquad Q_s\varphi \to Q_{t-}\varphi = E_0(\varphi|\mathfrak{B}_{t-}), \quad \text{as } s \nearrow t,$$

provided $t > 0$, and

$$Q_s\varphi \to Q_{t+}\varphi = E_0(\varphi|\mathfrak{B}_{t+}), \quad \text{as } s \searrow t, \tag{4.50}$$

where

$$\mathfrak{B}_{t-} = \sigma\Big(\bigcup_{s<t} \mathfrak{B}_s\Big), \quad \mathfrak{B}_{t+} = \bigcap_{s>t} \mathfrak{B}_s. \tag{4.51}$$

It is also easy to verify that $\mathfrak{B}_{t-} = \mathfrak{B}_t$, for $t > 0$, so $Q_s\varphi \to Q_t\varphi$ as $s \nearrow t$. On the other hand, it is not true that $\mathfrak{B}_{t+} = \mathfrak{B}_t$, so the continuity of $Q_t\varphi$ from above requires more work.

Suppose $t_j \in \mathbb{Q}^+$ and

$$0 \le t_1 < t_2 < \cdots < t_\ell \le t < t_{\ell+1} < \cdots < t_{\ell+k}. \tag{4.52}$$

Let $f_j \in C(\dot{\mathbb{R}}^n)$. Consider any function on $\mathfrak{P}$ of the form

$$\begin{aligned} \varphi(\omega) &= A_\ell(\omega)B_{k\ell}(\omega) \\ &= f_1\big(\omega(t_1)\big)\cdots f_\ell\big(\omega(t_\ell)\big)\cdot f_{\ell+1}\big(\omega(t_{\ell+1})\big)\cdots f_{\ell+k}\big(\omega(t_{\ell+k})\big). \end{aligned} \tag{4.53}$$

Denote by $\mathcal{C}\hat{}$ the linear span of the set of such functions. For φ of the form (4.53), we have

$$E_0(\varphi|\mathfrak{B}_t) = A_\ell(\omega)E_0(B_{k\ell}|\mathfrak{B}_t). \tag{4.54}$$

If $t_{\ell+\nu} = t + s_\nu,\ 1 \le \nu \le k$, we have, by (4.8),

$$E_0(B_{k\ell}|\mathfrak{B}_t) = E_{\omega(t)}\big(f_{\ell+1}(\omega(s_1))\cdots f_{\ell+k}(\omega(s_k))\big), \quad \text{a.e. on } \mathfrak{P}_0. \tag{4.55}$$

Now, if $t \le t+h < t_{\ell+1}$, we also have

$$\begin{aligned} E_0(\varphi|\mathfrak{B}_{t+h}) &= A_\ell(\omega)E_0(B_{k\ell}|\mathfrak{B}_{t+h}) \\ &= A_\ell(\omega)E_{\omega(t+h)}(\psi_\ell), \end{aligned} \tag{4.56}$$

where

$$\psi_\ell(\omega) = f_{\ell+1}\big(\omega(s_1-h)\big)\cdots f_{\ell+k}\big(\omega(s_k-h)\big). \tag{4.57}$$

Now, as in (1.9),

$$\begin{aligned} E_x(\psi_\ell) = \int\cdots\int\ & p(s_1-h,x_1)p(s_2-s_1,x_2-x_1) \\ & \cdots p(s_k-s_{k-1},x_k-x_{k-1}) \\ & \cdot f_{\ell+1}(x_1+x)\cdots f_{\ell+k}(x_k+x)\,dx_k\cdots dx_1. \end{aligned} \tag{4.58}$$

The continuity in (x, h) is clear. Since paths in $\mathfrak{P}_0$ are continuous, we have, by linearity, that

$$\varphi \in \mathcal{C}^{\wedge} \Longrightarrow E_0(\varphi|\mathfrak{B}_t) = \lim_{h \searrow 0} E_0(\varphi|\mathfrak{B}_{t+h}), \quad W_0\text{-a.e.} \tag{4.59}$$

Now the Stone-Weierstrass theorem implies that $\mathcal{C}^{\wedge}$ is dense in $C(\mathfrak{P})$, which is dense in $L^2(\mathfrak{P}, dW_0) = L^2(\mathfrak{P}_0, dW_0)$. Thus we have

$$E_0(\varphi|\mathfrak{B}_{t+}) = E_0(\varphi|\mathfrak{B}_t), \quad W_0\text{-a.e.,} \tag{4.60}$$

for every $\varphi \in L^2(\mathfrak{P}_0, dW_0)$, and the proposition is proved.

Exercises

1. Show that the martingale maximal inequality applied to $\mathfrak{x}_t(\omega) = \omega(t)$ yields
$$W_0\Big(\{\omega \in \mathfrak{P}_0 : \sup_{0 \le s \le t} \omega(s) > b\sqrt{4t/\pi}\}\Big) \le \frac{1}{b}.$$
Compare with the precise result in (3.29).
2. With $\mathfrak{B}_{t-}$ characterized by (4.51), show that $\mathfrak{B}_{t-} = \mathfrak{B}_t$, as stated in the proof of Proposition 4.11. (*Hint*: In the characterization (4.1) of $\mathfrak{B}_t$, one can restrict attention to E *open* in $\mathbb{R}^n$.)
3. Using (4.60), show that
$$S \in \mathfrak{B}_{0+} \Longrightarrow W_0(S) = 0 \text{ or } 1.$$
This is called Blumenthal's 01 law. If $E \in \mathbb{R}^n$ is a closed set, show that
$$\{\omega \in \mathfrak{P}_0 : \omega(t_\nu) \in E \text{ for some } t_\nu \searrow 0\}$$
is a set in $\mathfrak{B}_{0+}$. (*Hint*: Consider $\{\omega \in \mathfrak{P}_0 : \text{dist}(\omega(t), E) \ge \delta > 0 \text{ for } t \in [2^{-\nu}\varepsilon, \varepsilon] \cap \mathbb{Q}\} = S(E, \delta, \varepsilon, \nu)$.)
4. Let $\mathcal{N}$ be the collection of (W_0-outer measurable) subsets of $\mathfrak{P}_0$ with W_0-measure zero. Form the family of σ-algebras $\mathfrak{B}_t^{\#} = \mathfrak{B}_t \cup \mathcal{N}$, called the *augmentation* of $\mathfrak{B}_t$. Show that $\mathfrak{B}_t^{\#} \supset \mathfrak{B}_{t+}$ and, with notation parallel to (4.51),
$$\mathfrak{B}_{t-}^{\#} = \mathfrak{B}_t^{\#} = \mathfrak{B}_{t+}^{\#}.$$
Note: The augmentation of $\mathfrak{B}_t$ is bigger than the completion of $\mathfrak{B}_t$.
5. Let $\widetilde{\mathfrak{F}}_t$ be the σ-algebra of subsets of $\mathfrak{P}_0$ generated by sets of the form (4.1) for $s \ge t$, and set $\mathcal{A}_\infty = \bigcap_{t>0} \widetilde{\mathfrak{F}}_t$. Using Blumenthal's 01 law and Exercise 2 of §1, show that
$$S \in \mathcal{A}_\infty \Longrightarrow W_0(S) = 0 \text{ or } 1.$$
If $E \subset \mathbb{R}^n$ is a closed set, show that
$$\{\omega \in \mathfrak{P}_0 : \omega(t_\nu) \in E \text{ for some } t_\nu \nearrow \infty\}$$
is a set in $\mathcal{A}_\infty$.

5. First exit time and the Poisson integral

At the end of §3 we produced a formula for PI f, giving the solution u to

$$\Delta u = 0 \text{ in } \Omega, \quad u = f \text{ on } \partial\Omega, \tag{5.1}$$

at least in case Ω is a bounded domain in $\mathbb{R}^n$ with smooth boundary. Here we produce a formula that is somewhat neater than (3.36) and that is also amenable to extension to general bounded, open $\Omega \subset \mathbb{R}^n$, with no smoothness assumed on $\partial\Omega$. In the smooth case, the formula is

$$\text{PI}\, f(x) = E_x\big(f(\omega(\vartheta_{\overline{\Omega}}))\big), \quad x \in \overline{\Omega}, \tag{5.2}$$

where $\vartheta_{\overline{\Omega}}(\omega)$ is the first exit time defined by (3.24).

From an intuitive point of view, the formula (5.2) has a very easy and natural justification. To show that the right side of (5.2), which we denote by $u(x)$, is harmonic on Ω, it suffices to verify the mean-value property. Let $x \in \Omega$ be the center of a closed ball $B \subset \Omega$. We claim that $u(x)$ is equal to the mean value of $u|_{\partial B}$. Indeed, a continuous path ω starting from x and reaching $\partial\Omega$ must cross ∂B, say at a point $y = \omega(\vartheta_B)$. The future behavior of such paths is independent of their past, so the probability distribution of the first contact point $\omega(\vartheta_{\overline{\Omega}})$, when averaged over starting points in ∂B, should certainly coincide with the probability distribution of such a first contact point in $\partial\Omega$, for paths starting at x (the distribution of whose first contact point with ∂B must be constant, by symmetry).

The key to converting this into a mathematical argument is to note that the time $\vartheta_B(\omega)$ is not constant, so one needs to make use of the strong Markov property as a tool to establish the mean-value property of the function $u(x)$ defined by the right side of (5.2).

Let us first make some comments on the right side $u(x)$ of (5.2). By (3.40) we have

$$\Big| u(x) - \int_{\mathfrak{P}_{\overline{\Omega},N}} f\big(\omega(\vartheta_{\overline{\Omega}})\big)\, dW_x(\omega) \Big| \le C \|f\|_{L^\infty(\partial\Omega)}\, N^{-1}. \tag{5.3}$$

Let us extend $f \in C(\partial\Omega)$ to an element $f \in C_0(\mathbb{R}^n)$, without increasing the sup norm. By (3.41), we have

$$f\big(\omega(\vartheta_{\overline{\Omega}})\big) = \lim_{\nu\to\infty} f\big(\omega(\vartheta_{\nu,N}(\omega))\big), \ \text{ for } \omega \in \mathfrak{P}_{\overline{\Omega},N}, \tag{5.4}$$

where $\vartheta_{\nu,N}(\omega) = \min\big(N, \inf\{s \in 2^{-\nu}\mathbb{Z}^+ : \omega(s) \notin \overline{\Omega}\}\big)$. Thus, if the integral in (5.3) is denoted by $u_N(x)$, then

$$u_N(x) = \lim_{\nu\to\infty} u_{N\nu}(x) = \lim_{\nu\to\infty} \int_{\mathfrak{P}_{\overline{\Omega},N}} f\big(\omega(\vartheta_{\nu,N}(\omega))\big)\, dW_x(\omega). \tag{5.5}$$

Here the limit exists pointwise in $x \in \Omega$. Now each $u_{N\nu}$ is continuous on Ω, indeed on $\mathbb{R}^n$. Consequently, $u(x)$ given by the right side of (5.2) is at least a bounded, measurable function of x.

To continue the analysis, given $x \in \Omega$, we define a probability measure $\nu_{x,\Omega}$ on $\partial\Omega$ by

$$E_x\big(f(\omega(\vartheta_{\overline{\Omega}}))\big) = \int_{\Omega} f(y)\, d\nu_{x,\Omega}(y), \tag{5.6}$$

for $f \in C(\partial\Omega)$.

Lemma 5.1. *If $x \in \mathcal{O} \subset\subset \Omega$ and $\mathcal{O}$ and Ω are open, then*

$$\nu_{x,\Omega} = \int_{\partial\mathcal{O}} \nu_{y,\Omega}\, d\nu_{x,\mathcal{O}}(y). \tag{5.7}$$

Proof. The identity (5.4) is equivalent to the statement that, for $f \in C(\partial\Omega)$,

$$E_x\big(f(\omega(\vartheta_{\overline{\Omega}}))\big) = \int_{\partial\mathcal{O}} E_y\big(f(\omega(\vartheta_{\overline{\Omega}}))\big)\, d\nu_{x,\mathcal{O}}(y). \tag{5.8}$$

The right side is equal to

$$E_x\big(g(\omega(\vartheta_{\overline{\mathcal{O}}}))\big), \quad g(y) = E_y\big(f(\omega(\vartheta_{\overline{\Omega}}))\big). \tag{5.9}$$

In other words,

$$g(\omega(\vartheta_{\overline{\mathcal{O}}})) = E_{\omega(\vartheta_{\overline{\mathcal{O}}})}(\varphi), \quad \varphi(\omega) = f\big(\omega(\vartheta_{\overline{\Omega}}(\omega))\big). \tag{5.10}$$

Now we use the strong Markov property, in the form (4.32), namely,

$$E_{\omega(\tau)}(\varphi) = E_x\big(\varphi \circ \sigma_\tau \big| \mathfrak{B}_{\tau+}\big),$$

for W_x-almost all ω, where $(\sigma_\tau\omega)(t) = \omega(t+\tau)$ and τ is a stopping time. This implies

$$\int_{\mathfrak{P}_0} E_{\omega(\tau)}(\varphi)\, dW_x(\omega) = \int_{\mathfrak{P}_0} E_x\big(\varphi \circ \sigma_\tau \big| \mathfrak{B}_{\tau+}\big)\, dW_x(\omega) = E_x(\varphi \circ \sigma_\tau). \tag{5.11}$$

Applied to $\tau = \vartheta_{\overline{\mathcal{O}}}$, this shows that (5.9) is equal to $E_x(\varphi \circ \sigma_{\vartheta_{\overline{\mathcal{O}}}})$. Now, with $\widetilde{\omega}(t) = \sigma_{\vartheta_{\overline{\mathcal{O}}}}\omega(t) = \omega(t + \vartheta_{\overline{\mathcal{O}}}(\omega))$, we have, for $\mathcal{O} \subset\subset \Omega$, $\vartheta_{\overline{\Omega}}(\widetilde{\omega}) = \vartheta_{\overline{\Omega}}(\omega) - \vartheta_{\overline{\mathcal{O}}}(\omega)$, as long as ω is a continuous path starting in $\mathcal{O}$. Hence

$$\varphi(\widetilde{\omega}) = f\big(\widetilde{\omega}(\vartheta_{\overline{\Omega}}(\omega) - \vartheta_{\overline{\mathcal{O}}}(\omega))\big) = f\big(\omega(\vartheta_{\overline{\Omega}}(\omega))\big) = \varphi(\omega). \tag{5.12}$$

Thus (5.9) is equal to $E_x(\varphi)$, which is the left side of (5.6), and the lemma is proved.

Consequently, the right side $u(x)$ of (5.2) is a bounded, measurable function of x satisfying the mean-value property. An integration yields that such $u(x)$ is equal to the mean value of u over any *ball* $\mathcal{D} \subset \Omega$, centered at x, from which it follows that $u(x)$ is continuous in Ω. Then the mean-value property guarantees that u is harmonic on Ω. To verify (5.2), it remains to show that $u(x)$ has the correct boundary values.

Lemma 5.2. *Assume $\partial\Omega$ is smooth. Given $y \in \partial\Omega$, we have $u(y) = f(y)$, and u is continuous at $y \in \overline{\Omega}$.*

Proof. That $u(y) = f(y)$ follows from the fact that $\vartheta_{\overline{\Omega}}(\omega) = 0$ for W_y-almost all ω, according to Proposition 3.7. To show that $u(x) \to u(y)$ as $x \to y$ from within Ω, we argue as follows.

By (3.23), for $x \in \Omega,\ E_x(\vartheta_{\overline{\Omega}}) = -\Delta^{-1}1(x)$. Hence this quantity approaches 0 as $x \to y$. Thus, given $\varepsilon_j > 0$, there exists $\delta > 0$ such that

$$|x - y| \le \delta \Longrightarrow W_x\big(\{\omega : \vartheta_{\overline{\Omega}}(\omega) > \varepsilon_1\}\big) < \varepsilon_2. \tag{5.13}$$

Meanwhile, in a short time, $0 \le s \le \varepsilon_1$, a path $\omega(s)$ is not likely to wander far. In fact, by (3.28) plus a scaling argument,

$$\begin{aligned} \mathcal{W}_{\varepsilon_1} &= \{\omega \in \mathfrak{P}_0 : \sup_{0\le s\le\varepsilon_1} |\omega(s) - \omega(0)| \ge \varepsilon_1^{1/3}\} \\ &\Longrightarrow W_x(\mathcal{W}_{\varepsilon_1}) \le \psi(\varepsilon_1), \end{aligned} \tag{5.14}$$

where $\psi(\varepsilon) \to 0$ as $\varepsilon \to 0$.

Thus, if $|x - y| \le \delta$, with probability $> 1 - \varepsilon_2 - \psi(\varepsilon_1)$, a path starting at x will, within time ε_1, hit $\partial\Omega$, without leaving the ball $B_{\varepsilon_1^{1/3}}(x)$ of radius $\varepsilon_1^{1/3}$ centered at x. Now, a given $f \in C(\partial\Omega)$ varies only a little over $\{z \in \partial\Omega : |z - y| \le \varepsilon_1^{1/3} + \delta\}$ if ε_1 and δ are small enough. Therefore, indeed $u(x) \to u(y)$, as $x \to y$.

We have completed the demonstration of the following.

Proposition 5.3. *If Ω is a bounded region in $\mathbb{R}^n$ with smooth boundary and $f \in C(\partial\Omega)$, then PI f is given by* (5.2).

Recall from §5 of Chap. 5 the construction of

$$\text{PI} : C(\partial\Omega) \longrightarrow L^\infty(\Omega) \cap C^\infty(\Omega) \tag{5.15}$$

when Ω is an arbitrary bounded, open subset of $\mathbb{R}^n$, with perhaps a very nasty boundary. As shown there, we can take

$$\Omega_1 \subset\subset \Omega_2 \subset\subset \cdots \subset\subset \Omega_j \nearrow \Omega \tag{5.16}$$

such that each boundary $\partial\Omega_j$ is smooth, and, if f is extended from $\partial\Omega$ to an element of $C_o(\mathbb{R}^n)$, then

$$x \in \Omega \Longrightarrow \text{PI}\, f(x) = \lim_{j\to\infty} u_j(x), \tag{5.17}$$

where $u_j \in C(\overline{\Omega}_j)$ is the Poisson integral of $f|_{\partial\Omega_j}$. In (5.17) one has uniform convergence on compact sets $K \subset \Omega$, the right side being defined for $j \ge j_0$, where $K \subset \Omega_{j_0}$. The details were carried out in Chap. 5 for $f \in C^\infty(\mathbb{R}^n)$, but approximation by smooth functions plus use of the maximum principle readily extends this to $f \in C_o(\mathbb{R}^n)$.

If we apply Proposition 5.3 to Ω_j, we conclude that, for $f \in C_o(\mathbb{R}^n), x \in \Omega$,

$$\text{PI}\, f(x) = \lim_{j\to\infty} E_x\Big(f\big(\omega(\vartheta_{\overline{\Omega}_j})\big)\Big). \tag{5.18}$$

On the other hand, it is straightforward from the definitions that

$$\vartheta_{\overline{\Omega}_j}(\omega) \nearrow \vartheta_\Omega(\omega), \quad \text{for all } \omega \in \mathfrak{P}_0. \tag{5.19}$$

Therefore, via the dominated convergence theorem, we can pass to the limit in (5.18), proving the following.

Proposition 5.4. *If Ω is any bounded, open region in $\mathbb{R}^n$ and $f \in C(\partial\Omega)$, then*

$$PI\, f(x) = E_x\Big(f\big(\omega(\vartheta_\Omega)\big)\Big), \quad x \in \Omega. \tag{5.20}$$

We recall from Chap. 5 the notion of a *regular* boundary point. A point $y \in \partial\Omega$ is regular provided PI f is continuous at y, for all $f \in C(\partial\Omega)$. We discussed several criteria for a boundary point to be regular, particularly in Propositions 5.11–5.16 of Chap. 5. Here is another criterion.

Proposition 5.5. *If $\Omega \subset \mathbb{R}^n$ is a bounded open set, $y \in \partial\Omega$, then y is a regular boundary point if and only if*

$$E_x(\vartheta_\Omega) \to 0, \quad \textit{as } x \to y, \ x \in \Omega. \tag{5.21}$$

Proof. Recall from (3.20) that $E_x(\vartheta_\Omega) = -\Delta^{-1}1(x)$. Thus (5.21) holds if and only if this function is a weak barrier at $y \in \partial\Omega$, as defined in Chap. 5, right after (5.26). Therefore, (5.21) here implies y is a regular point. On the other hand, $\Delta^{-1}1(x)$ can be written as the sum $x_1^2/2 + u_0(x)$, where $u_0 = -(1/2)\,\text{PI}\,(x_1^2|_{\partial\Omega})$, so if (5.21) fails, y is not a regular point.

One might both compare and contrast this proof with that of Lemma 5.2. In that case, where $\partial\Omega$ was assumed smooth, the known regularity of each boundary point was exploited to guarantee that $E_x(\vartheta_{\overline{\Omega}}) \to 0$ as $x \to y \in \partial\Omega$, which then was exploited to show that $u(x) \to u(y)$ as $x \to y$.

In the next section, we will derive another criterion for y to be regular, in terms of "capacity."

Exercises

1. Explore connections between the formulas for PI $f(x)$, for $f \in C(\partial\Omega)$, when Ω is bounded and $\partial\Omega$ smooth, given by (3.36) and by (5.2), respectively.

6. Newtonian capacity

The (Newtonian) capacity of a set is a measure of size that is very important in potential theory and closely related to the probability of a Brownian path hitting that set. In our development here, we restrict attention to the case $n \geq 3$ and define the capacity of a compact set $K \subset \mathbb{R}^n$. We first assume that K is the closure of an open set with smooth boundary.

Proposition 6.1. *Assume $n \geq 3$. If $K \subset \mathbb{R}^n$ is compact with smooth boundary ∂K, then there exists a unique function U_K, harmonic on $\mathbb{R}^n \setminus K$, such that $U_K(x) \to 1$ as $x \to K$ and $U_K(x) \to 0$ as $|x| \to \infty$.*

Proof. We can assume that the origin $0 \in \mathbb{R}^n$ is in the interior of K. Then the inversion $\psi(x) = x/|x|^2$ interchanges 0 and the point at infinity, and the transformation

$$v(x) = |x|^{-(n-2)} w(|x|^{-2} x) \tag{6.1}$$

preserves harmonicity. We let w be the unique harmonic function on the bounded domain $\psi(\mathbb{R}^n \setminus K)$, with boundary value $w(x) = |x|^{-(n-2)}$ on $\psi(\partial K)$. Then v, defined by (6.1), is the desired solution. The uniqueness is immediate, via the maximum principle.

Note that the construction yields

$$|U_K(x)| \leq C|x|^{-(n-2)}, \quad |\partial_r U_K(x)| \leq C|x|^{-(n-1)}, \quad |x| \to \infty. \tag{6.2}$$

The $n = 3$ case of this result was done in §1 of Chap. 9.

Another approach to the proof of Proposition 6.1 would be to represent $U_K(x)$ as a single-layer potential, as in (11.44) of Chap. 7. This was noted in a remark after the proof of Proposition 11.5 in that chapter.

Now that we have established the existence of such U_K, Exercises 7–9 of §3 apply, to yield

$$U_K^t(x) \nearrow U_K(x), \quad \text{as } t \nearrow \infty, \tag{6.3}$$

where, for $x \in \mathcal{O} = \mathbb{R}^n \setminus K$,

$$\begin{aligned} U_K^t(x) &= 1 - e^{t\Delta_{\mathcal{O}}} 1(x) \\ &= W_x\big(\{\omega : \vartheta_{\mathcal{O}}(\omega) \leq t\}\big). \end{aligned} \tag{6.4}$$

Here, Δ_O is the Laplace operator on $\mathcal{O}$, with Dirichlet boundary condition. The last identity follows from (3.27). We can replace the first exit time $\vartheta_{\mathcal{O}}$ by the first hitting time:

$$\mathfrak{h}_K(\omega) = \vartheta_{\mathbb{R}^n \setminus K}(\omega). \tag{6.5}$$

Consequently,

$$U_K(x) = W_x\big(\{\omega : \mathfrak{h}_K(\omega) < \infty\}\big); \tag{6.6}$$

that is, for $x \in \mathcal{O}$, $U_K(x)$ is the probability that a Brownian path ω, starting at x, eventually hits K.

We set $U_K(x) = 1$ for $x \in K$. Then (6.6) holds for $x \in K$ also. It follows that $U_K \in C_o(\mathbb{R}^n)$, and ΔU_K is a distribution supported on ∂K. In fact, Green's formula yields, for $\varphi \in C_0^\infty(\mathbb{R}^n)$,

$$(U_K, \Delta\varphi) = -\int_{\partial K} \varphi(y)\, \frac{\partial}{\partial\nu} U_K(y)\, dS(y), \tag{6.7}$$

where ν is the unit normal to ∂K, pointing into K. By Zaremba's principle, $\partial_\nu U_K(y) > 0$, for all $y \in \partial K$, so we see that $\Delta U_K = -\mu_K$, where μ_K is a positive measure supported on ∂K. The total mass of μ_K is called the *capacity* of K:

$$\text{cap } K = \int_K d\mu_K(x). \tag{6.8}$$

Since, with $C_n = (n-2) \cdot \text{Area}(S^{n-1})$,

$$U_K(x) = -\Delta^{-1}\mu_K = C_n \int |x-y|^{-(n-2)}\, d\mu_K(y), \tag{6.9}$$

we have

$$C_n \iint \frac{d\mu_K(x)\, d\mu_K(y)}{|x-y|^{n-2}} = \int U_K(x)\, d\mu_K(x) = \text{ cap } K, \tag{6.10}$$

the left side being proportional to the potential energy of a collection of charged particles, with density $d\mu_K$, interacting by a repulsive force with potential $C_n|x-y|^{-(n-2)}$. The function $U_K(x)$ is called the *capacitary potential* of K. Note that we can also use Green's theorem to get

$$\|\nabla U_K\|^2_{L^2(\mathbb{R}^n)} = \int_K U_K(x)\, d\mu_K(x) = \text{ cap } K. \tag{6.11}$$

Note that if $K_1 \subset K_2$ have capacitary potentials $U_j,\ \Delta U_j = -\mu_j$, then $U_2 = 1$ on K_1, so

$$\text{(6.12)} \qquad \begin{aligned} \text{cap } K_1 &= \int U_2(x)\, d\mu_1(x) = -(U_2, \Delta U_1) \\ &= \int U_1(x)\, d\mu_2(x) \le \text{ cap } K_2, \end{aligned}$$

since $U_1(x) \le 1$. Thus capacity is a monotone set function.

Before establishing more formulas involving capacity, we extend it to general compact $K \subset \mathbb{R}^n$. We can write $K = \bigcap K_j$, where $K_1 \supset\supset K_2 \supset\supset \cdots \supset\supset K_j \searrow K$, each K_j being compact with smooth boundary. Clearly, $U_j = U_{K_j}$ is a decreasing sequence of functions ≤ 1, and by (6.11), ∇U_j is bounded in $L^2(\mathbb{R}^n)$. Furthermore, $\Delta U_j = -\mu_j$, where μ_j is a positive measure supported on ∂K_j, of total mass cap K_j, which is nonincreasing, by (6.12). Consequently, we have a limit:

$$\text{(6.13)} \qquad \lim_{j\to\infty} U_j = U_K,$$

defined a priori pointwise, but also holding in various topologies, such as the weak* topology of $L^\infty(\mathbb{R}^n)$. We have $U_K \in L^\infty(\mathbb{R}^n),\ 0 \le U_K(x) \le 1;\ \nabla U_K \in L^2(\mathbb{R}^n)$, and $\Delta U_K = -\mu$, where μ is a positive measure, supported on K. Furthermore, $\mu_j \to \mu$ in the weak* topology, and $U_K = -\Delta^{-1}\mu$. Any neighborhood of K contains some K_j. Thus, if $K_1' \supset\supset K_2' \supset\supset \cdots \supset\supset K_j' \searrow K$ is another choice, one is seen to obtain the same limit U_K, hence the same measure μ, which we denote as μ_K. We set

$$\text{(6.14)} \qquad \text{cap } K = \int d\mu_K(x).$$

Note that, as in (6.12), cap $K = \int U_j(x)\, d\mu_K(x)$, for each j. Thus, as before, cap $K = \int U_K(x)\, d\mu_K(x)$, this time by the monotone convergence theorem. Consequently,

$$\text{(6.15)} \qquad U_K(x) = 1 \quad \mu_K\text{-almost everywhere.}$$

Clearly, cap $K \le \inf$ cap K_j. In fact, we claim

$$\text{(6.16)} \qquad \text{cap } K = \inf \text{ cap } K_j.$$

This is easy to see; μ_j converges to μ_K pointwise on $C_o(\mathbb{R}^n)$; choose $g \in C_o(\mathbb{R}^n)$, equal to 1 on K_1; then

$$\text{(6.17)} \qquad \text{cap } K = (g, \mu_K) = \lim\, (g, \mu_j) = \lim\ \text{cap} K_j,$$

proving (6.16). We consequently extend the monotonicity property:

Proposition 6.2. *For general compact $K \subset L$, we have cap $K \le$ cap L.*

Proof. We can take compact approximants with smooth boundary, $K_j \searrow K$, $L_j \searrow L$, such that $K_j \subset L_j$. By (6.12) we have cap $K_j \le$ cap L_j, and this persists in the limit by (6.16). We also have $U_K(x) \le U_L(x)$ for all x. Using (6.15), we obtain

$$\text{cap } K = \int U_L(x)\, d\mu_K(x). \tag{6.18}$$

One possibility is that cap $K = 0$. This happens if and only if $\mu_K = 0$, thus if and only if $U_K = 0$ almost everywhere. If cap $K > 0$, we continue to call U_K the capacitary potential of K.

We record some more ways in which $U_j \to U_K$. First, it certainly holds in the weak* topology on $L^\infty(\mathbb{R}^n)$. Hence $\nabla U_j \to \nabla U_K$ in $\mathcal{D}'(\mathbb{R}^n)$. By (6.11), ∇U_j is bounded in $L^2(\mathbb{R}^n)$; hence $\nabla U_j \to \nabla U_K$ weakly in $L^2(\mathbb{R}^n)$. Since also $U_j \in C_o(\mathbb{R}^n)$, we have

$$\begin{aligned} \|\nabla U_K\|_{L^2}^2 &= \lim_{j\to\infty} (\nabla U_j, \nabla U_K) = \lim_{j\to\infty} -(U_j, \Delta U_K) \\ &= \lim_{j\to\infty} \int U_j(x)\, d\mu(x) = \text{cap } K, \end{aligned} \tag{6.19}$$

the last identity holding as in the derivation of (6.15). Thus (6.11) is extended to general compact K. Furthermore, this implies

$$\nabla U_j \longrightarrow \nabla U_K \quad \text{in } L^2(\mathbb{R}^n)\text{-norm}. \tag{6.20}$$

Hence

$$\mu_j \longrightarrow \mu_K \quad \text{in } H^{-1}(\mathbb{R}^n)\text{-norm}. \tag{6.21}$$

We now extend the identities (6.3) and (6.6) to general compact K, in reverse order.

Proposition 6.3. *The identity* (6.6) *holds for general compact $K \subset \mathbb{R}^n$.*

Proof. Since (6.6) has been established for the compact K_j with smooth boundary, we have

$$1 - U_j(x) = W_x(\mathfrak{A}_{K_j}), \quad \mathfrak{A}_{K_j} = \{\omega \in \mathfrak{P}_0 : \omega(\mathbb{R}^+) \subset \mathbb{R}^n \setminus K_j\}. \tag{6.22}$$

Clearly, if $K_j \searrow K$, $\mathfrak{A}_{K_1} \subset \mathfrak{A}_{K_2} \subset \cdots \subset \mathfrak{A}_{K_j} \nearrow \widetilde{\mathfrak{A}}_K$, where $\widetilde{\mathfrak{A}}_K$ is a proper subset of $\mathfrak{A}_K = \{\omega \in \mathfrak{P}_0 : \omega(\mathbb{R}^+) \subset \mathbb{R}^n \setminus K\}$. However, for $n \ge 3$, Brownian motion is nonrecurrent, as was established in Exercise 10 of §3. Thus $|\omega(t)| \to \infty$ as $t \to \infty$, for W_x-almost all ω, so in fact $W_x(\mathfrak{A}_K \setminus \widetilde{\mathfrak{A}}_K) = 0$, and hence $1 - U_K(x) = W_x(\mathfrak{A}_K)$, which is equivalent to (6.6).

Proposition 6.4. *The identity* (6.3) *holds for general compact* $K \subset \mathbb{R}^n$.

Proof. We define $U_K^t(x)$ to be $1 - e^{t\Delta_O}1(x)$, as in (6.4); the second identity in (6.4) continues to hold, by (3.27). Now, clearly, the family of sets $S_t = \{\omega \in \mathfrak{P}_0 : \mathfrak{h}_K(\omega) \le t\}$ is increasing as $t \nearrow \infty$, with union

$$\bigcup S_t = \{\omega \in \mathfrak{P}_0 : \mathfrak{h}_K(\omega) < \infty\},$$

and this gives (6.3).

We next establish the subadditivity of capacity.

Proposition 6.5. *If K and L are compact, then*

$$U_{K\cup L}(x) \le U_K(x) + U_L(x) \tag{6.23}$$

and

$$\mathit{cap}(K \cup L) \le \big(\mathit{cap}\ K\big) + \big(\mathit{cap}\ L\big). \tag{6.24}$$

Proof. The inequality (6.23) follows directly from (6.6) and the subadditivity of Wiener measure. Now, as in (6.12), we have

$$\begin{aligned}\int U_K(x)\, d\mu_{K\cup L}(x) &= -(U_K, \Delta U_{K\cup L})\\ &= \int U_{K\cup L}(x)\, d\mu_K(x)\\ &= \operatorname{cap} K,\end{aligned} \tag{6.25}$$

the last identity by (6.18), with L replaced by $K \cup L$. Hence

$$\operatorname{cap} K + \operatorname{cap} L = \int \big[U_K(x) + U_L(x)\big]\, d\mu_{K\cup L}(x),$$

so the estimate (6.23) implies (6.24).

Note that even if K and L are disjoint, typically there is inequality in (6.23), hence in (6.24). In fact, if K and L are disjoint compact sets,

$$\begin{aligned}(\operatorname{cap} K) + (\operatorname{cap} L) &= \operatorname{cap}(K \cup L) + R,\\ R = \int_L U_K(x)\, d\mu_{K\cup L}(x) &+ \int_K U_L(x)\, d\mu_{K\cup L}(x),\end{aligned} \tag{6.26}$$

the quantity R being > 0 unless either $\operatorname{cap} K = 0$ or $\operatorname{cap} L = 0$. Unlike measures, the capacity is *not* an additive set function on disjoint compact sets.

We began this section with the statement that the capacity of K is closely related to the probability of a Brownian path hitting K. We have directly tied $U_K(x)$ to this probability, via (6.6). We now provide a two-sided estimate on $U_K(x)$ in terms of cap K.

Proposition 6.6. *Let $\delta(x) = \sup\{|x-y| : y \in K\}$, and let $d(x)$ denote the distance of $x \in \mathbb{R}^n$ from K. Then*

$$\frac{C_n}{\delta(x)^{n-2}}\big(cap\ K\big) \le U_K(x) \le \frac{C_n}{d(x)^{n-2}}\big(cap\ K\big). \tag{6.27}$$

Proof. The formula $U_K(x) = C_n \int |x-y|^{-(n-2)} d\mu_K(y)$ represents $U_K(x)$ as $C_n(\text{cap } K)$ times a weighted average of $|x-y|^{-(n-2)}$ over K. Now, for $y \in K$, $d(x) \le |x-y| \le \delta(x)$, so (6.27) follows.

We want to compare this with the probability that a Brownian path hits ∂K in the interval $[0,t]$. It t is large, we know that $|\omega(t)|$ is probably large, given that $n \ge 3$, and hence $\omega(s)$ probably will not hit K for any $s > t$. Thus we expect this probability (which is equal to $U_K^t(x)$) to be close to $U_K(x)$. We derive a quantitative estimate as follows. Since $1 - U_K^t(x) = e^{t\Delta_{\mathcal{O}}}1(x)$, we have, for $s \ge 0$,

$$U_K^{t+s}(x) - U_K^t(x) = e^{t\Delta_{\mathcal{O}}}1(x) - e^{(t+s)\Delta_{\mathcal{O}}}1(x) = e^{t\Delta_{\mathcal{O}}}U_K^s(x), \tag{6.28}$$

and taking $s \nearrow \infty$, we get

$$U_K(x) - U_K^t(x) = e^{t\Delta_{\mathcal{O}}}U_K(x). \tag{6.29}$$

Hence, if we denote the heat kernel on $\mathcal{O} = \mathbb{R}^n \setminus K$ by $p_{\mathcal{O}}(t,x,y)$, and that on $\mathbb{R}^n$ by $p(t,x-y)$, as in (1.5),

$$\begin{aligned} U_K(x) &- U_K^t(x) \\ &= \int p_{\mathcal{O}}(t,x,y)U_K(y)\,dy \le \int p(t,x-y)U_K(y)\,dy \\ &= C_n \iint \frac{p(t,x-y)}{|y-z|^{n-2}}\,dy\,d\mu_K(z) \le (\text{cap } K)\sigma_K(t,x), \end{aligned} \tag{6.30}$$

where

$$\sigma_K(t,x) = C_n \sup_{z\in K} \int \frac{p(t,x-y)}{|y-z|^{n-2}}\,dy = \sup_{z\in K}\int_t^\infty p(s,x-z)\,ds, \tag{6.31}$$

the last integral being another way of writing $e^{t\Delta}(-\Delta)^{-1}\delta(x-z)$ when $n \ge 3$. An upper bound on $\sigma_K(t,x)$ is $\int_t^\infty (4\pi s)^{-n/2}\,ds$, so we have

$$0 \le U_K(x) - U_K^t(x) \le \frac{2}{n-2}\big(4\pi\big)^{-n/2} t^{-n/2+1}\big(\text{cap } K\big). \tag{6.32}$$

There is an interesting estimate on the smallest eigenvalue of $-\Delta$ on the complement of a compact set K, in terms of cap K, which we now describe. Let $Q = \{x \in \mathbb{R}^n : 0 \le x_j \le 1\}$ be the closed unit cube in $\mathbb{R}^n$, and let $K \subset Q$ be compact. We consider the boundary condition on functions on $Q \setminus K$:

$$u = 0 \text{ on } \partial K, \quad \frac{\partial u}{\partial \nu} = 0 \text{ on } \partial Q \setminus \partial K. \tag{6.33}$$

To define this precisely, let $H^1(Q, K)$ denote the closure in $H^1(Q)$ of the set of functions in $C^\infty(Q)$ vanishing on a neighborhood of K. Then the quadratic form $(du, dv)_{L^2}$ restricted to $H^1(Q, K) \times H^1(Q, K)$ defines an unbounded, self-adjoint operator L, which we denote $-\Delta_{Q,K}$, with $\mathcal{D}(L^{1/2}) = H^1(Q, K) \subset H^1(Q)$. Hence $-\Delta_{Q,K}$ has compact resolvent and thus a discrete spectrum. Let $\lambda_0(K)$ be its smallest eigenvalue.

Proposition 6.7. *The smallest eigenvalue $\lambda_0(K)$ of $-\Delta$ on $Q \setminus K$, with boundary condition (6.33), satisfies the estimate*

$$\lambda_0(K) \ge \gamma_n \text{ cap } K, \tag{6.34}$$

for some $\gamma_n > 0$.

Proof. Let $p_{Q,K}(t, x, y)$ denote the heat kernel of $\Delta_{Q,K}$. With $\mathcal{O} = \mathbb{R}^n \setminus K$, let $p_{\mathcal{O}}(t, x, y)$ denote the heat kernel of Δ on $\mathcal{O}$, with Dirichlet boundary condition, as in (6.30). We claim that

$$\int_Q p_{Q,K}(t, x, y)\, dy \le \int_{\mathbb{R}^n} p_{\mathcal{O}}(t, x, y)\, dy, \quad x \in Q. \tag{6.35}$$

To see this, define $\widetilde{K}$ by the method of images, so in each unit cube with integer vertices we have a reflected image of K, and, with $\widetilde{\mathcal{O}} = \mathbb{R}^n \setminus \widetilde{K}$,

$$p_{Q,K}(t, x, y) = \sum_j p_{\widetilde{\mathcal{O}}}(t, x, R_j y), \quad x, y \in Q, \tag{6.36}$$

where the transformations R_j are appropriate reflections. Then (6.35) follows from the obvious pointwise estimate $p_{\widetilde{\mathcal{O}}}(t, x, y) \le p_{\mathcal{O}}(t, x, y)$. Now, if we set

$$M(t) = \sup_{x \in Q} \int_{\mathbb{R}^n} p_{\mathcal{O}}(t, x, y)\, dy, \tag{6.37}$$

it follows that

$$\sup_x \int_Q p_{Q,K}(t, x, y)\, dy \le M(t), \quad \sup_y \int_Q p_{Q,K}(t, x, y)\, dx \le M(t), \tag{6.38}$$

the latter by symmetry. It is well known that the operator norm of $e^{t\Delta_{Q,K}}$ is bounded by the quantities (6.38). (See Proposition 5.1 in Appendix A.) Thus

$$\|e^{t\Delta_{Q,K}}\| \le M(t). \tag{6.39}$$

To relate this to capacity, note that

$$M(t) = \sup_{x\in Q}\bigl(1 - U_K^t(x)\bigr). \tag{6.40}$$

Now, applying the first estimate of (6.27), in concert with the estimate (6.32), we have

$$M(t) \le 1 - C_n n^{-n/2+1}\bigl(\mathrm{cap}\ K\bigr) + \frac{2}{n-2}(4\pi)^{-n/2}t^{-n/2+1}\bigl(\mathrm{cap}\ K\bigr). \tag{6.41}$$

In particular, there exists a finite $T = T_n$ and $\kappa > 0$ such that

$$M(T) \le 1 - \kappa(\mathrm{cap}\ K) \le e^{-\kappa\ \mathrm{cap}\ K}. \tag{6.42}$$

Since this is an upper bound on $\|e^{T\Delta_{Q,K}}\|$, we have $\lambda_0(K) \ge (\kappa/T)\ \mathrm{cap}\ K$, proving (6.34).

As an application of this, we establish the following result of Molchanov on a class of Dirichlet problems with compact resolvent.

Proposition 6.8. *Let Ω be an unbounded, open subset of $\mathbb{R}^n$, with complement S. Suppose that there exists $\psi(a) \nearrow \infty$ as $a \searrow 0$, such that, for each $a \in (0,1]$, if $\mathbb{R}^n$ is tiled by cubes Q_{aj} of edge a, we have*

$$cap(Q_{aj} \cap S) \ge \psi(a)a^{2(n-2)}, \tag{6.43}$$

for all but finitely many j. Then the Laplace operator Δ on Ω, with Dirichlet boundary condition, has compact resolvent.

Proof. By scaling Q_{aj} to a unit cube, we see that if (6.43) holds, then $-\Delta$ on $Q_{aj} \setminus S$, with Dirichlet boundary condition on ∂S, Neumann on $\partial Q_{aj} \setminus S$, has smallest eigenvalue $\ge \gamma_n(\mathrm{cap}\ Q_{aj} \cap S)a^{-2(n-2)}$, which, by hypothesis (6.43) is $\ge \gamma_n\psi(a)$ for all but finitely many j. The variational characterization of the spectrum implies that the spectral subspace of $L^2(\Omega)$ on which $-\Delta$ has spectrum in $[0, \gamma_n\psi(a)]$ is finite-dimensional, for each $a > 0$, and this implies that Δ has compact resolvent.

In our continued study of which boundary points of a region Ω are regular, it will be useful to have the following variant of Proposition 6.6. Here, B_r is the ball of radius r centered at the origin in $\mathbb{R}^n$; see Fig. 6.1.

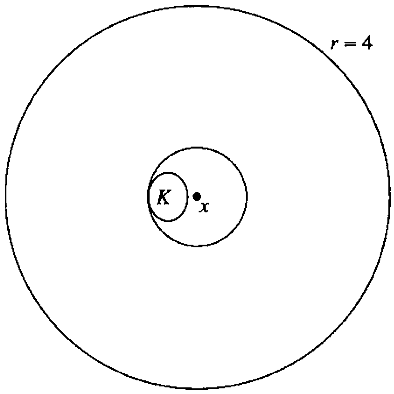

FIGURE 6.1 The Set K

Proposition 6.9. *Let K be a compact subset of the ball B_1. Let $V_K(x)$ denote the probability that a Brownian path, starting at $x \in \mathbb{R}^n$, hits K before hitting the shell $\partial B_4 = \{x : |x| = 4\}$. Then there is a constant $\widetilde{\gamma}_n > 0$ such that*

$$x \in B_1 \Longrightarrow V_K(x) \geq \widetilde{\gamma}_n \big(cap\ K\big). \tag{6.44}$$

Proof. Note that, by (5.20), V_K is also defined by

$$\Delta V_K = 0 \text{ on } B_4 \setminus K, \quad V_K = 1 \text{ on } K, \quad V_K = 0 \text{ on } \partial B_4. \tag{6.45}$$

We will compare $V_K(x)$ with $U_K(x)$. By (6.27), we have

$$x \in B_1 \Longrightarrow U_K(x) \geq 2^{-(n-2)} C_n \big(\text{cap } K\big) \tag{6.46}$$

and

$$x \in \partial B_4 \Longrightarrow U_K(x) \leq 3^{-(n-2)} C_n \big(\text{cap } K\big). \tag{6.47}$$

By (6.47) and the maximum principle, we have, for $x \in B_4 \setminus K$,

$$V_K(x) \geq \frac{U_K(x) - q(K)}{1 - q(K)}, \quad q(K) = 3^{-(n-2)} C_n \big(\text{cap } K\big). \tag{6.48}$$

Now $C_n(\text{cap } K) \leq C_n(\text{cap } B_1) = 1$ (compare with Exercise 1 at the end of this section), so using (6.46) we readily obtain (6.44), with

$$\widetilde{\gamma}_n = \Big(1 - 3^{-(n-2)}\Big)^{-1} \Big(2^{-(n-2)} - 3^{-(n-2)}\Big) C_n. \tag{6.49}$$

In particular, $\widetilde{\gamma}_3 = C_3/4 = \pi$.

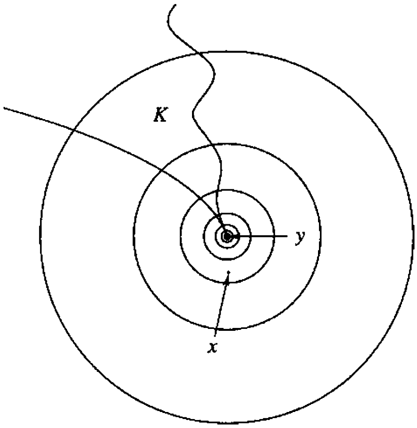

FIGURE 6.2 Setup for the Wiener Test

Of course, since $V_K(x) \le U_K(x)$, we also have

(6.50) $$x \in B_4,\ \mathrm{dist}(x, K) \ge \rho \Longrightarrow V_K(x) \le C_n \rho^{-(n-2)} \big(\mathrm{cap}\ K\big).$$

This upper bound is valid for $K \subset B_4$; we don't need $K \subset B_1$.

Now suppose $y \in K$ is the center of concentric balls B_j, of radius $2^{-j} r$, where $r > 0$ is fixed, $0 \le j \le \nu$. See Fig. 6.2. Pick $x \in B_\nu$. We want to estimate the probability that a Brownian path starting at x will exit B_0 before hitting K. Let's call the probability $p_{\mathrm{miss}}(x, K)$. Using Proposition 6.9 and scaling, we see that, given $x \in B_j$, the probability that it hits ∂B_{j-2} before hitting $K \cap B_j$ is $\le 1 - \widetilde{\gamma}_n r_j^{-(n-2)} \cdot \mathrm{cap}(K \cap B_j)$, where $r_j = 2^{-j} r$. Using the independence of this event and of the event that, given $x \in \partial B_{j-2}$, the path will hit ∂B_{j-4} before hitting $K \cap B_{j-2}$, which follows from the strong Markov property, we have an upper bound

(6.51) $$p_{\mathrm{miss}}(x, K) \le \prod_{j \in \mathcal{S}_\nu} \Big(1 - \widetilde{\gamma}_n r^{-(n-2)} 2^{(n-2)j} \cdot \mathrm{cap}\big(K \cap B_j\big)\Big),$$

where $\mathcal{S}_\nu = \{j : 0 \le j \le \nu,\ j = \nu \bmod 2\}$. A similar argument dominates $p_{\mathrm{miss}}(x, K)$ by a product over $\{1, \dots \nu\} \setminus \mathcal{S}_\nu$, so

(6.52) $$p_{\mathrm{miss}}(x, K)^2 \le \prod_{j=0}^{\nu} \Big(1 - \widetilde{\gamma}_n r^{-(n-2)} 2^{(n-2)j} \cdot \mathrm{cap}\big(K \cap B_j\big)\Big).$$

Note that, as $\nu \to \infty$, the right side of (6.52) tends to zero, precisely when the sum

(6.53) $$\sum_{j=0}^{\infty} 2^{(n-2)j} \cdot \mathrm{cap}\big(K \cap B_j\big)$$

is *infinite*. We are now ready to state the Wiener criterion for regular points.

Proposition 6.10. *Let Ω be a bounded, open set in $\mathbb{R}^n$, and let $y \in \partial\Omega$. If Ω is inside a ball $\widetilde{B}$, set $K = \widetilde{B} \setminus \Omega$. Then y is a regular point for Ω if and only if the infinite series (6.53) is divergent, where $B_j = \{x \in \mathbb{R}^n : |x - y| \le 2^{-j}\}$.*

Proof. First suppose (6.53) is divergent. Fix $f \in C(\partial\Omega)$, and look at

$$u(x) = \mathrm{PI}\, f(x) = E_x\Big(f\big(\omega(\mathfrak{h}_K)\big)\Big). \tag{6.54}$$

Given $\varepsilon > 0$, fix $r > 0$ so that f varies by less than ε on $\{z \in \partial\Omega : |z - y| \le r\}$. By (6.52), if $\delta > 0$ is small enough and $|x - y| \le \delta$, then the probability that a Brownian path $\omega(t)$, starting at x, crosses $\partial B_0 = \{z : |z - y| = r\}$ before hitting K is $< \varepsilon$. Consequently,

$$|x - y| \le \delta \Longrightarrow \Big|E_x\Big(f\big(\omega(\mathfrak{h}_K)\big)\Big) - f(y)\Big| \le \varepsilon + \varepsilon \cdot \sup\, |f|. \tag{6.55}$$

This shows that $PI\ f(x) \to f(y)$ as $x \to y$, for any $f \in C(\partial\Omega)$, so y is regular.

For the converse, if (6.53) converges, we claim there is a $J < \infty$ such that there exist points in $\Omega \cap B_J$, arbitrarily close to y, which are starting points of Brownian paths whose probability of hitting K before exiting B_J is $\le 1/2$.

Consider the shells $A_j = \{x : 2^{-j-1} \le |x - y| \le 2^{-j}\}$; $B_j = \bigcup_{\ell \ge j} A_\ell$. We will estimate the probability that a point picked at random in A_ℓ is the starting point of a Brownian path that hits K before exiting B_J, where ℓ is chosen $> J$. Since we are assuming $n \ge 3$, by the analysis behind nonrecurrence in Exercises 7–10 of §3, the probability that a path starting in A_ℓ *ever* hits $B_{\ell+3}$ is $\le 1/4$. Thus if we alter K to $K_\ell = K \setminus B_{\ell+3}$, the probability that a Brownian path starting in A_ℓ hits K_ℓ before ∂B_J is not decreased by more than $1/4$. We aim to show that this new probability is $\le 1/4$ if J is chosen large enough.

Now there is no further decrease in probability that the path hits K_ℓ before ∂B_J if we instead have it start at a random point in $B_{\ell+5}$, since almost all such paths will pass into A_ℓ, in a uniformly distributed fashion through its inner boundary. So we deal with the modified problem of estimating the probability $\widetilde{p}$ that a Brownian path, starting at a random point in $B_{\ell+5}$, hits $K_\ell = K \setminus B_{\ell+3}$ before exiting B_J.

We partition the set $\{j : J \le j \le \ell + 3\}$ into two sets, where j is even or odd; call these subsets $\mathcal{J}_0$ and $\mathcal{J}_1$, respectively. Then form

$$\mathcal{A}_0 = \bigcup_{j \in \mathcal{J}_0} A_j, \quad \mathcal{A}_1 = \bigcup_{j \in \mathcal{J}_1} A_j. \tag{6.56}$$

We estimate the probability p_μ that a path starting in $B_{\ell+5}$ hits $K_\ell \cap \mathcal{A}_\mu$ before hitting ∂B_J. We have

$$p_\mu(x) \le \sum_{j \in \mathcal{J}_\mu} p_{\mu j}, \tag{6.57}$$

where $p_{\mu j}$ is the probability that, given $|x - y| = (3/4) \cdot 2^{-j-1}$ (i.e., x is on a shell S_{j+1} halfway between the two boundary components of A_{j+1}), then a path starting at x hits $K \cap A_j$ before hitting S_{j-1}. By (6.50) and a dilation argument, we have an estimate of the form

$$p_{\mu j} \le \gamma_n' 2^{(n-2)j} \operatorname{cap}(K \cap A_j). \tag{6.58}$$

Thus the probability $\widetilde{p}$ that we want to estimate satisfies

$$\widetilde{p} \le \gamma_n' \sum_{j=J}^{\ell+3} 2^{(n-2)j} \operatorname{cap}(K \cap A_j). \tag{6.59}$$

Of course, $\operatorname{cap}(K \cap A_j) \le \operatorname{cap}(K \cap B_j)$, so if (6.53) is assumed to converge, we can pick J sufficiently large that the right side of (6.59) is guaranteed to be $\le 1/4$.

From here it is easy to pick $f \in C(\partial\Omega)$ such that $f(y) = 1$ but (6.54) does not converge to 1 as $x \to y$. This completes the proof of Proposition 6.10 and also shows that the hypothesis of convergence or divergence of (6.53) can be replaced by such a hypothesis on

$$\sum_{j=0}^{\infty} 2^{(n-2)j} \cdot \operatorname{cap}\big(K \cap A_j\big). \tag{6.60}$$

We can extend capacity to arbitrary sets $S \subset \mathbb{R}^n$. The *inner capacity* $\operatorname{cap}^-(S)$ is defined by

$$\operatorname{cap}^-(S) = \sup\,\{\operatorname{cap} K : K \text{ compact}, K \subset S\}. \tag{6.61}$$

Clearly, $\operatorname{cap}^-(K) = \operatorname{cap} K$ for compact K. If $U \subset \mathbb{R}^n$ is open, we also set $\operatorname{cap} U = \operatorname{cap}^-(U)$. Now the *outer capacity* $\operatorname{cap}^+(S)$ is defined by

$$\operatorname{cap}^+(S) = \inf\,\{\operatorname{cap} U : U \text{ open}, S \subset U\}. \tag{6.62}$$

It is easy to see that $\operatorname{cap}^+(S) \ge \operatorname{cap}^-(S)$ for all S. If $\operatorname{cap}^+(S) = \operatorname{cap}^-(S)$, then S is said to be capacitable, and the common quantity is denoted $\operatorname{cap} S$. The analysis leading to (6.16) shows that every compact set is capacitable; also, by definition, every open set is capacitable. G. Choquet proved that every Borel set is capacitable; in fact, his capacitability theorem extends to a more general class of sets, known as Souslin sets. We refer to [Mey] for a detailed presentation of this result.

The outer capacity can be shown to satisfy the property that, for any increasing sequence of sets $S_j \subset \mathbb{R}^n$,

$$S_j \nearrow S \Longrightarrow \operatorname{cap}^+(S_j) \nearrow \operatorname{cap}^+(S).$$

We establish a useful special case of this.

Proposition 6.11. *If U_j and U are open and $U_j \nearrow U$, then*

$$\text{cap}\, U_j \nearrow \text{cap}\, U.$$

Proof. Given $\varepsilon > 0$, pick a compact $K \subset U$ such that cap $K \geq$ cap $U - \varepsilon$. Then $K \subset U_j$ for large j, so cap $U_j \geq$ cap $U - \varepsilon$ for large j.

We next present a result, due to M. Brelot, to the effect that the set of irregular boundary points of a given bounded, open set is rather small.

Proposition 6.12. *If $\Omega \subset \mathbb{R}^n$ is open and bounded, the set I of irregular boundary points in $\partial\Omega$ has inner capacity zero.*

Proof. The claim is that if $K \subset I$ is compact, then cap $K = 0$. By subadditivity, it suffices to show the following: Given $y \in \partial\Omega$, there is a neighborhood B of y in $\mathbb{R}^n$ such that any compact $K \subset I \cap B$ has capacity zero.

We prove the result in the case that Ω is connected. Let $L = \overline{B} \setminus \Omega$, and consider the capacitary potential $U_L(x)$. In this case, $\mathbb{R}^n \setminus L$ is connected. The function $1 - U_L(x)$ is a weak barrier at any $z \in L \cap \partial\Omega$ with the property that $U_L(x) \to 1$ as $x \to z$, $x \in \mathbb{R}^n \setminus L$. Thus it suffices to show that the set $J = \{z \in L : U_L(z) < 1\}$ has inner capacity zero.

Let $K \subset J$ be compact. We know that $U_K(x) \leq U_L(x)$ for all $x \in \mathbb{R}^n$. Thus $U_K(x) < 1$ on K. Now, by (6.15), $U_K(x) = 1$ for μ_K-almost all x, so we conclude that $\mu_K = 0$, hence cap $K = 0$. This completes the proof when Ω is connected.

The general case can be done as follows. If Ω is not connected, it has at most countably many connected components. One can connect the various components via little tubes whose total (inner) capacity can be arranged, via Proposition 6.11, to be arbitrarily small, say $< \varepsilon$. Then the set of irregular points is decreased by a set of inner capacity $< \varepsilon$. The reader is invited to supply the details.

As noted in Proposition 5.5, the set of irregular points of $\partial\Omega$ can be characterized as the set of points of discontinuity of a function E, defined on $\overline{\Omega}$ to be $-\Delta^{-1}1(x)$ for $x \in \Omega$ and to be 0 on $\partial\Omega$. Such a set of points of discontinuity is a Borel subset of Ω, in fact an $\mathcal{F}_{\sigma\delta}$-set. Thus the capacitability theorem applies: If $\Omega \subset \mathbb{R}^n$ is a bounded open set, the set of irregular points of $\partial\Omega$ has capacity zero. This sharpening of Proposition 6.12 was first established by H. Cartan.

As we stated at the beginning of this section, we have been working under the assumption that $n \geq 3$. Two phenomena that we have exploited fail when $n = 2$. One is that Δ has a fundamental solution ≤ 0 on all of $\mathbb{R}^n$. The other is that Brownian motion is nonrecurrent. (Of course, these two phenomena are related.) There is a theory of logarithmic capacity of planar sets. One way to approach things is to consider capacities only of subsets of some fixed disk, of large radius R, and use the Laplace operator on this disk, with the Dirichlet boundary condition. Then one looks at Brownian paths only up to the first exit time from this disk. The results of this section extend. In particular, the Wiener criterion for $n = 2$ is the convergence or divergence of

$$\sum_{j=1}^{\infty} j \cdot \operatorname{cap}\big(K \cap A_j\big). \tag{6.63}$$

Exercises

1. If $K \subset \mathbb{R}^n$ is compact, show that

$$\lim_{|x|\to\infty} |x|^{n-2} U_K(x) = C_n \operatorname{cap} K.$$

If $K = B_a$ is a ball of radius a, show that $\operatorname{cap} B_a = a^{n-2}/C_n$.
Show generally that if $a > 0$ and $K_a = \{ax : x \in K\}$, then $\operatorname{cap} K_a = a^{n-2} \operatorname{cap} K$.

2. Show that $\operatorname{cap} K = \operatorname{cap} \partial K$. Show that the identity $\operatorname{cap} \partial B_a = a^{n-2}/C_n$ follows from (6.27), with x the center of B_a.
3. Let C_{ar} be the union of two balls of radius a, with centers separated by a distance r. Show that

$$\operatorname{cap} C_{ar} \nearrow 2 \operatorname{cap} B_a, \quad \text{as } r \to \infty.$$

Estimate the rate of convergence.

4. The task here is to estimate the capacity of a cylinder in $\mathbb{R}^n$, of height b and radius a. Suppose $\mathcal{C}(a, b) = \{x \in \mathbb{R}^n : 0 \le x_n \le b,\ x_1^2 + \cdots + x_{n-1}^2 \le a^2\}$. Show that there are positive constants α_n and β_n such that

$$\operatorname{cap} \mathcal{C}(a, 1) \sim \alpha_n a^{n-3}, \quad a \to 0,\ n \ge 4,$$
$$\operatorname{cap} \mathcal{C}(a, 1) \sim \beta_n a^{n-2}, \quad a \to \infty,\ n \ge 3.$$

Derive an appropriate result for $n = 3,\ a \to 0$.

5. Let ν be a positive measure supported on a compact set $K \subset \mathbb{R}^n$, such that

$$U_\nu(x) = -\Delta^{-1}\nu(x) = C_n \int \frac{d\nu(x)}{|x - y|^{n-2}} \le 1.$$

Show that $U_\nu(x) \le U_K(x)$ for all $x \in \mathbb{R}^n$. Taking the limit as $|x| \to \infty$, deduce from the asymptotic behavior of $U_\nu(x)$ and $U_K(x)$ (as in Exercise 1) that $\int d\nu(x) \le \operatorname{cap} K$.

6. Show that, for compact $K \subset \mathbb{R}^n$,

$$\operatorname{cap} K = \inf\Big\{\int |\nabla f(x)|^2\, dx : f \in C_0^\infty(\mathbb{R}^n),\ f = 1 \text{ on nbd of } K\Big\}. \tag{6.64}$$

(*Hint*: Show that a minimizing sequence f_j approaches U_K.)
Show that the condition $f = 1$ on a neighborhood of K can be replaced by $f \ge 1$ on K. Show that if $f \in C_0^1(\mathbb{R}^n),\ \lambda > 0$,

$$\operatorname{cap}\big(\{x \in \mathbb{R}^n : |f(x)| \ge \lambda\}\big) \le \lambda^{-2} \|\nabla f\|_{L^2}^2. \tag{6.65}$$

7. Show that, for compact $K \subset \mathbb{R}^n$,

$$\frac{1}{\operatorname{cap} K} = \inf\Big\{C_n \iint \frac{d\lambda(x)\, d\lambda(y)}{|x - y|^{n-2}} : \lambda \in \mathcal{P}_K^+\Big\}, \tag{6.66}$$

where $\mathcal{P}_K^+$ denotes the space of probability measures supported on K.
(*Hint*: Consider the sesquilinear form

$$\gamma(\mu,\lambda) = C_n \iint |x-y|^{-n+2}\, d\mu(x)\, d\overline{\lambda}(y) = -(\Delta^{-1}\mu,\lambda)$$

as a (positive-definite) inner product on the Hilbert space $H_K^{-1}(\mathbb{R}^n) = \{u \in H^{-1}(\mathbb{R}^n) : \operatorname{supp} u \subset K\}$. Thus

$$|\gamma(\mu,\lambda)| \le \gamma(\mu,\mu)^{1/2}\gamma(\lambda,\lambda)^{1/2}.$$

Take $\mu = (\text{cap } K)^{-1}\mu_K \in \mathcal{P}_K^+$, where μ_K is the measure in (6.8)–(6.10). Show that (at least, when ∂K is smooth),

$$\lambda \in \mathcal{P}_K^+ \cap H_K^{-1}(\mathbb{R}^n) \Longrightarrow \gamma(\mu,\lambda) = \frac{1}{\text{cap } K}\int U_K(y)\, d\lambda(y) = \frac{1}{\text{cap } K},$$

and conclude that $\gamma(\lambda,\lambda) \ge 1/(\text{cap } K)$. Then use some limiting arguments.)

8. If $K \subset \mathbb{R}^3$ is compact, relate cap K to the zero frequency limit of the scattering amplitude, defined in Chap. 9, §1.
9. Try to establish directly the equivalence between the regularity criteria given by Propositions 5.5 and 6.10.
10. In Chap. 5, §5, a compact set $K \subset \mathbb{R}^n$ was called "negligible" provided there is no nonzero $u \in H^{-1}(\mathbb{R}^n)$ supported on K. Show that if K is negligible, then cap $K = 0$. Try to prove the converse.
11. Sharpen the subadditivity result (6.24) to

$$\text{cap}(K\cup L) + \text{cap}(K\cap L) \le (\text{cap } K) + (\text{cap } L),$$

for compact sets K and L. This property is called "strong subadditivity."
(*Hint*: By (6.6), $U_K(x) = W_x(S_K)$, where $S_K = \{\omega : \mathfrak{h}_K(\omega) < \infty\}$. Show that $S_{K\cup L} = S_K \cup S_L$ and $S_{K\cap L} = S_K \cap S_L$, and deduce that

$$U_{K\cup L}(x) + U_{K\cap L}(x) \le U_K(x) + U_L(x).$$

Extending the reasoning used in the proof of Proposition 6.5, deduce that

$$\begin{aligned}\text{cap } K + \text{cap } L &= \int \Big[U_K(x) + U_L(x)\Big]\, d\mu_{K\cup L}(x) \\ &\ge \int \Big[U_{K\cup L}(x) + U_{K\cap L}(x)\Big]\, d\mu_{K\cup L}(x) \\ &= \text{cap}(K\cup L) + \text{cap}(K\cap L).)\end{aligned}$$

7. Stochastic integrals

We will motivate the introduction of the stochastic integral by modifying the Feynman–Kac formula, to produce a formula for the solution operator $e^{t(\Delta+X)}$ to

$$\frac{\partial u}{\partial t} = \Delta u + Xu, \quad u(0) = f; \quad Xu = \sum X_j(x) \frac{\partial u}{\partial x_j}. \tag{7.1}$$

As in (2.2), we use the Trotter product formula to write

$$e^{t(\Delta+X)} f = \lim_{k\to\infty} \Big(e^{(t/k)X} e^{(t/k)\Delta} \Big)^k f. \tag{7.2}$$

If we assume that each coefficient X_j of the vector field X is bounded and uniformly Lipschitz, then Proposition A.2 applies to (7.2), given $f \in L^p(\mathbb{R}^n)$, $1 \le p < \infty$, or $f \in C_o(\mathbb{R}^n)$, in view of Proposition 9.13 in Appendix A. Now, for any k, $\big(e^{(t/k)X} e^{(t/k)\Delta}\big)^k f$ can be expressed as a k-fold integral:

$$\begin{aligned} &\Big(e^{(t/k)X} e^{(t/k)\Delta} \Big)^k f(x) \\ &\quad = \int \cdots \int f\big(x_k\big) p\big(\frac{t}{k}, x_k - x_{k-1} - \frac{t}{k}\xi_{k-1}\big) \\ &\qquad \cdots p\big(\frac{t}{k}, x_2 - x_1 - \frac{t}{k}\xi_1\big) p\big(\frac{t}{k}, x_1 - x - \frac{t}{k}\xi_0\big)\, dx_1 \cdots dx_k, \end{aligned} \tag{7.3}$$

where (with $x_0 = x$)

$$\xi_j = X(x_j) + r_j, \quad r_j = O(k^{-1}). \tag{7.4}$$

Now we can write

$$p\big(\frac{t}{k}, x_{j+1} - x_j - \frac{t}{k}\xi_j\big) = p\big(\frac{t}{k}, x_{j+1} - x_j\big)\, e^{\xi_j\cdot(x_{j+1}-x_j)/2-(t/k)|\xi_j|^2/4}. \tag{7.5}$$

Consequently, parallel to (2.4),

$$\Big(e^{(t/k)X} e^{(t/k)\Delta} \Big)^k f(x) = E_x(\varphi_k), \tag{7.6}$$

where

$$\varphi_k(\omega) = f\big(\omega(t)\big) e^{A_k(\omega)-B_k(\omega)}, \tag{7.7}$$

with

$$\begin{aligned} A_k(\omega) &= \frac{1}{2} \sum_{j=0}^{k-1} \Big[X\big(\omega\big(\frac{j}{k}t\big)\big) + r_j \Big] \cdot \Big[\omega\big(\frac{j+1}{k}t\big) - \omega\big(\frac{j}{k}t\big) \Big] \\ B_k(\omega) &= \frac{1}{4}\frac{t}{k} \sum_{j=0}^{k-1} \Big[X\big(\omega\big(\frac{j}{k}t\big)\big) + r_j \Big]^2. \end{aligned} \tag{7.8}$$

Thus we expect to establish a formula of the form

$$e^{t(\Delta+X)} f(x) = E_x\Big(f\big(\omega(t)\big)e^{A(t,\omega)-B(t,\omega)}\Big), \tag{7.9}$$

where

$$B(t,\omega) = \frac{1}{4}\int_0^t X\big(\omega(s)\big)^2\, ds, \tag{7.10}$$

and

$$A(t,\omega) = \frac{1}{2}\lim_{k\to\infty}\sum_{j=0}^{k-1} X\big(\omega(\tfrac{j}{k}t)\big)\cdot\Big[\omega\big(\frac{j+1}{k}t\big) - \omega\big(\frac{j}{k}t\big)\Big]. \tag{7.11}$$

In (7.10), $X(\omega)^2$ denotes $\sum X_j(\omega)^2$. If the coefficients X_j are real-valued, this is equal to $|X(\omega)|^2$.

Certainly $B_k(\omega) \to B(t,\omega)$ nicely for all $\omega \in \mathfrak{P}_0$. The limit we now need to investigate is (7.11), which we would like to write as

$$A(t,\omega) = \frac{1}{2}\int_0^t X\big(\omega(s)\big)\cdot d\omega(s). \tag{7.12}$$

However, $\omega(s)$ has unbounded variation for W_x-almost all ω, so there remains some analysis to be done on this object, which is a prime example of a stochastic integral.

We aim to make sense out of stochastic integrals of the form

$$\int_0^t g\big(s,\omega(s)\big)\cdot d\omega(s), \tag{7.13}$$

beginning with

$$\int_0^t g(s)\cdot d\omega(s) = \lim_{k\to\infty}\sum_{j=0}^{k-1} g\big(\frac{j}{k}t\big)\cdot\Big[\omega\big(\frac{j+1}{k}t\big) - \omega\big(\frac{j}{k}t\big)\Big]. \tag{7.14}$$

This is readily seen to be well defined in $L^2(\mathfrak{P}_0, dW_x)$, in view of the fact that the terms $\theta_j(\omega) = \omega\big((j+1)t/k\big) - \omega\big(jt/k\big)$ satisfy

$$\|\theta_j\|^2_{L^2(\mathfrak{P}_0,dW_x)} = 2\frac{t}{k}, \quad (\theta_j,\theta_\ell)_{L^2(\mathfrak{P}_0,dW_x)} = 0, \ \text{ for } j \neq \ell, \tag{7.15}$$

the first by (1.38). Thus

$$\Big\| \sum_{j=0}^{k-1} g\big(\frac{j}{k}t\big)\Big[\omega\big(\frac{j+1}{k}t\big) - \omega\big(\frac{j}{k}t\big)\Big]\Big\|^2_{L^2(\mathfrak{P}_0, dW_x)} = 2\sum_{j=0}^{k-1} \frac{t}{k}\big|g\big(\frac{j}{k}t\big)\big|^2. \tag{7.16}$$

For continuous g, this is a Riemann sum approximating $\int_0^t |g(s)|^2\, ds$, as $k \to \infty$. Thus we obtain the following:

Proposition 7.1. *Given $g \in C\big([0,t]\big)$, the right side of (7.14) converges in $L^2(\mathfrak{P}_0, dW_x)$. The resulting correspondence*

$$g \mapsto \int_0^t g(s)\, d\omega(s)$$

extends uniquely to $\sqrt{2}$ times an isometry of $L^2\big([0,t], dt\big)$ into $L^2(\mathfrak{P}_0, dW_x)$.

We next consider

$$S_k(\omega) = \sum_{j=0}^{k-1} g\big(t_j, \omega(t_j)\big)\cdot\big[\omega(t_{j+1}) - \omega(t_j)\big] = \sum_{j=0}^{k-1} g_j(\omega)\cdot\theta_j(\omega), \tag{7.17}$$

where $\theta_j(\omega) = \omega(t_{j+1}) - \omega(t_j)$, $t_j = (j/k)t$. Following [Si], Chap. 5, we compute

$$\|S_k\|^2_{L^2(\mathfrak{P}_0, dW_x)} = \sum_{j,\ell} E_x\Big(g_j(\omega)\theta_j(\omega)g_\ell(\omega)\theta_\ell(\omega)\Big). \tag{7.18}$$

If $\ell > j$, $\theta_\ell(\omega) = \omega(t_{\ell+1}) - \omega(t_\ell)$ is independent of the other factors in parentheses on the right side of (7.18), so the expectation of the product is equal to $E_x(g_j\theta_j g_\ell)E_x(\theta_\ell) = 0$ since $E_x(\theta_\ell) = 0$. Similarly the terms in the sum in (7.18) vanish when $\ell < j$, so

$$\begin{aligned} \|S_k\|^2_{L^2(\mathfrak{P}_0, dW_x)} &= \sum_j E_x\big(|g_j(\omega)|^2\big)E_x\big(|\theta_j|^2\big) \\ &= 2\sum_j E_x\big(|g(t_j, \omega(t_j))|^2\big)(t_{j+1} - t_j). \end{aligned} \tag{7.19}$$

If g and ω are continuous, this is a Riemann sum approximating the integral $2\int_0^t E_x\big(|g(s,\omega(s))|^2\big)\, ds$, and we readily obtain the following result.

Proposition 7.2. *Given $g \in BC\big([0,t]\times\mathbb{R}^n\big)$, the expression (7.17) converges as $k \to \infty$, in $L^2(\mathfrak{P}_0, dW_x)$, to a limit we denote by (7.13). Furthermore, the map*

$$g \mapsto \int_0^t g\big(s, \omega(s)\big)\cdot d\omega(s)$$

is $\sqrt{2}$ times an isometry into $L^2(\mathfrak{P}_0, dW_x)$, when g has the square norm

$$Q_x(g) = \int_0^t E_x\Big(\big|g\big(s, \omega(s)\big)\big|^2\Big)\, ds. \tag{7.20}$$

Note that $Q_x(g) = \int_0^t \int_{\mathbb{R}^n} |g(s,y)|^2\, p(s, x-y)\, dy\, ds$. In case $g = g(\omega(s))$, we have $Q_x(g)$ given as the square of a weighted L^2-norm:

$$Q_x(g) = \int\limits_{\mathbb{R}^n} |g(y)|^2\, r_t(x-y)\, dy = R_t(D)|g|^2(x), \tag{7.21}$$

where

$$R_t(D) = \Delta^{-1}(e^{t\Delta} - I), \quad r_t(x) = R_t(D)\delta(x). \tag{7.22}$$

We see that $R_t(D) \in OPS^{-2}(\mathbb{R}^n)$. The convolution kernel $r_t(x)$ is smooth on $\mathbb{R}^n \setminus 0$ and rapidly decreasing as $|x| \to \infty$. More precisely, one easily verifies that

$$r_t(x) \le C(n,t)|x|^{-2} e^{-|x|^2/4t}, \ \text{ for } |x| \ge \frac{1}{2}, \tag{7.23}$$

and

$$r_t(x) \le C(n,t)|x|^{2-n}, \ \text{ for } |x| \le \frac{1}{2}, \quad n \ge 3, \tag{7.24}$$

with $|x|^{2-n}$ replaced by $\log 1/|x|$ for $n = 2$ and by 1 for $n = 1$. Of course, $r_t(x) > 0$ for all $t > 0$, $x \in \mathbb{R}^n \setminus 0$.

In particular, the integral in (7.21) is absolutely convergent and $Q_x(g)$ is a continuous function of x provided

$$g \in L^p_{\text{loc}}(\mathbb{R}^n), \ \text{for some } p > n, \ \text{ and } \ g \in L^2\big(\mathbb{R}^n, \langle x\rangle^{-2} e^{-|x|^2/4t}\, dx\big). \tag{7.25}$$

Proposition 7.2 is adequate to treat the case where the coefficients X_j are in $BC(\mathbb{R}^n)$ and purely imaginary. Since $A_k(\omega) \to A(t,\omega)$ in $L^2(\mathfrak{P}_0, dW_x)$,

$$e^{A_k(\omega)} \longrightarrow e^{A(t,\omega)} \text{ in measure,} \tag{7.26}$$

and boundedly, since the terms in (7.26) all have absolute value 1. Then convergence of (7.6) follows from the dominated convergence theorem. In such a case, $X(\omega)^2$ in (7.10) is equal to $-|X(\omega)|^2$. We have the following.

Proposition 7.3. *If $X = iY$ is a vector field on $\mathbb{R}^n$ with coefficients that are bounded, continuous, and purely imaginary, then*

(7.27)
$$e^{t(\Delta+iY)}f(x)=E_x\Bigl(f\bigl(\omega(t)\bigr)e^{(i/2)\int_0^t Y(\omega(s))\cdot d\omega(s)+(1/4)\int_0^t |Y(\omega(s))|^2\,ds}\Bigr).$$

One final ingredient is required to prove Proposition 7.3, since in this case e^{tX} is not a semigroup of bounded operators, so we cannot apply Proposition A.2. However, we can apply Proposition A.3, with

$$S(t)f(x)=\int f(y)p\bigl(t,y-x-tX(x)\bigr)\,dy.$$

If $X=iY$ is purely imaginary, then, parallel to (7.5), we have

$$p\bigl(t,y-x-itY(x)\bigr)=p(t,y-x)e^{iY(x)\cdot(y-x)/2+t|Y(x)|^2/4}.$$

If V is bounded and continuous, a simple modification of the analysis above, combining techniques of §2, yields

$$e^{t(\Delta+X-V)}f(x)=E_x\Bigl(f\bigl(\omega(t)\bigr)e^{A(t,\omega)/2-B(t,\omega)/4-\int_0^t V(\omega(s))\,ds}\Bigr) \tag{7.28}$$

when X is purely imaginary. For another interpretation of this, consider

$$\begin{aligned} H&=\sum_j\Bigl(-i\frac{\partial}{\partial x_j}-A_j(x)\Bigr)^2+V\\ &=-\sum_j\Bigl(\frac{\partial^2}{\partial x_j^2}-2iA_j\frac{\partial}{\partial x_j}-i\frac{\partial A_j}{\partial x_j}-A_j^2\Bigr)+V.\end{aligned} \tag{7.29}$$

Assume each A_j is real-valued, and $A_j,\partial A_j/\partial x_j\in BC(\mathbb{R}^n)$. Then

$$\begin{aligned} e^{-tH}f(x)&=E_x\Bigl(f\bigl(\omega(t)\bigr)e^{S(t,\omega)}\Bigr),\\ S(t,\omega)&=i\int_0^t A\bigl(\omega(s)\bigr)\cdot d\omega(s)\\ &\qquad -i\int_0^t\bigl(\operatorname{div}A\bigr)\bigl(\omega(s)\bigr)\,ds-\int_0^t V\bigl(\omega(s)\bigr)\,ds.\end{aligned} \tag{7.30}$$

Compare with the derivation in [Si], Chap. 5.

If the coefficients of X are not assumed to be purely imaginary, we need some more estimates. More generally, we will derive further estimates on the approximants $S_k(\omega)$ to $\int_0^t g(s,\omega(s))\cdot d\omega(s)$, defined by (7.17).

Lemma 7.4. *If g is bounded and continuous, then*

$$E_x\bigl(e^{S_k}\bigr)\le e^{t\gamma^2},\quad \gamma=\|g\|_{L^\infty}, \tag{7.31}$$

and

$$E_x\big(e^{\lambda|S_k|}\big) \le 2e^{t\lambda^2\gamma^2}. \tag{7.32}$$

Proof. The left side of (7.31) is

$$\begin{aligned} E_x\big(e^{g_0\theta_0}\cdots e^{g_{k-1}\theta_{k-1}}\big) &= \sum_{\nu=0}^{\infty} \frac{1}{\nu!} E_x\Big(e^{g_0\theta_0}\cdots e^{g_{k-2}\theta_{k-2}} g_{k-1}^{\nu}\theta_{k-1}^{\nu}\Big) \\ &\le \sum_{\nu=0}^{\infty} \frac{\gamma^\nu}{\nu!} E_x\big(e^{g_0\theta_0}\cdots e^{g_{k-2}\theta_{k-2}}\big) E_x\big(\theta_{k-1}^{\nu}\big) \\ &= E_x\big(e^{g_0\theta_0}\cdots e^{g_{k-2}\theta_{k-2}}\big) E_x\big(e^{\gamma\theta_{k-1}}\big), \end{aligned} \tag{7.33}$$

by independence arguments such as used in the analysis of (7.18). Note that the sums over ν above have terms that vanish for odd ν. Now $E_x(e^{\gamma\theta_j}) = e^{(t_{j+1}-t_j)\gamma^2}$. An inductive argument leads to (7.31), and (7.32) follows from this plus $e^{|u|} \le e^u + e^{-u}$.

We next estimate the $L^2(\mathfrak{P}_0, dW_x)$-norm of $S_{2k} - S_k$. Another calculation, parallel to (7.18)–(7.19), yields

$$\begin{aligned} &\|S_{2k} - S_k\|^2_{L^2(\mathfrak{P}_0,dW_x)} \\ &= \sum_j E_x\Big(\big|g\big(t_{j+1/2},\omega(t_{j+1/2})\big) - g\big(t_j,\omega(t_j)\big)\big|^2\Big)(t_{j+1}-t_{j+1/2}), \end{aligned} \tag{7.34}$$

where $t_j = jt/k$ as in (7.17), and $t_{j+1/2} = (j+1/2)t/k$. If we assume a Lipschitz condition on g, we obtain the following estimate.

Lemma 7.5. *Assume that*

$$|g(t,x) - g(s,y)|^2 \le C_0|t-s|^2 + C_1|x-y|^2. \tag{7.35}$$

Then

$$\|S_{2k} - S_k\|^2_{L^2(\mathfrak{P}_0,dW_x)} \le C_0\frac{t^2}{k^2} + 2C_1\frac{t}{k}. \tag{7.36}$$

Proof. This follows from (7.34) plus $E_x\big(|\omega(t)-\omega(s)|^2\big) = 2|t-s|$.

We can now make an estimate directly relevant to the limiting behavior of (7.7).

Lemma 7.6. *Given the bound $\|g\|_{L^\infty} \le \gamma$, we have*

$$\|e^{S_{2k}} - e^{S_k}\|_{L^1(\mathfrak{P}_0,dW_x)} \le \sqrt{2}\|S_{2k} - S_k\|_{L^2(\mathfrak{P}_0,dW_x)} e^{32t\gamma^2}. \tag{7.37}$$

Proof. Using $e^u - e^v = (u-v)\Phi(u,v)$, with $|\Phi(u,v)| \le e^{2|u|+2|v|}$, we have

$$\|e^{S_{2k}} - e^{S_k}\|_{L^1(\mathfrak{P}_0)} \le \|S_{2k} - S_k\|_{L^2(\mathfrak{P}_0)} \cdot \|e^{4|S_{2k}|+4|S_k|}\|^{1/2}_{L^1(\mathfrak{P}_0)}; \tag{7.38}$$

and the estimate (7.32), plus $2e^{u+v} \le e^{2u} + e^{2v}$, then yields (7.37).

With these estimates, we can pass to the limit in (7.6)–(7.7), obtaining the following result.

Proposition 7.7. *If X is a real vector field on $\mathbb{R}^n$ whose coefficients are bounded and uniformly Lipschitz, and if $f \in C_0^\infty(\mathbb{R}^n)$, then*

$$e^{t(\Delta+X)}f(x) = E_x\Big(f\big(\omega(t)\big)e^{(1/2)\int_0^t X(\omega(s))\cdot d\omega(s) - (1/4)\int_0^t |X(\omega(s))|^2\,ds}\Big). \tag{7.39}$$

Now that the identity (7.39) is established for X and f such as described above, one can use limiting arguments to extend the identity to more general cases. Such extensions are left to the reader.

We now evaluate the stochastic integral $\int_0^t \omega(s)\,d\omega(s)$ in the case of one-dimensional Brownian motion. One might anticipate that it should be $\omega(t)^2/2 - \omega(0)^2/2$. However, guesses based on what should happen if ω had bounded variation can be misleading, and the truth is a little stranger. Let us begin with

$$\begin{aligned}\omega(t)^2 - \omega(0)^2 &= \sum_{j=0}^{k-1}\big[\omega(t_{j+1})^2 - \omega(t_j)^2\big] \\ &= \sum_j \big[\omega(t_{j+1}) + \omega(t_j)\big]\cdot\big[\omega(t_{j+1}) - \omega(t_j)\big],\end{aligned} \tag{7.40}$$

where $t_j = (j/k)t$, as in (7.17). We also use $\theta_j(\omega) = \omega(t_{j+1}) - \omega(t_j)$ below. Recalling that $\int_0^t \omega(s)\,d\omega(s)$ is the limit of $\sum \omega(t_j)[\omega(t_{j+1}) - \omega(t_j)]$, we write (7.40) as

$$\omega(t)^2 - \omega(0)^2 = 2\sum_{j=0}^{k-1}\omega(t_j)\theta_j(\omega) + \sum_{j=0}^{k-1}\theta_j(\omega)^2. \tag{7.41}$$

The next result is the key to the computation.

Lemma 7.8. *Given $t > 0$,*

$$\Theta_k(\omega) = \sum_{j=0}^{k-1}\Big[\omega\big(\frac{j+1}{k}t\big) - \omega\big(\frac{j}{k}t\big)\Big]^2 \longrightarrow 2t \ \text{ in } L^2(\mathfrak{P}_0, dW_x), \tag{7.42}$$

as $k \to \infty$.

Proof. We have

$$
\begin{aligned}
E_x\big(|\Theta_k - 2t|^2\big) &= E_x\Big(\Big|\sum_j \Big[\theta_j(\omega)^2 - 2\frac{t}{k}\Big]\Big|^2\Big) \\
&= \sum_j E_x\Big(\Big[\theta_j(\omega)^2 - 2\frac{t}{k}\Big]^2\Big),
\end{aligned}
\tag{7.43}
$$

the last identity by independence of the different θ_j. Now we know that $E_x(\theta_j^2) = 2t/k$; furthermore, generally $E_x\big([F - E_x(F)]^2\big) \le E_x(F^2)$, so it follows that

$$
E_x\big(|\Theta_k - t|^2\big) \le \sum_j E_x(\theta_j^4) = 12\frac{t^2}{k}.
\tag{7.44}
$$

This proves the lemma.

Thus, as $k \to \infty$, the right side of (7.41) converges in $L^2(\mathfrak{P}_0, dW_x)$ to $\int_0^t \omega(s)\, d\omega(s) + t$. This gives the identity

$$
\int_0^t \omega(s)\, d\omega(s) = \frac{1}{2}\big[\omega(t)^2 - \omega(0)^2 - 2t\big],
\tag{7.45}
$$

for W_x-almost all ω.

More generally, for sufficiently smooth f, we can write

$$
f\big(\omega(t)\big) - f\big(\omega(0)\big) = \sum_{j=0}^{k-1} \big[f\big(\omega(t_{j+1})\big) - f\big(\omega(t_j)\big)\big]
\tag{7.46}
$$

and use the expansion

$$
\begin{aligned}
&f\big(\omega(t_{j+1})\big) - f\big(\omega(t_j)\big) \\
&\quad = \theta_j(\omega) f'\big(\omega(t_j)\big) + \frac{1}{2}\theta_j(\omega)^2 f''\big(\omega(t_j)\big) + O\big(|\theta_j(\omega)|^3\big)
\end{aligned}
\tag{7.47}
$$

to generalize (7.45) to Ito's fundamental identity:

$$
f\big(\omega(t)\big) - f\big(\omega(0)\big) = \int_0^t f'\big(\omega(s)\big)\, d\omega(s) + \int_0^t f''\big(\omega(s)\big)\, ds,
\tag{7.48}
$$

for one-dimensional Brownian motion. For n-dimensional Brownian motion and functions of the form $f = f(t, x)$, this generalizes to

$$
\begin{aligned}
&f\big(t,\omega(t)\big) - f\big(0,\omega(0)\big) \\
(7.49)\qquad &\quad = \int_0^t \big(\nabla_x f\big)\big(s,\omega(s)\big)\cdot d\omega(s) \\
&\qquad + \int_0^t \big(\Delta f\big)\big(s,\omega(s)\big)\,ds + \int_0^t f_t\big(s,\omega(s)\big)\,ds.
\end{aligned}
$$

Another way of writing this is

$$
(7.50)\qquad df\big(t,\omega(t)\big) = (\nabla_x f)\cdot d\omega + (\Delta f)\,dt + f_t\,dt.
$$

We remind the reader that our choice of $e^{t\Delta}$ rather than $e^{t\Delta/2}$ to define the transition probabilities for Brownian paths leads to formulas that sometimes look different from those arising from the latter convention, which for example would replace $(\Delta f)\,dt$ by $(1/2)(\Delta f)\,dt$ in (7.50).

Note in particular that

$$
d\big(e^{\lambda\omega(t)-\lambda^2 t}\big) = \lambda\, e^{\lambda\omega(t)-\lambda^2 t}\, d\omega(t);
$$

in other words, we have a solution to the "stochastic differential equation":

$$
(7.51)\qquad d\mathfrak{X} = \lambda\mathfrak{X}\,d\omega(t),\quad \mathfrak{X}(t) = e^{\lambda\omega(t)-\lambda^2 t},
$$

for W_0-almost all ω. Recall from (4.16) that this is the martingale $\mathfrak{z}_t(\omega)$.

We now discuss a dynamical theory of Brownian motion due to Langevin, whose purpose was to elucidate Einstein's work on the motion of a Brownian particle. Langevin produced the following equation for the *velocity* of a small particle suspended in a liquid, undergoing the sort of random motion investigated by R. Brown:

$$
(7.52)\qquad \frac{dv}{dt} = -\beta v + \omega'(t),\quad v(0) = v_0.
$$

Here, the term $-\beta v$ represents the frictional force, tending to slow down the particle as it moves through the fluid. The term $\omega'(t)$, which contributes to the force, is due to "white noise," a random force whose statistical properties identify it with the time derivative of ω, which is defined, not classically, but through Propositions 7.1 and 7.2. Thus we rewrite (7.52) as the stochastic differential equation

$$
(7.53)\qquad dv = -\beta v\,dt + d\omega,\quad v(0) = v_0.
$$

As in the case of ODE, we have $d(e^{\beta t}v) = e^{\beta t}(dv + \beta v\,dt)$, so (7.50) yields $d(e^{\beta t}v) = e^{\beta t}d\omega$, which integrates to

$$v(t) = v_0 e^{-\beta t} + \int_0^t e^{-\beta(t-s)}\, d\omega(s)$$
$$(7.54) \qquad = v_0 e^{-\beta t} + \omega(t) - \beta \int_0^t e^{-\beta(t-s)} \omega(s)\, ds.$$

The actual path of such a particle is given by

$$x(t) = x_0 + \int_0^t v(s)\, ds. \tag{7.55}$$

In the case $x_0 = 0, v_0 = 0$, we have

$$x(t) = \int_0^t \int_0^s e^{-\beta(s-r)}\, d\omega(r)\, ds$$
$$(7.56) \qquad = \frac{1}{\beta} \int_0^t \bigl[1 - e^{-\beta(t-s)}\bigr]\, d\omega(s).$$

Via the identity in (7.54), we have

$$x(t) = \int_0^t e^{-\beta(t-s)} \omega(s)\, ds. \tag{7.57}$$

Of course, the path $x(t)$ taken by such a particle is not the same as the "Brownian path" $\omega(t)$ we have been studying, but it is approximated by $\omega(t)$ in the following sense. It is observed experimentally that the frictional force component in (7.52) acts to slow down a particle in a very short time ($\sim 10^{-8}$ sec.). In other words, the dimensional quantity β in (7.52) is, in terms of units humans use to measure standard macroscopic quantities, "large." Now (7.57) implies

$$\lim_{\beta \to \infty} \beta x_\beta(t) = \omega(t), \tag{7.58}$$

where $x_\beta(t)$ denotes the path (7.57).

There has been further work on the dynamics of Brownian motion, particularly by L. Ornstein and G. Uhlenbeck [UO]. See [Nel3] for more on this, and references to other work.

Exercises

1. If $g \in C^1([0,t])$, show that the integral of Proposition 7.1 is given by

$$\int_0^t g(s)\, d\omega(s) = g(t)\omega(t) - g(0)\omega(0) - \int_0^t g'(s)\omega(s)\, ds.$$

Show that this yields the second identity in (7.54) and the implication (7.56) $\Rightarrow$ (7.57).

2. With θ_j as in (7.15), show that

$$E_x\Big(\sum_{j=0}^{k-1} |\theta_j(\omega)|^3\Big) \to 0, \quad \text{as } k \to \infty.$$

(*Hint*: Use $2|\theta_j|^3 \le \varepsilon|\theta_j|^2 + \varepsilon^{-1}|\theta_j|^4$ and (7.44).)

3. Making use of Exercise 2, give a detailed proof of Ito's formula (7.48). Assume $f \in C^2(\mathbb{R})$ and

$$|D^\alpha f(x)| \le C_\varepsilon e^{\varepsilon|x|^2}, \quad \forall\, \varepsilon > 0, \ |\alpha| \le 2.$$

More generally, establish (7.49).
Warning: The estimate of the remainder term in (7.47) is valid only when $|\omega(t_{j+1} - \omega(t_j)|$ is bounded (say $\le K$). But the probability that $|\omega(t_{j+1}) - \omega(t_j)|$ is $\ge K$ is very small.

4. Show that (7.42) implies that W_x-almost all paths ω have locally unbounded variation, on any interval $[s,t] \subset [0,\infty)$.
5. If $\psi(t,\omega) = \int_0^t g\big(s,\omega(s)\big)\cdot d\omega(s)$ is a stochastic integral given by Proposition 7.2, show that

$$E_x\big(\psi(t,\cdot)\big) = 0.$$

Show that $\psi(t,\cdot)$ is a martingale, that is, $E_x\big(\psi(t,\cdot)|\mathfrak{B}_s\big) = \psi(s,\cdot)$, for $s \le t$. Compare Exercise 2 of §8.

8. Stochastic integrals, II

In §7 we considered stochastic integrals of the form

$$h(t,\omega) = \int_0^t g\big(s,\omega(s)\big)\cdot d\omega(s), \tag{8.1}$$

where g is defined on $[0,\infty)\times\mathbb{R}^n$. This is a special case of integrals of the form

$$\psi(t,\omega) = \int_0^t \varphi(s,\omega)\cdot d\omega(s), \tag{8.2}$$

where φ is defined on $[0,\infty)\times\mathfrak{P}_0$. There are important examples of such φ which are not of the form $\varphi(s,\omega) = g\big(s,\omega(s)\big)$, such as the function h in (8.1), typically. It is important to be able to handle more general integrals of the form (8.2), for a certain class of functions φ on $[0,\infty)\times\mathfrak{P}_0$ called "adapted," which will be defined below.

To define (8.2), we extend the analysis in (7.17)–(7.19). Thus we consider

$$S_k(t,\omega) = \sum_{j=0}^{k-1} \varphi(t_j,\omega)\cdot\big[\omega(t_{j+1}) - \omega(t_j)\big] = \sum_{j=0}^{k-1} \varphi_j(\omega)\cdot\theta_j(\omega), \tag{8.3}$$

where, as before, $\theta_j(\omega) = \omega(t_{j+1}) - \omega(t_j)$, $t_j = (j/k)t$. As in (7.18), we want to compute

$$\|S_k(t,\cdot)\|^2_{L^2(\mathfrak{P}_0,dW_x)} = \sum_{j,\ell} E_x(\varphi_j\theta_j\varphi_\ell\theta_\ell). \tag{8.4}$$

Following the analysis of (7.18), we want θ_ℓ to be independent of the other factors in the parentheses on the right side of (8.4) when $\ell > j$. Thus we demand of φ that

$$\varphi(s,\cdot) \text{ is independent of } \omega(t+h)-\omega(t), \quad \forall\, t \geq s,\ \ h > 0. \tag{8.5}$$

Granted this, we see that the terms in the sum in (8.4) vanish when $j \neq \ell$, and

$$\begin{aligned} \|S_k(t,\cdot)\|^2_{L^2(\mathfrak{P}_0,dW_x)} &= \sum_j E_x\big(|\varphi_j|^2\big)E_x\big(|\theta_j|^2\big) \\ &= 2\sum_j E_x\big(|\varphi(t_j,\cdot)|^2\big)(t_{j+1}-t_j). \end{aligned} \tag{8.6}$$

If $\varphi \in C\big(\mathbb{R}^+, L^2(\mathfrak{P}_0, dW_x)\big)$, this is a Riemann sum approximating

$$2\int_0^t E_x\big(|\varphi(s,\cdot)|^2\big)\,ds = 2\|\varphi\|^2_{L^2([0,t]\times\mathfrak{P}_0)}.$$

We use the following spaces:

$$\begin{aligned} C\big(I,\mathcal{R}(Q)\big) &= \big\{\varphi \in C\big(I, L^2(\mathfrak{P}_0, dW_x)\big) : \varphi(t) = Q_t\varphi(t), \forall t \in I\big\}, \\ L^2\big(I,\mathcal{R}(Q)\big) &= \big\{\varphi \in L^2\big(I, L^2(\mathfrak{P}_0, dW_x)\big) : \varphi(t) = Q_t\varphi(t), \forall t \in I\big\}, \end{aligned} \tag{8.7}$$

where $I = [0,T]$, and, as in §4, $Q_t\varphi = E_x(\varphi|\mathfrak{B}_t)$. Elements of these spaces satisfy (8.5), by Corollary 4.4.

Proposition 8.1. *Given $\varphi \in C\big(I,\mathcal{R}(Q)\big)$, the expression (8.3) converges as $k = 2^\nu \to \infty$, in the space $C\big(I,\mathcal{R}(Q)\big)$, to a limit we denote (8.2). Furthermore, $\psi = \mathfrak{I}(\varphi)$ extends uniquely to a linear map*

$$\mathfrak{I} : L^2\big(I,\mathcal{R}(Q)\big) \to C\big(I,\mathcal{R}(Q)\big), \tag{8.8}$$

satisfying

$$\|\mathfrak{I}(\varphi)(t,\cdot)\|_{L^2(\mathfrak{P}_0,dW_x)} = \sqrt{2}\,\|\varphi\|_{L^2([0,t)\times\mathfrak{P}_0,dt\ dW_x)}. \tag{8.9}$$

Regarding continuity, note that

$$\|\mathfrak{I}(\varphi)(t+h,\cdot)-\mathfrak{I}(\varphi)(t,\cdot)\|_{L^2(\mathfrak{P}_0,dW_x)} = \sqrt{2}\,\|\varphi\|_{L^2([t,t+h]\times\mathfrak{P}_0,dt\ dW_x)}. \tag{8.10}$$

We need to verify that $\mathfrak{I}(\varphi)(t,\cdot) \in \mathcal{R}(Q_t)$. But clearly, each term $\varphi(t_j,\omega)\cdot[\omega(t_{j+1})-\omega(t_j)]$ in (8.3) belongs to $\mathcal{R}(Q_t)$ in this case, so we have the desired result.

We mention an approach to (8.8) just slightly different from that described above. Define a simple function to be a function $\varphi(t,\omega)$ that is constant in t for t in intervals of the form $[\ell 2^{-\nu}, (\ell+1)2^{-\nu})$, with values in $\mathcal{R}(Q_s),\ s = \ell 2^{-\nu}$, for some $\nu \in \mathbb{Z}^+$. For a simple function φ, the stochastic integral has a form similar to (8.3), namely,

$$\begin{aligned}\int_0^t \varphi(s,\omega)\cdot d\omega(s) = \sum_{j=0}^{\ell-1} &\varphi(t_j,\omega)\cdot\big[\omega(t_{j+1})-\omega(t_j)\big] \\ &+ \varphi(t_\ell,\omega)\cdot\big[\omega(t)-\omega(t_\ell)\big],\end{aligned} \tag{8.11}$$

where $t_j = j2^{-\nu}$ and $t \in \big[\ell 2^{-\nu}, (\ell+1)2^{-\nu}\big)$. An identity similar to (8.6), together with the denseness of the set of simple functions in $L^2(I,\mathcal{R}(Q))$, yields (8.8).

There is the following generalization of Ito's formula (7.49)–(7.50). Suppose

$$\mathfrak{X}(t) = \mathfrak{X}_0 + \int_{t_0}^t u(s,\omega)\,ds + \int_{t_0}^t v(s,\omega)\,d\omega(s), \tag{8.12}$$

where $u, v \in L^2\big(I,\mathcal{R}(Q)\big)$. Then $\mathfrak{X} \in C\big(I,\mathcal{R}(Q)\big)$. We write

$$d\mathfrak{X} = u\,dt + v\,d\omega. \tag{8.13}$$

We might assume $\mathfrak{X}, u$, and ω take values in $\mathbb{R}^n$ and v is $n\times n$ matrix-valued. More generally, let ω take values in $\mathbb{R}^n$, $\mathfrak{X}$ and u in $\mathbb{R}^m$, and v in $\mathrm{Hom}(\mathbb{R}^n,\mathbb{R}^m)$.

If $\mathfrak{Y}(t) = g\big(t,\mathfrak{X}(t)\big)$, with $g(t,x)$ real-valued and smooth in its arguments, then

$$\begin{aligned}d\mathfrak{Y}(t) = {}&(\nabla_x g)\big(t,\mathfrak{X}(t)\big)\cdot d\mathfrak{X}(t) \\ &+ (D^2 g)\big(t,\mathfrak{X}(t)\big)\big(d\mathfrak{X}(t),d\mathfrak{X}(t)\big) + g_t\big(t,\mathfrak{X}(t)\big)\,dt,\end{aligned} \tag{8.14}$$

where $(D^2 g)(d\mathfrak{X},d\mathfrak{X}) = \sum(\partial^2 g/\partial x_j\partial x_k)\,d\mathfrak{X}_j\cdot d\mathfrak{X}_k$ is computed, via (8.13), by the rules

$$dt\cdot dt = dt\cdot d\omega_j = d\omega_j\cdot dt = 0, \quad d\omega_j\cdot d\omega_k = \delta_{jk}\,dt. \tag{8.15}$$

There is also an integral formula for $g\big(t,\mathfrak{X}(t)\big) - g(t_0,\mathfrak{X}_0)$, parallel to (7.49):

$$\begin{aligned}g\big(t,\mathfrak{X}(t)\big) = g(t_0,\mathfrak{X}_0) &+ \int_{t_0}^t \Big(\frac{\partial^2 g}{\partial x_j\partial x_k}\Big) v_{j\ell}v_{k\ell}\,ds \\ &+ \int_{t_0}^t g_t\big(s,\mathfrak{X}(s)\big)\,ds + \int_{t_0}^t \frac{\partial g}{\partial x_j}\big(u_j\,ds + v_{j\ell}\,d\omega_\ell\big).\end{aligned} \tag{8.16}$$

Here, we sum over repeated indices. The formulas (7.49) and (7.50) cover the special case $u = 0,\ v = I$. The proof of (8.16) is parallel to that of (7.49).

If we apply (8.14) to $g(x) = e^{\lambda x},\ m = 1$, we obtain for

$$\text{(8.17)}\qquad \begin{aligned} \mathfrak{Y}(t) &= \exp\Big(\lambda \mathfrak{X}(t) - \lambda^2 \int_{t_0}^{t} |v(s,\omega)|^2\, ds\Big), \\ \mathfrak{X}(t) &= \int_{t_0}^{t} v(s,\omega)\cdot d\omega(s), \end{aligned}$$

the stochastic differential equation

$$\text{(8.18)}\qquad d\mathfrak{Y} = \lambda \mathfrak{Y}\, v\cdot d\omega,$$

generalizing the identity (7.51).

There is another important property that $\mathfrak{Y}(t)$, defined by (8.17), has in common with $\mathfrak{z}_t(\omega) = e^{\lambda\omega(t)-\lambda^2 t}$.

Proposition 8.2. *Given $v \in L^2\big(I, \mathcal{R}(Q)\big)$, with values in $\mathbb{R}^n$, the function $\mathfrak{Y}(t)$ defined by (8.17) is a supermartingale; that is, for $s \le t$,*

$$\text{(8.19)}\qquad E_x\big(\mathfrak{Y}(t)\big|\mathfrak{B}_s\big) \le \mathfrak{Y}(s), \quad W_x\text{-a.e. on } \mathfrak{P}_0.$$

Proof. We treat the case $t_0 = 0$. First suppose v_ν is a simple function, constant as a function of t on intervals of the form $[\ell 2^{-\nu}, (\ell+1)2^{-\nu})$, with values in $\mathcal{R}(Q_{\ell 2^{-\nu}})$, and $\mathfrak{Y}_\nu$ is given by (8.17), with $v = v_\nu$. We claim that $\mathfrak{Y}_\nu$ is a martingale, that is,

$$\text{(8.20)}\qquad E_x\big(\mathfrak{Y}_\nu(t)\big|\mathfrak{B}_s\big) = \mathfrak{Y}_\nu(s), \ \ \text{for } s \le t.$$

Suppose, for example, that $0 \le t < 2^{-\nu}$, so $v_\nu(s) = v_\nu(0)$, for $s \le t$. Now $v_\nu(0)$ is independent of $\omega(t) - \omega(s)$, so in this case

$$\begin{aligned} E_x\big(\mathfrak{Y}_\nu(t)\big|\mathfrak{B}_s\big) &= E_x\big(e^{\lambda v_\nu(0)[\omega(t)-x]-\lambda^2 t|v_\nu(0)|^2}\big|\mathfrak{B}_s\big) \\ &= e^{\lambda v_\nu(0)[\omega(s)-x]-\lambda^2 s|v_\nu(0)|^2}\cdot E_x\Big(e^{\lambda v_\nu(0)[\omega(t)-\omega(s)]-\lambda^2(t-s)|v_\nu(0)|^2}\big|\mathfrak{B}_s\Big), \end{aligned}$$

and the last conditional expectation is 1. A similar argument in the case $\ell 2^{-\nu} \le s \le t \le (\ell+1)2^{-\nu}$, using (8.11), gives

$$E_x\big(\mathfrak{Y}_\nu(t)\big|\mathfrak{B}_s\big) = \mathfrak{Y}_\nu(t_{\nu\ell}) E_x\Big(e^{\lambda v_{\nu\ell}[\omega(t)-\omega(t_{\nu\ell})]-\lambda^2(t-t_{\nu\ell})|v_{\nu\ell}|^2}\big|\mathfrak{B}_s\Big) = \mathfrak{Y}_\nu(s),$$

where $t_{\nu\ell} = \ell 2^{-\nu},\ v_{\nu\ell} = v_\nu(t_{\nu\ell})$. The identity (8.20), for general $s \le t$, follows easily from this.

For general $v \in L^2\big(I, \mathcal{R}(Q)\big)$, we can take simple v_ν converging to v in the norm of this space, and then $\mathfrak{X}_\nu \to \mathfrak{X}$ in $C\big(I, \mathcal{R}(Q)\big)$, where $\mathfrak{X}_\nu(t) = \int_0^t v_\nu(s,\omega)\cdot d\omega(s)$. Passing to a subsequence, we can assume (for fixed s, t)

that $\mathfrak{X}_\nu(s) \to \mathfrak{X}(s)$ and $\mathfrak{X}_\nu(t) \to \mathfrak{X}(t)$, W_x-a.e.; hence $\mathfrak{Y}_\nu(s) \to \mathfrak{Y}(s)$ and $\mathfrak{Y}_\nu(t) \to \mathfrak{Y}(t)$, W_x-a.e. Then (8.19) follows, by Fatou's lemma.

The case of general $t_0 \ge 0$ is easily obtained from this; one can extend $v(s,\omega)$ to be 0 for $0 \le s < t_0$.

Note in particular that $s = 0$ in (8.19) implies

$$E_x\Big(e^{\lambda \mathfrak{X}(t) - \lambda^2 \int_{t_0}^t |v(s,\cdot)|^2\,ds}\Big) \le 1. \tag{8.21}$$

Using Cauchy's inequality, we deduce that

$$E_x\Big(e^{\lambda \mathfrak{X}(t)/2}\Big) \le E_x\Big(e^{\lambda^2 \int_{t_0}^t |v(s,\cdot)|^2\,ds}\Big)^{1/2}. \tag{8.22}$$

We get a similar estimate upon replacing $v(s,\omega)$ by $-v(s,\omega)$, which converts $\mathfrak{X}(t)$ to $-\mathfrak{X}(t)$. Since $e^{|x|} \le e^x + e^{-x}$, we have (replacing λ by 2λ)

$$E_x\Big(e^{\lambda |\mathfrak{X}(t)|}\Big) \le 2 E_x\Big(e^{4\lambda^2 \int_{t_0}^t |v(s,\cdot)|^2\,ds}\Big)^{1/2}. \tag{8.23}$$

Compare with Lemma 7.4. Note that the convexity of the exponential function implies

$$E_x\Big(e^{t^{-1}\int_0^t F(s,\cdot)\,ds}\Big) \le \frac{1}{t}\int_0^t E_x\Big(e^{F(s,\cdot)}\Big)\,ds. \tag{8.24}$$

Therefore, (8.23) implies

$$\begin{aligned} E_x\big(e^{\lambda|\mathfrak{X}(t)|}\big) &\le 2\left[\frac{1}{t-t_0}\int_{t_0}^t E_x\Big(e^{4\lambda^2 t |v(s\cdot)|^2}\Big) ds\right]^{1/2} \\ &\le 2 \max_{t_0 \le s \le t} E_x\Big(e^{4\lambda^2 t |v(s,\cdot)|^2}\Big)^{1/2}. \end{aligned} \tag{8.25}$$

If we expand $\mathfrak{Y}_\nu(t) = e^{\lambda \mathfrak{X}_\nu(t) - \lambda^2 \int_{t_0}^t |v_\nu(s,\cdot)|^2\,ds}$ in powers of λ, the coefficient of each λ^j is a martingale. The coefficient of λ^4, for example, is

$$\frac{1}{24}|\mathfrak{X}_\nu(t)|^4 - \frac{1}{2}\mathfrak{X}_\nu(t)^2\Big(\int_{t_0}^t |v_\nu(s,\omega)|^2\,ds\Big) + \frac{1}{2}\Big(\int_{t_0}^t |v_\nu(s,\omega)|^2\,ds\Big)^2. \tag{8.26}$$

This has expectation zero; hence

$$\begin{aligned} \frac{1}{24}E_x\big(|\mathfrak{X}_\nu(t)|^4\big) &\le \frac{1}{2}E_x\Big(\mathfrak{X}_\nu(t)^2\Big(\int_{t_0}^t |v_\nu(s,\cdot)|^2\,ds\Big)\Big) \\ &\le \frac{1}{48}E_x\big(|\mathfrak{X}_\nu(t)|^4\big) + 48 E_x\Big(\Big(\int_{t_0}^t |v_\nu(s,\cdot)|^2\,ds\Big)^2\Big), \end{aligned} \tag{8.27}$$

so

$$
\begin{aligned}
E_x\big(|\mathfrak{X}_\nu(t)|^4\big) &\le 48^2 E_x\Big(\Big(\int_{t_0}^t |v_\nu(s,\cdot)|^2\,ds\Big)^2\Big) \\
&\le \big(48|t-t_0|\big)^2 \frac{1}{t-t_0}\int_{t_0}^t E_x\big(|v_\nu(s,\cdot)|^4\big)\,ds \\
&\le \big(48|t-t_0|\big)^2 \max_{t_0\le s\le t} E_x\big(|v_\nu(s,\cdot)|^4\big),
\end{aligned}
\tag{8.28}
$$

where the second inequality here uses convexity, as in (8.24). Again a use of Fatou's lemma yields for

$$
\mathfrak{X}(t) = \int_{t_0}^t v(s,\omega)\cdot d\omega(s)
\tag{8.29}
$$

the estimate

$$
\|\mathfrak{X}(t)\|_{L^4(\mathfrak{P}_0)} \le \big(48|t-t_0|\big)^{1/2} \max_{t_0\le s\le t} \|v(s,\cdot)\|_{L^4(\mathfrak{P}_0)}.
\tag{8.30}
$$

Similarly we obtain, for $t_1 < t_2$,

$$
\|\mathfrak{X}(t_1)-\mathfrak{X}(t_2)\|_{L^4(\mathfrak{P}_0)} \le C_1|t_1-t_2|^{1/2} \max_{t_1\le s\le t_2} \|v(s,\cdot)\|_{L^4(\mathfrak{P}_0)},
\tag{8.31}
$$

with $C_1 = \sqrt{48}$, when $\mathfrak{X}(t)$ is given by (8.29). If $\mathfrak{X}(t)$ is given more generally by (8.12), we have

$$
\begin{aligned}
\|\mathfrak{X}(t_1)-\mathfrak{X}(t_2)\|_{L^4(\mathfrak{P}_0)} \le C_0|t_1-t_2| &\max_{t_1\le s\le t_2} \|u(s,\cdot)\|_{L^4(\mathfrak{P}_0)} \\
&+ C_1|t_1-t_2|^{1/2} \max_{t_1\le s\le t_2} \|v(s,\cdot)\|_{L^4(\mathfrak{P}_0)}.
\end{aligned}
\tag{8.32}
$$

The martingale maximal inequality of Proposition 4.7 extends to submartingales, but it is not obvious that it applies to the supermartingale $\mathfrak{Y}(t)$. However, it does apply to $\mathfrak{Y}_\nu(t)$, so, for each $\nu \in \mathbb{Z}^+$, we have

$$
\begin{aligned}
W_x\Big(\Big\{\omega\in\mathfrak{P}_0 : \sup_{t\in I(t_0,t_1)} \mathfrak{X}_\nu(t)-\mathfrak{X}_\nu(t_0) - \lambda\int_{t_0}^t |v_\nu(s,\omega)|^2\,ds > \beta\Big\}\Big) \\
\le e^{-\lambda\beta},
\end{aligned}
\tag{8.33}
$$

where $I(t_0,t_1) = [t_0,t_1]\cap\mathbb{Q}$. It follows that

(8.34)

$$
\begin{aligned}
W_x\Big(\Big\{\omega\in\mathfrak{P}_0 : \sup_{t\in I(t_0,t_1)} |\mathfrak{X}_\nu(t)-\mathfrak{X}_\nu(t_0)| > \lambda\int_{t_0}^{t_1} |v_\nu(s,\omega)|^2\,ds + \beta\Big\}\Big) \\
\le 2e^{-\lambda\beta}.
\end{aligned}
$$

Thus, if we have

$$\int_{t_0}^{t_1} |v_\nu(s,\omega)|^2\, ds < \frac{\beta}{\lambda}, \quad \text{for } \omega \in S, \tag{8.35}$$

then

$$W_x\Big(S \cap \big\{\omega \in \mathfrak{P}_0 : \sup_{t\in I(t_0,t_1)} |\mathfrak{X}_\nu(t) - \mathfrak{X}_\nu(t_0)| > 2\beta\big\}\Big) \le 2e^{-\lambda\beta}. \tag{8.36}$$

Now

$$\begin{aligned} &W_x\Big(\Big\{\omega \in \mathfrak{P}_0 : \int_{t_0}^{t_1} |v_\nu(s,\omega)|^2\, ds \ge \frac{\beta}{\lambda}\Big\}\Big) \\ &\qquad \le \frac{\lambda}{\beta}\int_{t_0}^{t_1} E_x\big(|v_\nu(s,\cdot)|^2\big)\, ds. \end{aligned} \tag{8.37}$$

Taking $\beta = \delta,\ \lambda = 1/\delta^2$, we deduce that if

$$\int_{t_0}^{t_1} \|v_\nu(s,\cdot)\|^2_{L^2(\mathfrak{P}_0)}\, ds < \delta^3\varepsilon, \tag{8.38}$$

then

$$W_x\Big(\big\{\omega \in \mathfrak{P}_0 : \sup_{t\in I(t_0,t_1)} |\mathfrak{X}_\nu(t) - \mathfrak{X}_\nu(t_0)| > 2\delta\big\}\Big) \le \varepsilon + e^{-1/\delta}. \tag{8.39}$$

Since $\mathfrak{X}_\nu(t)$ converges to $\mathfrak{X}(t)$ in measure, locally uniformly in t, we have

$$W_x\Big(\big\{\omega \in \mathfrak{P}_0 : \sup_{t\in I(t_0,t_1)} |\mathfrak{X}(t) - \mathfrak{X}(t_0)| > 2\delta\big\}\Big) \le \varepsilon + e^{-1/\delta} \tag{8.40}$$

whenever

$$\int_{t_0}^{t} \|v(s,\cdot)\|^2_{L^2(\mathfrak{P}_0)}\, ds < \delta^3\varepsilon. \tag{8.41}$$

The estimate (8.40) enables us to establish the following important result.

Proposition 8.3. *Let $I = [0,T]$. Given $v \in L^2\big(I,\mathcal{R}(Q)\big)$, so $\int_0^t v(s,\omega)\cdot d\omega(s) = \mathfrak{X}(t)$ belongs to $C\big(I,\mathcal{R}(Q)\big)$, you can define $\mathfrak{X}(t,\omega)$ so that $t \mapsto \mathfrak{X}(t,\omega)$ is continuous in t, for W_x-a.e. ω.*

Proof. Start with any measurable function on $I \times \mathfrak{P}_0$ representing $\mathfrak{X}(t)$; call it $\mathfrak{X}^b(t,\omega)$, so for each $t \in I,\ \mathfrak{X}^b(t,\cdot) = \mathfrak{X}(t),\ W_x$-a.e. on $\mathfrak{P}_0$. Set $\mathfrak{X}(t,\omega) = \mathfrak{X}^b(t,\omega)$, for $t \in I \cap \mathbb{Q}$. From (8.40)–(8.41) it follows that there is a set $N \subset \mathfrak{P}_0$ such that $W_x(N) = 0$ and $\sigma_\omega(t) = \mathfrak{X}(t,\omega)$ is uniformly continuous in $t \in I \cap \mathbb{Q}$ for each $\omega \in \mathfrak{P}_0 \setminus N$. Then, for $\omega \in \mathfrak{P}_0,\ t \in I \setminus \mathbb{Q}$, define $\mathfrak{X}(t,\omega)$ by continuity:

$$\mathfrak{X}(t,\omega) = \lim_{I\cap\mathbb{Q}\ni t_\nu\to t} \mathfrak{X}^b(t_\nu,\omega), \quad \omega \in \mathfrak{P}_0 \setminus N. \tag{8.42}$$

If $\omega \in N$, define $\mathfrak{X}(t,\omega)$ arbitrarily.

To show that this works, it remains to check that, for each $t \in I$,

$$\mathfrak{X}(t,\cdot) = \mathfrak{X}(t), \quad W_x\text{-a.e. on } \mathfrak{P}_0. \tag{8.43}$$

Indeed, since $\mathfrak{X}^b(t_\nu,\cdot) \to \mathfrak{X}(t)$ in L^2-norm, passing to a subsequence we have $\mathfrak{X}^b(t_{\nu_j},\cdot) \to \mathfrak{X}(t)$ W_x-a.e. Comparing with (8.42), we have (8.43).

Exercises

1. Generalize (8.30) to show that $\mathfrak{X}(t) = \int_{t_0}^t v(s,\omega)\cdot d\omega(s)$ satisfies

$$\|\mathfrak{X}(t)\|^{2k}_{L^{2k}(\mathfrak{P}_0)} \le C_k|t-t_0|^{k-1}\int_{t_0}^t \|v(s,\cdot)\|^{2k}_{L^{2k}(\mathfrak{P}_0)}\,ds,$$

for $k \in \mathbb{Z}^+$.

2. Given $\varphi \in L^2\big([0,\infty),\mathcal{R}(Q)\big)$, show that, for $t \ge s$,

$$E_x\Big(\int_s^t \varphi(\tau,\omega)\cdot d\omega(\tau)\big|\mathfrak{B}_s\Big) = 0.$$

Deduce that the stochastic integral $\psi(t,\omega) = \int_0^t \varphi(s,\omega)\cdot d\omega(s)$ is a martingale, so that, for $t \ge s$,

$$E_x\big(\psi(t,\cdot)\big|\mathfrak{B}_s\big) = \psi(s,\cdot).$$

3. Show that if $v(s,\omega)$ satisfies the hypotheses of Proposition 8.2, then the supermartingale $\mathfrak{Y}(t)$ in (8.17) is a martingale if and only if

$$E_x\,\mathfrak{Y}(t) = 1, \quad \forall\, t \ge 0.$$

9. Stochastic differential equations

In this section we treat stochastic differential equations of the form

$$d\mathfrak{X} = b(t,\mathfrak{X})\,dt + \sigma(t,\mathfrak{X})\,d\omega, \quad \mathfrak{X}(t_0) = \mathfrak{X}_0. \tag{9.1}$$

The function $\mathfrak{X}$ is an unknown function on $I \times \mathfrak{P}_0$, where $I = [t_0, T]$. We assume $t_0 \ge 0$. As in the case of ordinary differential equations, we will use the Picard iteration method, to obtain the solution $\mathfrak{X}$ as the limit of a sequence of approximate solutions to (8.1), which we write as a stochastic integral equation:

$$\mathfrak{X}(t) = \mathfrak{X}_0 + \int_{t_0}^t b\big(s,\mathfrak{X}(s)\big)\,ds + \int_{t_0}^t \sigma\big(s,\mathfrak{X}(s)\big)\,d\omega(s) = \Phi\mathfrak{X}(t). \tag{9.2}$$

The last identity defines the transformation Φ, and we look for a fixed point of Φ. As usual, $\mathfrak{X}(t)$ is shorthand for $\mathfrak{X}(t,\omega)$. If ω is a Brownian path in $\mathbb{R}^n$, we can let $\mathfrak{X}$ and $b(t,x)$ take values in $\mathbb{R}^m$ and let $\sigma(t,x)$ be an $m \times n$ matrix-valued function.

Let us assume that $\sigma(t,x)$ and $b(t,x)$ are continuous in their arguments and satisfy

$$
\begin{aligned}
&|b(t,x)| \le K_0(1+|x|), \qquad |b(t,x)-b(t,y)| \le L_0|x-y|,\\
&|\sigma(t,x)| \le K_1(1+|x|^2)^{1/2}, \quad |\sigma(t,x)-\sigma(t,y)| \le L_1|x-y|.
\end{aligned} \tag{9.3}
$$

We will use results of §8 to show that

$$
\Phi : L^2\big(I,\mathcal{R}(Q)\big) \longrightarrow C\big(I,\mathcal{R}(Q)\big), \tag{9.4}
$$

where, as in (8.7),

$$
C\big(I,\mathcal{R}(Q)\big) = \big\{\varphi \in C\big(I, L^2(\mathfrak{P}_0, dW_0)\big) : \varphi(t) \in \mathcal{R}(Q_t),\ \forall t \in I\big\},
$$

and $L^2\big(I,\mathcal{R}(Q)\big)$ is similarly defined. Note that $\mathfrak{X}(s)$ belongs to $\mathcal{R}(Q_s)$ if and only if $\mathfrak{X}(s)$ is (equal W_0-a.e. to) a $\mathfrak{B}_s$-measurable function on $\mathfrak{P}_0$, so if $\mathfrak{X}(s) \in \mathcal{R}(Q_s)$, then also $\sigma\big(s,\mathfrak{X}(s)\big)$ and $b\big(s,\mathfrak{X}(s)\big)$ belong to $\mathcal{R}(Q_s)$. Thus Proposition 8.1 applies to the second integral in (9.2), and if $\mathfrak{X}_0 \in \mathcal{R}(Q_{t_0})$, we have (9.4).

Applying (8.9) to estimate the second integral in (9.2), we have

$$
\begin{aligned}
\|\Phi\mathfrak{X}(t)-\mathfrak{X}_0\|^2_{L^2(\mathfrak{P}_0)} \le{}& 2K_0^2\left(\int_{t_0}^t \big(1+\|\mathfrak{X}(s)\|_{L^2(\mathfrak{P}_0)}\big)\,ds\right)^2\\
&+4K_1^2\int_{t_0}^t \big(1+\|\mathfrak{X}(s)\|^2_{L^2(\mathfrak{P}_0)}\big)\,ds.
\end{aligned} \tag{9.5}
$$

Also (8.9) applies to an estimate of the second integral in

$$
\begin{aligned}
\Phi\mathfrak{X}(t)-\Phi\mathfrak{Y}(t) ={}& \int_{t_0}^t \big[b\big(s,\mathfrak{X}(s)\big)-b\big(s,\mathfrak{Y}(s)\big)\big]\,ds\\
&+\int_{t_0}^t \big[\sigma\big(s,\mathfrak{X}(s)\big)-\sigma\big(s,\mathfrak{Y}(s)\big)\big]\,ds.
\end{aligned} \tag{9.6}
$$

We get

$$
\begin{aligned}
\|\Phi\mathfrak{X}(t)-\Phi\mathfrak{Y}(t)\|^2_{L^2(\mathfrak{P}_0)} \le{}& 2L_0^2\left(\int_{t_0}^t \|\mathfrak{X}(s)-\mathfrak{Y}(s)\|_{L^2(\mathfrak{P}_0)}\,ds\right)^2\\
&+4L_1^2\int_{t_0}^t \|\mathfrak{X}(s)-\mathfrak{Y}(s)\|^2_{L^2(\mathfrak{P}_0)}\,ds.
\end{aligned} \tag{9.7}
$$

To solve (9.2), we take $\mathfrak{X}_0(t,\omega) = \mathfrak{X}_0(\omega)$, the given initial value, and inductively define $\mathfrak{X}_{j+1} = \Phi\mathfrak{X}_j$. Note that

$$\mathfrak{X}_1(t,\omega) = \mathfrak{X}_0(\omega) + \int_{t_0}^t b\big(s,\mathfrak{X}_0(\omega)\big)\,ds + \int_{t_0}^t \sigma\big(s,\mathfrak{X}_0(\omega)\big)\,d\omega(s) \tag{9.8}$$

contains a stochastic integral of the form (7.14), provided $\mathfrak{X}_0(\omega)$ is constant. On the other hand, the stochastic integral yielding $\mathfrak{X}_2(t,\omega)$ is usually not even of the form (7.13), but rather of the more general form (8.2). The following estimate will readily yield convergence of the sequence $\mathfrak{X}_j$.

Lemma 9.1. *For some $M = M(T) < \infty$, we have*

$$\|\mathfrak{X}_{j+1}(t) - \mathfrak{X}_j(t)\|^2_{L^2(\mathfrak{P}_0)} \le \frac{(M|t-t_0|)^{j+1}}{(j+1)!}, \quad t_0 \le t \le T. \tag{9.9}$$

Proof. We establish this estimate inductively. For $j = 0$, we can use (9.5), with $\mathfrak{X} = \mathfrak{X}_1$, and the $j = 0$ case of (9.9) follows. Assume that (9.9) holds for $j = 0,\dots,k-1$; we need to get it for $j = k$. To do this, apply (9.7) with $\mathfrak{X} = \mathfrak{X}_k$, $\mathfrak{Y} = \mathfrak{X}_{k-1}$, to get

$$\begin{aligned}\|\mathfrak{X}_{k+1}(t) - \mathfrak{X}_k(t)\|^2_{L^2(\mathfrak{P}_0)} \le{}& \frac{2L_0^2M^k}{k!}\left(\int_{t_0}^t |s-t_0|^{k/2}\,ds\right)^2 \\ &+ \frac{4L_1^2M^k}{k!}\int_{t_0}^t |s-t_0|^k\,ds.\end{aligned} \tag{9.10}$$

This is $\le \big(M|t-t_0|\big)^{k+1}/(k+1)!$ as long as M is sufficiently large for (9.9) to hold for $j = 0$ and also $M \ge 2L_0^2\max(1,T) + 4L_1^2$.

These estimates immediately yield an existence theorem:

Theorem 9.2. *Given $0 \le t_0 < T < \infty$, $I = [t_0,T]$, if b and σ are continuous on $I \times \mathbb{R}^n$ and satisfy the estimates* (9.3)*, and if $\mathfrak{X}_0 \in \mathcal{R}(Q_{t_0})$, then the equation* (9.2) *has a unique solution $\mathfrak{X} \in C\big(I,\mathcal{R}(Q)\big)$.*

Only the uniqueness remains to be demonstrated. But if $\mathfrak{X}$ and $\mathfrak{Y}$ are two such solutions, we have $\Phi\mathfrak{X} = \mathfrak{X}$ and $\Phi\mathfrak{Y} = \mathfrak{Y}$, so (9.7) implies

$$\|\mathfrak{X}(t) - \mathfrak{Y}(t)\|^2_{L^2(\mathfrak{P}_0)} \le \text{ right side of (9.7)},$$

and a Gronwall argument implies $\|\mathfrak{X}(t) - \mathfrak{Y}(t)\|_{L^2} = 0$, for all $t \in I$.

Of course, the hypothesis that b and σ are continuous in t can be weakened in ways that are obvious from an examination of (9.4)–(9.7). Allowing b and σ to be piecewise continuous in t, still satisfying (9.3), we can reduce (9.1) to the case $t_0 = 0$, by setting $b(t,x) = 0$ and $\sigma(t,x) = 0$ for $0 \le t < t_0$.

If $\mathfrak{X}_0$ has higher integrability, so does the solution $\mathfrak{X}(t)$. To see this, in case $\mathfrak{X}_0 \in L^4(\mathfrak{P}_0)$, we can exploit (8.26)–(8.30) to produce the following estimate, parallel to (9.7):

$$\begin{aligned}
&\|\Phi\mathfrak{X}(t) - \Phi\mathfrak{Y}(t)\|^4_{L^4(\mathfrak{P}_0)} \le \\
&\quad 8L_0^4\Bigl(\int_{t_0}^t \|\mathfrak{X}(s) - \mathfrak{Y}(s)\|_{L^4(\mathfrak{P}_0)}\, ds\Bigr)^4 \\
&\quad + 8(48^2)L_1^4|t - t_0| \int_{t_0}^t \|\mathfrak{X}(s) - \mathfrak{Y}(s)\|^4_{L^4(\mathfrak{P}_0)}\, ds.
\end{aligned} \tag{9.11}$$

Using this, assuming $\mathfrak{X}_0 \in L^4(\mathfrak{P}_0, dW_0)$, we can obtain the following analogue of (9.9):

$$\|\mathfrak{X}_{j+1}(t) - \mathfrak{X}_j(t)\|^4_{L^4(\mathfrak{P}_0)} \le \frac{\bigl(M|t - t_0|^2\bigr)^{j+1}}{(j+1)!}, \tag{9.12}$$

for $M = M(T)$, on any interval $t \in [t_0, T]$. We have the following:

Proposition 9.3. *Under the hypotheses of Theorem 9.2, if also $\mathfrak{X}_0 \in L^4(\mathfrak{P}_0, dW_0)$, then $\mathfrak{X} \in C\bigl(I, L^4(\mathfrak{P}_0, dW_0)\bigr)$.*

More generally, one can establish that $\mathfrak{X} \in C\bigl(I, L^{2k}(\mathfrak{P}_0)\bigr)$, provided $\mathfrak{X}_0 \in L^{2k}(\mathfrak{P}_0)$, $k \ge 1$. The case $2k = 4$ enables us to prove part of the following important result.

Proposition 9.4. *The solution $\mathfrak{X}(t)$ to (9.2) given by Theorem 9.2 can be represented as $\mathfrak{X}(t, \omega)$ such that, for W_0-a.e. $\omega \in \mathfrak{P}_0$, the map $t \mapsto \mathfrak{X}(t, \omega)$ is continuous in t.*

Proof. First we assume $\mathfrak{X}_0 \in L^4(\mathfrak{P}_0, dW_0)$ and give a demonstration that is somewhat parallel to that of Theorem 1.1. Given $\varepsilon > 0, \delta > 0$, and $s, t \in \mathbb{R}^+$ such that $|t - s| < \delta$, we estimate the probability that $|\mathfrak{X}(t) - \mathfrak{X}(s)| > \varepsilon$. We use the estimate

$$\|\mathfrak{X}(t) - \mathfrak{X}(s)\|^4_{L^4(\mathfrak{P}_0)} \le C|t - s|^2, \tag{9.13}$$

$C = C(T)$, for $s, t \in [0, T]$, which follows (when $t > s$) from

$$\begin{aligned}
\|\mathfrak{X}(t) - \mathfrak{X}(s)\|^4_{L^4(\mathfrak{P}_0)} &\le C\Bigl(\int_s^t \|b(\tau, \mathfrak{X}(\tau))\|_{L^4}\, d\tau\Bigr)^4 \\
&\quad + C\int_s^t \|\sigma(\tau, \mathfrak{X}(\tau))\|^4_{L^4}\, d\tau,
\end{aligned} \tag{9.14}$$

together with the estimate $\|\mathfrak{X}(s)\|_{L^4} \le C(\tau)$. Consequently, given $s, t \in \mathbb{R}^+$,

$$W_0\Big(\{\omega \in \mathfrak{P}_0 : |\mathfrak{X}(t,\omega) - \mathfrak{X}(s,\omega)| > \varepsilon\}\Big) \le \frac{C}{\varepsilon^4}|t-s|^2. \tag{9.15}$$

Now an argument parallel to that of Lemma 1.2 gives

$$\begin{aligned} W_0\Big(\{\omega \in \mathfrak{P}_0 : |\mathfrak{X}(t_1,\omega) - \mathfrak{X}(t_j,\omega)| > \varepsilon, \text{ for some } j = 2,\dots,\nu\}\Big) \\ \le C r\Big(\frac{\varepsilon}{2},\delta\Big), \end{aligned} \tag{9.16}$$

when $\{t_1,\dots,t_\nu\}$ is any finite set of numbers in $\mathbb{Q}^+$ such that $0 \le t_1 < \cdots < t_\nu$ and $t_\nu - t_1 \le \delta$, where

$$r(\varepsilon,\delta) = \min\big(1, C\delta^2\varepsilon^{-4}\big). \tag{9.17}$$

The function $r(\varepsilon,\delta)$ takes the place of $\rho(\varepsilon,\delta)$ in (1.23); as in (1.21), we have

$$\frac{r(\varepsilon,\delta)}{\delta} \to 0, \quad \text{as } \delta \to 0, \tag{9.18}$$

for each $\varepsilon > 0$. From here, one shows just as in the proof of Theorem 1.1 that, for some $Z \subset \mathfrak{P}_0$ such that $W_0(Z) = 0$, the map $t \mapsto \mathfrak{X}(t,\omega)$ is uniformly continuous on $t \in \mathbb{Q}^+$, for each $\omega \in \mathfrak{P}_0 \setminus Z$. the rest of the proof of Proposition 9.4 can be carried out just like the proof of Proposition 8.3.

We now give another demonstration of Proposition 9.4, not requiring $\mathfrak{X}_0$ to be in $L^4(\mathfrak{P}_0)$, but only in $L^2(\mathfrak{P}_0)$. In such a case, under the hypotheses, and conclusions, of Theorem 9.2, we have $\sigma\big(t,\mathfrak{X}(t)\big) \in C\big(I,\mathcal{R}(Q)\big)$. Hence Proposition 8.3 applies to the second integral in (9.2), so $\mathfrak{A}(t,\omega) = \int_{t_0}^t \sigma\big(s,\mathfrak{X}(s)\big)\,d\omega(s)$ can be represented as a continuous function of t, for W_0-a.e. $\omega \in \mathfrak{P}_0$. Furthermore, we have $b\big(t,\mathfrak{X}(t)\big) \in C\big(I, L^2(\mathfrak{P}_0)\big) \subset C\big(I, L^1(\mathfrak{P}_0)\big)$. Thus, by Fubini's theorem, the first integral in (9.2) is absolutely integrable, hence continuous in t, for W_0-a.e. ω. This establishes the desired property for the left side of (9.2).

We next investigate the dependence of the solution to (9.2) on the initial data $\mathfrak{X}_0$, in a fashion roughly parallel to the method used in §6 of Chap. 1. Thus, let $\mathfrak{Y}$ solve

$$\mathfrak{Y}(t) = \mathfrak{Y}_0 + \int_{t_0}^t b\big(s,\mathfrak{Y}(s)\big)\,ds + \int_{t_0}^t \sigma\big(s,\mathfrak{Y}(s)\big)\,d\omega(s). \tag{9.19}$$

Proposition 9.5. *Assume that $b(t,x)$ and $\sigma(t,x)$ satisfy the hypotheses of Theorem 9.2 and are also C^1 in x. If $\mathfrak{X}(t)$ and $\mathfrak{Y}(t)$ solve (9.2) and (9.19), respectively, then*

$$\|\mathfrak{X}(t) - \mathfrak{Y}(t)\|_{L^2(\mathfrak{P}_0)} \le C(t, L_0, L_1)\|\mathfrak{X}_0 - \mathfrak{Y}_0\|_{L^2(\mathfrak{P}_0)}. \tag{9.20}$$

Proof. Consider $\mathcal{Z}(t) = \mathfrak{X}(t) - \mathfrak{Y}(t)$, which satisfies the identity

$$\begin{aligned}\mathcal{Z}(t) = \mathcal{Z}_0 &+ \int_{t_0}^t b'\big(s, \mathfrak{X}(s), \mathfrak{Y}(s)\big)\mathcal{Z}(s)\, ds \\ &+ \int_{t_0}^t \sigma'\big(s, \mathfrak{X}, \mathfrak{Y}(s)\big)\mathcal{Z}(s)\, d\omega(s),\end{aligned} \tag{9.21}$$

with $\mathcal{Z}_0 = \mathfrak{X}_0 - \mathfrak{Y}_0$. Here

$$b'(s, x, y) = \int_0^1 D_x b\big(s, ux + (1-u)y\big)\, du, \tag{9.22}$$

so $b'(s, x, y)(x - y) = b(s, x) - b(s, y)$, and similarly

$$\sigma'(s, x, y) = \int_0^1 D_x \sigma\big(s, ux + (1-u)y\big)\, du. \tag{9.23}$$

We estimate the right side of (9.21) in $L^2(\mathfrak{P}_0)$. By (9.3), $|b'(s, x, y)| \le L_0$, so

$$\left\| \int_{t_0}^t b'\big(s, \mathfrak{X}(s), \mathfrak{Y}(s)\big)\mathcal{Z}(s)\, ds \right\|_{L^2} \le L_0 \int_{t_0}^t \|\mathcal{Z}(s)\|_{L^2}\, ds. \tag{9.24}$$

Since $|\sigma'(s, x, y)| \le L_1$ and $\sigma'(s, \mathfrak{X}(s), \mathfrak{Y}(s))\mathcal{Z}(s) \in \mathcal{R}(Q_s)$, we have

$$\left\| \int_{t_0}^t \sigma'\big(s, \mathfrak{X}(s), \mathfrak{Y}(s)\big)\mathcal{Z}(s)\, d\omega(s) \right\|_{L^2}^2 \le L_1^2 \int_{t_0}^t \|\mathcal{Z}(s)\|_{L^2}^2\, ds. \tag{9.25}$$

Thus the identity (9.21) implies

$$\|\mathcal{Z}(t)\|_{L^2}^2 \le 3\|\mathfrak{X}_0 - \mathfrak{Y}_0\|_{L^2}^2 + 3\big[L_0^2(t - t_0)^2 + L_1^2\big] \int_{t_0}^t \|\mathcal{Z}(s)\|_{L^2}^2\, ds. \tag{9.26}$$

Now Gronwall's inequality applied to this estimate yields (9.20).

Note that (9.21) is a linear stochastic equation for $\mathcal{Z}(t)$, of a form a little different from (9.2), if $\mathfrak{X}(s)$ and $\mathfrak{Y}(s)$ are regarded as given. On the other hand, we can regard $\mathfrak{X}, \mathfrak{Y}$, and $\mathcal{Z}$ as solving together a system of stochastic equations, of the same form as (9.2).

An important special case of (9.2) is the case $\mathfrak{X}_0 = x$, a given point of $\mathbb{R}^m$, so let us look at $\mathfrak{X}^{x,s}(t)$, defined for $t \ge s$ as the solution to

$$\mathfrak{X}^{x,s}(t) = x + \int_s^t b\big(r, \mathfrak{X}(r)\big)\, dr + \int_s^t \sigma\big(r, \mathfrak{X}(r)\big)\, d\omega(r). \tag{9.27}$$

In this case we have the following useful property, which is basically the Markov property. Let $\mathfrak{B}_s^t$ denote the σ-algebra of subsets of $\mathfrak{P}_0$ generated by all sets of the form

$$\{\omega \in \mathfrak{P}_0 : \omega(t_1) - \omega(s_1) \in A\}, \quad s \le s_1 \le t_1 \le t,\ A \subset \mathbb{R}^m \text{ Borel}, \tag{9.28}$$

plus all sets of W_0-measure zero.

Proposition 9.6. *For any fixed $t \ge s$, the solution $\mathfrak{X}^{x,s}(t)$ to (9.27) is $\mathfrak{B}_s^t$-measurable.*

Proof. By the proof of Theorem 9.2, we have $\mathfrak{X}^{x,s}(t) = \lim_{k\to\infty} \mathfrak{X}_k(t)$, where $\mathfrak{X}_0(t) = x$ and, for $k \ge 0$,

$$\mathfrak{X}_{k+1}(t) = x + \int_s^t b\big(r, \mathfrak{X}_k(r)\big)\, dr + \int_s^t \sigma\big(r, \mathfrak{X}_k(r)\big)\, d\omega(r).$$

It follows inductively that each $\mathfrak{X}_k(t)$ is $\mathfrak{B}_s^t$-measurable, so the limit also has this property.

The behavior of $\mathfrak{X}^{x,s}(t)$ will be important for the next section. We derive another useful property here.

Proposition 9.7. *For $s \le \tau \le t$, we have*

$$\mathfrak{X}^{x,s}(t,\omega) = \mathfrak{X}^{q,\tau}(t,\omega), \quad q = \mathfrak{X}^{x,s}(\tau,\omega), \tag{9.29}$$

for W_0-a.e. $\omega \in \mathfrak{P}_0$.

Proof. Let $\mathfrak{Y}(t)$ denote the right side of (9.29). Thus $\mathfrak{Y}(\tau) = \mathfrak{X}^{x,s}(\tau)$. The stochastic equation satisfied by $\mathfrak{X}^{x,s}(t)$ then implies

$$\mathfrak{Y}(t) = \mathfrak{X}^{x,s}(\tau) + \int_\tau^t b\big(r, \mathfrak{Y}(r)\big)\, dr + \int_\tau^t \sigma\big(r, \mathfrak{Y}(r)\big)\, d\omega(r).$$

Now (9.27) implies that $\mathfrak{X}^{x,s}(t)$ satisfies this same stochastic equation, for $t \ge \tau$. The identity $\mathfrak{Y}(t) = \mathfrak{X}^{x,s}(t)$ a.e. on $\mathfrak{P}_0$ follows from the uniqueness part of Theorem 9.2.

Exercises

1. Show that the solution to

$$d\mathfrak{X} = a(t)\mathfrak{X}(t)\, dt + b(t)\mathfrak{X}(t)\, d\omega(t),$$

in case $m = n = 1$, is given by

$$\mathfrak{X}(t) = \mathfrak{X}(0)\ \exp\Big\{\int_0^t [a(s) - b(s)^2]\, ds + \int_0^t b(s)\, d\omega(s)\Big\} = \mathfrak{X}(0)e^{\mathfrak{Z}(t)}. \tag{9.30}$$

In this problem and the following one, $\mathfrak{X}(t)$ depends on ω, but $a(t)$ and $b(t)$ do not depend on ω, nor do $f(t)$ and $g(t)$ below.

2. Show that the solution to

$$d\mathfrak{X}(t) = \big[f(t) + a(t)\mathfrak{X}(t)\big]\,dt + \big[g(t) + b(t)\mathfrak{X}(t)\big]\,d\omega(t),$$

in case $m = n = 1$, is given by $\mathfrak{X}(t) = e^{\mathfrak{Z}(t)}\mathfrak{Y}(t)$, where $e^{\mathfrak{Z}(t)}$ is as in (9.30) and

$$\mathfrak{Y}(t) = \mathfrak{X}(0) + \int_0^t \Big[e^{-\mathfrak{Z}(s)} f(s) - g(s)b(s)\Big]\,ds + \int_0^t g(s)e^{-\mathfrak{Z}(s)}\,d\omega(s).$$

3. Consider the system

$$d\mathfrak{X}(t) = \big[A(t)\mathfrak{X}(t) + f(t)\big]\,dt + g(t)\,d\omega(t), \tag{9.31}$$

where $A(t) \in \mathrm{End}(\mathbb{R}^m)$, $f(t) \in \mathbb{R}^m$, and $g(t) \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$. Suppose $S(t,s)$ is the solution operator to the linear $m \times m$ system of differential equations

$$\frac{dy}{dt} = A(t)y, \quad S(t,t) = I,$$

as considered in Chap. 1, §5. Show that the solution to (9.31) is

$$\mathfrak{X}(t) = S(t,0)\mathfrak{X}(0) + \int_0^t S(t,s)f(s)\,ds + \int_0^t S(t,s)g(s)\,d\omega(s).$$

4. The following Langevin equation is more general than (7.52):

$$x''(t) = -\nabla V\big(x(t)\big) - \beta x'(t) + \omega'(t). \tag{9.32}$$

Rewrite this as a first-order system of the form (9.1). Using Exercise 3, solve this equation when $V(x)$ is the harmonic oscillator potential, $V(x) = ax^2$.

10. Application to equations of diffusion

Let $\mathfrak{X}^{x,s}(t)$ solve the stochastic equation

$$\mathfrak{X}^{x,s}(t) = x + \int_s^t b\big(\mathfrak{X}^{x,s}(r)\big)\,dr + \int_s^t \sigma\big(\mathfrak{X}^{x,s}(r)\big)\,d\omega. \tag{10.1}$$

As in (9.2), x and b can take values in $\mathbb{R}^m$ and σ values in $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$. We want to study the transformations on functions on $\mathbb{R}^m$ defined by

$$\Phi_s^t f(x) = E_0\, f\big(\mathfrak{X}^{x,s}(t)\big), \quad 0 \le s \le t. \tag{10.2}$$

Clearly, $\mathfrak{X}^{x,s}(s) = x$, so

$$\Phi_t^t f(x) = f(x). \tag{10.3}$$

We assume $b(x)$ and $\sigma(x)$ are bounded and satisfy the Lipschitz conditions of (9.3). For simplicity we have taken b and σ to be independent of t in (10.1). We claim this implies the following:

$$\Phi_0^t f(x) = \Phi_s^{t+s} f(x), \tag{10.4}$$

for $s, t \geq 0$. In fact, it is clear that

$$\mathfrak{X}^{x,s}(t+s,\omega) = \mathfrak{X}^{x,0}(t,\vartheta_s\omega), \tag{10.5}$$

where $\vartheta_s\omega(\tau) = \omega(\tau+s) - \omega(s)$, as in (4.11). The measure-preserving property of the map $\vartheta_s : \mathfrak{P}_0 \to \mathfrak{P}_0$ then implies

$$E_0\, f\big(\mathfrak{X}^{x,0}(t,\vartheta_s\omega)\big) = E_0\, f\big(\mathfrak{X}^{x,0}(t)\big) = \Phi_0^t f(x),$$

so we have established (10.4). Let us set

$$P^t f = \Phi_0^t f = E_0\, f\big(\mathfrak{X}^x(t)\big), \tag{10.6}$$

where for notational convenience we have set $\mathfrak{X}^x(t) = \mathfrak{X}^{x,0}(t)$.

We will study the action of P^t on the Banach space $C_o(\mathbb{R}^m)$ of continuous functions on $\mathbb{R}^m$ that vanish at infinity.

Proposition 10.1. *For each $t \geq 0$,*

$$P^t : C_o(\mathbb{R}^m) \longrightarrow C_o(\mathbb{R}^m), \tag{10.7}$$

and P^t forms a strongly continuous semigroup of operators on $C_o(\mathbb{R}^m)$.

Proof. If $f \in C_o(\mathbb{R}^m)$, then f is uniformly continuous, that is, it has a modulus of continuity:

$$|f(x) - f(y)| \leq \omega_f\big(|x-y|\big), \tag{10.8}$$

where $\omega_f(\delta)$ is a bounded, continuous function of δ such that $\omega_f(\delta) \to 0$ as $\delta \to 0$. Then

$$\begin{aligned} \big|P^t f(x) - P^t f(y)\big| &\leq E_0\, \big|f\big(\mathfrak{X}^x(t)\big) - f\big(\mathfrak{X}^y(t)\big)\big| \\ &\leq E_0\, \omega_f\big(\big|\mathfrak{X}^x(t) - \mathfrak{X}^y(t)\big|\big). \end{aligned} \tag{10.9}$$

Now if x is fixed and $y = x_\nu \to x$, then, for each $t \geq 0$, $\mathfrak{X}^x(t) - \mathfrak{X}^{x_\nu}(t) \to 0$ in $L^2(\mathfrak{P}_0)$, by Proposition 9.5. Hence $\mathfrak{X}^x(t) - \mathfrak{X}^{x_\nu}(t) \to 0$ in measure on $\mathfrak{P}_0$, so the Lebesgue dominated convergence theorem implies that (10.9) tends to 0 as $y \to x$. This shows that $P^t f \in C(\mathbb{R}^m)$ if $f \in C_o(\mathbb{R}^m)$.

To show that $P^t f(x)$ vanishes at infinity, for each $t \geq 0$, we note that, for most $\omega \in \mathfrak{P}_0$ (in a sense that will be quantified below), $|\mathfrak{X}^x(t) - x| \leq C\langle t\rangle$ if C is large, so if $f \in C_o(\mathbb{R}^m)$ and $|x|$ is large, then $f\big(\mathfrak{X}^x(t,\omega)\big)$ is small for most $\omega \in \mathfrak{P}_0$.

In fact, subtracting x from both sides of (10.1) and estimating L^2-norms, we have

$$\|\mathfrak{X}^x(t) - x\|^2_{L^2(\mathfrak{P}_0)} \leq 2B^2t^2 + 2S^2t, \quad B = \sup\,|b|,\ S = \sup\,|\sigma|. \tag{10.10}$$

Hence

$$W_0\Big(\big\{\omega \in \mathfrak{P}_0 : |\mathfrak{X}^x(t,\omega) - x| > \lambda\big\}\Big) \leq \frac{2B^2t^2 + 2S^2t}{\lambda^2}. \tag{10.11}$$

The mapping property (10.7) follows.

We next examine continuity in t. In fact, parallel to (10.9), we have

$$\big|P^t f(x) - P^s f(x)\big| \leq E_0\, \omega_f\big(\big|\mathfrak{X}^x(t) - \mathfrak{X}^x(s)\big|\big). \tag{10.12}$$

We know from §9 that $\mathfrak{X}^x(t) \in C\big(\mathbb{R}^+, L^2(\mathfrak{P}_0)\big)$, and estimates from there readily yield that the modulus of continuity can be taken to be independent of x. Then the vanishing of (10.12), uniformly in x, as $s \to t$, follows as in the analysis of (10.9).

There remains the semigroup property, $P^sP^{t-s} = P^t$, for $0 \leq s \leq t$. By (10.4), this is equivalent to $\Phi_0^s\Phi_s^t = \Phi_0^t$. To establish this, we will use the identity

$$E_0\Big(f\big(\mathfrak{X}^{x,s}(t)\big)\big|\mathfrak{B}_s\Big) = E_0\ f\big(\mathfrak{X}^{x,s}(t)\big) = \Phi_s^t f(x), \tag{10.13}$$

which is an immediate consequence of Proposition 9.6. If we replace s by τ in (10.13), and then replace x by $\mathfrak{X}^{x,s}(\tau)$, with $s \leq \tau \leq t$, and use the identity

$$\mathfrak{X}^{q,\tau}(t) = \mathfrak{X}^{x,s}(t), \quad q = \mathfrak{X}^{x,s}(\tau), \tag{10.14}$$

established in Proposition 9.7, we obtain

$$E_0\Big(f\big(\mathfrak{X}^{x,s}(t)\big)\big|\mathfrak{B}_\tau\Big) = \Phi_\tau^t f\big(\mathfrak{X}^{x,s}(\tau)\big). \tag{10.15}$$

We thus have, for $s \leq \tau \leq t$,

$$\begin{aligned} \Phi_s^\tau\Phi_\tau^t f(x) &= E_0\Big(\Phi_\tau^t f\big(\mathfrak{X}^{x,s}(\tau)\big)\big|\mathfrak{B}_s\Big) \\ &= E_0\Big(E_0\Big(f\big(\mathfrak{X}^{q,\tau}(t)\big)\big|\mathfrak{B}_\tau\Big)\Big|\mathfrak{B}_s\Big) \\ &= E_0\Big(f\big(\mathfrak{X}^{q,\tau}(t)\big)\big|\mathfrak{B}_s\Big), \end{aligned} \tag{10.16}$$

and again using (10.14) we see that this is equal to the left side of (10.13), hence to $\Phi_s^t f(x)$, as desired. This completes the proof of Proposition 10.1.

We want to identify the infinitesimal generator of P^t. Assume now that $D^\alpha f$, for $|\alpha| \le 2$, are bounded and continuous on $\mathbb{R}^m$. Then Ito's formula implies

$$\begin{aligned} f\big(\mathfrak{X}^x(t)\big) = f(x) &+ \int_0^t \Big(\frac{\partial^2 f}{\partial x_j \partial x_k}\Big)\sigma_{j\ell}\sigma_{k\ell}\, dr \\ &+ \int_0^t \frac{\partial f}{\partial x_j}\big(b_j\, dr + \sigma_{j\ell}\, d\omega_\ell\big), \end{aligned} \tag{10.17}$$

using the summation convention. Let us apply E_0 to both sides. Now

$$E_0\Big(\int_0^t \frac{\partial f}{\partial x_j}\sigma_{j\ell}\, d\omega_\ell\Big) = 0, \tag{10.18}$$

so we have

$$\begin{aligned} E_0\big(f(\mathfrak{X}^x(t))\big) = f(x) &+ \int_0^t E_0\Big(\frac{\partial^2 f}{\partial x_j \partial x_k} A_{jk}\Big)\, dr \\ &+ \int_0^t E_0\Big(\frac{\partial f}{\partial x_j} b_j\Big)\, dr, \end{aligned} \tag{10.19}$$

where A_{jk} in the first integral is given by

$$A_{jk}(y) = \sum_\ell \sigma_{j\ell}(y)\sigma_{k\ell}(y), \quad y = \mathfrak{X}^x(r). \tag{10.20}$$

In matrix notation,

$$A = \sigma\sigma^t. \tag{10.21}$$

We can take the t-derivative of the right side of (10.16), obtaining

$$\begin{aligned} &\frac{\partial}{\partial t} P^t f(x) = \\ &E_0\Big(A_{jk}\big(\mathfrak{X}^x(t)\big)\,\partial_j\partial_k f\big(\mathfrak{X}^x(t)\big) + b_j\big(\mathfrak{X}^x(t)\big)\,\partial_j f\big(\mathfrak{X}^x(t)\big)\Big). \end{aligned} \tag{10.22}$$

In particular,

$$\frac{\partial}{\partial t} P^t f(x)\Big|_{t=0} = \sum_{j,k} A_{jk}(x)\,\partial_j\partial_k f(x) + \sum_j b_j(x)\,\partial_j f(x) = Lf(x), \tag{10.23}$$

where the last identity defines the second-order differential operator L, acting on functions of x. This is known as Kolmogorov's diffusion equation. We have shown that the infinitesimal generator of the semigroup P^t, acting on $C_o(\mathbb{R}^m)$, is a closed extension of the operator

$$L = \sum A_{jk}(x)\,\partial_j\partial_k + \sum b_j(x)\,\partial_j, \tag{10.24}$$

defined initially, let us say, on $C_0^2(\mathbb{R}^m)$.

It is clear from (10.6) that $\|P^t f\|_{L^\infty} \le \|f\|_{L^\infty}$ for each $f \in C_o(\mathbb{R}^m)$, so P^t is a contraction semigroup on $C_o(\mathbb{R}^m)$. It is also clear that

$$f \ge 0 \Longrightarrow P^t f \ge 0 \ \text{ on } \mathbb{R}^m, \tag{10.25}$$

that is, P^t is "positivity preserving." For given $x \in \mathbb{R}^n,\ t \ge 0,\ f \mapsto P^t f(x)$ is a positive linear functional on $C_o(\mathbb{R}^m)$. Hence there is a uniquely defined positive Borel measure $\mu_{x,t}$ on $\mathbb{R}^m$, of mass ≤ 1, such that

$$P^t f(x) = \int f(y)\, d\mu_{x,t}(y). \tag{10.26}$$

In fact, by the construction (10.6),

$$\mu_{x,t} = F_{(x,t)*}W_0, \tag{10.27}$$

where $F_{(x,t)}(\omega) = \mathfrak{X}^x(t,\omega)$, and (10.27) means $\mu_{x,t}(U) = W_0\big(F^{-1}_{(x,t)}(U)\big)$ for a Borel set $U \subset \mathbb{R}^m$. This implies that, for each $x, t,\ \mu_{x,t}$ is a probability measure on $\mathbb{R}^m$, since $|\mathfrak{X}^x(t)|$ is finite for W_0-a.e. $\omega \in \mathfrak{P}_0$.

We will use the notation

$$P(s,x,t,U) = \mu_{x,t-s}(U), \quad 0 \le s \le t,\ U \subset \mathbb{R}^m,\ \text{Borel}. \tag{10.28}$$

We can identify $P(s,x,t,U)$ with the probability that $\mathfrak{X}^{x,s}(t)$ is in U. We can rewrite (10.26) as

$$P^t f(x) = \int f(y)\ P(0,x,t,dy) \tag{10.29}$$

or

$$\Phi_s^t f(x) = \int f(y)\ P(s,x,t,dy). \tag{10.30}$$

The semigroup property on P^t implies

$$P(s,x,t,U) = \int P(s,x,\tau,dy)\ P(\tau,y,t,U), \quad 0 \le s \le \tau \le t, \tag{10.31}$$

which is known as the Chapman–Kolmogorov equation.

Let us denote by $\mathcal{L}$ the extension of (10.24) that is the infinitesimal generator of P^t. If V is a bounded, continuous function on $\mathbb{R}^m$, then $\mathcal{L} - V$ generates a semigroup on $C_o(\mathbb{R}^m)$, and an application of the Trotter product formula similar to that done in §2 yields

$$e^{t(\mathcal{L}-V)} f(x) = E_0\left(f\left(\mathfrak{X}^x(t)\right) e^{-\int_0^t V(\mathfrak{X}^x(s))\,ds}\right). \tag{10.32}$$

This furnishes an existence result for weak solutions to the initial-value problem

$$\begin{aligned} \frac{\partial u}{\partial t} &= \sum A_{jk}(x)\, \partial_j \partial_k u + \sum b_j(x)\, \partial_j u - Vu, \\ u(0) &= f \in C_o(\mathbb{R}^m), \end{aligned} \tag{10.33}$$

under the hypotheses that V is bounded and continuous, the coefficients b_j are bounded and uniformly Lipschitz, and A_{jk} has the form (10.20), with $\sigma_{j\ell}$ bounded and uniformly Lipschitz. As for the last property, we record the following fact:

Proposition 10.2. *If $A(x)$ is a C^2 positive-semidefinite, matrix-valued function on $\mathbb{R}^m$ with $D^\alpha A(x)$ bounded on $\mathbb{R}^m$ for $|\alpha| \le 2$, then there exists a bounded, uniformly Lipschitz, matrix-valued function $\sigma(x)$ on $\mathbb{R}^m$ such that $A(x) = \sigma(x)\sigma(x)^t$.*

This result is quite easy to prove in the elliptic case, that is, when for certain $\lambda_j \in (0, \infty)$,

$$\lambda_0 |\xi|^2 \le \sum A_{jk}(x) \xi_j \xi_k \le \lambda_1 |\xi|^2, \tag{10.34}$$

but a careful argument is required if $A(x)$ is allowed to degenerate. See the exercises for more on this.

If $A_{jk}(x)$ has bounded, continuous derivatives of order ≤ 2, we can form the formal adjoint of (10.24):

$$L^t f = \sum \partial_j \partial_k \big(A_{jk}(x) f\big) - \sum \partial_j \big(b_j(x) f\big) = \widetilde{L} f - Vf, \tag{10.35}$$

where $\widetilde{L}$ has the same second-order derivatives as L, though perhaps a different first-order part, and $V(x) = -\sum \partial_j \partial_k A_{jk}(x) + \sum \partial_k b_j(x)$. Thus $\widetilde{L}$ has an extension, which we denote as $\widetilde{\mathcal{L}}$, generating a contraction semigroup on $C_o(\mathbb{R}^m)$, with the positivity-preserving property. Furthermore, $\widetilde{\mathcal{L}} - V$ generates a semigroup on $C_o(\mathbb{R}^m)$, and there is a formula for $e^{t(\widetilde{\mathcal{L}}-V)} f$ parallel to (10.32). Thus we obtain a weak solution to the initial-value problem

$$\frac{\partial u}{\partial t} = \sum \partial_j \partial_k \big(A_{jk}(x) u\big) - \sum \partial_j \big(b_j(x) u\big), \quad u(0) = f \in C_o(\mathbb{R}^m), \tag{10.36}$$

provided that $A_{jk}(x)$ satisfies the conditions of Proposition 10.2, and that each b_j is bounded, with bounded, continuous first derivatives. Equation (10.36) is called the Fokker-Planck equation.

To continue, we shall make a further simplifying hypothesis, namely that the ellipticity condition (10.34) hold. We will also assume $A_{jk}(x)$ and $b_j(x)$ are C^∞, and that $D^\alpha A_{jk}(x)$ and $D^\alpha b_j(x)$ are bounded for all α. In such a case, $(g_{jk}) = (A_{jk})^{-1}$ defines a Riemannian metric on $\mathbb{R}^m$, and if Δ_g denotes its Laplace operator, we have

$$Lf = \Delta_g f + Xf, \tag{10.37}$$

for some smooth vector field $X = \sum \xi_j(x)\,\partial_j$, such that $D^\alpha \xi_j(x)$ is bounded for $|\alpha| \le 1$. Note that if we use the inner product

$$(f, g) = \int f(x)\overline{g(x)}\, dV(x), \tag{10.38}$$

where dV is the Riemannian volume element determined by the Riemannian metric g_{jk}, then this puts the same topology on $L^2(\mathbb{R}^m)$ as the inner product $\int f(x)\overline{g(x)}\, dx$. We prefer the inner product (10.38), since Δ_g is then self-adjoint.

Now consider the closed operator $\mathcal{L}_2$ on $L^2(\mathbb{R}^m)$ defined by

$$\mathcal{L}_2 f = Lf \quad \text{on } \mathcal{D}(\mathcal{L}_2) = H^2(\mathbb{R}^m). \tag{10.39}$$

It follows from results on Chap. 6, §2, that $\mathcal{L}_2$ generates a strongly continuous semigroup $e^{t\mathcal{L}_2}$ on $L^2(\mathbb{R}^n)$. To relate this semigroup to the semigroup $P^t = e^{t\mathcal{L}}$ on $C_o(\mathbb{R}^m)$ described above, we claim that

$$e^{t\mathcal{L}_2} f = e^{t\mathcal{L}} f, \quad \text{for } f \in C_0^\infty(\mathbb{R}^m). \tag{10.40}$$

To see this, let $u_0(t, x)$ and $u_1(t, x)$ denote the left and right sides, respectively. These are both weak solutions to $\partial_t u_j = Lu_j$, for which one has regularity results. Also, estimates discussed in §2 of Chap. 6 imply that $u_0(t, x)$ vanishes as $|x| \to \infty$, locally uniformly in $t \in [0, \infty)$. Thus the maximum principle applies to $u_0(t, x) - u_1(t, x)$, and we have (10.40). From here a simple limiting argument yields

$$e^{t\mathcal{L}_2} f = e^{t\mathcal{L}} f, \quad \text{for } f \in C_o(\mathbb{R}^m) \cap L^2(\mathbb{R}^m). \tag{10.41}$$

Now the dual semigroup $(e^{t\mathcal{L}_2})^*$ is a strongly continuous semigroup on $L^2(\mathbb{R}^n)$, with infinitesimal generator $\mathcal{L}_2^t$ defined by

$$\mathcal{L}_2^t f = L^t f \quad \text{on } \mathcal{D}(\mathcal{L}_2^t) = H^2(\mathbb{R}^m), \tag{10.42}$$

where L^t is given by (10.35). An argument parallel to that used to establish (10.41) shows that

$$(10.43)\qquad \left(e^{t\mathcal{L}_2}\right)^* f = e^{t\mathcal{L}_2^t} f = e^{t(\widetilde{\mathcal{L}}-V)} f, \quad \text{for } f \in C_o(\mathbb{R}^m) \cap L^2(\mathbb{R}^m).$$

On the other hand, $(P^t)^* = (e^{t\mathcal{L}})^*$ is a weak*-continuous semigroup of operators on $\mathfrak{M}(\mathbb{R}^m)$, the space of finite Borel measures on $\mathbb{R}^m$; it is not strongly continuous. Using (10.43), we see that

$$(10.44)\qquad \left(f, e^{t\mathcal{L}_2} g\right) = \left(e^{t(\widetilde{\mathcal{L}}-V)} f, g\right), \quad \text{for } f, g \in C_0^\infty(\mathbb{R}^m),$$

and bringing in (10.40) we have

$$(10.45)\qquad \left(e^{t\mathcal{L}}\right)^* f = e^{t(\widetilde{\mathcal{L}}-V)} f,$$

for $f \in C_0^\infty(\mathbb{R}^m)$, hence for $f \in C_0(\mathbb{R}^m) \cap L^1(\mathbb{R}^m)$. From here one can deduce that $\left(e^{t\mathcal{L}}\right)^*$ preserves $L^1(\mathbb{R}^m)$ and acts as a strongly continuous semigroup on this space.

Let us return to the family of measures $P(s, x, t, \cdot)$. Under our current hypotheses, regularity results for parabolic PDE imply that, for $s < t$, there is a smooth function $p(s, x, t, y)$ such that

$$(10.46)\qquad P(s, x, t, U) = \int_U p(s, x, t, y)\, dy.$$

We have

$$(10.47)\qquad \Phi_s^t f(x) = \int f(y)\, p(s, x, t, y)\, dy, \quad s < t,$$

and

$$(10.48)\qquad \left(\Phi_s^t\right)^* f(y) = \int f(x)\, p(s, x, t, y)\, dx, \quad s < t.$$

Furthermore, we have for $p(s, x, t, y)$ the "backward" Kolmogorov equation

$$(10.49)\qquad \frac{\partial p}{\partial s} = -\sum_{j,k} A_{jk}(x) \frac{\partial^2 p}{\partial x_j \partial x_k} - \sum_j b_j(x) \frac{\partial p}{\partial x_j}$$

and the Fokker-Planck equation

$$(10.50)\qquad \frac{\partial p}{\partial t} = \sum_{j,k} \frac{\partial^2}{\partial y_j \partial y_k}\Big(A_{jk}(y) p\Big) - \sum_j \frac{\partial}{\partial y_j}\Big(b_j(y) p\Big).$$

While we have restricted attention to the smooth elliptic case for the last set of results, it is also interesting to relax the regularity required on the coefficients as much as possible, and to let the coefficients depend on t, and also to allow degeneracy. See [Fdln] and [StV] for more on this. Exercise 5 below illustrates the natural occurence of degenerate L.

We mention that, working with (10.32), we can obtain the solution to

$$\frac{\partial u}{\partial t} = Lu, \quad \text{for } t \geq 0,\ x \in \Omega, \qquad u(t,x) = 0, \quad \text{for } x \in \partial\Omega, \quad u(0,x) = f(x), \tag{10.51}$$

by considering a sequence $V_\nu \to \infty$ on $\mathbb{R}^m \setminus \Omega$, as in the analysis in §3, when Ω is an open domain in $\mathbb{R}^m$, with smooth boundary, or at least with the regularity property used in Proposition 3.3. In analogy with (3.8), we get

$$u(t) = E_0\Big(f\big(\mathfrak{X}^x(t)\big)\psi_{\overline{\Omega}}(\mathfrak{X}^x, t)\Big), \tag{10.52}$$

where

$$\psi_{\overline{\Omega}}(\mathfrak{X}^x, t) = 1 \quad \text{if } \mathfrak{X}^x\big([0,t]\big) \subset \overline{\Omega}, \qquad 0 \quad \text{otherwise.} \tag{10.53}$$

The proof can be carried out along the same lines as in the proof of Proposition 3.3, provided $\mathcal{L}_2$ (defined in (10.39)) is self-adjoint. Otherwise a different approach is required. Also, when $\mathcal{L}_2$ is self-adjoint, the analysis leading to Proposition 3.5 extends to (10.51), for any open $\Omega \subset \mathbb{R}^m$, with no boundary regularity required. For other approaches to these matters, and also to the Dirichlet problem for $Lu = f$ on Ω, in both the elliptic and degenerate cases, see [Fdln] and [Fr].

We end this section with a look at a special case of (10.1), namely when $\sigma = I$, so we solve

$$\mathfrak{X}^x(t) = x + \omega(t) + \int_0^t b\big(\mathfrak{X}^x(r)\big)\, dr. \tag{10.54}$$

Assume as before that b is bounded and uniformly Lipschitz. Then the analysis of (10.6) done above implies

$$e^{t(\Delta+X)} f(x) = E_0\, f\big(\mathfrak{X}^x(t)\big), \quad X = \sum b_j(x)\,\partial_j. \tag{10.55}$$

On the other hand, in §7 we derived the formula

$$e^{t(\Delta+X)} f(x) = E_x\Big(f\big(\omega(t)\big)e^{\mathcal{Z}(t)}\Big), \tag{10.56}$$

where

$$\mathcal{Z}(t) = \frac{1}{2}\int_0^t b\big(\omega(s)\big)\cdot d\omega(s) - \frac{1}{4}\int_0^t \big|b\big(\omega(s)\big)\big|^2\, ds.$$

We conclude that the right-hand sides of (10.55) and (10.56) coincide. We can restate this identity as follows. Given $x \in \mathbb{R}^n$, we have a map

$$\Xi^x : \mathfrak{P}_0 \to \mathfrak{P}_0, \quad \Xi^x(\omega)(t) = \mathfrak{X}^x(t). \tag{10.57}$$

Then Wiener measure W_0 on $\mathfrak{P}_0$ gives rise to a measure $\Xi^x_* W_0$ on $\mathfrak{P}_0$, by

$$\Xi^x_* W_0(S) = W_0\big((\Xi^x)^{-1}(S)\big). \tag{10.58}$$

For example, if $0 \le t_1 < \cdots < t_k$,

$$\int\limits_{\mathfrak{P}_0} F\big(\omega(t_1), \dots, \omega(t_k)\big)\, d\,\Xi^x_* W_0 = E_0\ F\big(\mathfrak{X}^x(t_1), \dots, \mathfrak{X}^x(t_k)\big). \tag{10.59}$$

Thus the identity of (10.55) and (10.56) can be written as

$$\int\limits_{\mathfrak{P}_0} f\big(\omega(t)\big)\, d\,\Xi^x_* W_0 = \int\limits_{\mathfrak{P}_0} f\big(\omega(t)\big) e^{\mathcal{Z}(t)}\, dW_x. \tag{10.60}$$

This is a special case of the following result of Cameron-Martin and Girsanov:

Proposition 10.3. *Given $t \in (0, \infty)$, $\Xi^x_* W_0\big|_{\mathfrak{B}_t}$ is absolutely continuous with respect to $W_x\big|_{\mathfrak{B}_t}$, with Radon-Nikodym derivative*

$$\frac{d\,\Xi^x_* W_0}{d W_x} = e^{\mathcal{Z}(t)}. \tag{10.61}$$

Note that by taking $f_\nu \nearrow 1$ in (10.56), we have $E_x\big(e^{\mathcal{Z}(t)}\big) = 1$, so the supermartingale $e^{\mathcal{Z}(t)}$ is actually a martingale in this case.

To prove the proposition, it suffices to show that, for $0 \le t_1 < \cdots < t_k \le t$, and a sufficiently large class of continuous functions f_j,

$$\begin{aligned} &E_0\Big(f_1\big(\mathfrak{X}^x(t_1)\big)\cdots f_k\big(\mathfrak{X}^x(t_k)\big)\Big) \\ &= E_x\Big(f_1\big(\omega(t_1)\big)\cdots f_k\big(\omega(t_k)\big)e^{\mathcal{Z}(t)}\Big). \end{aligned} \tag{10.62}$$

We will get this by extending (10.55) and (10.56) to formulas for the solution operators to time-dependent equations of the form

$$\frac{\partial u}{\partial t} = (\Delta + X)u - V(t,x)u, \quad u(0) = f. \tag{10.63}$$

Only the coefficient $V(t,x)$ depends on t; X does not. Parallel to (2.16), we can extend (10.55) to

$$u(t) = E_0\Big(f\big(\mathfrak{X}^x(t)\big)e^{-\int_0^t V(s,\mathfrak{X}^x(s))\,ds}\Big), \tag{10.64}$$

and we can extend (10.56) to

$$u(t) = E_x\Big(e^{\mathcal{Z}(t)} f\big(\omega(t)\big)e^{-\int_0^t V(s,\omega(s))\,ds}\Big). \tag{10.65}$$

Now we can pick $V(s,x)$ to be highly peaked, as a function of s, near $s = t_1,\dots,t_k$, in such a way as to get

$$e^{-\int_0^t V(s,\omega(s))\,ds} \approx e^{-V_1\left(\omega(t_1)\right)}\cdots e^{-V_k\left(\omega(t_k)\right)}. \tag{10.66}$$

Thus having the identity of (10.64) and (10.65) for a sufficiently large class of functions $V(s,x)$ can be seen to yield (10.62). We leave the final details to the reader.

For further material on the Cameron-Martin-Girsanov formula (10.61), see [Fr], [Kal], [McK], and [Øk].

Exercises

1. As an alternative derivation of (10.13), namely,

$$E_0\Big(f\big(\mathfrak{X}^{x,s}(t)\big)\big|\mathfrak{B}_s\Big) = P^{t-s}f(x),$$

via the Markov property, show that in light of the identity (10.5), it follows by applying (4.12) to $E_0\Big(f\big(\mathfrak{X}^x(t-s,\vartheta_s\omega)\big)\big|\mathfrak{B}_s\Big)$.

2. Under the hypotheses of Proposition 10.1, show that, for $\lambda > 0$,

$$E_0\Big(e^{\lambda|\mathfrak{X}^x(t)-x|}\Big) \le 2e^{2\lambda^2S^2t+\lambda Bt}.$$

(*Hint*: If $\mathcal{Z}(t)$ denotes the last integral in (10.1), use (8.23) to estimate the quantity $E_0\big(e^{\lambda|\mathcal{Z}(t)|}\big)$.) Using this estimate in place of (10.10), get as strong a bound as you can on the behavior of $P^tf(x)$, for fixed $t \in \mathbb{R}^+$, as $|x| \to \infty$, given $f \in C_0(\mathbb{R}^n)$, that is, f continuous with compact support.

3. Granted the hypotheses under which the identity $\big(e^{t\mathcal{L}}\big)^* = e^{t(\widetilde{\mathcal{L}}-V)}$ on the space $C_o(\mathbb{R}^m)\cap L^1(\mathbb{R}^m)$ was established in (10.45), show that if $\widetilde{P}(t)$ denotes $\big(e^{t\mathcal{L}}\big)^*$ restricted to $L^1(\mathbb{R}^m)$, then $\mathcal{P}(t) = \widetilde{P}(t)^* : L^\infty(\mathbb{R}^m) \to L^\infty(\mathbb{R}^m)$ is given by the same formula as (10.6):

$$\mathcal{P}(t)f(x) = E_0\, f\big(\mathfrak{X}^x(t)\big), \quad f \in L^\infty(\mathbb{R}^m).$$

Show that

$$P(s,x,t,U) = \mathcal{P}(t-s)\chi_U(x).$$

4. Assume $A(x)$ is real-valued, $A \in C^2(\mathbb{R}^m)$, and $A(x) \ge 0$ for all x. Show that

$$|\nabla A(x)|^2 \le 4A(x) \ \sup \Big\{ |D^2 A(y)| : |x-y| < \frac{2A(x)}{|\nabla A(x)|} \Big\}.$$

Use this to show that $\sqrt{A(x)}$ is uniformly Lipschitz on $\mathbb{R}^m$, establishing the scalar case of Proposition 10.2. (*Hint*: Reduce to the case $m = 1$; show that if $A'(c) > 0$, then A' must change by at least $A'(c)/2$ on an interval of length $\le 2A(c)/A'(c)$, to prevent A from changing sign. Use the mean-value theorem to deduce $|A''(\zeta)| \ge |A'(c)|^2/4A(c)$ for some ζ in this interval.) For the general case of Proposition 10.2, see [Fdln], p. 189.
5. Suppose (10.1) is the system arising in Exercise 4 of §9, for $\mathfrak{X} = (x, v)$. Show that the generator L for P^t is given by

$$L = \frac{\partial^2}{\partial v^2} - \big[\beta v + V'(x)\big]\frac{\partial}{\partial v} + v\frac{\partial}{\partial x}. \tag{10.67}$$

6. Using methods produced in Chap. 8, §6, to derive Mehler's formula, compute the integral kernel for e^{tL} when L is given by (10.67), with $V(x) = ax^2$.
Remark: This integral kernel is smooth for $t > 0$, reflecting the hypoellipticity of $\partial_t - L$. This is a special case of a general phenomenon analyzed in [Ho]. A discussion of this work can also be found in Chap. 15 of [T3].

A. The Trotter product formula

It is often of use to analyze the solution operator to an evolution equation of the form

$$\frac{\partial u}{\partial t} = Au + Bu$$

in terms of the solution operators e^{tA} and e^{tB}, which individually might have fairly simple behavior. The case where A is the Laplace operator and B is multiplication by a function is used in §2 to establish the Feynman–Kac formula, as a consequence of Proposition A.4 below.

The following result, known as the Trotter product formula, was established in [Tro].

Theorem A.1. *Let A and B generate contraction semigroups e^{tA} and e^{tB}, on a Banach space X. If $\overline{A+B}$ is the generator of a contraction semigroup $R(t)$, then*

$$R(t)f = \lim_{n\to\infty} \big(e^{(t/n)A}\, e^{(t/n)B}\big)^n f, \tag{A.1}$$

for all $f \in X$.

Here, $\overline{A+B}$ denotes the closure of $A+B$. A simplified proof in the case where $A+B$ itself is the generator of $R(t)$ is given in an appendix to [Nel2]. We will give that proof.

Proposition A.2. *Assume that A, B, and $A+B$ generate contraction semigroups $P(t)$, $Q(t)$, and $R(t)$ on X, respectively, where $\mathcal{D}(A+B) = \mathcal{D}(A)\cap\mathcal{D}(B)$. Then* (A.1) *holds for all $f \in X$.*

Proof. It suffices to prove (A.1) for $f \in \mathcal{D} = \mathcal{D}(A+B)$. In such a case, we have

$$P(h)Q(h)f - f = h(A+B)f + o(h), \tag{A.2}$$

since $P(h)Q(h)f - f = (P(h)f - f) + P(h)(Q(h)f - f)$. Also, $R(h)f - f = h(A+B) + o(h)$, so

$$P(h)Q(h)f - R(h)f = o(h) \text{ in } X, \quad \text{for } f \in \mathcal{D}.$$

Since $A+B$ is a closed operator, $\mathcal{D}$ is a Banach space in the norm $\|f\|_{\mathcal{D}} = \|(A+B)f\| + \|f\|$. For each $f \in \mathcal{D}$, $h^{-1}\big(P(h)Q(h) - R(h)\big)f$ is a bounded set in X. By the uniform boundedness principle, there is a constant C such that

$$\frac{1}{h}\big\|P(h)Q(h)f - R(h)f\big\| \le C\|f\|_{\mathcal{D}},$$

for all $h > 0$ and $f \in \mathcal{D}$. In other words, $\big\{h^{-1}\big(P(h)Q(h) - R(h)\big) : h > 0\big\}$ is bounded in $\mathcal{L}(\mathcal{D}, X)$, and the family tends strongly to 0 as $h \to 0$. Consequently,

$$\frac{1}{h}\big\|P(h)Q(h)f - R(h)f\big\| \longrightarrow 0$$

uniformly for f is a compact subset of $\mathcal{D}$.

Now, with $t \ge 0$ fixed, for any $f \in \mathcal{D}$, $\{R(s)f : 0 \le s \le t\}$ is a compact subset of $\mathcal{D}$, so

$$\big\|\big(P(h)Q(h) - R(h)\big)R(s)f\big\| = o(h), \tag{A.3}$$

uniformly for $0 \le s \le t$. Set $h = t/n$. We need to show that $\big(P(h)Q(h)\big)^n f - R(hn)f \to 0$, as $n \to \infty$. Indeed, adding and subtracting terms of the form $(P(h)Q(h))^j R(hn - hj)$, and using $\|P(h)Q(h)\| \le 1$, we have

$$\begin{aligned} \big\|\big(P(h)Q(h)\big)^n f - R(hn)f\big\| &\\ \le \big\|\big(P(h)Q(h) - R(h)\big)R(h(n-1))f\big\| &\\ + \big\|\big(P(h)Q(h) - R(h)\big)R(h(n-2))f\big\| &\\ + \cdots + \big\|\big(P(h)Q(h) - R(h)\big)f\big\|. & \end{aligned} \tag{A.4}$$

This is a sum of n terms that are uniformly $o(t/n)$, by (A.3), so the proof is done.

Note that the proof of Proposition A.2 used the contractivity of $P(t)$ and of $Q(t)$, but not that of $R(t)$. On the other hand, the contractivity of $R(t)$ follows from (A.1). Furthermore, the hypothesis that $P(t)$ and $Q(t)$ are contraction semigroups can be generalized to $\|P(t)\| \le e^{at}$, $\|Q(t)\| \le e^{bt}$. If $C = A + B$ generates a semigroup $R(t)$, we conclude that $\|R(t)\| \le e^{(a+b)t}$.

We also note that only certain properties of $S(h) = P(h)Q(h)$ play a role in the proof of Proposition A.2. We use

$$S(h)f - f = hCf + o(h), \quad f \in \mathcal{D} = \mathcal{D}(C), \tag{A.5}$$

where C is the generator of the semigroup $R(h)$, to get

$$S(h)f - R(h)f = o(h), \quad f \in \mathcal{D}. \tag{A.6}$$

As above, we have $h^{-1}\|S(h)f - R(h)f\| \le C\|f\|_{\mathcal{D}}$ in this case, and consequently $h^{-1}\|S(h)f - R(h)f\| \to 0$ uniformly for f in a compact subset of $\mathcal{D}$, such as $\{R(s)f : 0 \le s \le t\}$. Thus we have analogues of (A.3) and (A.4), with $P(h)Q(h)$ everywhere replaced by $S(h)$, proving the following.

Proposition A.3. *Let $S(t)$ be a strongly continuous, operator-valued function of $t \in [0,\infty)$, such that the strong derivative $S'(0)f = Cf$ exists, for $f \in \mathcal{D} = \mathcal{D}(C)$, where C generates a semigroup on a Banach space X. Assume $\|S(t)\| \le 1$ or, more generally, $\|S(t)\| \le e^{ct}$. Then, for all $f \in X$,*

$$e^{tC}f = \lim_{n\to\infty} S(n^{-1}t)^n f. \tag{A.7}$$

This result was established in [Chf], in the more general case where $S'(0)$ has *closure* C, generating a semigroup.

Proposition A.2 applies to the following important family of examples. Let $X = L^p(\mathbb{R}^n)$, $1 \le p < \infty$, or let $X = C_o(\mathbb{R}^n)$, the space of continuous functions vanishing at infinity. Let $A = \Delta$, the Laplace operator, and $B = -M_V$, that is, $Bf(x) = -V(x)f(x)$. If V is bounded and continuous on $\mathbb{R}^n$, then B is bounded on X, so $\Delta - V$, with domain $\mathcal{D}(\Delta)$, generates a semigroup, as shown in Proposition 9.12 of Appendix A. Thus Proposition A.2 applies, and we have the following:

Proposition A.4. *If $X = L^p(\mathbb{R}^n)$, $1 \le p < \infty$, or $X = C_o(\mathbb{R}^n)$, and if V is bounded and continuous on $\mathbb{R}^n$, then, for all $f \in X$,*

$$e^{t(\Delta - V)}f = \lim_{n\to\infty} \left(e^{(t/n)\Delta}e^{-(t/n)V}\right)^n f. \tag{A.8}$$

This is the result used in §2. If $X = L^p(\mathbb{R}^n)$, $p < \infty$, we can in fact take $V \in L^\infty(\mathbb{R}^n)$. See the exercises for other extensions of this proposition.

It will be useful to extend Proposition A.2 to solution operators for time-dependent evolution equations:

$$\frac{\partial u}{\partial t} = Au + B(t)u, \quad u(0) = f. \tag{A.9}$$

We will restrict attention to the special case that A generates a contraction semigroup and $B(t)$ is a continuous family of *bounded* operators on a Banach space X. The solution operator $S(t,s)$ to (A.9), satisfying $S(t,s)u(s) = u(t)$, can be constructed via the integral equation

$$u(t) = e^{tA} f + \int_0^t e^{(t-s)A} B(s)u(s)\,ds, \tag{A.10}$$

parallel to the proof of Proposition 9.12 in Appendix A on functional analysis. We have the following result.

Proposition A.5. *If A generates a contraction semigroup and $B(t)$ is a continuous family of bounded operators on X, then the solution operator to (A.9) satisfies*

$$S(t,0)f = \lim_{t\to\infty} \Big(e^{(t/n)A}\, e^{(t/n)B((n-1)t/n)}\Big)\cdots\Big(e^{(t/n)A}\, e^{(t/n)B(0)}\Big) f, \tag{A.11}$$

for each $f \in X$.

There are n factors in parentheses on the right side of (A.11), the jth from the right being $e^{(t/n)A}e^{(t/n)B((j-1)t/n)}$.

The proof has two parts. First, in close parallel to the derivation of (A.4), we have, for any $f \in \mathcal{D}(A)$, that the difference between the right side of (A.11) and

$$e^{(t/n)(A+B((n-1)t/n))}\cdots e^{(t/n)(A+B(0))} f \tag{A.12}$$

has norm $\le n\cdot o(1/n)$, tending to zero as $n\to\infty$, for t in any bounded interval $[0,T]$. Second, we must compare (A.12) with $S(t,0)f$. Now, for any fixed $t>0$, define $v(s)$ on $0\le s\le t$ by

$$\frac{\partial v}{\partial s} = Av + B\Big(\frac{j-1}{n}t\Big)v, \quad \frac{j-1}{n}t \le s < \frac{j}{n}t; \quad v(0) = f. \tag{A.13}$$

Thus (A.12) is equal to $v(t)$. Now we can write

$$\frac{\partial v}{\partial s} = Av + B(s)v + R(s)v, \quad v(0) = f, \tag{A.14}$$

where, for n large enough, $\|R(s)\| \le \varepsilon$, for $0\le s\le t$. Thus

$$v(t) = S(t,0)f + \int_0^t S(t,s)R(s)v(s)\,ds, \tag{A.15}$$

and the last term in (A.15) is small. This establishes (A.11).

Thus we have the following extension of Proposition A.4. Denote by $BC(\mathbb{R}^n)$ the space of bounded, continuous functions on $\mathbb{R}^n$, with the sup norm.

Proposition A.6. *If* $X = L^p(\mathbb{R}^n)$, $1 \le p < \infty$, *or* $X = C_o(\mathbb{R}^n)$, *and if* $V(t)$ *belongs to* $C\big([0,\infty), BC(\mathbb{R}^n)\big)$, *then the solution operator* $S(t,0)$ *to*

$$\frac{\partial u}{\partial t} = \Delta u - V(t)u$$

satisfies

$$\text{(A.16)}\quad S(t,0)f = \lim_{n\to\infty} \Big(e^{(t/n)\Delta} e^{-(t/n)V((n-1)t/n)}\Big) \cdots \Big(e^{(t/n)\Delta} e^{-(t/n)V(0)}\Big) f,$$

for all $f \in X$.

To end this appendix, we give an alternative proof of the Trotter product formula when $Au = \Delta u$ and $Bu(x) = V(x)u(x)$, which, while valid for a more restricted class of functions $V(x)$ than the proof of Proposition A.4, has some desirable features. Here, we define $v_k = \Big(e^{(1/n)\Delta} e^{-(1/n)V}\Big)^k f$ and set

$$\text{(A.17)}\qquad v(t) = e^{s\Delta} e^{-sV} v_k, \quad \text{for } t = \frac{k}{n} + s,\ 0 \le s \le \frac{1}{n}.$$

We use Duhamel's principle to compare $v(t)$ with $u(t) = e^{t(\Delta - V)} f$. Note that $v(t) \to v_{k+1}$ as $t \nearrow (k+1)/n$, and for $k/n < t < (k+1)/n$,

$$\text{(A.18)}\qquad \begin{aligned} \frac{\partial v}{\partial t} &= \Delta v - e^{s\Delta} V e^{-sV} v_k \\ &= (\Delta - V)v + [V, e^{s\Delta}] e^{-sV} v_k. \end{aligned}$$

Thus, by Duhamel's principle,

$$\text{(A.19)}\qquad v(t) = e^{t(\Delta - V)} f + \int_0^t e^{(t-s)(\Delta - V)} R(s)\, ds,$$

where

$$\text{(A.20)}\qquad R(s) = [V, e^{\sigma\Delta}] e^{-\sigma V} v_k, \quad \text{for } s = \frac{k}{n} + \sigma,\ 0 \le \sigma < \frac{1}{n}.$$

We can write $[V, e^{\sigma\Delta}] = [V, e^{\sigma\Delta} - 1]$, and hence

$$\text{(A.21)}\qquad R(s) = V(e^{\sigma\Delta} - 1)e^{-\sigma V} v_k - (e^{\sigma\Delta} - 1)V e^{-\sigma V} v_k.$$

Now, as long as

$$\text{(A.22)}\qquad \mathcal{D}(\Delta - V) = \mathcal{D}(\Delta) = H^2(\mathbb{R}^n),$$

we have, for $0 \le \gamma \le 1$,

$$\|e^{t(\Delta-V)}\|_{\mathcal{L}(H^{-2\gamma},L^2)} = \|e^{t(\Delta-V)}\|_{\mathcal{L}(L^2,H^{2\gamma})} \le C(T)t^{-\gamma}, \tag{A.23}$$

for $0 < t \le T$. Thus, if we take $\gamma \in (0,1)$ and $t \in (0,T]$, we have for

$$F(t) = \int_0^t e^{(t-s)(\Delta-V)} R(s)\, ds, \tag{A.24}$$

the estimate

$$\|F(t)\|_{L^2} \le C \int_0^t (t-s)^{-\gamma}\, \|R(s)\|_{H^{-2\gamma}}\, ds. \tag{A.25}$$

We can estimate $\|R(s)\|_{H^{-2\gamma}}$ using (A.21), together with the estimate

$$\|e^{\sigma\Delta} - 1\|_{\mathcal{L}(L^2,H^{-2\gamma})} \le C\, \sigma^{\gamma}, \quad 0 \le \gamma \le 1. \tag{A.26}$$

Since $\sigma \in [0, 1/n]$ in (A.21), we have

$$\begin{aligned} &\|R(s)\|_{H^{-2\gamma}} \le C n^{-\gamma} \varphi(V) \|f\|_{L^2}, \\ &\varphi(V) = \Big(\|V\|_{\mathcal{L}(H^{2\gamma})} + \|V\|_{L^\infty}\Big) e^{s\|V\|_{L^\infty}}. \end{aligned} \tag{A.27}$$

Thus, estimating $v(t) = u(t)$ at $t = 1$, we have

$$\Big\| \Big(e^{(1/n)\Delta} e^{-(1/n)V}\Big)^n f - e^{(\Delta-V)} f \Big\|_{L^2} \le C_\gamma \varphi(V) \|f\|_{L^2} \cdot n^{-\gamma}, \tag{A.28}$$

for $0 < \gamma < 1$, provided multiplication by V is a bounded operator on $H^{2\gamma}(\mathbb{R}^n)$. Note that this holds if $D^\alpha V \in L^\infty(\mathbb{R}^n)$ for $|\alpha| \le 2$, and

$$\|V\|_{\mathcal{L}(H^{2\gamma})} \le C \sup_{|\alpha| \le 2} \|D^\alpha V\|_{L^\infty}. \tag{A.29}$$

One can similarly establish the estimate

$$\Big\| \Big(e^{(t/n)\Delta} e^{-(t/n)V}\Big)^n f - e^{t(\Delta-V)} f \Big\|_{L^2} \le C(t) \varphi(V) \|f\|_{L^2} \cdot n^{-\gamma}. \tag{A.30}$$

Exercises

1. Looking at Exercises 2–4 of §2, Chap. 8, extend Proposition A.4 to any V, continuous on $\mathbb{R}^n$, such that Re $V(x)$ is bounded from below and |Im $V(x)$| is bounded.
   (*Hint*: First apply those exercises directly to the case where V is smooth, real-valued, and bounded from below.)

2. Let $H = L^2(\mathbb{R})$, $Af = df/dx$, $Bf = ixf(x)$, so $e^{tA}f(x) = f(x+t)$, $e^{tB}f(x) = e^{itx}f(x)$. Show that Theorem A.1 applies to this case, but not Proposition A.2. Compute both sides of

$$e^{pA+qB}f = \lim_{n\to\infty} \big(e^{(p/n)A}e^{(q/n)B}\big)^n f,$$

and verify this identity directly.
Compare with the discussion of the Heisenberg group, in §14 of Chap. 7.
3. Suppose A and B are *bounded* operators. Show that

$$\big\| e^{t(A+B)} - \big(e^{(t/n)A}e^{(t/n)B}\big)^n \big\| \le \frac{Ct}{n}$$

and that

$$\big\| e^{t(A+B)} - \big(e^{(t/2n)A}e^{(t/n)B}e^{(t/2n)A}\big)^n \big\| \le \frac{ct}{n^2}.$$

(*Hint*: Use the power series expansions for $e^{(t/n)A}$, and so forth.)

References

[Az] R. Azencott, Behavior of diffusion semi-groups at infinity, *Bull. Soc. Math. France* 102(1974), 193–240

[B] D. Bell, *The Malliavin Calculus*, Longman, Essex, 1987

[BG] R. Blumenthal and R. Getoor, *Markov Processes and Potential Theory*, Academic, New York, 1968

[CM] R. Cameron and W. Martin, Evaluation of various Wiener integrals by use of certain Sturm-Liouville differential equations, *Bull. AMS* 51(1945), 73–90

[Chf] P. Chernoff, Note on product formulas for operator semigroups, *J. Func. Anal.* 2(1968), 238–242

[CW] K. Chung and R. Williams, *Introduction to Stochastic Integration*, Birkhauser, Boston, 1990

[Db1] J. Doob, The Brownian movements and stochastic equations, *Ann. Math*. 43(1942), 351–369

[Db2] J. Doob, *Stochastic Processes*, Wiley, New York, 1953

[Db3] J. Doob, *Classical Potential Theory and its Probabilistic Counterpart*, Springer, New York, 1984

[DS] N. Dunford and J. Schwartz, *Linear Operators*, Wiley, New York, 1958

[Dur] R. Durrett, *Brownian Motion and Martingales in Analysis*, Wadsworth, Belmont, CL, 1984

[Ein] A. Einstein, *Investigations on the Theory of the Brownian Movement*, Dover, New York, 1956

[El] K. Elworthy, *Stochastic Differential Equations on Manifolds*, LMS Lecture Notes #70, Cambridge University Press, Cambridge, 1982

[Em] M. Emery, *Stochastic Calculus in Manifolds*, Springer, New York, 1989

[Fdln] M. Freidlin, *Functional Integration and Partial Differential Equations*, Princeton University Press, Princeton, NJ, 1985

[Fr] A. Friedman, *Stochastic Differential Equations and Applications*, Vols. 1 & 2, Academic, New York, 1975

[HP] E. Hille and R. Phillips, *Functional Analysis and Semi-groups*, Colloq. Publ. AMS, Providence, RI, 1957

[Ho] L. Hörmander, Hypoelliptic second order differential equations, *Acta Math.* 119 (1967), 147–171

[IkW] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North Holland, Amsterdam, 1981

[Ito] K. Ito, *On Stochastic Differential Equations*, Memoirs AMS #4, 1951

[IMc] K. Ito and H. McKean, *Diffusion Processes and Their Sample Paths*, Springer, New York, 1974

[Kac] M. Kac, *Probability and Related Topics in Physical Sciences*, Wiley, New York, 1959

[Kal] G. Kallianpur, Stochastic Filtering Theory, Springer, New York, 1980

[KS] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, New York, 1988

[K] T. Kato, *Perturbation Theory for Linear Operators*, Springer, New York, 1966

[Kol] A. Kolmogorov, Uber die analytishen Methoden in Wahrscheinlichkeitsrechnung, *Math. Ann.* 104(1931), 415–458

[Lam] J. Lamperti, *Stochastic Processes*, Springer, New York, 1977

[Lev] P. Lévy, Random functions, *Univ. of Calif. Publ. in Statistics* I(12)(1953), 331–388

[Mal] P. Malliavin, Stochastic calculus of variations and hypoelliptic operators, *Proc. Intern. Symp. SDE, Kyoto* (1976), 195–263

[McK] H. McKean, *Stochastic Integrals*, Academic, New York, 1969

[Mey] P. Meyer, *Probability and Potentials*, Blaisdell, Waltham, MA, 1966

[Nel] E. Nelson, *Operator Differential Equations*, Graduate Lecture Notes, Princeton University, Princeton, NJ, 1965

[Nel2] E. Nelson, Feynman integrals and the Schrödinger equation, *J. Math. Phys.* 5(1964), 332–343

[Nel3] E. Nelson, *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton, NJ, 1967

[Øk] B. Øksendal, *Stochastic Differential Equations*, Springer, New York, 1989

[Par] E. Pardoux, Stochastic partial differential equations, a review, *Bull. des Sciences Math.* 117(1993), 29–47

[Pet] K. Petersen, *Brownian Motion, Hardy Spaces, and Bounded Mean Oscillation*, LMS Lecture Notes #28, Cambridge University Press, Cambridge, 1977

[PS] S. Port and C. Stone, *Brownian Motion and Classical Potential Theory*, Academic, New York, 1979

[RaT] J. Rauch and M. Taylor, Potential and scattering theory on wildly perturbed domains, *J. Func. Anal.* 18(1975), 27–59

[RS] M. Reed and B. Simon, *Methods of Mathematical Physics*, Academic, New York, Vols. 1,2, 1975; Vols. 3,4, 1978

[Sch] Z. Schuss, *Theory and Applications of Stochastic Differential Equations*, Wiley, New York, 1980

[Si] B. Simon, *Functional Integration and Quantum Physics*, Academic, New York, 1979

[Stk] D. Stroock, The Kac approach to potential theory I, *J. Math. Mech.* 16(1967), 829–852

[Stk2] D. Stroock, The Malliavin calculus, a functional analytic approach, *J. Func. Anal.* 44(1981), 212–257

[StV] D. Stroock and S. Varadhan, *Multidimensional Diffusion Processes*, Springer, New York, 1979

[T] M. Taylor, Scattering length and perturbations of $-\Delta$ by positive potentials, *J. Math. Anal. Appl*. 53(1976), 291–312

[T2] M. Taylor, Estimate on the fundamental frequency of a drum, *Duke Math. J.* 46(1979), 447–453

[T3] M. Taylor, *Pseudodifferential Operators*, Princeton Univ. Press, Princeton, NJ, 1981

[Tro] H. Trotter, On the product of semigroups of operators, *Proc. AMS* 10(1959), 545–551

[Ts] M. Tsuji, *Potential Theory and Modern Function Theory*, Chelsea, New York, 1975

[UO] G. Uhlenbeck and L. Ornstein, On the theory of Brownian motion, *Phys. Rev*. 36(1930), 823–841

[Wal] J. Walsch, An introduction to stochastic partial differential equations, pp. 265–439 in *Ecole d'été de Probabilité de Saint-Fleur XIV*, LNM #1180, Springer, New York, 1986

[Wie] N. Wiener, Differential space, *J. Math. Phys*. 2(1923), 131–174

[Yo] K. Yosida, *Functional Analysis*, Springer, New York, 1965

12

The $\overline{\partial}$-Neumann Problem

Introduction

Here we study a boundary problem arising in the theory of functions of several complex variables. A function u on an open domain $\Omega \subset \mathbb{C}^n$ is holomorphic if $\overline{\partial} u = 0$, where

$$\overline{\partial} u = \sum_j \frac{\partial u}{\partial \overline{z}_j} \, d\overline{z}_j, \tag{0.1}$$

with $d\overline{z}_j = dx_j - i\,dy_j$ and

$$\frac{\partial u}{\partial \overline{z}_j} = \frac{1}{2}\left(\frac{\partial u}{\partial x_j} + i\,\frac{\partial u}{\partial y_j}\right). \tag{0.2}$$

In the study of complex function theory on Ω, one is led to consider the equation

$$\overline{\partial} u = f, \tag{0.3}$$

with $f = \sum f_j \, d\overline{z}_j$. More generally, one studies (0.3) as an equation for a $(0, q)$-form u, given a $(0, q+1)$-form f; definitions of these terms are given in §1. One is led to a study of a boundary problem for the second-order operator

$$\square = \overline{\partial}\,\overline{\partial}^* + \overline{\partial}^* \overline{\partial}, \tag{0.4}$$

the $\overline{\partial}$-Neumann boundary problem, which will also be specified in §1. While the operator $\square$ is elliptic, the boundary condition does not satisfy the regularity condition dealt with in Chap. 5. The solution to this boundary problem by J. J. Kohn [K1] thus marked an important milestone in the theory of linear PDE, as well as a significant advance in complex function theory.

The way that (0.3) leads to the $\overline{\partial}$-Neumann problem is somewhat parallel to the way the deRham complex leads to the boundary problems for the Hodge

M.E. Taylor, *Partial Differential Equations II: Qualitative Studies of Linear Equations*, Applied Mathematical Sciences 116, DOI 10.1007/978-1-4419-7052-7_6,

Laplacian discussed in §9 of Chap. 5. Appendix A to this chapter puts the deRham complex in a general context. Though appendices to chapters in this book are almost always put at the end of their chapters, we put this one at the beginning, since its role is to link the previous material on the Hodge Laplacian, particularly with absolute boundary conditions, to the material of this chapter.

In §1 we introduce the $\overline{\partial}$ complex and the $\overline{\partial}$-Neumann problem, and show that it is not regular. §§2–4 are devoted to establishing replacements for the regular elliptic estimates established in Chap. 5, for (0, 1)-forms, under a "geometrical" hypothesis on Ω, namely that it be "strongly pseudoconvex." This notion is defined in §2, in the course of establishing an estimate of C. B. Morrey. In §3 we show how this leads to a "1/2-estimate," to wit, an estimate of the form

$$Q(u,u) \geq C\|u\|^2_{H^{1/2}}, \tag{0.5}$$

in a situation where a regular elliptic boundary problem would yield an estimate on $\|u\|^2_{H^1}$. We then define the Friedrichs extension $\mathcal{L}$ of $\Box$, and show that it has compact resolvent. In §4 we produce higher-order a priori estimates, of the form

$$\|u\|_{H^{k+1}} \leq C\|\mathcal{L}u\|_{H^k}, \tag{0.6}$$

assuming $u \in \mathcal{D}(\mathcal{L})$ is smooth on $\overline{\Omega}$. In §5 we establish the associated regularity theorem, that $\mathcal{L}^{-1} : H^k(\Omega, \Lambda^{0,1}) \to H^{k+1}(\Omega, \Lambda^{0,1})$. Following [KN], we use the method of elliptic regularization to accomplish this.

In §6 we apply the results established in §§2–5 to solve (0.3), when f is a (0, 1)-form satisfying $\overline{\partial} f = 0$ and smooth on $\overline{\Omega}$ (assumed to be strongly pseudoconvex). We obtain a solution $u \in C^\infty(\overline{\Omega})$ under these hypotheses. As a consequence, we show that such Ω is a "domain of holomorphy"; that is, there exist holomorphic functions on Ω that cannot be extended beyond any point of $\partial\Omega$.

In §7 we derive a formula for the orthogonal projection B of $L^2(\Omega)$ onto the subspace $\mathcal{H}(\Omega)$ of L^2-holomorphic functions on Ω, in terms of $\mathcal{L}^{-1}$ acting on (0, 1)-forms, and we establish some consequences. We consider Toeplitz operators, of the form $T_f = BM_f$, on $\mathcal{H}(\Omega)$, for $f \in C(\overline{\Omega})$. We show that T_f is Fredholm if $f\big|_{\partial\Omega}$ is invertible, and we briefly discuss the problem of computing the index of T_f; this index problem is related to index problems considered in Chap. 10.

In §§2–6 we concentrate on (0, 1)-forms, making use of this theory to study (0, 0)-forms in §7. In §8 we study the $\overline{\partial}$-Neumann problem on $(0, q)$-forms for general $q \geq 1$. The main point is to extend Morrey's inequality. Once this is done, it is routine to extend the arguments of §§3–6. We also have in §8 an extension of results of §7 regarding the compactness of commutators of M_f, for $f \in C(\overline{\Omega})$, with certain projections, namely the orthogonal projections onto the positive, negative, and zero spectral subspaces of the relevant closed extension of $\overline{\partial} + \overline{\partial}^*$ on $\bigoplus_{q\geq 0} L^2(\Omega, \Lambda^{0,q})$.

In §9 we discuss a method that provides an alternative to the sort of energy estimates done in §§2–5, namely reduction of the problem to an equation on the boundary, of the form

$$\Box^+ g = f, \tag{0.7}$$

where $\Box^+$ is a first-order pseudodifferential operator on $\partial\Omega$. The operator $\Box^+$ is not elliptic; its principal symbol is ≥ 0 and vanishes to second order on a ray bundle over $\partial\Omega$. We show that this operator is hypoelliptic when Ω is strongly pseudoconvex. In fact, we do this via energy estimates that are not completely different from those arising in earlier sections, though alternative approaches to the analysis of $\Box^+$ are mentioned.

Up through §9, our attention is confined to the $\overline{\partial}$-complex on domains in $\mathbb{C}^n$. In §10 we analyze the $\overline{\partial}$-complex on a strongly pseudoconvex, complex manifold; in fact, we consider manifolds with an almost complex structure, satisfying an integrability condition, that can be stated as

$$\overline{\partial}^2 = 0. \tag{0.8}$$

By doing the estimates in this context, one is able to prove the Newlander–Nirenberg theorem, that an integrable, almost complex structure actually is a complex structure, admitting local holomorphic coordinates. (In fact, all this was done by Kohn in [K1].)

At the end of this chapter are two additional appendices. Appendix B gives some complementary results on the Levi form, introduced in §2 in the course of deriving Morrey's inequality and defining strong pseudoconvexity. Appendix C derives a result on the Neumann operator $\mathcal{N}$ for the Dirichlet problem (for the Laplace operator), useful for the analysis in §9. Namely, we specify the principal symbol of $\mathcal{N} + \sqrt{-\Delta_X} \in OPS^0(\partial\Omega)$, where Δ_X is the Laplace operator on $X = \partial\Omega$, in terms of the second fundamental form of $\partial\Omega \hookrightarrow \overline{\Omega}$.

Other methods have been applied to the $\overline{\partial}$-complex. We mention particularly the method of weighted L^2-estimates, such as done in [AV, Ho1], and [Ho3]. These methods also apply directly to general pseudoconvex domains (i.e., one can omit the "strong"), though they ignore detailed boundary behavior. Another approach to the $\overline{\partial}$-equation is given in [EMM]. There has also been considerable work on the $\overline{\partial}$-Neumann problem on various classes of weakly pseudoconvex domains, including particularly [Cat, Chr, FeK, K2], and [NRSW].

There is a very different approach to the $\overline{\partial}$-equation, making use of explicit integral kernels; see the survey article of [Kh], or [HP1].

As another offshoot of the study of the $\overline{\partial}$-complex, we mention the study of "CR manifolds" (of which the boundary of a complex domain is the simplest example); surveys of this are given in [Bog] and [Tai]. Also, there are studies of general overdetermined systems; see [Sp] for a survey.

A. Elliptic complexes

We give a brief discussion of a setting in which the study of boundary problems for the Hodge Laplacian in §9 of Chap. 5 can be included. The Hodge Laplacian arises from the deRham complex, a sequence of maps

$$\cdots \xrightarrow{d} \Lambda^k \xrightarrow{d} \Lambda^{k+1} \xrightarrow{d} \cdots, \tag{A.1}$$

where d is the exterior derivative. Key properties are that $d^2 = 0$ and that the symbol sequence is exact, for each nonzero $\xi \in T_x^*$:

$$\cdots \xrightarrow{\wedge\xi} \Lambda^k T_x^* \xrightarrow{\wedge\xi} \Lambda^{k+1} T_x^* \xrightarrow{\wedge\xi} \cdots. \tag{A.2}$$

This implies that $d^*d + dd^*$ is elliptic.

More generally, consider a sequence of first-order differential operators between sections of vector bundles $F_j \to \overline{M}$. For notational simplicity, we will use script (e.g., $\mathcal{F}_j$) to denote spaces of smooth sections of various vector bundles. Suppose we have maps

$$\vartheta : \mathcal{F}_k \longrightarrow \mathcal{F}_{k+1}, \quad \vartheta^2 = 0. \tag{A.3}$$

Suppose the symbol sequence is exact; we are said to have an *elliptic complex*. Set

$$E_0 = \bigoplus_j F_{2j}, \quad E_1 = \bigoplus_j F_{2j+1}, \quad E = E_0 \oplus E_1, \tag{A.4}$$

so we also have spaces of sections, $\mathcal{E}_0, \mathcal{E}_1$, and $\mathcal{E}$. Using metrics on F_j and $\overline{M}$ to define $\vartheta^* : \mathcal{F}_k \to \mathcal{F}_{k-1}$, we have

$$D_0 : \mathcal{E}_0 \longrightarrow \mathcal{E}_1, \quad D_1 : \mathcal{E}_1 \longrightarrow \mathcal{E}_0, \quad D_j = \vartheta + \vartheta^*, \tag{A.5}$$

and we fit these together to form $D : \mathcal{E} \to \mathcal{E}$. Since $\vartheta^2 = 0$, we have

$$D^*D = D^2 = \vartheta\vartheta^* + \vartheta^*\vartheta : \mathcal{F}_k \longrightarrow \mathcal{F}_k. \tag{A.6}$$

Now the general Green formula implies

$$(D^*Du, v) = (Du, Dv) + \beta(u, v), \tag{A.7}$$

with boundary term

$$\beta(u, v) = \frac{1}{i} \int_{\partial M} \Big[\langle \sigma_{D_1}(x, \nu) D_0 u_0, v_0 \rangle + \langle \sigma_{D_0}(x, \nu) D_1 u_1, v_1 \rangle \Big] \, dS, \tag{A.8}$$

where we write $u \in \mathcal{E}$ as $u = u_0 + u_1,\ u_j \in \mathcal{E}_j$. For the sake of definiteness, let us take

$$u \in \mathcal{F}_\ell, \quad v \in \mathcal{F}_\ell, \tag{A.9}$$

for a fixed ℓ. Then

$$\begin{aligned}\beta(u,v) &= \frac{1}{i}\int\limits_{\partial M} \Big\langle (\vartheta + \vartheta^*)u, \sigma_{(\vartheta+\vartheta^*)}(x,\nu)v \Big\rangle \, dS \\ &= \frac{1}{i}\int\limits_{\partial M} \Big[\langle \vartheta u, \sigma_\vartheta(x,\nu)v\rangle + \langle \vartheta^* u, \sigma_{\vartheta^*}(x,\nu)v\rangle \Big] \, dS.\end{aligned} \tag{A.10}$$

We rewrite this in two different ways, parallel to (9.1) and (9.2) of Chap. 5, respectively, namely

$$\beta(u,v) = \frac{1}{i}\int\limits_{\partial M} \Big[\langle \vartheta u, \sigma_\vartheta(x,\nu)v\rangle + \langle \sigma_\vartheta(x,\nu)\vartheta^* u, v\rangle \Big] \, dS \tag{A.11}$$

and

$$\beta(u,v) = \frac{1}{i}\int\limits_{\partial M} \Big[\langle \sigma_{\vartheta^*}(x,\nu)\vartheta u, v\rangle + \langle \vartheta^* u, \sigma_{\vartheta^*}(x,\nu)v\rangle \Big] \, dS. \tag{A.12}$$

Thus there arise two boundary problems for D^*D on $\mathcal{F}_\ell$, the generalization of the "relative" boundary condition (9.4):

$$\sigma_\vartheta(x,\nu)u = 0, \quad \sigma_\vartheta(x,\nu)\vartheta^* u = 0 \text{ on } \partial M, \tag{A.13}$$

and the generalization of the "absolute" boundary condition (9.5) of Chap. 5:

$$\sigma_{\vartheta^*}(x,\nu)u = 0, \quad \sigma_{\vartheta^*}(x,\nu)\vartheta u = 0 \text{ on } \partial M. \tag{A.14}$$

In each case we have $\beta(u,v) = 0$ provided u and v satisfy the boundary condition. We remark that the "absolute" boundary condition (A.14) is often called the "abstract Neumann boundary condition."

Define Sobolev spaces $H^j_b(M, F_\ell)$ in analogy with (9.11) of Chap. 5, with $b = R$ or A; namely, $u \in H^1(M, F_\ell)$ belongs to $H^1_b(M, F_\ell)$ if and only if the zero-order boundary condition in (A.13) (for $b = R$) or (A.14) (for $b = A$) is satisfied, and $u \in H^2(M, F_\ell)$ belongs to $H^2_b(M, F_\ell)$ if and only if both boundary conditions, in either (A.13) or (A.14), are satisfied.

Lemma A.1. *Given $u \in H_b^1(M, F_\ell)$ and $D^*Du \in L^2(M, F_\ell)$, then $\beta(u, v) = 0$ for all $v \in H_b^1(M, F_\ell)$ if and only if all the appropriate boundary data for u vanish (e.g., $\sigma_\vartheta(x, \nu)\vartheta^*u = 0$ on ∂M, in case $b = R$).*

Proof. We need to establish the "only if" part. Take the case $b = R$. To start the argument, pick $\sigma \in C^\infty\Big(\overline{M}, \text{Hom}(F_{\ell-1}, F_\ell)\Big)$ such that $\sigma(x) = \sigma_\vartheta(x, \nu)$ for $x \in \partial M$. Then, for any $w \in \mathcal{F}_{\ell-1}$, we have $v = \sigma w \in H_R^1(M, F_\ell)$, and hence, for any $u \in H_R^1(M, F_\ell)$,

$$\begin{aligned}\beta(u, v) &= \frac{1}{i}\int_{\partial M} \Big\langle \sigma_\vartheta(x, \nu)\vartheta^*u, \sigma_\vartheta(x, \nu)w\Big\rangle\, dS \\ &= \frac{1}{i}\int_{\partial M} \Big\langle \sigma_\vartheta(x, \nu)^*\sigma_\vartheta(x, \nu)\Big(\vartheta^*u\Big), w\Big\rangle\, dS.\end{aligned} \tag{A.15}$$

This vanishes for all $w \in \mathcal{F}_{\ell-1}$ if and only if $\sigma_\vartheta(x, \nu)^*\sigma_\vartheta(x, \nu)(\vartheta^*u) = 0$ on ∂M, which in turn occurs if and only if $\sigma_\vartheta(x, \nu)(\vartheta^*u) = 0$ on ∂M. This establishes the lemma for $b = R$; the case $b = A$ is similar.

Of course, the method of proof of the existence and regularity results in Propositions 9.4–9.7 of Chap. 5, via Lemma 9.2, does not extend to this more general situation. It is conceivable that one of the boundary conditions, (A.13) or (A.14), for $L = D^*D$, could be regular for all ℓ, for some ℓ, or for no ℓ. Since L is strongly elliptic, Proposition 11.13 of Chap. 5 can be used to examine regularity. We will now investigate consequences of the hypothesis that one of these boundary conditions is regular, for L acting on sections of F_ℓ. We will call this hypothesis $\text{Reg}(\vartheta, \ell, b)$, with $b = R$ or A.

Let us define the unbounded operator $D_{\ell b}$ on $L^2(M, F_\ell) \to L^2(M, E)$ to be the closure of D acting on $H_b^1(M, F_\ell)$. Let $D_{\ell b}^*$ denote the Hilbert space adjoint of $D_{\ell b}$, an unbounded operator on $L^2(M, E)$. Then $\mathcal{L} = D_{\ell b}^* D_{\ell b}$ is an unbounded, self-adjoint operator on $L^2(M, F_\ell)$, with dense domain $\mathcal{D}(\mathcal{L})$. Since for all $u \in \mathcal{D}(\mathcal{L})$, $v \in H_b^1(M, F_\ell)$, we have $(\mathcal{L}u, v) = (Du, Dv)$, taking $v \in \mathcal{F}_\ell$ compactly supported in the interior M implies $\mathcal{L}u = D^*Du$ in $\mathcal{D}'(M)$. Hence u has well-defined boundary data, in (A.13) or (A.14), and, by Lemma A.1, the appropriate boundary data vanish. Therefore, the regularity result of Proposition 11.14 in Chap. 5 is applicable; we have $\mathcal{D}(\mathcal{L}) \subset H_b^2(M, F_\ell)$, under the hypothesis $\text{Reg}(\vartheta, \ell, b)$. The reverse inclusion is easy. If we define $\mathcal{L}_b$ to be D^*D on $H_b^2(M, F_\ell)$, it follows that $\mathcal{L}_b$ is a symmetric extension of $\mathcal{L}$, but a self-adjoint operator cannot have a proper symmetric extension. Thus

$$\mathcal{D}(\mathcal{L}) = H_b^2(M, F_\ell),$$

granted the hypothesis $\text{Reg}(\vartheta, \ell, b)$. We restate this as follows:

Proposition A.2. *Under the hypothesis $\text{Reg}(\vartheta, \ell, b)$, the operator $\mathcal{L}$ defined by*

$$\mathcal{D}(\mathcal{L}) = H_b^2(M, F_\ell), \quad \mathcal{L}u = D^* Du \text{ on } \mathcal{D}(\mathcal{L}) \tag{A.16}$$

is self-adjoint.

It follows then from $\text{Reg}(\vartheta, \ell, b)$ that $\text{Ker}\, \mathcal{L}$ is a finite-dimensional subspace of $\mathcal{F}_\ell$; call it $\mathcal{H}_\ell^b$. Parallel to (9.38) of Chap. 5, we have

$$u \in \mathcal{H}_\ell^b \Longleftrightarrow u \in H_b^1(M, F_\ell) \text{ and } \vartheta u = \vartheta^* u = 0. \tag{A.17}$$

Denote by P_h^b the orthogonal projection of $L^2(M, F_\ell)$ onto $\mathcal{H}_\ell^b$. As in (9.38)–(9.39) of Chap. 5, we have continuous maps

$$G^b : L^2(M, F_\ell) \longrightarrow H_b^2(M, F_\ell) \tag{A.18}$$

such that G^b annihilates $\mathcal{H}_\ell^b$ and inverts L on the orthogonal complement of $\mathcal{H}_\ell^b$, so

$$LG^b u = (I - P_h^b)u, \quad \text{for } u \in L^2(M, F_\ell) \tag{A.19}$$

and, by elliptic regularity, $G^b : H^j(M, F_\ell) \to H^{j+2}(M, F_\ell)$. The following result generalizes Proposition 9.8 of Chap. 5.

Proposition A.3. *Granted $\text{Reg}(\vartheta, \ell, b)$, then given $u \in H^j(M, F_\ell),\ j \ge 0$, we have*

$$u = \vartheta\vartheta^* G^b u + \vartheta^*\vartheta G^b u + P_h^b u = P_\vartheta^b u + P_{\vartheta^*}^b u + P_h^b u. \tag{A.20}$$

The three terms on the right side are mutually orthogonal in $L^2(M, F_\ell)$. Furthermore,

$$P_\vartheta^b, P_{\vartheta^*}^b, P_h^b : H^j(M, F_\ell) \longrightarrow H^j(M, F_\ell).$$

Proof. Only the orthogonality remains to be checked. As in the proof of Proposition 9.8 of Chap. 5, we use

$$(\vartheta u, v) = (u, \vartheta^* v) + \gamma(u, v), \tag{A.21}$$

for sections u of F_{j-1} and v of F_j, with

$$\gamma(u, v) = \frac{1}{i} \int_{\partial M} \langle \sigma_\vartheta(x, \nu)u, v\rangle \, dS = \frac{1}{i} \int_{\partial M} \langle u, \sigma_{\vartheta^*}(x, \nu)v\rangle \, dS. \tag{A.22}$$

Note that $\gamma(u, v) = 0$ if either $u \in H_R^1(M, F_{j-1})$ or $v \in H_A^1(M, F_j)$. In particular, we see that

$$u \in H^1_R(M, F_{j-1}) \Longrightarrow \vartheta u \perp \ \ker \vartheta^* \cap H^1(M, F_j), \tag{A.23}$$

$$v \in H^1_A(M, F_j) \Longrightarrow \vartheta^* v \perp \ \ker \vartheta \cap H^1(M, F_{j-1}). \tag{A.24}$$

From the definitions, we have

$$\begin{aligned} \vartheta^* : H^2_R(M, F_j) &\longrightarrow H^1_R(M, F_{j-1}), \\ \vartheta : H^2_A(M, F_j) &\longrightarrow H^1_A(M, F_{j+1}), \end{aligned} \tag{A.25}$$

so

$$\vartheta \vartheta^* H^2_R(M, F_\ell) \perp \ \ker \vartheta^* \cap H^1(M, F_\ell) \tag{A.26}$$

and

$$\vartheta^* \vartheta H^2_A(M, F_\ell) \perp \ \ker \vartheta \cap H^1(M, F_\ell). \tag{A.27}$$

Now (A.26) and (A.27) imply, respectively, for the ranges,

$$\mathcal{R}(P^R_\vartheta) \perp \mathcal{R}(P^R_{\vartheta^*}) + \mathcal{R}(P^R_h) \ \text{ and } \ \mathcal{R}(P^A_{\vartheta^*}) \perp \mathcal{R}(P^A_\vartheta) + \mathcal{R}(P^A_h). \tag{A.28}$$

Now, if $u \in \mathcal{H}^R_\ell$ and $v = \vartheta G^R w$, then $\gamma(u, v) = 0$, so $(u, \vartheta^* v) = (\vartheta u, v) = 0$. Similarly, if $v \in \mathcal{H}^A_\ell$ and $u = \vartheta^* G^A w$, then $\gamma(u, v) = 0$, so $(\vartheta u, v) = (u, \vartheta^* v) = 0$. Thus

$$\mathcal{R}(P^R_{\vartheta^*}) \perp \mathcal{R}(P^R_h) \ \text{ and } \ \mathcal{R}(P^A_\vartheta) \perp \mathcal{R}(P^A_h). \tag{A.29}$$

The proof is complete.

Even though the proof of Proposition A.3 is perfectly parallel to that of Proposition 9.11 of Chap. 5, we have included the details, as they will be needed for an argument below that is *not* parallel to one of §9 in Chap. 5.

The application made to relative cohomology in (9.51)–(9.55) of Chap. 5 does not have a straightforward extension to the general setting. The natural generalization of $C^\infty_r(\overline{M}, \Lambda^k)$ in (9.51) is

$$\mathcal{F}^R_k = \Big\{ u \in \mathcal{F}_k : \sigma_\vartheta(x, \nu) u = 0 \text{ on } \partial M \Big\}, \tag{A.30}$$

but in contrast to (9.52), we cannot expect in general to have

$$\vartheta : \mathcal{F}^R_k \longrightarrow \mathcal{F}^R_{k+1}. \tag{A.31}$$

Of course, we do have $\vartheta : \mathcal{F}_k \to \mathcal{F}_{k+1}$. We can define $\mathcal{E}^k \subset \mathcal{C}^k \subset \mathcal{F}_k$ as the image and kernel of ϑ, respectively, and then we have cohomology groups

$$\mathcal{H}^k(\vartheta) = \mathcal{C}^k / \mathcal{E}^k. \tag{A.32}$$

The argument in Proposition 9.11 of Chap. 5, relating $\mathcal{H}^k(\overline{M})$ to the space $\mathcal{H}_k^A$ of harmonic forms, used a homotopy argument, which has no analogue in the general case. However, another approach works, to give the following:

Proposition A.4. *Under the hypothesis Reg*(ϑ, ℓ, A)*, there is a natural isomorphism*

$$\mathcal{H}_\ell^A \approx \mathcal{H}^\ell(\vartheta). \tag{A.33}$$

Proof. Let $u \in \mathcal{F}_\ell$, $\vartheta u = 0$. Use the orthogonal decomposition (A.20), with $b = A$, to write $u = P_\vartheta^A u + P_{\vartheta *}^A u + P_h^A u$. Now (A.27) implies $P_{\vartheta *}^A u = 0$, so $u = \vartheta(\vartheta^* G^A u) + P_h^A u$, hence every $u \in \mathcal{C}^\ell$ is cohomologous to an element of $\mathcal{H}_\ell^A$. Thus the natural homomorphism arising from $\mathcal{H}_\ell^A \subset \mathcal{C}^\ell$,

$$\tilde{\kappa} : \mathcal{H}_\ell^A \longrightarrow \mathcal{H}^\ell(\vartheta),$$

is surjective. The proof that $\tilde{\kappa}$ is injective is parallel to the argument used in Proposition 9.11 of Chap. 5. If $v \in \mathcal{H}_\ell^A$ and $v = \vartheta u$, $u \in \mathcal{F}_{\ell-1}$, then $\gamma(u, v) = 0$, so $(v, v) = (\vartheta u, v) = (u, \vartheta^* v) = 0$. Hence $\tilde{\kappa}$ is injective, and the proof is complete.

With this sketch of elliptic complexes done, it is time to deliver the bad news. The regularity hypothesis is rarely satisfied, other than for the deRham complex. The most fundamental complex that arises next is the $\overline{\partial}$-complex, for which the regularity hypothesis does not hold. However, for a certain class of domains $\overline{M}$, one has "subelliptic estimates," from which useful variants of Propositions A.3 and A.4 follow. We will explore this in the rest of this chapter.

1. The $\overline{\partial}$-complex

To begin, let us assume Ω is an open subset of $\mathbb{C}^n$. Standard complex coordinates on $\mathbb{C}^n$ are $(z_1, \dots, z_n)$, with $z_j = x_j + iy_j$. We identify $\mathbb{C}^n \approx \mathbb{R}^{2n}$, with coordinates $(x_1, \dots, x_n, y_1, \dots, y_n)$. A (p, q)-form on Ω is by definition a section of $\mathbb{C} \otimes \Lambda^{p+q} T^* \Omega$ of the form

$$u = \sum_{\beta,\gamma} u_{\beta\gamma}(z)\, dz^\beta \wedge d\overline{z}^\gamma, \tag{1.1}$$

where

$$dz^\beta = dz_{\beta_1} \wedge \cdots \wedge dz_{\beta_p}, \quad d\overline{z}^\gamma = d\overline{z}_{\gamma_1} \wedge \cdots \wedge d\overline{z}_{\gamma_q}, \tag{1.2}$$

with

$$dz_j = dx_j + i\, dy_j, \quad d\overline{z}_j = dx_j - i\, dy_j. \tag{1.3}$$

We impose the same anticommutation relations on wedge products as before, so

$$dz_j \wedge dz_k = -dz_k \wedge dz_j, \quad d\overline{z}_j \wedge d\overline{z}_k = -d\overline{z}_k \wedge d\overline{z}_j, \quad dz_j \wedge d\overline{z}_k = -d\overline{z}_k \wedge dz_j. \tag{1.4}$$

If the coefficients $u_{\beta\gamma}$ in (1.1) belong to $C^\infty(\Omega)$, we write $u \in \Lambda^{p,q}(\Omega)$; if they belong to $C^\infty(\overline{\Omega})$, we write $u \in \Lambda^{p,q}(\overline{\Omega})$. There is a differential operator

$$\overline{\partial} : \Lambda^{p,q}(\overline{\Omega}) \longrightarrow \Lambda^{p,q+1}(\overline{\Omega}) \tag{1.5}$$

defined by

$$\begin{aligned} \overline{\partial} u &= \sum_{\beta,\gamma,j} \frac{\partial u_{\beta\gamma}}{\partial \overline{z}_j}\, d\overline{z}_j \wedge dz^\beta \wedge d\overline{z}^\gamma \\ &= (-1)^p \sum_{\beta,\gamma,j} \frac{\partial u_{\beta\gamma}}{\partial \overline{z}_j}\, dz^\beta \wedge d\overline{z}_j \wedge d\overline{z}^\gamma. \end{aligned} \tag{1.6}$$

Here, we define $\partial/\partial\overline{z}_j$ by

$$\frac{\partial v}{\partial \overline{z}_j} = \frac{1}{2}\left(\frac{\partial v}{\partial x_j} + i\frac{\partial v}{\partial y_j}\right), \tag{1.7}$$

so that a complex-valued function $v \in C^\infty(\Omega)$ is holomorphic if and only if $\partial v/\partial\overline{z}_j = 0,\ 1 \le j \le n$. Equivalently, $v \in \Lambda^{0,0}(\Omega)$ is holomorphic if and only if $\overline{\partial} v = 0$. The operator $\overline{\partial}$ has some properties in common with the exterior derivative d. For example, just as $d^2 = 0$, we have

$$\overline{\partial}^2 = 0, \tag{1.8}$$

by virtue of the identity $\partial^2 u_{\beta\gamma}/\partial\overline{z}_k\partial\overline{z}_j = \partial^2 u_{\beta\gamma}/\partial\overline{z}_j\partial\overline{z}_k$ and the relation $d\overline{z}_j \wedge d\overline{z}_k = -d\overline{z}_k \wedge d\overline{z}_j$. Thus we have, for each p, a complex:

$$\cdots \xrightarrow{\overline{\partial}} \Lambda^{p,q}(\overline{\Omega}) \xrightarrow{\overline{\partial}} \Lambda^{p,q+1}(\overline{\Omega}) \xrightarrow{\overline{\partial}} \cdots. \tag{1.9}$$

As in (A.6), we form the second-order operator

$$\Box = \overline{\partial}\,\overline{\partial}^* + \overline{\partial}^*\overline{\partial} : \Lambda^{p,q}(\overline{\Omega}) \longrightarrow \Lambda^{p,q}(\overline{\Omega}). \tag{1.10}$$

When $\overline{\partial}$ is given by (1.6) and we take $dz^\beta \wedge d\overline{z}^\gamma$ to be orthogonal to the form $dz^{\beta'} \wedge d\overline{z}^{\gamma'}$ when the indices satisfy

$$\begin{gathered}\beta_1 < \cdots < \beta_p, \quad \beta'_1 < \cdots < \beta'_p, \\ \gamma_1 < \cdots < \gamma_q, \quad \gamma'_1 < \cdots < \gamma'_q, \\ (\beta, \gamma) \neq (\beta', \gamma'),\end{gathered}$$

and we take $dz^\beta \wedge d\overline{z}^\gamma$ to have square norm 2^{p+q}, we obtain

$$\Box u = -\frac{1}{2} \sum_{\beta,\gamma} \sum_j \left(\frac{\partial^2 u_{\beta\gamma}}{\partial x_j^2} + \frac{\partial^2 u_{\beta\gamma}}{\partial y_j^2} \right) dz^\beta \wedge d\overline{z}^\gamma, \tag{1.11}$$

when u has the form (1.1). In other words, $\Box u = -(1/2)\Delta u$, where Δu is computed componentwise. The "absolute" boundary condition (A.14) becomes

$$\sigma_{\overline{\partial}^*}(x, \nu) u = 0, \quad \sigma_{\overline{\partial}^*}(x, \nu) \overline{\partial} u = 0 \quad \text{on } \partial\Omega. \tag{1.12}$$

This is the (homogeneous) $\overline{\partial}$-Neumann boundary condition.

Now the system (1.11)–(1.12) does not generally yield a regular elliptic boundary problem. If it did, the frozen-coefficient boundary problem on any region $\mathcal{O} \subset \mathbb{C}^n$ bounded by a hyperplane would also be regular. We can investigate such a boundary problem as follows.

First, applying a rotation by a unitary matrix acting on $\mathbb{C}^n$, we can take $\mathcal{O}$ to be $\{z \in \mathbb{C}^n : \operatorname{Im} z_n > 0\}$. Let us consider the case $(p, q) = (0, 1)$, so

$$u = \sum_{j=1}^n u_j \, d\overline{z}_j. \tag{1.13}$$

Then, since $\overline{\partial} u = \sum_{j,k} (\partial u_j / \partial \overline{z}_k) \, d\overline{z}_k \wedge d\overline{z}_j$, we have

$$\sigma_{\overline{\partial}^*}(x, \nu) \overline{\partial} u = \sum_{j=1}^{n-1} \frac{\partial u_j}{\partial \overline{z}_n} \, d\overline{z}_j - \sum_{k=1}^{n-1} \frac{\partial u_n}{\partial \overline{z}_k} \, d\overline{z}_k, \tag{1.14}$$

so the boundary condition (1.12) says that, for $z = (z', x_n, 0) \in \partial\mathcal{O}$, we have

$$u_n(z', x_n, 0) = 0, \quad \frac{\partial u_j}{\partial \overline{z}_n}(z', x_n, 0) = 0, \quad 1 \leq j \leq n-1. \tag{1.15}$$

Thus, in this case the $\overline{\partial}$-Neumann problem decouples into n boundary problems for the Laplace operator Δ acting on complex-valued functions. One is the Dirichlet problem, which of course is regular. The other $n-1$ are all of the form

$$\Delta v = f \text{ on } \mathcal{O}, \quad \frac{\partial v}{\partial \bar{z}_n}(z', x_n, 0) = 0. \tag{1.16}$$

Equivalently, we can investigate regularity for

$$\Delta v = 0 \text{ on } \mathcal{O}, \quad \frac{\partial v}{\partial \bar{z}_n}(z', x_n, 0) = g(z', x_n). \tag{1.17}$$

If we attempt to write $v = PI\ h$, then g and h are related by

$$\frac{1}{2}\left(\frac{\partial}{\partial x_n} + i\mathcal{N}\right) h = g, \tag{1.18}$$

where $\mathcal{N}$ is the Neumann operator for Δ, given by

$$(\mathcal{N}h)\widehat{\ }(\xi, \eta') = -(|\xi|^2 + |\eta'|^2)^{1/2}\hat{h}(\xi, \eta'), \tag{1.19}$$

where $\xi = (\xi_1, \dots, \xi_n)$ and $\eta' = (\eta_1, \dots, \eta_{n-1})$ are variables dual to $x = (x_1, \dots, x_n)$ and to $y = (y_1, \dots, y_{n-1})$, respectively. Thus,

$$\left(\frac{\partial}{\partial x_n} + i\mathcal{N}\right)h\widehat{\ }(\xi, \eta') = -i\left(\xi_n + \sqrt{|\xi|^2 + |\eta'|^2}\right)\hat{h}(\xi, \eta'). \tag{1.20}$$

We see that the pseudodifferential operator $\partial/\partial x_n + i\mathcal{N}$ is not elliptic. The ray (ξ, η') on which $\xi_1 = \cdots = \xi_{n-1} = 0 = \eta_1 = \cdots = \eta_{n-1}$ but $\xi_n < 0$ is characteristic for this operator. Since this operator is not elliptic, the boundary problem (1.17) is not regular. Consequently, if $n \geq 2$, the $\overline{\partial}$-Neumann problem is never a regular elliptic boundary problem for $(0, 1)$-forms.

Exercises

1. Define $\partial : \Lambda^{p,q}(\overline{\Omega}) \to \Lambda^{p+1,q}(\overline{\Omega})$ by

$$\partial u = \sum_{\beta,\gamma,j} \frac{\partial u_{\beta\gamma}}{\partial z_j}\, dz_j \wedge dz^\beta \wedge d\bar{z}^\gamma \tag{1.21}$$

when u is given by (1.1) and we set

$$\frac{\partial v}{\partial z_j} = \frac{1}{2}\left(\frac{\partial v}{\partial x_j} - i\frac{\partial v}{\partial y_j}\right), \tag{1.22}$$

parallel to (1.7). Show that

$$\partial^2 = 0.$$

2. If $u = \sum u_j\, d\bar{z}_j$, show that

$$\overline{\partial}^* u = -2\sum_j \frac{\partial u_j}{\partial z_j}. \tag{1.23}$$

More generally, calculate $\overline{\partial}^*$ on (p,q)-forms. Then verify the formula (1.11) for $\Box = \overline{\partial}\,\overline{\partial}^* + \overline{\partial}^*\overline{\partial}$.

3. When $\dim \Omega = n$, show that the $\overline{\partial}$-Neumann problem for $(0,n)$-forms is equivalent to the Dirichlet problem for Δ acting on scalar functions and consequently is coercive.

2. Morrey's inequality, the Levi form, and strong pseudoconvexity

The following estimate of C. B. Morrey provides the first useful handle on the $\overline{\partial}$-Neumann problem.

Proposition 2.1. *If Ω is a smoothly bounded region in $\mathbb{C}^n$ that is strongly pseudoconvex, then, for some $C > 0$,*

$$\|\overline{\partial}u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2 \geq C \int_{\partial\Omega} |u|^2\, dS, \quad \forall\, u \in \mathcal{D}^{0,1}. \tag{2.1}$$

Here, $\mathcal{D}^{0,1}$ consists of smooth $(0,1)$-forms on $\overline{\Omega}$ satisfying the zero-order part of the $\overline{\partial}$-Neumann boundary condition (1.12). More generally, we set

$$\mathcal{D}^{p,q} = \{u \in \Lambda^{p,q}(\overline{\Omega}) : \sigma_{\overline{\partial}^*}(x,\nu)u = 0 \text{ on } \partial\Omega\}. \tag{2.2}$$

We will define "strongly pseudoconvex" below, after deriving an *identity* that leads to (2.1) once the appropriate definition is made.

We prepare to work on the left side of (2.1). Writing $u = \sum u_j\, d\overline{z}_j$, we have

$$\overline{\partial}u = \sum_{j<k}\Big(\frac{\partial u_j}{\partial \overline{z}_k} - \frac{\partial u_k}{\partial \overline{z}_j}\Big)\, d\overline{z}_k \wedge d\overline{z}_j, \quad \overline{\partial}^* u = -2\sum_j \frac{\partial u_j}{\partial z_j}, \tag{2.3}$$

and if $\rho \in C^\infty(\overline{\Omega})$ is a real-valued, defining function for Ω, so $\rho = 0$ on $\partial\Omega$ and $\rho < 0$ on Ω, while $|\nabla\rho| = 1$ on $\partial\Omega$, then

$$u \in \mathcal{D}^{0,1} \Longleftrightarrow \sum u_j \frac{\partial\rho}{\partial z_j} = 0 \quad \text{on } \partial\Omega. \tag{2.4}$$

Thus, for $u \in \mathcal{D}^{0,1}$,

$$\begin{aligned} \|\overline{\partial}u\|_{L^2}^2 &= 4\sum_{j<k}\Big\|\frac{\partial u_j}{\partial \overline{z}_k} - \frac{\partial u_k}{\partial \overline{z}_j}\Big\|_{L^2}^2 \\ &= 4\sum_{j,k}\Big\|\frac{\partial u_j}{\partial \overline{z}_k}\Big\|_{L^2}^2 - 4\sum_{j,k}\Big(\frac{\partial u_j}{\partial \overline{z}_k}, \frac{\partial u_k}{\partial \overline{z}_j}\Big)_{L^2}. \end{aligned} \tag{2.5}$$

Integration by parts yields

$$
\begin{aligned}
\Big(\frac{\partial u_j}{\partial \bar z_k}, \frac{\partial u_k}{\partial \bar z_j}\Big)_{L^2} &= -\Big(\frac{\partial^2 u_j}{\partial z_j \partial \bar z_k}, u_k\Big)_{L^2} + \int_{\partial\Omega} \frac{\partial \rho}{\partial z_j}\frac{\partial u_j}{\partial \bar z_k}\bar u_k \, dS \\
&= \Big(\frac{\partial u_j}{\partial z_j}, \frac{\partial u_k}{\partial z_k}\Big)_{L^2} + \int_{\partial\Omega} \Big[\frac{\partial \rho}{\partial z_j}\frac{\partial u_j}{\partial \bar z_k}\bar u_k - \frac{\partial \rho}{\partial \bar z_k}\frac{\partial u_j}{\partial z_j}\bar u_k\Big] \, dS.
\end{aligned} \tag{2.6}
$$

The condition (2.4) implies that $\sum_k (\partial\rho/\partial\bar z_k)\bar u_k = 0$ on $\partial\Omega$, so the last term on the right side of (2.6) vanishes after being summed over k. Also, (2.4) implies that $\sum_k \bar u_k \partial/\partial \bar z_k = \overline{Z}$ is a *tangential derivative* on $\partial\Omega$. Hence

$$
\sum_k \bar u_k \frac{\partial}{\partial \bar z_k}\left(\sum_j u_j \frac{\partial \rho}{\partial z_j}\right) = 0 \quad \text{on } \partial\Omega, \tag{2.7}
$$

so

$$
\sum_{j,k} \frac{\partial \rho}{\partial z_j}\frac{\partial u_j}{\partial \bar z_k}\bar u_k = -\sum_{j,k} \frac{\partial^2 \rho}{\partial z_j \partial \bar z_k} u_j \bar u_k \quad \text{on } \partial\Omega. \tag{2.8}
$$

Thus (2.6) becomes

$$
\sum_{j,k}\Big(\frac{\partial u_j}{\partial \bar z_k}, \frac{\partial u_k}{\partial \bar z_j}\Big)_{L^2} = \sum_{j,k}\Big(\frac{\partial u_j}{\partial z_j}, \frac{\partial u_k}{\partial z_k}\Big)_{L^2} - \sum_{j,k}\int_{\partial\Omega} \mathcal{L}_{jk} u_j \bar u_k \, dS, \tag{2.9}
$$

where

$$
\mathcal{L}_{jk} = \frac{\partial^2 \rho}{\partial z_j \partial \bar z_k}. \tag{2.10}
$$

Since the first term on the right side of (2.9) is equal to $(1/4)\|\overline{\partial}^* u\|_{L^2}^2$, we have from (2.5) the identity

$$
\|\overline{\partial} u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2 = 4\sum_{j,k}\Big\|\frac{\partial u_j}{\partial \bar z_k}\Big\|_{L^2}^2 + 4\int_{\partial\Omega}\Big(\sum_{j,k}\mathcal{L}_{jk} u_j \bar u_k\Big)\, dS. \tag{2.11}
$$

The integrand in the last integral involves the *Levi form*, a sesquilinear form defined as follows on the “holomorphic tangent space” of $\partial\Omega$. If $p \in \partial\Omega$, we set

$$
\mathcal{L}_p(a,b) = \sum_{j,k}\mathcal{L}_{jk}(p) a_j \bar b_k = \sum_{j,k}\frac{\partial^2 \rho}{\partial z_j \partial \bar z_k} a_j \bar b_k, \quad \forall\, a,b \in \mathfrak{H}_p(\partial\Omega), \tag{2.12}
$$

where

$$\mathfrak{H}_p(\partial\Omega) = \Big\{a \in \mathbb{C}^n : \sum a_j \frac{\partial\rho}{\partial z_j}(p) = 0\Big\}. \tag{2.13}$$

Note that $\mathfrak{H}_p(\partial\Omega)$ is precisely the maximal $\mathbb{C}$-linear subspace of the tangent space $T_p\partial\Omega \subset \mathbb{R}^{2n} = \mathbb{C}^n$. It is readily verified that (2.12) is unchanged if ρ is replaced by another defining function $\tilde{\rho}$, satisfying the conditions specified above for ρ. (This fact is also an immediate consequence of the formula (2.21) below.)

By definition, a smoothly bounded domain $\Omega \subset \mathbb{C}^n$ is *strongly pseudoconvex* if and only if its Levi form is a positive-definite Hermitian form on $\mathfrak{H}_p(\partial\Omega)$, for all $p \in \partial\Omega$.

In view of the fact that, for $u \in \mathcal{D}^{0,1}$, the n-tuple $\big(u_j(p)\big)$ belongs to $\mathfrak{H}_p(\partial\Omega)$ for each $p \in \partial\Omega$, we see from (2.10) that if Ω is strongly pseudoconvex, then (2.1) holds. In fact, we have a stronger estimate:

$$\|\overline{\partial}u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2 \ge 4\sum_{j,k}\Big\|\partial u_j/\partial\overline{z}_k\Big\|_{L^2}^2 + C\int_{\partial\Omega}|u|^2\,dS, \quad \forall\, u \in \mathcal{D}^{0,1}. \tag{2.14}$$

Exercises

1. A smooth function $f : \mathbb{C}^n \to \mathbb{R}$ is called "strongly plurisubharmonic" if

$$\Big(\frac{\partial^2 f}{\partial z_j\,\partial\overline{z}_k}\Big) \quad \text{is positive-definite.} \tag{2.15}$$

   Given such a function, show that $\Omega = \{z \in \mathbb{C}^n : f(z) < 0\}$ is strongly pseudoconvex if it is not empty.
2. Show that any strongly convex, smooth $f : \mathbb{C}^n \to \mathbb{R}$ is strongly plurisubharmonic, and deduce that any strongly convex, bounded $\Omega \subset \mathbb{C}^n$ is strongly pseudoconvex.
   (*Hint*: See (B.13).)
3. Suppose Ω is a bounded domain in $\mathbb{C}^n$, $\mathcal{O}$ a neighborhood of $\partial\Omega$, and $f : \mathcal{O} \to \mathbb{R}$ a smooth function such that $f = c =$ const. on $\partial\Omega$ and (2.15) holds on $\mathcal{O}$, while $f < c$ in Ω. Deduce that Ω is strongly pseudoconvex.
4. Suppose conversely that Ω is strongly pseudoconvex, with defining function ρ as in (2.10). Show that, for sufficiently large $\lambda > 0$, $f = e^{\lambda\rho}$ satisfies (2.15) on a neighborhood of $\partial\Omega$.
   (*Hint*: Use the identity

$$\frac{\partial^2 f}{\partial z_j\,\partial\overline{z}_k} = \lambda e^{\lambda\rho}\frac{\partial^2\rho}{\partial z_j\,\partial\overline{z}_k} + \lambda^2 e^{\lambda\rho}\frac{\partial\rho}{\partial z_j}\frac{\partial\rho}{\partial\overline{z}_k}.)$$

5. Given $f : \mathcal{O} \to \mathbb{R}$ such that $f = c =$ const. on $\partial\Omega$ and (2.15) holds on $\mathcal{O}$, a neighborhood of $\partial\Omega$, while $f < c$ in Ω, and given $p \in \partial\Omega$, consider the function

$$g(z) = \sum_j \frac{\partial f}{\partial z_j}(p)\,(z_j - p_j) + \frac{1}{2}\sum_{j,k}\frac{\partial^2 f}{\partial z_j\,\partial z_k}(p)\,(z_j - p_j)(z_k - p_k). \tag{2.16}$$

Show that p has a neighborhood U such that

$$S = \{z \in U : g(z) = 0\} \Longrightarrow S \cap \overline{\Omega} = \{p\}.$$

(*Hint*: Write out the power series of $f(p + hz)$, to $O(h^3)$, using $\partial/\partial z_j, \partial/\partial \overline{z}_j$, etc., rather than $\partial/\partial x_j,\ \partial/\partial y_j$, etc., to see that

$$f(p + hz) = f(p) + 2\,\mathrm{Re}\, g(p + hz) + h^2 \sum_{j,k} \frac{\partial^2 f}{\partial z_j \partial \overline{z}_k} z_j \overline{z}_k + O(|hz|^3).)$$

3. The $\frac{1}{2}$-estimate and some consequences

Here we will derive a "1/2-estimate" from Morrey's inequality, and discuss a few consequences, before establishing higher-order a priori estimates and regularity in the next two sections. Throughout this section we assume that Ω is a bounded, strongly pseudoconvex domain in $\mathbb{C}^n$.

Proposition 3.1. *For some $C > 0$,*

$$\|\overline{\partial} u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2 \geq C \|u\|_{H^{1/2}}^2, \quad \forall\, u \in \mathcal{D}^{0,1}. \tag{3.1}$$

Proof. From (1.10) and (1.11) we have

$$K\|\Delta u\|_{H^{-1}}^2 \leq \|\overline{\partial} u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2, \tag{3.2}$$

and together with (2.1) this yields (for various $K > 0$)

$$\|\overline{\partial} u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2 \geq K\|\Delta u\|_{H^{-1}}^2 + K \int_{\partial\Omega} |u|^2\, dS, \quad \forall\, u \in \mathcal{D}^{0,1}. \tag{3.3}$$

Now we claim that regularity for the *Dirichlet problem* implies

$$\|u\|_{H^{\frac{1}{2}}}^2 \leq K\|\Delta u\|_{H^{-1}}^2 + K \int_{\partial\Omega} |u|^2\, dS, \tag{3.4}$$

and this yields (3.1).

To see (3.4), suppose $\Box u = -(1/2)\Delta u = f,\ u\big|_{\partial\Omega} = g$. Write $u = u_1 + u_2$, where

$$\Delta u_1 = -2f,\ u_1\big|_{\partial\Omega} = 0; \quad \Delta u_2 = 0,\ u_2\big|_{\partial\Omega} = g. \tag{3.5}$$

Then results of Chap. 5, §1 imply

$$\|u_1\|_{H^1}^2 \leq C\|f\|_{H^{-1}}^2, \tag{3.6}$$

while Propositions 11.14 and 11.15 of Chap. 5 imply

$$\|u_2\|^2_{H^{1/2}} \le C\,\|g\|^2_{L^2(\partial\Omega)}. \tag{3.7}$$

More precisely, using the spaces $H_{(k,s)}(\mathcal{C})$ defined in §11 of Chap. 5, where $\mathcal{C}$ is a collar neighborhood of $\partial\Omega$, we have

$$g \in H^{2+s-1/2}(\partial\Omega) \Longrightarrow u_2 \in H_{(2,s)}(\mathcal{C}), \tag{3.8}$$

and, in particular, if $\|\cdot\|_{(k,s)}$ denotes the norm in $H_{(k,s)}(\mathcal{C})$, $\|u_2\|^2_{(2,3/2)} \le C\,\|g\|^2_{L^2(\partial\Omega)}$; hence

$$\|u\|^2_{(1,-1/2)} \le C\,\|\overline{\partial} u\|^2_{L^2} + C\,\|\overline{\partial}^* u\|^2_{L^2}, \quad \forall\, u \in \mathcal{D}^{0,1}. \tag{3.9}$$

Recall from (11.95) of Chap. 5 that if $\mathcal{C}$ is identified with $[0, 1) \times \partial\Omega$, then

$$\|u\|^2_{(k,s)} = \sum_{j=0}^{k} \int_0^1 \|D_y^j u(y,\cdot)\|^2_{H^{k+s-j}(\partial\Omega)}\, dy. \tag{3.10}$$

Note that (3.7) is basically equivalent to the statement that the Poisson integral has the property

$$\mathrm{PI} : L^2(\partial\Omega) \longrightarrow H^{1/2}(\Omega).$$

This also follows from results in §12 of Chap. 7.

We next define a self-adjoint extension of $\Box = \overline{\partial}\,\overline{\partial}^* + \overline{\partial}^*\overline{\partial}$ on (0, 1)-forms, satisfying $\overline{\partial}$-Neumann boundary conditions. Let $\mathcal{H}_1$ be the Hilbert space completion of $\mathcal{D}^{0,1}$ with respect to the square norm

$$Q(u,u) = \|\overline{\partial} u\|^2_{L^2} + \|\overline{\partial}^* u\|^2_{L^2}. \tag{3.11}$$

We can identify $\mathcal{H}_1$ with the closure of $\mathcal{D}^{0,1}$ in $\{u \in L^2(\Omega, \Lambda^{0,1}) : \overline{\partial} u, \overline{\partial}^* u \in L^2(\Omega)\}$. Then we have a natural, continuous, dense injection $\mathcal{H}_1 \hookrightarrow \mathcal{H}_0 = L^2(\Omega, \Lambda^{0,1})$. Thus, the Friedrichs extension method (discussed in §1 of Chap. 8 and in §8 of Appendix A) yields an unbounded, self-adjoint operator $\mathcal{L}$ on $\mathcal{H}_0$ such that

$$\begin{aligned} \mathcal{D}(\mathcal{L}) &= \{u \in \mathcal{H}_1 : v \mapsto (v,u)_{\mathcal{H}_1} \text{ is } \mathcal{H}_0\text{-continuous}\}, \\ (\mathcal{L}u, v) &= (u, v)_{\mathcal{H}_1}. \end{aligned} \tag{3.12}$$

Note that

$$\mathcal{D}^{0,1} \subset \mathcal{D}(\mathcal{L}) \subset \mathcal{D}(\mathcal{L}^{1/2}) = \mathcal{H}_1, \tag{3.13}$$

the latter identity by Proposition 1.10 of Chap. 8. The estimate (3.1) implies the inclusion $H^{1/2}(\Omega, \Lambda^{0,1}) \supset \mathcal{H}_1$, so

$$\mathcal{D}(\mathcal{L}) \subset \mathcal{D}(\mathcal{L}^{1/2}) \subset H^{1/2}(\Omega, \Lambda^{0,1}). \tag{3.14}$$

The characterization (3.12) implies that, for all $u \in \mathcal{D}(\mathcal{L})$,

$$\mathcal{L}u = \Box u = -\frac{1}{2}\Delta u \quad \text{in } \mathcal{D}'(\Omega). \tag{3.15}$$

Thus, interior elliptic regularity implies

$$\mathcal{D}(\mathcal{L}) \subset H^2_{\text{loc}}(\Omega) \cap H^{1/2}(\Omega). \tag{3.16}$$

We see that $\mathcal{L}$ has compact resolvent. Since $\|\mathcal{L}^{1/2}u\|^2_{L^2} = (u,u)_{\mathcal{H}_1} \geq C\|u\|^2_{H^{1/2}}$ for all $u \in \mathcal{D}(\mathcal{L}^{1/2})$, zero is in its resolvent set, so $\mathcal{L}^{-1}$ is a compact, self-adjoint operator on $L^2(\Omega, \Lambda^{0,1})$.

Our next goal is to demonstrate that elements of $\mathcal{D}(\mathcal{L})$ do indeed satisfy the $\overline{\partial}$-Neumann boundary conditions. First, if $u \in \mathcal{D}(\mathcal{L}^{1/2})$, then since $\overline{\partial}u \in L^2(\Omega)$ and $\overline{\partial}^* u \in L^2(\Omega)$, it follows that $u\big|_{\partial\Omega}$ is well defined in $\mathcal{D}'(\partial\Omega)$. Indeed, since u is a limit of a sequence $u_j \in \mathcal{D}^{0,1}$ in $\mathcal{H}_1$-norm, we can deduce from (2.1) that $u\big|_{\partial\Omega} \in L^2(\partial\Omega)$ and $u_j\big|_{\partial\Omega} \to u\big|_{\partial\Omega}$ in $L^2(\partial\Omega)$. It follows that

$$u \in \mathcal{D}(\mathcal{L}^{1/2}) \Longrightarrow \sigma_{\overline{\partial}^*}(x,\nu)u = 0 \quad \text{on } \partial\Omega. \tag{3.17}$$

Furthermore, if $u \in \mathcal{D}(\mathcal{L})$, so $\mathcal{L}u = f \in L^2(\Omega, \Lambda^{0,1})$, we can write $u = u_1 + u_2$ where $u_1 \in H^2(\Omega) \cap H^1_0(\Omega)$ solves $\Delta u_1 = -2f$ and $u_2 \in H^{1/2}(\Omega)$ is harmonic. It follows that

$$u_1\big|_{\partial\Omega} \in H^{3/2}(\partial\Omega), \quad \overline{\partial}u_1\big|_{\partial\Omega} \in H^{1/2}(\partial\Omega).$$

Since u_2 is harmonic, $u_2\big|_{\partial\Omega}$ and $\overline{\partial}u_2\big|_{\partial\Omega}$ are well defined, in $\mathcal{D}'(\partial\Omega)$. Hence $u\big|_{\partial\Omega}$ and $\overline{\partial}u\big|_{\partial\Omega}$ are well defined. The same argument also applies to $\overline{\partial}^* u$.

We now establish the following.

Proposition 3.2. *If $u \in \mathcal{D}(\mathcal{L})$, then u satisfies the boundary conditions* (1.12), *namely,*

$$\sigma_{\overline{\partial}^*}(x,\nu)u = 0, \quad \sigma_{\overline{\partial}^*}(x,\nu)\,\overline{\partial}u = 0 \quad \textit{on } \partial\Omega. \tag{3.18}$$

Proof. The first identity in (3.18) follows from (3.17). To get the second identity, note that if $\mathcal{L}u = f$, we have

$$(\bar\partial u, \bar\partial v)_{L^2} + (\bar\partial^* u, \bar\partial^* v)_{L^2} = (f, v)_{L^2}, \quad \forall\, v \in \mathcal{D}^{0,1}. \tag{3.19}$$

We have already noted that $\bar\partial\bar\partial^* u + \bar\partial^*\bar\partial u = f$ in Ω. Furthermore, the comments above imply that, for all $v \in \Lambda^{0,1}(\overline{\Omega})$,

$$(\bar\partial u, \bar\partial v)_{L^2} + (\bar\partial^* u, \bar\partial^* v)_{L^2} = (\Box u, v)_{L^2} + \beta(u, v), \tag{3.20}$$

where, in parallel with (A.12),

$$\beta(u, v) = \frac{1}{i} \int\limits_{\partial\Omega} \Big[\langle \sigma_{\bar\partial^*}(x, \nu)\, \bar\partial u, v \rangle + \langle \bar\partial^* u, \sigma_{\bar\partial^*}(x, \nu) v \rangle \Big]\, dS. \tag{3.21}$$

The last term in the integrand vanishes if $v \in \mathcal{D}^{0,1}$, so we deduce that

$$u \in \mathcal{D}(\mathcal{L}) \Longrightarrow \int\limits_{\partial\Omega} \langle \sigma_{\bar\partial^*}(x, \nu)\, \bar\partial u, v \rangle\, dS = 0, \quad \forall\, v \in \mathcal{D}^{0,1}. \tag{3.22}$$

In particular, (3.22) holds for $v = \sigma_{\bar\partial^*}(x, \nu)\varphi$ on $\partial\Omega$, for any $\varphi \in \Lambda^{0,2}(\overline{\Omega})$, so

$$u \in \mathcal{D}(\mathcal{L}) \Longrightarrow \int\limits_{\partial\Omega} \langle \sigma^* \sigma \bar\partial u, \varphi \rangle\, dS = 0, \quad \forall\, \varphi \in \Lambda^{0,2}(\overline{\Omega}), \tag{3.23}$$

where σ is short for $\sigma_{\bar\partial^*}(x, \nu)$. This implies that $\sigma^*\sigma$ annihilates $\bar\partial u$ on $\partial\Omega$. Since $u\big|_{\partial\Omega}$ has been shown only to be in $\mathcal{D}'(\partial\Omega)$, we need a little care in deducing that σ annihilates $\bar\partial u$ on $\partial\Omega$, but since $\sigma(x)^*\sigma(x)$ is a smooth, *projection-valued* function on $\partial\Omega$, this implication follows, and Proposition 3.2 is proved.

For a converse of sorts, suppose $u \in \mathcal{H}_1$ and $\Box u = f \in L^2(\Omega)$. The argument below (3.17) implies that $u\big|_{\partial\Omega}$ and $\bar\partial u\big|_{\partial\Omega}$ are well defined in $\mathcal{D}'(\partial\Omega)$. Also, (3.20)–(3.21) hold for such u and for any $v \in \mathcal{D}^{0,1}$. Hence, as long as $\sigma_{\bar\partial^*}(x, \nu) u = 0 = \sigma_{\bar\partial^*}(x, \nu)\, \bar\partial u$ on $\partial\Omega$, we have

$$Q(v, u) \le C(u) \|v\|_{L^2}, \quad \forall\, v \in \mathcal{D}^{0,1}.$$

In view of the characterization (3.12), we have the following result:

Proposition 3.3. *The domain of $\mathcal{L}$ is specified by*

$$\mathcal{D}(\mathcal{L}) = \Big\{ u \in \mathcal{H}_1 : \Box u \in L^2(\Omega), \sigma_{\bar\partial^*}(x, \nu) u = 0, \\ \sigma_{\bar\partial^*}(x, \nu) \bar\partial u = 0 \text{ on } \partial\Omega \Big\}. \tag{3.24}$$

We record another regularity estimate:

Proposition 3.4. *If* $u \in \mathcal{D}(\mathcal{L})$, *then* $u \in H_{(1,-1/2)}(\mathcal{C})$ *and*

$$\|u\|^2_{(1,-1/2)} \le C\,\|\mathcal{L}u\|_{L^2}\,\|u\|_{L^2}. \tag{3.25}$$

Proof. If $u \in \mathcal{D}(\mathcal{L}) \subset \mathcal{D}(\mathcal{L}^{1/2})$, then the estimate (3.9) holds. Hence

$$\|u\|^2_{(1,-1/2)} \le C(u,u)_{\mathcal{H}_1} = C(\mathcal{L}u,u)_{L^2}.$$

Applying Cauchy's inequality yields (3.25).

Exercises

1. Consider the space

$$\mathcal{H}_1^\# = \Big\{u \in L^2(\Omega,\Lambda^{0,1}) : \overline{\partial}u \in L^2,\ \overline{\partial}^* u \in L^2,\ \sigma_{\overline{\partial}^*}(x,\nu)u\Big|_{\partial\Omega} = 0\Big\},$$

with square norm $\|\overline{\partial}u\|^2_{L^2} + \|\overline{\partial}^* u\|^2_{L^2} + \|u\|^2_{L^2}$. Try to show that $\mathcal{D}^{0,1}$ is dense in $\mathcal{H}_1^\#$ and hence that $\mathcal{H}_1^\# = \mathcal{H}_1$.

2. For small $s \ge 0$, let

$$\Sigma_s = \{z \in \overline{\Omega} : \rho(z) = -s\},$$

so $\Sigma_0 = \partial\Omega$. Assume that, for $0 \le s \le b,\ \Sigma_s$ is a smooth surface on which $\nabla\rho \ne 0$. Show that

$$\sup_{0\le s\le b} \|u\|^2_{L^2(\Sigma_s)} \le C\Big(\|\overline{\partial}u\|^2_{L^2} + \|\overline{\partial}^* u\|^2_{L^2}\Big), \quad u \in \mathcal{D}^{0,1}. \tag{3.26}$$

(*Hint*: Follow the argument using (3.5)–(3.7), but replace (3.7) by

$$\mathrm{PI} : L^2(\partial\Omega) \longrightarrow L^2(\Sigma_s),$$

with an appropriate norm estimate.)

3. Show that (3.26), together with the fact that

$$(\mathcal{L}+I)^{-1} : L^2(\Omega,\Lambda^{0,1}) \longrightarrow H^2_{\mathrm{loc}}(\Omega,\Lambda^{0,1}),$$

implies that $(\mathcal{L}+I)^{-1}$ is compact on $L^2(\Omega,\Lambda^{0,1})$, without making use of (3.14). Compare [Mor], p. 336.

4. Higher-order subelliptic estimates

We want to extend the estimates (3.9) and (3.25) to estimates on higher derivatives of $u \in \Lambda^{0,1}(\overline{\Omega}) \cap \mathcal{D}(\mathcal{L})$, in terms of estimates on $\mathcal{L}u$. The associated regularity results will be established in §5. As in §3, we make the standing assumption that Ω is a bounded, strongly pseudoconvex domain in $\mathbb{C}^n$. We begin with the following improvement of (3.25).

Proposition 4.1. *For* $u \in \Lambda^{0,1}(\overline{\Omega}) \cap \mathcal{D}(\mathcal{L})$*, we have*

$$\|u\|_{H^1}^2 \le C\|\mathcal{L}u\|_{L^2}^2. \tag{4.1}$$

Proof. It suffices to consider the case where u is supported on the collar neighborhood $\mathcal{C}$ of $\partial\Omega$ introduced in §3. As there, we identify $\mathcal{C}$ with $[0, 1) \times \partial\Omega$.

Let $\Lambda^{1/2} \in OPS^{1/2}(\partial\Omega)$ be an elliptic self-adjoint operator, with scalar principal symbol acting on sections of $\Lambda^{0,1}\big|_{\partial\Omega}$. Arrange that $\Lambda^{1/2}$ commutes with $P(x) = \sigma(x)^*\sigma(x)$, where $\sigma(x) = \sigma_{\overline{\partial}^*}(x, \nu)$. Note that

$$\|u\|_{H^1(\mathcal{C})}^2 \approx \|\Lambda^{1/2}u\|_{(1,-1/2)}^2. \tag{4.2}$$

Now, if $u \in \Lambda^{0,1}(\overline{\Omega}) \cap \mathcal{D}(\mathcal{L})$, then $\Lambda^{1/2}u \in \mathcal{D}^{0,1}$ and, by (3.9),

$$\|\Lambda^{1/2}u\|_{(1,-1/2)}^2 \le CQ(\Lambda^{1/2}u, \Lambda^{1/2}u). \tag{4.3}$$

Below, we will show that, for a certain smooth family of operators $A(y) \in OPS^1(\partial\Omega)$, we have $Au \in \mathcal{D}^{0,1}$ and

$$Q(\Lambda^{1/2}u, \Lambda^{1/2}u) = Q(u, Au) + R, \quad |R| \le C\|u\|_{(1,-1/2)}^2. \tag{4.4}$$

Granted this, we have (4.3) dominated by

$$\begin{aligned} C \text{ Re } Q(u, Au) + C\|u\|_{(1,-1/2)}^2 &= C \text{ Re } (\mathcal{L}u, Au)_{L^2} + C\|u\|_{(1,-1/2)}^2 \\ &\le C\|\mathcal{L}u\|_{L^2}\|u\|_{(0,1)} + C\|u\|_{(1,-1/2)}^2. \end{aligned} \tag{4.5}$$

Writing $C\|\mathcal{L}u\|_{L^2}\|u\|_{(0,1)} \le (C^2/\varepsilon)\|\mathcal{L}u\|_{L^2}^2 + \varepsilon\|u\|_{H^1}^2$ and absorbing the latter term on the left, we have

$$\|u\|_{H^1}^2 \le C\|\mathcal{L}u\|_{L^2}^2 + C\|u\|_{(1,-1/2)}^2. \tag{4.6}$$

If we use (3.25) to estimate the last term and recall that zero is not in the spectrum of $\mathcal{L}$, we have (4.1).

Our next step is to obtain higher-order a priori estimates in the tangential directions.

Proposition 4.2. *For* $u \in \Lambda^{0,1}(\overline{\Omega}) \cap \mathcal{D}(\mathcal{L}),\ k \ge 1$*, we have*

$$\|u\|_{(1,k/2-1/2)}^2 \le C\|\mathcal{L}u\|_{(0,k/2-1/2)}^2 + C\|u\|_{L^2}^2. \tag{4.7}$$

Proof. We will prove (4.7) by induction on k; the case $k = 1$ is implied by (4.1). To begin, we have

$$\|u\|^2_{(1,k/2-1/2)} \approx \|\Lambda^{k/2}u\|^2_{(1,-1/2)} \le C\; Q\Big(\Lambda^{k/2}u, \Lambda^{k/2}u\Big), \tag{4.8}$$

the latter inequality by (3.9), since $\Lambda^{k/2}u \in \mathcal{D}^{0,1}$. Now, extending (4.4), we have

$$Q(\Lambda^{k/2}u, \Lambda^{k/2}u) = Q(u, A_k u) + R_k, \quad |R_k| \le C\,\|u\|^2_{(1,k/2-1)}, \tag{4.9}$$

for a certain smooth family of operators $A_k(y) \in OPS^k(\partial\Omega)$, for which $A_k u \in \mathcal{D}^{0,1}$, as will be demonstrated below. Thus (4.8) is dominated by

$$\begin{aligned} &C\;\mathrm{Re}\; Q(u, A_k u) + C\,\|u\|^2_{(1,k/2-1)} \\ &\quad = C\;\mathrm{Re}\,(\mathcal{L}u, A_k u)_{L^2} + C\,\|u\|^2_{(1,k/2-1)} \\ &\quad \le C\,\|\mathcal{L}u\|_{(0,k/2-1/2)}\|u\|_{(0,k/2+1/2)} + C\,\|u\|^2_{(1,k/2-1)}. \end{aligned} \tag{4.10}$$

As in the passage from (4.5) to (4.6), this implies

$$\|u\|^2_{(1,k/2-1/2)} \le C\,\|\mathcal{L}u\|^2_{(0,k/2-1/2)} + C\,\|u\|^2_{(1,k/2-1)}, \tag{4.11}$$

which by induction on k yields the desired estimate (4.7).

We now take up the task of proving (4.4) and (4.9). It will be convenient to assume that the diffeomorphism $\mathcal{C} \approx [0,1) \times \partial\Omega$ has the property that Lebesgue measure on $\mathcal{C}$, induced from that on $\mathbb{C}^n$, coincides with the product measure on $[0,1) \times \partial\Omega$, up to a constant factor, a matter that can be arranged. We retain the fiber metric on $\Lambda^{0,1}$; on $\{y\} \times \partial\Omega$ this fiber metric depends on y. Then $\Lambda^{k/2}$, originally specified to be self-adjoint on $L^2(\partial\Omega, \Lambda^{0,1})$, has the property

$$(\Lambda^{k/2}u, v)_{L^2(C)} = (u, \widetilde{\Lambda}_{k/2}v)_{L^2}, \tag{4.12}$$

where

$$\begin{aligned} \widetilde{\Lambda}_{k/2}u(y) &= \Lambda^{k/2}u(y) + B_k(y)u(y), \\ B_k(y) \in OPS^{k/2-1}(\partial\Omega), &\quad B_k(0) = 0. \end{aligned} \tag{4.13}$$

Then we take

$$A_k(y) = \widetilde{\Lambda}_{k/2}\Lambda^{k/2} = \Lambda^k + B_k(y)\Lambda^{k/2}. \tag{4.14}$$

Clearly, A_k preserves $\mathcal{D}^{0,1}$.

Now, if we also let $\Lambda^{k/2}$ denote an elliptic self-adjoint operator in the class $OPS^{k/2}(\partial\Omega)$, acting on on sections of $\Lambda^{0,0}$ and $\Lambda^{0,2}$, having the same scalar principal symbol as the one acting on sections of $\Lambda^{0,1}$, we can write

$$(\bar\partial\Lambda^{k/2}u, \bar\partial\Lambda^{k/2}u)_{L^2} = (\bar\partial u, \widetilde{\Lambda}_{k/2}\bar\partial\Lambda^{k/2}u)_{L^2} + ([\bar\partial, \Lambda^{k/2}]u, \bar\partial\Lambda^{k/2}u)_{L^2}. \tag{4.15}$$

Further commutator pushing, plus use of the fact that the left side of (4.15) is real-valued, yields

$$(\bar\partial\Lambda^{k/2}u, \bar\partial\Lambda^{k/2}u)_{L^2} = \operatorname{Re}(\bar\partial u, \bar\partial Au)_{L^2} + R_{k1}, \tag{4.16}$$

where

$$\begin{aligned} R_{k1} &= \operatorname{Re}\left\{\left([\bar\partial, \Lambda^{k/2}]u, \bar\partial\Lambda^{k/2}u\right)_{L^2} + \left(\bar\partial u, [\widetilde{\Lambda}_{k/2}, \bar\partial]\Lambda^{k/2}u\right)_{L^2}\right\} \\ &= \operatorname{Re}(R_{k2} + R_{k3}). \end{aligned} \tag{4.17}$$

Now

$$\begin{aligned} R_{k2} + \overline{R}_{k3} &= \left([\bar\partial, \Lambda^{k/2}]u, \bar\partial\Lambda^{k/2}u\right)_{L^2} - \left([\bar\partial, \widetilde{\Lambda}_{k/2}]\Lambda^{k/2}u, \bar\partial u\right)_{L^2} \\ &= \left([\bar\partial, \Lambda^{k/2}]u, [\bar\partial, \Lambda^{k/2}]u\right)_{L^2} + \left(B_k[\bar\partial, \Lambda^{k/2}]u, \bar\partial u\right)_{L^2} \\ &\quad + \left([\Lambda^{k/2}, [\bar\partial, \Lambda^{k/2}]]u, \bar\partial u\right)_{L^2} + \left([B_k, \bar\partial]\Lambda^{k/2}u, \bar\partial u\right)_{L^2} \\ &= R_{k4} + R_{k5} + R_{k6} + R_{k7}, \end{aligned} \tag{4.18}$$

and standard pseudodifferential operator estimates yield

$$|R_{k4}| + |R_{k5}| + |R_{k6}| + |R_{k7}| \le C\,\|u\|^2_{(1,k/2-1)}, \tag{4.19}$$

which consequently bounds $\operatorname{Re}(R_{k2} + R_{k3})$. The term

$$(\bar\partial^*\Lambda^{k/2}u, \bar\partial^*\Lambda^{k/2}u)_{L^2}$$

has a similar analysis, so the estimates in (4.4) and (4.9) follow, and the proofs of Propositions 4.1 and 4.2 are complete.

The following is our main a priori estimate.

Proposition 4.3. *For $u \in \Lambda^{0,1}(\overline{\Omega}) \cap \mathcal{D}(\mathcal{L})$ and $j, k \ge 1$, we have*

$$\|u\|^2_{(j,k/2-1/2)} \le C\,\|\mathcal{L}u\|^2_{(j-1,k/2-1/2)} + C\,\|u\|^2_{L^2}, \tag{4.20}$$

and hence

$$\|u\|^2_{H^j} \le C\,\|\mathcal{L}u\|^2_{H^{j-1}}. \tag{4.21}$$

Proof. It suffices to prove (4.20) since the $k = 1$ case of this plus the invertibility of $\mathcal{L}$ implies (4.21). Note that the $j = 1$ case of (4.20) is precisely the conclusion of Proposition 4.2. We will give an inductive proof for $j \ge 2$. Note that if $j \ge 2$, (4.22) $$\|u\|^2_{(j,k/2-1/2)} \approx \|D^2_y\|^2_{(j-2,k/2-1/2)} + \|u\|^2_{(j-1,k/2+1/2)}.$$

Now since $\Box$ is elliptic, we can use the standard trick of writing $D^2_y u$ in terms of $\Box u$, second-order tangential derivatives of u, and first-order tangential derivatives of $D_y u$, to obtain

(4.23) $$\|u\|^2_{(j,k/2-1/2)} \leq C\,\|\mathcal{L}u\|^2_{(j-2,k/2-1/2)} + C\,\|u\|^2_{(j-1,k/2+1/2)}.$$

The inductive hypothesis dominates the last term by $C\,\|\mathcal{L}u\|^2_{(j-2,k/2+1/2)} + C\,\|u\|^2_{L^2}$, and this implies (4.20).

Note that if the $\overline{\partial}$-Neumann boundary condition were regular, we would have the estimate $\|u\|^2_{H^{j+1}} \leq C\,\|\mathcal{L}u\|^2_{H^{j-1}}$ in place of (4.21). The estimate (4.21) is called a *subelliptic estimate*. One also says that the $\overline{\partial}$-Neumann problem on a strongly pseudoconvex domain is subelliptic, with loss of one derivative.

Exercises

1. Sharpen the estimate (4.20) to

$$\|u\|^2_{(j,k/2-1/2)} \leq C\,\|\mathcal{L}u\|^2_{(j-2,k/2+1/2)} + C\,\|u\|^2_{L^2},$$

for all $u \in \Lambda^{0,1}(\overline{\Omega}) \cap \mathcal{D}(\mathcal{L})$, provided $k \geq 1$ and $j \geq 2$. In particular,

$$\|u\|^2_{H^2} \leq C\,\|\mathcal{L}u\|^2_{(0,1)} + C\,\|u\|^2_{L^2}.$$

2. Verify (4.19), namely, that $|R_{kj}| \leq C\,\|u\|^2_{(1,k/2-1)}$ for $4 \leq j \leq 7$.
(*Hint*: For example, *part* of the desired estimate on $|R_{k4}|$ follows from an estimate

$$\left\|[X, \Lambda^{k/2}]u\right\|_{L^2(\partial\Omega)} \leq C\,\|u\|_{H^{k/2}(\partial\Omega)},$$

for any first-order differential operator X on $\partial\Omega$. This in turn follows since

$$[X, \Lambda^{k/2}] \in OPS^{k/2}(\partial\Omega).$$

Similarly, part of the desired estimate on $|R_{k6}|$ follows because

$$[\Lambda^{k/2}, [X, \Lambda^{k/2}]] \in OPS^{k-1}(\partial\Omega).$$

5. Regularity via elliptic regularization

Our main goal here is to go from the a priori estimate that $\|u\|^2_{H^j} \leq C\,\|\mathcal{L}u\|^2_{H^{j-1}}$ provided $u \in \mathcal{D}(\mathcal{L})$ is smooth on $\overline{\Omega}$ to the regularity result that whenever

$u \in \mathcal{D}(\mathcal{L})$ and $\mathcal{L}u = f \in H^{j-1}(\Omega)$, then $u \in H^j(\Omega)$. Following [KN], we use the method of elliptic regularization, which is the following. For $\delta > 0$, consider the quadratic form

$$Q_\delta(u,u) = Q(u,u) + \delta \sum_j \|\partial_j u\|_{L^2}^2, \quad u \in \mathcal{D}^{0,1}, \tag{5.1}$$

where $Q(u,u) = \|\overline{\partial} u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2$ as in §3, and $\partial_j = \partial/\partial x_j$, $\partial_{n+j} = \partial/\partial y_j$, $1 \le j \le n$, applied to u componentwise. We take $\mathcal{H}_{1\delta}$ to be the completion of $\mathcal{D}^{0,1}$ with respect to the square norm Q_δ. Due to the last term in (5.1), we obviously have

$$\mathcal{H}_{1\delta} = \{u \in \mathcal{H}^1(\Omega, \Lambda^{0,1}) : \sigma_{\overline{\partial}^*}(x,\nu)u = 0 \text{ on } \partial\Omega\}, \quad \forall\, \delta > 0. \tag{5.2}$$

Note that $\mathcal{H}_{1\delta} \subset \mathcal{H}_1$, for $\delta > 0$, and $Q_\delta(u,u) \ge Q(u,u)$, for $u \in \mathcal{H}_{1\delta}$. Thus Morrey's inequality and the proof of Proposition 3.1 apply, yielding

$$Q_\delta(u,u) \ge C\|u\|_{H^{1/2}}^2 + C\int_{\partial\Omega} |u|^2\, dS + C\delta\|u\|_{H^1}^2, \quad u \in \mathcal{H}_{1\delta}. \tag{5.3}$$

We will define the self-adjoint operator $\mathcal{L}_\delta$ by the Friedrichs extension method, so $\mathcal{D}(\mathcal{L}_\delta^{1/2}) = \mathcal{H}_{1\delta}$ and

$$(\mathcal{L}_\delta u, v)_{L^2} = Q_\delta(u,v), \quad u \in \mathcal{D}(\mathcal{L}_\delta),\ v \in \mathcal{H}_{1\delta}. \tag{5.4}$$

Thus $\mathcal{L}_\delta^{-1}$ is a compact, self-adjoint operator on $L^2(\Omega, \Lambda^{0,1})$. Note that if $u \in \mathcal{D}(\mathcal{L}_\delta)$, the argument used in the proof of Proposition 3.2 shows that $\overline{\partial} u\big|_{\partial\Omega}$ is well defined in $\mathcal{D}'(\partial\Omega)$, and, for $v \in \mathcal{D}^{0,1}$ we have

$$Q_\delta(u,v) = \Bigl([\Box - \delta\Delta]u, v\Bigr)_{L^2} + \beta_\delta(u,v), \tag{5.5}$$

where

$$\beta_\delta(u,v) = \int_{\partial\Omega} \Bigl[\frac{1}{i}\Bigl\langle \sigma_{\overline{\partial}^*}(x,\nu)\,\overline{\partial} u, v\Bigr\rangle - \delta\Bigl\langle \frac{\partial u}{\partial \nu}, v\Bigr\rangle\Bigr]\, dS. \tag{5.6}$$

If we set $v = \sigma_{\overline{\partial}^*}(x,\nu)\varphi$ on $\partial\Omega$, we deduce that

$$u \in \mathcal{D}(\mathcal{L}_\delta) \Longrightarrow \left\{ \begin{array}{c} \sigma_{\overline{\partial}^*}(x,\nu)u = 0 \\ \sigma_{\overline{\partial}}(x,\nu)\Bigl[\frac{1}{i}\sigma_{\overline{\partial}^*}(x,\nu)\,\overline{\partial} u + \delta\frac{\partial u}{\partial \nu}\Bigr] = 0 \end{array} \right\} \text{ on } \partial\Omega. \tag{5.7}$$

For any $\delta > 0$, (5.3) is a coercive estimate. Such arguments as used in §7 of Chap. 5, for the Neumann boundary problem, produce higher-order estimates of the form

$$\|u\|_{H^{j+2}} \le C_{j\delta} \|\mathcal{L}_\delta u\|_{H^j}, \quad u \in \mathcal{D}(\mathcal{L}_\delta), \tag{5.8}$$

plus associated regularity theorems. Alternatively, the boundary condition (5.7) for the operator $\Box_\delta = \Box - \delta\Delta = -(1/2+\delta)\Delta$ is seen to be a regular boundary condition, and the results of §11 in Chap. 5 apply. Thus, for each $\delta > 0$,

$$\mathcal{L}_\delta^{-1} : H^j(\Omega) \longrightarrow H^{j+2}(\Omega), \quad j \ge 0. \tag{5.9}$$

The estimates in (5.9) depend crucially on δ of course, and one loses control as $\delta \searrow 0$. However, the analysis of §4 applies to $\mathcal{L}_\delta$, and one obtains

$$\|u\|_{H^{j+1}} \le C_j \|\mathcal{L}_\delta u\|_{H^j}, \quad u \in \mathcal{D}(\mathcal{L}_\delta) \cap \Lambda^{0,1}(\overline{\Omega}), \tag{5.10}$$

with C_j independent of $\delta \in (0,1]$. Using this, we will establish the following:

Proposition 5.1. *The operator $\mathcal{L}$ has the property that*

$$\mathcal{L}^{-1} : H^j(\Omega) \longrightarrow H^{j+1}(\Omega), \quad j \ge 0, \tag{5.11}$$

and

$$\mathcal{L}^{-1} : \Lambda^{0,1}(\overline{\Omega}) \longrightarrow \Lambda^{0,1}(\overline{\Omega}). \tag{5.12}$$

Of course, (5.12) follows from (5.11), but it will be technically convenient to prove these results together, completing the proof of (5.12) shortly before that of (5.11).

To begin, take $f \in \Lambda^{0,1}(\overline{\Omega})$ (so f is smooth on $\overline{\Omega}$). Then, for each $\delta > 0$, $\mathcal{L}_\delta^{-1} f = u_\delta \in \Lambda^{0,1}(\overline{\Omega})$. Hence (5.10) is applicable; we have $\{u_\delta : \delta \in (0,1]\}$ bounded in $H^j(\Omega)$, for each j. Thus this set is relatively compact in $H^{j-1}(\Omega)$ for each j, so there is a limit point

$$u_0 \in \bigcap_{j>0} H^j(\Omega, \Lambda^{0,1}) = \Lambda^{0,1}(\overline{\Omega});$$

$u_{\delta_\nu} \to u_0$ in the C^∞-topology while $\delta_\nu \searrow 0$. Now

$$(\Box - \delta\Delta) u_\delta = f \Longrightarrow \Box u_0 = f. \tag{5.13}$$

Also,

$$\sigma_{\overline{\partial}^*}(x,\nu) u_\delta \Big|_{\partial\Omega} = 0 \Longrightarrow \sigma_{\overline{\partial}^*}(x,\nu) u_0 \Big|_{\partial\Omega} = 0 \tag{5.14}$$

and

$$\sigma_{\overline{\partial}}(x,\nu) \Big[\frac{1}{i} \sigma_{\overline{\partial}^*}(x,\nu)\, \overline{\partial} u + \delta \frac{\partial u}{\partial \nu} \Big] \Big|_{\partial\Omega} = 0 \Longrightarrow \sigma_{\overline{\partial}^*}(x,\nu)\, \overline{\partial} u_0 \Big|_{\partial\Omega} = 0. \tag{5.15}$$

Therefore, $u_0 \in \mathcal{D}(\mathcal{L})$, so

$$\mathcal{L}^{-1} f = u_0 \in \Lambda^{0,1}(\overline{\Omega}). \tag{5.16}$$

This proves (5.12).

To prove (5.11), if $f \in H^j(\Omega, \Lambda^{0,1})$, take $f_\nu \in \Lambda^{0,1}(\overline{\Omega})$ so that $f_\nu \to f$ in $H^j(\Omega)$. We have $u_\nu = \mathcal{L}^{-1} f_\nu \in \Lambda^{0,1}(\overline{\Omega})$ and, by (4.21),

$$\|u_\nu - u_\mu\|_{H^{j+1}} \le C \|f_\nu - f_\mu\|_{H^j}. \tag{5.17}$$

Hence (u_ν) is Cauchy in $H^{j+1}(\Omega)$, so $\mathcal{L}^{-1} f = \lim_{\nu\to\infty} u_\nu \in H^{j+1}(\Omega)$.

Exercises

1. Verify that the boundary condition described in (5.7) is a regular boundary condition for $\mathcal{L}_\delta = C_\delta \Delta$, as defined in §11 of Chap. 5.
2. As an approach to Exercise 1, show that the analogue of the boundary condition (1.15) in this case, for the region $\{\operatorname{Im} z_n > 0\}$, is

$$u_n(z', x_n, 0) = 0, \quad \frac{\partial u_j}{\partial \overline{z}_n} + \delta \frac{\partial u_j}{\partial y_n}(z', x_n, 0) = 0, \quad 1 \le j \le n-1.$$

Show that the pseudodifferential equation arising in parallel with (1.18) is

$$\Big(\frac{\partial}{\partial x_n} + (i + 2\delta) N\Big) h = g$$

and that, for any $\delta > 0$, the pseudodifferential operator acting on h is elliptic.

6. The Hodge decomposition and the $\overline{\partial}$-equation

We begin with the following Hodge decomposition theorem.

Theorem 6.1. *If Ω is a bounded, strongly pseudoconvex domain in $\mathbb{C}^n$, then, given $u \in \Lambda^{0,1}(\overline{\Omega})$, we have*

$$u = \overline{\partial}\,\overline{\partial}^* \mathcal{L}^{-1} u + \overline{\partial}^* \overline{\partial} \mathcal{L}^{-1} u = P_{\overline{\partial}} u + P_{\overline{\partial}^*} u. \tag{6.1}$$

The two terms on the right side are mutually orthogonal in $L^2(\Omega, \Lambda^{0,1})$. Furthermore,

$$P_{\overline{\partial}}, P_{\overline{\partial}^*} : H^j(\Omega, \Lambda^{0,1}) \longrightarrow H^{j-1}(\Omega, \Lambda^{0,1}), \quad j \ge 1. \tag{6.2}$$

Proof. The first identity in (6.1) is equivalent to $u = \mathcal{L}\mathcal{L}^{-1} u$, and the second is simply the definition of $P_{\overline{\partial}}$ and $P_{\overline{\partial}^*}$. That (6.2) holds follows from (5.11). Only the orthogonality remains to be checked.

Following the proof of Proposition A.3, we use

$$(\overline{\partial} v, w)_{L^2} = (v, \overline{\partial}^* w)_{L^2} + \frac{1}{i} \int_{\partial\Omega} \langle v, \sigma_{\overline{\partial}^*}(x,\nu) w\rangle \, dS, \tag{6.3}$$

valid for $v \in \Lambda^{0,q}(\overline{\Omega}),\ w \in \Lambda^{0,q+1}(\overline{\Omega})$. Thus

$$w \in \mathcal{D}^{0,q+1} \Longrightarrow \overline{\partial}^* w \perp \ \ker \overline{\partial} \cap \Lambda^{0,q}(\overline{\Omega}), \tag{6.4}$$

where $\mathcal{D}^{0,q+1}$ is defined as in (2.2). Results established in previous sections imply

$$\overline{\partial}\mathcal{L}^{-1} : \Lambda^{0,1}(\overline{\Omega}) \longrightarrow \mathcal{D}^{0,2}, \tag{6.5}$$

so we can apply (6.4) to $w = \overline{\partial}\mathcal{L}^{-1}u$ to get

$$\overline{\partial}^*\overline{\partial}\mathcal{L}^{-1}u \perp \ \ker \overline{\partial} \cap \Lambda^{0,1}(\overline{\Omega}). \tag{6.6}$$

Hence

$$u_j \in \Lambda^{0,1}(\overline{\Omega}) \Longrightarrow P_{\overline{\partial}} u_1 \perp P_{\overline{\partial}^*} u_2 \quad \text{in } L^2(\Omega, \Lambda^{0,1}). \tag{6.7}$$

This finishes the proof of the theorem. It also implies that $P_{\overline{\partial}}$ and $P_{\overline{\partial}^*}$ extend uniquely to bounded operators (in fact, to complementary orthogonal projections) acting on $L^2(\Omega, \Lambda^{0,1})$.

The most significant application of this Hodge decomposition is to the equation

$$\overline{\partial} u = f, \tag{6.8}$$

given $f \in \Lambda^{0,1}(\overline{\Omega})$, for some $u \in \Lambda^{0,0}(\overline{\Omega}) = C^\infty(\overline{\Omega})$. Since $\overline{\partial}^2 = 0$, a necessary condition for solvability of (6.8) is

$$\overline{\partial} f = 0. \tag{6.9}$$

For strongly pseudoconvex domains, this is sufficient:

Theorem 6.2. *If Ω is a bounded, strongly pseudoconvex domain in $\mathbb{C}^n$, and $f \in \Lambda^{0,1}(\overline{\Omega})$ satisfies* (6.9)*, then there exists $u \in C^\infty(\overline{\Omega})$ satisfying* (6.8)*.*

Proof. With $g = \mathcal{L}^{-1} f \in \Lambda^{0,1}(\overline{\Omega})$, we have

$$f = \overline{\partial}\overline{\partial}^* g + \overline{\partial}^*\overline{\partial} g = P_{\overline{\partial}} f + P_{\overline{\partial}^*} f. \tag{6.10}$$

However, (6.4) applied to $w = \overline{\partial} g$ implies $P_{\overline{\partial}^*} f \perp f$, so in fact $P_{\overline{\partial}^*} f = 0$ and

$$f = \overline{\partial}(\overline{\partial}^* g). \tag{6.11}$$

Thus we have (6.8), with $u = \overline{\partial}^* g$.

We will use Theorem 6.2 to establish the following important result concerning function theory on a bounded, strongly pseudoconvex domain.

Proposition 6.3. *Let Ω be a bounded, strongly pseudoconvex domain in $\mathbb{C}^n$, and fix $p \in \partial\Omega$. Then there is a function u, holomorphic on Ω, such that $u \in C^\infty(\overline{\Omega} \setminus \{p\})$, but u blows up at p.*

Proof. It is shown in the exercises for §2 that there are a neighborhood $\mathcal{O}$ of p and a holomorphic function g, given by (2.16), such that $\{z \in \mathcal{O} : g(z) = 0\} \cap \overline{\Omega} = \{p\}$. Now the function

$$v = \frac{1}{g(z)} \tag{6.12}$$

is holomorphic on $\mathcal{O} \cap \Omega$ and C^∞ on $\mathcal{O} \cap \overline{\Omega} \setminus \{p\}$, and it blows up at p.

Pick $\psi \in C_0^\infty(\mathcal{O})$ such that $\psi = 1$ on a neighborhood $\mathcal{O}_2$ of p, and set

$$w = \psi v \tag{6.13}$$

on $\mathcal{O}$, extended to be 0 on the complement of $\mathcal{O}$. Now consider

$$f = \overline{\partial} w \tag{6.14}$$

on $\overline{\Omega}$; we take $f = 0$ on $\mathcal{O}_2 \cap \overline{\Omega}$. Thus $f \in \Lambda^{0,1}(\overline{\Omega})$ and $\overline{\partial} f = 0$, so by Theorem 6.2 there exists

$$w_2 \in C^\infty(\overline{\Omega}), \quad \overline{\partial} w_2 = f. \tag{6.15}$$

Now we set

$$u = w - w_2. \tag{6.16}$$

We have $\overline{\partial} u = f - f = 0$ on Ω, so u is holomorphic on Ω. The construction of w and the smoothness of w_2 on $\overline{\Omega}$ imply that $u \in C^\infty(\overline{\Omega} \setminus \{p\})$ and that u blows up at p, so the proof is complete.

Assuming that Ω is a bounded, strongly pseudoconvex domain in $\mathbb{C}^n$, we construct another special holomorphic function on Ω, as follows. Let $\{p_j : j \in \mathbb{Z}^+\}$ be a dense set of points in $\partial\Omega$, and for each j let u_j be a holomorphic function on Ω such that $u_j \in C^\infty(\Omega \setminus \{p_j\})$, constructed as above. Then we can produce mutually disjoint line segments γ_j lying in $\overline{\Omega}$, normal to $\partial\Omega$ at p_j, such that $u_j\big|_{\gamma_j}$ blows up at p_j. Now consider

$$u = \sum_{j \ge 0} c_j u_j, \tag{6.17}$$

where $c_j \in \mathbb{C}$ are all nonzero, but picked so small that

$$\text{(i)} \qquad |c_j u_j(z)| < 2^{-j} \quad \text{on } \mathcal{O}_j = \{z \in \Omega : \operatorname{dist}(z, \partial\Omega) \geq 2^{-j}\},$$

$$\text{(ii)} \qquad |c_j u_j(z)| < 2^{-j}, \quad \text{for } z \in \bigcup_{\ell < j} \gamma_\ell.$$

Condition (i) implies that (6.17) is uniformly convergent on compact subsets of Ω, hence u is holomorphic on Ω. Condition (ii) implies that, for each $k \in \mathbb{Z}^+$, $v_k = \sum_{j \neq k} c_j u_j$ is bounded on γ_k; hence $u = v_k + c_k u_k$ is *unbounded* on γ_k. This produces a holomorphic function on Ω with the following property:

Proposition 6.4. *If Ω is a bounded, strongly pseudoconvex domain in $\mathbb{C}^n$, then there is a holomorphic function u on Ω that is unbounded on each open set $\mathcal{O} \cap \Omega$, for any open $\mathcal{O}$ such that $\mathcal{O} \cap \partial\Omega \neq \emptyset$. Hence u does not extend holomorphically past any point in $\partial\Omega$.*

A domain $\Omega \subset \mathbb{C}^n$ having such an inextensible holomorphic function u is called a *domain of holomorphy*. Domains of holomorphy play an important role in the theory of holomorphic functions of several complex variables; we refer to [GR, Ho3, Kr1], and [Lel] for material on this.

We mention that, for the solution to (6.8) given by

$$u = Sf = \overline{\partial}^* \mathcal{L}^{-1} f, \tag{6.18}$$

we have $S : H^j(\Omega) \to H^j(\Omega)$, as a consequence of (5.11). In fact, one can do better:

$$S : H^j(\Omega) \longrightarrow H^{j+1/2}(\Omega). \tag{6.19}$$

One method of proving this is sketched in the exercises after §9.

Exercises

1. Interpolate (6.2) with the L^2-boundedness of $P_{\overline{\partial}}$ and $P_{\overline{\partial}^*}$ to show that

$$P_{\overline{\partial}}, P_{\overline{\partial}^*} : H^j(\Omega, \Lambda^{0,1}) \longrightarrow H^{j-\varepsilon}(\Omega, \Lambda^{0,1}), \quad \forall\, \varepsilon > 0,\ j \geq 1.$$

(*Hint*: Replace j by Nj in (6.2).)
Can you get rid of the ε?

7. The Bergman projection and Toeplitz operators

We use the operator $\mathcal{L}^{-1}$ on $(0,1)$-forms to produce the following Hodge decomposition for $(0,0)$-forms. Throughout this section we assume that Ω is a bounded strongly pseudoconvex domain in $\mathbb{C}^n$.

Proposition 7.1. *For all* $u \in \Lambda^{0,0}(\overline{\Omega})$,

$$u = Bu + \overline{\partial}^* \mathcal{L}^{-1} \overline{\partial} u \tag{7.1}$$

is an orthogonal decomposition in $L^2(\Omega)$*. The operator* B*, extended to* $L^2(\Omega)$*, coincides with the orthogonal projection onto*

$$\mathcal{H}(\Omega) = \{u \in L^2(\Omega) : \overline{\partial} u = 0\}. \tag{7.2}$$

Here, we take (7.1) as the definition of B. Thus, by (5.12), $B : \Lambda^{0,0}(\overline{\Omega}) \to \Lambda^{0,0}(\overline{\Omega})$. We need to prove that the decomposition (7.1) is orthogonal and that B, extended to $L^2(\Omega)$, is indeed the stated projection.

We first note that

$$(v, \overline{\partial}^* \mathcal{L}^{-1} \overline{\partial} u)_{L^2} = (\overline{\partial} v, \mathcal{L}^{-1} \overline{\partial} u)_{L^2}, \quad \forall\, u, v \in \Lambda^{0,0}(\overline{\Omega}), \tag{7.3}$$

since the two sides differ by the integral over $\partial\Omega$ of $\langle v, \sigma_{\overline{\partial}^*}(x,\nu) \mathcal{L}^{-1} \overline{\partial} u \rangle$, which vanishes. This identity shows that, given $v \in \Lambda^{0,0}(\overline{\Omega})$,

$$\overline{\partial} v = 0 \Longrightarrow v \perp \overline{\partial}^* \mathcal{L}^{-1} \overline{\partial} u, \quad \forall\, u \in \Lambda^{0,0}(\overline{\Omega}). \tag{7.4}$$

Next we claim that

$$\overline{\partial} B v = 0, \quad \forall\, v \in \Lambda^{0,0}(\overline{\Omega}). \tag{7.5}$$

This is equivalent to the statement that

$$\overline{\partial}\,\overline{\partial}^* \mathcal{L}^{-1} \overline{\partial} v = \overline{\partial} v. \tag{7.6}$$

Now, if we apply the decomposition (6.1) to $\overline{\partial} v$, we see that the two sides of (7.6) differ by $\overline{\partial}^* \overline{\partial} \mathcal{L}^{-1} \overline{\partial} v$; but this vanishes, by (6.6), so we have (7.5).

Combining (7.4) and (7.5), we have

$$Bv \perp \overline{\partial}^* \mathcal{L}^{-1} \overline{\partial} u, \quad \forall\, u, v \in \Lambda^{0,0}(\overline{\Omega}), \tag{7.7}$$

so the decomposition (7.1) is orthogonal. Thus B does extend to an orthogonal projection on $L^2(\Omega)$ and, by (7.5), $\mathcal{R}(B) \subset \mathcal{H}(\Omega)$. If we apply (7.1) to an element u of

$$\mathcal{H}(\overline{\Omega}) = \{u \in \Lambda^{0,0}(\overline{\Omega}) : \overline{\partial} u = 0\}, \tag{7.8}$$

we get $u = Bu$, so we have

$$\overline{\mathcal{H}}(\overline{\Omega}) \subset \mathcal{R}(B) \subset \mathcal{H}(\Omega),$$

where $\overline{\mathcal{H}}(\overline{\Omega})$ denotes the closure of $\mathcal{H}(\overline{\Omega})$ in $L^2(\Omega)$.

In fact, since $B : \Lambda^{0,0}(\overline{\Omega}) \to \Lambda^{0,0}(\overline{\Omega})$ and $\Lambda^{0,0}(\overline{\Omega})$ is dense in $L^2(\Omega)$, it is now clear that $\mathcal{R}(B) = \overline{\mathcal{H}}(\overline{\Omega})$. We could stop here (rephrasing the statement of Proposition 7.1), but it is of intrinsic interest to equate this space with $\mathcal{H}(\Omega)$, which we now do.

Lemma 7.2. *If Ω is a strongly pseudoconvex domain in $\mathbb{C}^n$, then $\mathcal{H}(\overline{\Omega})$ is dense in $\mathcal{H}(\Omega)$.*

Proof. It suffices to show that

$$\mathcal{H}(\Omega) \perp \overline{\partial}^* \mathcal{L}^{-1} \overline{\partial} u, \quad \forall\, u \in \Lambda^{0,0}(\overline{\Omega}). \tag{7.9}$$

Now, if $v \in \mathcal{H}(\Omega)$ and $u \in \Lambda^{0,0}(\overline{\Omega})$, so $w = \mathcal{L}^{-1}\overline{\partial} u \in \mathcal{D}^{0,1} \subset \Lambda^{0,1}(\overline{\Omega})$, then

$$(v, \overline{\partial}^* \mathcal{L}^{-1} \overline{\partial} u)_{L^2(\Omega)} = \lim_{s \to 0} \int_{\Omega_s} \langle v, \overline{\partial}^* w \rangle \, dV, \tag{7.10}$$

where $\Omega_s = \{z \in \overline{\Omega} : \rho(z) \le -s\}$. We have

$$\int_{\Omega_s} \langle v, \overline{\partial}^* w \rangle \, dV = \int_{\Omega_s} \langle \overline{\partial} v, w \rangle \, dV + \int_{\partial\Omega_s} \langle v, \sigma_{\overline{\partial}^*}(x, \nu) w \rangle \, dS. \tag{7.11}$$

Of course, the first term on the right side of (7.11) vanishes if $v \in \mathcal{H}(\Omega)$. Now, we can take a collar neighborhood of $\partial\Omega$ and identify $\partial\Omega_s$ with $\partial\Omega$, for s small. Then, for each $v \in \mathcal{H}(\Omega)$, $v(s) = v\big|_{\partial\Omega_s}$ provides a bounded family in $\mathcal{D}'(\partial\Omega)$ as $s \to 0$. Meanwhile, for any $w \in \mathcal{D}^{0,1}$,

$$\sigma w(s) = \sigma_{\overline{\partial}^*}(x, \nu) w\big|_{\partial\Omega_s} \to 0 \quad \text{in } C^\infty(\partial\Omega), \tag{7.12}$$

as $s \to 0$. Thus the second term on the right side of (7.11) vanishes as $s \to 0$, so we have (7.9), and the lemma is proved.

The orthogonal projection B is called the *Bergman projection*. If we take as its defining property that B projects $L^2(\Omega)$ onto $\mathcal{H}(\Omega)$, then the content of Proposition 7.1 is that we have a formula for B:

$$Bu = u - \overline{\partial}^* \mathcal{L}^{-1} \overline{\partial} u, \tag{7.13}$$

at least for $u \in H^1(\Omega)$. The mapping property (5.11) implies $B : H^j(\Omega) \to H^{j-1}(\Omega)$, for $j \geq 1$. If we interpolate this with $B : L^2(\Omega) \to L^2(\Omega)$, we deduce that

$$B : H^j(\Omega) \longrightarrow H^{j-\varepsilon}(\Omega), \quad \forall\, \varepsilon > 0,\ j \geq 1. \tag{7.14}$$

Compare with Exercise 1 in §6. In [K3] it is proved that actually $B : H^j(\Omega) \to H^j(\Omega)$.

Since Bu is holomorphic for each $u \in L^2(\Omega)$, the evaluation at any $z \in \Omega$ is a continuous linear functional on $L^2(\Omega)$, so there exists a unique element of $L^2(\Omega)$, which we denote as k_z, such that

$$Bu(z) = (u, k_z)_{L^2}, \quad \forall\, u \in L^2(\Omega). \tag{7.15}$$

Since holomorphic functions are harmonic, the mean-value property implies that whenever $\varphi_z \in C_0^\infty(\Omega)$ is real-valued and radially symmetric about $z \in \Omega$, with total integral 1, then

$$Bu(z) = (Bu, \varphi_z)_{L^2} = (u, B\varphi_z)_{L^2}, \tag{7.16}$$

so, for each $z \in \Omega$,

$$k_z = B\varphi_z \in C^\infty(\overline{\Omega}). \tag{7.17}$$

Also, one can clearly choose $\varphi_z(\zeta)$ depending smoothly on z and ζ, so the map $z \mapsto k_z$ is C^∞ on Ω, with values in $C^\infty(\overline{\Omega})$. Thus we can write

$$k_z(\zeta) = \overline{K(z, \zeta)}, \quad K \in C^\infty(\Omega \times \overline{\Omega}). \tag{7.18}$$

Then we can rewrite (7.15) as

$$Bu(z) = \int_\Omega u(\zeta) K(z, \zeta)\, dV(\zeta). \tag{7.19}$$

The function $K(z, \zeta)$ is called the *Bergman kernel function*. Since $B = B^*$, we have

$$K(z, \zeta) = \overline{K(\zeta, z)}; \tag{7.20}$$

hence (7.18) implies

$$K \in C^\infty(\Omega \times \overline{\Omega}) \cap C^\infty(\overline{\Omega} \times \Omega). \tag{7.21}$$

This regularity result is due to [Ker2].

In [F] an analysis was made of the asymptotic behavior of $K(z, z)$ as z approaches $\partial\Omega$. It was used there as a tool to prove that if Ω_1 and Ω_2 are two bounded, strongly pseudoconvex domains with smooth boundary and $\Phi : \Omega_1 \to \Omega_2$ is a biholomorphism, then Φ extends to a diffeomorphism $\Phi : \overline{\Omega}_1 \to \overline{\Omega}_2$. Later, S. Bell and E. Ligocka [BL] found a simpler proof of this mapping result, relying on the property that $B : C^\infty(\overline{\Omega}) \to C^\infty(\overline{\Omega})$ (which follows from (7.14). Nevertheless, the asymptotic analysis of $K(z, z)$ has substantial intrinsic interest. A discussion of a number of aspects of this study is given in the survey [BFG]. In [BSj] the analysis of $K(z, z)$ is related to an analysis of the Szegö projection, a projection analogous to the Bergman projection but defined on $L^2(\partial\Omega)$. Alternative approaches to the analysis of the Szegö projection are given in [KS] and in [Tay].

We turn now to a study of Toeplitz operators, defined as follows. Given $f \in L^\infty(\Omega)$, we denote $M_f u = fu$ and set

$$T_f u = B(fu), \quad u \in \mathcal{H}(\Omega). \tag{7.22}$$

Thus $T_f : \mathcal{H}(\Omega) \to \mathcal{H}(\Omega)$. We call T_f a Toeplitz operator. Note that since $\|B\| = 1$, we have an L^2-operator norm bound on T_f:

$$\|T_f\| \le \|f\|_{L^\infty}. \tag{7.23}$$

Toeplitz operators have a number of interesting properties, some of which we derive here. In the statements below, $\mathcal{L}$ denotes the space of bounded operators and $\mathcal{K}$ the space of compact operators, acting on the relevant Hilbert space, usually $L^2(\Omega)$ or $\mathcal{H}(\Omega)$.

Proposition 7.3. *If $f, g \in C(\overline{\Omega})$, then*

$$T_f T_g - T_{fg} \in \mathcal{K}. \tag{7.24}$$

Also, if $f \in C(\overline{\Omega})$,

$$f\Big|_{\partial\Omega} = 0 \Longrightarrow T_f \in \mathcal{K}. \tag{7.25}$$

Thus $f \mapsto T_f$ produces a homomorphism of Banach algebras

$$\tau : C(\partial\Omega) \longrightarrow \mathcal{L}/\mathcal{K}. \tag{7.26}$$

In view of (7.23), it suffices to prove (7.24) for $f, g \in C^\infty(\overline{\Omega})$. Also, it suffices to prove (7.25) for $f \in C_0^\infty(\Omega)$. In fact,

$$f \in C_0^\infty(\Omega) \Longrightarrow T_f : \mathcal{H}(\Omega) \to C^\infty(\overline{\Omega}), \tag{7.27}$$

so the compactness of T_f in this case is obvious.

To prove (7.24), note that

$$T_f T_g u - T_{fg} u = BfBgu - Bfgu = -Bf(I-B)gu,$$

so (7.24) follows if we show that

$$f \in C^\infty(\overline{\Omega}) \Longrightarrow BM_f(I-B) \in \mathcal{K}. \tag{7.28}$$

It is more convenient to show that

$$(I-B)M_f B \in \mathcal{K}, \tag{7.29}$$

which implies (7.28) upon taking adjoints. To see this, let us use (7.13) to write

$$(I-B)M_f Bu = \overline{\partial}^* \mathcal{L}^{-1}\, \overline{\partial}(fBu), \tag{7.30}$$

at least for $u \in C^\infty(\overline{\Omega})$. Since Bu is holomorphic, $\overline{\partial}(fBu) = (\overline{\partial} f)Bu$, so we have

$$(I-B)M_f Bu = \overline{\partial}^* \mathcal{L}^{-1}\Big((\overline{\partial} f)Bu\Big), \tag{7.31}$$

an identity that extends to $L^2(\Omega)$ since both sides are bounded on $L^2(\Omega)$. Thus (7.29) will be established, and the proof of Proposition 7.3 will be complete, when we establish the following:

Lemma 7.4. *We have*

$$\overline{\partial}^* \mathcal{L}^{-1} : L^2(\Omega, \Lambda^{0,1}) \longrightarrow L^2(\Omega) \quad \textit{compact}. \tag{7.32}$$

Proof. For $v \in \mathcal{H}_1$, we have $\|\overline{\partial}^* v\|_{L^2} \le \|\mathcal{L}^{1/2} v\|_{L^2}$, so, for $u \in L^2(\Omega, \Lambda^{0,1})$,

$$\|\overline{\partial}^* \mathcal{L}^{-1} u\|_{L^2} \le \|\mathcal{L}^{-1/2} u\|_{L^2}. \tag{7.33}$$

Since $\mathcal{L}^{-1/2}$ is compact on L^2, (7.32) easily follows.

Proposition 7.3 extends to the case when f and g take values in End$(\mathbb{C}^m)$, and T_f, T_g act on m-tuples of elements of $\mathcal{H}(\Omega)$. We then have the following.

Proposition 7.5. *If $f \in C\left(\overline{\Omega}, End(\mathbb{C}^m)\right)$ and $f(z)$ is invertible for each $z \in \partial\Omega$, then T_f is Fredholm.*

Proof. Let $g \in C\left(\overline{\Omega}, \text{End}(\mathbb{C}^m)\right)$ satisfy $fg = gf = I$ on $\partial\Omega$. Then Proposition 7.3 implies

$$I - T_f T_g \in \mathcal{K}, \quad I - T_g T_f \in \mathcal{K},$$

so T_g is a Fredholm inverse of T_f.

It is natural to ask what is the index of T_f, which clearly depends only on the homotopy class of $f : \partial\Omega \to Gl(m, \mathbb{C})$, by general results on Fredholm operators established in §7 of Appendix A, on functional analysis. A formula for Index T_f is given by [Ven] in case Ω is a ball in $\mathbb{C}^n$. The case of a general, strongly pseudoconvex domain is treated in [B2]. The formula given there is equivalent to an identity of the form

$$\text{Index}\, T_f = \text{Index}\, P_f, \tag{7.34}$$

where P_f is an elliptic pseudodifferential operator on $\partial\Omega$, constructed as follows. The manifold $\partial\Omega$ possesses a spin^c structure and associated Dirac operator $D_{\partial\Omega}$ (objects defined in Chap. 10). The operator $D_{\partial\Omega}$ is a self-adjoint operator on $L^2(\partial\Omega, S)$, where $S \to \partial\Omega$ is a certain spinor bundle. Denote by H_+ the closed linear span of the positive eigenspaces of $D_{\partial\Omega}$ and by P_+ the orthogonal projection onto H_+. If f takes values in End$(\mathbb{C}^m)$, let P_+ also denote the orthogonal projection of $L^2(\partial\Omega, S \otimes \mathbb{C}^m)$ onto $H_+ \otimes \mathbb{C}^m$. Then we set

$$P_f = P_+ M_f P_+ + (I - P_+), \ \text{acting on} \ L^2(\partial\Omega, S \otimes \mathbb{C}^m). \tag{7.35}$$

We see that $P_f \in OPS^0(\partial\Omega)$ if f is smooth, and P_f is elliptic if $f\big|_{\partial\Omega}$ is invertible. The index of P_f is given by the Atiyah–Singer formula; see (8.22)–(8.25) in Chap. 10.

We note that the correspondence $f \mapsto P_f$ has properties like those established for $f \mapsto T_f$ in Proposition 7.3. That is, if $f, g \in C(\partial\Omega)$,

$$P_f P_g - P_{fg} \in \mathcal{K}, \tag{7.36}$$

so $f \mapsto P_f$ produces a homomorphism of Banach algebras

$$\pi : C(\partial\Omega) \longrightarrow \mathcal{L}/\mathcal{K}. \tag{7.37}$$

In fact, (7.36) is established more easily than (7.24); if $f, g \in C^\infty(\partial\Omega)$, we have $P_f P_g - P_{fg} \in OPS^{-1}(\partial\Omega)$. In addition, one can also show that

$$\tau(\overline{f}) = \tau(f)^*, \quad \pi(\overline{f}) = \pi(f)^*. \tag{7.38}$$

The maps τ and π are said to produce extensions of $C(\partial\Omega)$ by $\mathcal{K}$. There are certain equivalence relations among such extensions, first specified by [BDF], and the resulting equivalence classes define elements of the K-homology group $K_1(\partial\Omega)$. In [BDT] it is proved that τ and π define the *same* element of $K_1(\partial\Omega)$, a result that implies (7.34) and hence refines Boutet de Monvel's index theorem.

Exercises

1. Let $\{u_j\}$ be an orthonormal basis of $\mathcal{H}(\Omega)$. Show that

$$K(z, \zeta) = \sum_j u_j(z)\overline{u_j(\zeta)},$$

the series converging in $C^\infty(\Omega \times \Omega)$.

2. Show that

$$K(z, z) \geq \frac{1}{\text{vol } \Omega}, \quad \forall\, z \in \Omega.$$

(*Hint*: Take a orthonormal basis $\{u_j\}$ of $\mathcal{H}(\Omega)$ with $u_1 = \text{const.}$)

3. Show that $\Psi(z) = \log\ K(z, z)$ is strongly plurisubharmonic on Ω, in the sense defined in (2.15). Deduce that

$$h_{jk}(z) = \frac{\partial^2 \Psi}{\partial z_j \partial \overline{z}_k}$$

defines a positive-definite Hermitian metric on Ω. This is called the Bergman metric on Ω.

4. Suppose $F : \Omega_1 \to \Omega_2$ is a biholomorphic diffeomorphism between two strongly pseudoconvex domains Ω_1 and Ω_2. Compute the relation between the Bergman kernel functions, and deduce that F *preserves* the Bergman metric.

5. Let $\mathbb{B}^n$ be the unit ball in $\mathbb{C}^n$. Show that an orthonormal basis for $\mathcal{H}(\mathbb{B}^n)$ is given by

$$u_\alpha(z) = b_\alpha z^\alpha, \quad b_\alpha = \sqrt{\frac{(n + |\alpha|)!}{\alpha!}}.$$

Deduce that the Bergman kernel function for $\mathbb{B}^n$ is given by

$$K(z, \zeta) = V_n^{-1}(1 - z \cdot \overline{\zeta})^{-(n+1)},$$

where $V_n = \text{ Vol } \mathbb{B}^n$. Compute the Bergman metric for the ball.

8. The $\overline{\partial}$-Neumann problem on $(0,q)$-forms

So far, we have analyzed the $\overline{\partial}$-Neumann problem for $(0,1)$-forms, but it was formulated for (p,q)-forms in §1. Here we extend the analysis of §§2–6 to $(0,q)$-forms. Our first order of business is to try to extend Morrey's inequality. We try to parallel the computation in (2.5)–(2.11). It is convenient to perform the computation in a more invariant way, using (1.10)–(1.11), that is,

$$\overline{\partial}\overline{\partial}^* u + \overline{\partial}^*\overline{\partial} u = \Box u = -\tfrac{1}{2}\Delta u, \tag{8.1}$$

where Δ acts on u componentwise, for $u \in \Lambda^{0,q}(\overline{\Omega}),\ \Omega \subset \mathbb{C}^n$. We have, as in (3.20)–(3.21),

$$\|\overline{\partial}u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2 = (\Box u, u)_{L^2} - \frac{1}{i}\int\limits_{\partial\Omega} \Big\langle \sigma_{\overline{\partial}^*}(x,\nu)\,\overline{\partial}u, u\Big\rangle\, dS, \tag{8.2}$$

for $u \in \mathcal{D}^{0,q}$, the other boundary integrand $\Big\langle \overline{\partial}^* u, \sigma_{\overline{\partial}^*}(x,\nu)u\Big\rangle$ vanishing in this case. Also, we have

$$(\Box u, u)_{L^2} = 2\sum_k \Big\| \frac{\partial u}{\partial \overline{z}_k}\Big\|_{L^2}^2 + \frac{2}{i}\int\limits_{\partial\Omega}\sum_k \Big\langle \sigma_{\partial/\partial z_k}(x,\nu)\frac{\partial u}{\partial \overline{z}_k}, u\Big\rangle\, dS. \tag{8.3}$$

Hence, for $u \in \mathcal{D}^{0,q}$,

$$\|\overline{\partial}u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2 = 2\sum_k \Big\| \frac{\partial u}{\partial \overline{z}_k}\Big\|_{L^2}^2 + \gamma(u,u), \tag{8.4}$$

where

$$\gamma(u,u) = -\frac{1}{i}\int\limits_{\partial\Omega} \Big\langle \sigma_{\overline{\partial}^*}(x,\nu)\,\overline{\partial}u - 2\sum_k \sigma_{\partial/\partial z_k}(x,\nu)\frac{\partial u}{\partial \overline{z}_k}, u\Big\rangle\, dS. \tag{8.5}$$

Note that when $q = 1$, the first term on the right side of (8.4) is equal to the first term on the right side of (2.11), since $|dz_j|^2 = 2$.

Let us write the integrand in (8.5) as $\alpha(u,u) + \beta(u,u)$, with

$$\alpha(u,u) = -\frac{1}{i}\Big\langle \sigma_{\overline{\partial}^*}(x,\nu)\,\overline{\partial}u, u\Big\rangle = -\langle \overline{\partial}u, \overline{\partial}\rho \wedge u\rangle \tag{8.6}$$

and

$$\beta(u,u) = \frac{2}{i}\sum_k \Big\langle \sigma_{\partial/\partial z_k}(x,\nu)\frac{\partial u}{\partial \overline{z}_k}, u\Big\rangle = 2\sum_k \frac{\partial \rho}{\partial z_k}\Big\langle \frac{\partial u}{\partial \overline{z}_k}, u\Big\rangle. \tag{8.7}$$

Note that in the case $q = 1$, when $u = \sum u_j \, d\overline{z}_j$, so $\overline{\partial} u$ is given by (2.3), we have

$$\overline{\partial}\rho \wedge u = \sum_{j<k} \left(\frac{\partial \rho}{\partial \overline{z}_j} u_k - \frac{\partial \rho}{\partial \overline{z}_k} u_j \right) \, d\overline{z}_j \wedge d\overline{z}_k, \tag{8.8}$$

so

$$\begin{aligned} \alpha(u, u) &= -4 \sum_{j<k} \left(\frac{\partial u_j}{\partial \overline{z}_k} - \frac{\partial u_k}{\partial \overline{z}_j} \right) \left(\frac{\partial \rho}{\partial z_j} \overline{u}_k - \frac{\partial \rho}{\partial z_k} \overline{u}_j \right) \\ &= 4 \sum_{j,k} \left(\frac{\partial \rho}{\partial z_k} \frac{\partial u_j}{\partial \overline{z}_k} \overline{u}_j - \frac{\partial \rho}{\partial z_j} \frac{\partial u_j}{\partial \overline{z}_k} \overline{u}_k \right). \end{aligned} \tag{8.9}$$

Here, the first part of the last sum cancels $\beta(u, u)$, and the rest is what appears on the left side of (2.8) (multiplied by -4). Upon applying the identity (2.8), we thus recover the identity (2.11), for $q = 1$.

More generally, if u is a $(0, q)$-form:

$$u = \sum u_\alpha \, d\overline{z}^\alpha, \tag{8.10}$$

summed over $\alpha_1 < \cdots < \alpha_q$, then

$$\begin{aligned} \alpha(u, u) &= -\sum \left\langle \frac{\partial u_\alpha}{\partial \overline{z}_j} \, d\overline{z}_j \wedge d\overline{z}^\alpha, \frac{\partial \rho}{\partial \overline{z}_k} u_\beta \, d\overline{z}_k \wedge d\overline{z}^\beta \right\rangle \\ &= -2^{q+1} \sum \operatorname{sgn} \begin{pmatrix} j\alpha \\ k\beta \end{pmatrix} \frac{\partial \rho}{\partial z_k} \frac{\partial u_\alpha}{\partial \overline{z}_j} \overline{u}_\beta, \end{aligned} \tag{8.11}$$

where $\operatorname{sgn} \binom{j\alpha}{k\beta}$ is $+1$ if $j, \alpha_1, \dots, \alpha_q$ are distinct and are an even permutation of $k, \beta_1, \dots, \beta_q$, is -1 if an odd permutation, and is zero otherwise. We also have

$$\beta(u, u) = 2^{q+1} \sum_{j,\alpha} \frac{\partial \rho}{\partial z_j} \frac{\partial u_\alpha}{\partial \overline{z}_j} \overline{u}_\alpha, \tag{8.12}$$

which cancels out the part of the last sum in (8.11) for which $j = k$ and $\alpha = \beta$.

We next want to extend the identity (2.8), so we look for some derivatives tangent to $\partial\Omega$, arising from $u \in \mathcal{D}^{0,q}$. A calculation gives

$$\sigma_{\overline{\partial}^*}(x, \nu) u = \sum_\gamma \Bigl(\sum_{j,\alpha} \operatorname{sgn} \begin{pmatrix} \alpha \\ j\gamma \end{pmatrix} u_\alpha \frac{\partial \rho}{\partial z_j} \Bigr) d\overline{z}^\gamma, \tag{8.13}$$

summed over $\gamma_1 < \cdots < \gamma_{q-1}$. Thus, if $u \in \mathcal{D}^{0,q}$, then

$$\sum_{j,\alpha} \operatorname{sgn}\begin{pmatrix}\alpha\\ j\gamma\end{pmatrix} u_\alpha \frac{\partial}{\partial z_j}\rho = 0, \quad \forall\, \gamma_1 < \cdots < \gamma_{q-1}. \tag{8.14}$$

Hence, extending (2.7), we have

$$\sum_{j,k,\alpha,\beta} \operatorname{sgn}\begin{pmatrix}\beta\\ j\gamma\end{pmatrix} \overline{u}_\gamma \frac{\partial}{\partial \overline{z}_j}\Bigl(\operatorname{sgn}\begin{pmatrix}\alpha\\ k\sigma\end{pmatrix} u_\alpha \frac{\partial \rho}{\partial z_k}\Bigr) = 0, \tag{8.15}$$

for all $\gamma_1 < \cdots < \gamma_{q-1}$ and $\sigma_1 < \cdots < \sigma_{q-1}$. Hence

$$\sum_{j,k,\alpha,\beta} \operatorname{sgn}\begin{pmatrix}\beta\\ j\gamma\end{pmatrix} \operatorname{sgn}\begin{pmatrix}\alpha\\ k\sigma\end{pmatrix} \Bigl\{\overline{u}_\beta \frac{\partial u_\alpha}{\partial \overline{z}_j}\frac{\partial \rho}{\partial z_k} + \overline{u}_\beta u_\alpha \frac{\partial^2 \rho}{\partial \overline{z}_j \partial z_k}\Bigr\} = 0, \tag{8.16}$$

for all such γ and σ. In this sum we also require $\beta_1 < \cdots < \beta_q$.

Now we can put (8.11) and (8.12) together with (8.16), to establish the following:

Proposition 8.1. *If $u \in \mathcal{D}^{0,q}$, then*

$$\begin{aligned}\|\overline{\partial}u\|^2_{L^2} &+ \|\overline{\partial}^* u\|^2_{L^2} \\ &= 2\sum_k \Bigl\|\frac{\partial u}{\partial \overline{z}_k}\Bigr\|^2_{L^2} + 2^{q+1}\sum_\gamma \int_{\partial\Omega} \sum_{j,k} \frac{\partial^2 \rho}{\partial \overline{z}_j \partial z_k} \overline{W}_{j\gamma} W_{k\gamma}\, dS,\end{aligned} \tag{8.17}$$

with

$$W_{k\gamma} = \sum_\alpha \operatorname{sgn}\begin{pmatrix}\alpha\\ k\gamma\end{pmatrix} u_\alpha. \tag{8.18}$$

Proof. It suffices to show that $\alpha(u,u) + \beta(u,u)$, given by (8.11) and (8.12), is equal to

$$-\sum \operatorname{sgn}\begin{pmatrix}\beta\\ j\gamma\end{pmatrix} \operatorname{sgn}\begin{pmatrix}\alpha\\ k\gamma\end{pmatrix} \frac{\partial \rho}{\partial z_k}\frac{\partial u_\alpha}{\partial \overline{z}_j}\overline{u}_\beta, \tag{8.19}$$

where we sum over $j, k, \alpha, \beta, \gamma$, with $\alpha_1 < \cdots < \alpha_q$, and so on.

To establish the identity at a given point $p \in \partial\Omega$, rotate coordinates so $\nabla\rho(p) = \partial/\partial y_n$, and hence $\partial\rho/\partial z_k = -(i/2)\delta_{kn}$. Then, at p, the quantity (8.19) is equal to i times

$$\frac{1}{2}\sum_{j,\alpha,\beta,\gamma} \operatorname{sgn}\begin{pmatrix}\beta\\ j\gamma\end{pmatrix} \operatorname{sgn}\begin{pmatrix}\alpha\\ n\gamma\end{pmatrix} \frac{\partial u_\alpha}{\partial \overline{z}_j}\overline{u}_\beta. \tag{8.20}$$

That $u \in \mathcal{D}^{0,q}$ implies that $u_\beta = 0$ at p whenever n occurs in β, so we can take the sum in (8.20) over $j \neq n$. Meanwhile, (8.11) and (8.12) sum to i times

$$\frac{1}{2} \sum_{j,\alpha,\beta} \operatorname{sgn} \binom{j\alpha}{n\beta} \frac{\partial u_\alpha}{\partial \overline{z}_j} \overline{u}_\beta - \frac{1}{2} \sum_{\alpha} \frac{\partial u_\alpha}{\partial \overline{z}_n} \overline{u}_\alpha \tag{8.21}$$

at p; this is equal to

$$\frac{1}{2} \sum_{\alpha,\beta} \operatorname{sgn} \binom{\alpha}{\beta} \frac{\partial u_\alpha}{\partial \overline{z}_n} \overline{u}_\beta - \frac{1}{2} \sum_{\alpha} \frac{\partial u_\alpha}{\partial \overline{z}_n} \overline{u}_\alpha + \frac{1}{2} \sum_{j \neq n} \operatorname{sgn} \binom{j\alpha}{n\beta} \frac{\partial u_\alpha}{\partial \overline{z}_j} \overline{u}_\beta . \tag{8.22}$$

Now the first two sums in (8.22) cancel and the last sum is equal to (8.20). This proves the proposition.

Note that (8.13) is equivalent to $\sum_j W_{j\gamma}\, \partial\rho/\partial z_j = 0$ for all γ, so the hypothesis that the Levi form be positive-definite implies

$$\sum_{j,k} \frac{\partial^2 \rho}{\partial \overline{z}_j \partial z_k} \overline{W}_{j\gamma} W_{k\gamma} \geq C \sum_j |W_{j\gamma}|^2, \quad \forall\, \gamma. \tag{8.23}$$

Hence the last term in (8.17) is

$$\geq C \sum_{j,\gamma} \int_{\partial\Omega} |W_{j\gamma}|^2 \, dS, \tag{8.24}$$

when Ω is strongly pseudoconvex. On the other hand, the map $u \mapsto (W_{k\gamma})$ is clearly injective, so $\sum_{j,\gamma} |W_{j\gamma}|^2 \geq C|u|^2$. We hence have the following:

Corollary 8.2. *If Ω is strongly pseudoconvex and $u \in \mathcal{D}^{0,q}$, $q \geq 1$, then*

$$\|\overline{\partial} u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2 \geq 2 \sum_k \Big\| \frac{\partial u}{\partial \overline{z}_k} \Big\|_{L^2}^2 + C \int_{\partial\Omega} |u|^2 \, dS. \tag{8.25}$$

In the rest of this section, we assume that Ω is a bounded, strongly pseudoconvex domain in $\mathbb{C}^n$.

From here, we can use the argument from Proposition 3.1 to show that

$$\|\overline{\partial} u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2 \geq C \|u\|_{H^{1/2}}^2, \quad \forall\, u \in \mathcal{D}^{0,q}, \tag{8.26}$$

as long as $q \geq 1$, and more precisely we can use $\|u\|_{(1,-1/2)}^2$ on the right, as in (3.9).

As in (3.12)–(3.16), we can define an unbounded, self-adjoint operator $\mathcal{L}$ on $L^2(\Omega, \Lambda^{0,q})$, for each $q \geq 1$, such that $\mathcal{D}(\mathcal{L}^{1/2})$ is the completion of $\mathcal{D}^{0,q}$ with respect to the square norm $Q(u,u) = \|\overline{\partial} u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2$, and

$$\|\mathcal{L}^{1/2}u\|_{L^2}^2 = Q(u,u), \quad \forall\, u \in \mathcal{D}(\mathcal{L}^{1/2}). \tag{8.27}$$

By (8.25) and (8.26), $\mathcal{L}^{1/2}$ has compact resolvent, and zero is not in the spectrum of $\mathcal{L}^{1/2}$, so $\mathcal{L}^{-1}$ exists and is a compact operator on $L^2(\Omega, \Lambda^{0,q})$, for $1 \le q \le n$. Furthermore, Proposition 3.2 extends, so (3.18) holds for $u \in \mathcal{D}(\mathcal{L})$.

Then the higher-order a priori estimates of §4 and the regularity results of §5 extend, to yield the following:

Proposition 8.3. *For* $1 \le q \le n$,

$$\mathcal{L}^{-1} : H^j(\Omega, \Lambda^{0,q}) \longrightarrow H^{j+1}(\Omega, \Lambda^{0,q}), \quad j \ge 0, \tag{8.28}$$

and

$$\mathcal{L}^{-1} : \Lambda^{0,q}(\overline{\Omega}) \longrightarrow \Lambda^{0,q}(\overline{\Omega}). \tag{8.29}$$

Thus the material of §6 extends. We have the next result:

Proposition 8.4. *Given* $u \in \Lambda^{0,q}(\overline{\Omega}),\ q \ge 1$, *we have*

$$u = \overline{\partial}\,\overline{\partial}^*\mathcal{L}^{-1}u + \overline{\partial}^*\overline{\partial}\mathcal{L}^{-1}u = P_{\overline{\partial}}u + P_{\overline{\partial}^*}u. \tag{8.30}$$

The two terms on the right side are mutually orthogonal in $L^2(\Omega, \Lambda^{0,q})$. *Furthermore, for* $j \ge 1$,

$$P_{\overline{\partial}}, P_{\overline{\partial}^*} : H^j(\Omega, \Lambda^{0,q}) \longrightarrow H^{j-\varepsilon}(\Omega, \Lambda^{0,q}), \quad \forall\, \varepsilon > 0. \tag{8.31}$$

Corollary 8.5. *If* $q \ge 1$ *and* $f \in \Lambda^{0,q}(\overline{\Omega})$ *satisfies* $\overline{\partial} f = 0$, *then there exists* $u \in \Lambda^{0,q-1}(\overline{\Omega})$ *satisfying* $\overline{\partial}u = f$.

Note that there is no "cohomology" here. In the more general case of strongly pseudoconvex complex *manifolds*, which will be discussed in §10, there can perhaps be cohomology, arising from a nontrivial null space of $\mathcal{L}$ on $\Lambda^{0,q}(\overline{\Omega}),\ q \ge 1$.

We next echo some constructions of §A. We define vector bundles $E_j \to \overline{\Omega}$ by

$$E_0 = \bigoplus_{j\ge 0} \Lambda^{0,2j}, \quad E_1 = \bigoplus_{j \ge 0} \Lambda^{0,2j+1}, \quad E = E_0 \oplus E_1. \tag{8.32}$$

We then define the unbounded operator D_N on $L^2(\Omega, E)$ to be the closure of $\overline{\partial} + \overline{\partial}^*$, acting on $\bigoplus_{q\ge 0} \mathcal{D}^{0,q}$. As usual, $\mathcal{D}^{0,q}$ is as defined in (2.2); in particular, we have $\mathcal{D}^{0,0} = \Lambda^{0,0}(\overline{\Omega})$. Note that, for $q \ge 1$, $\mathcal{D}(D_N) \cap L^2(\Omega, \Lambda^{0,q})$ coincides with $\mathcal{D}(\mathcal{L}^{1/2})$, as defined in §3 for $q = 1$ and in this section for general $q \ge 1$. Also, the orthogonality relations imply that

$$\mathcal{D}(D_N) = \bigoplus_{q \ge 0} \mathcal{D}(D_N) \cap L^2(\Omega, \Lambda^{0,q}). \tag{8.33}$$

It is easily verified from Green's formula that D_N is symmetric; in fact, such arguments as needed for Exercise 1 in §3 imply that D_N is self-adjoint.

General considerations imply that the Friedrichs extension method, applied to the quadratic form $Q(u,u) = \|D_N u\|_{L^2}^2$ on $\mathcal{H}_1 = \mathcal{D}(D_N)$, gives rise precisely to the positive, self-adjoint operator $D_N^* D_N = D_N^2$. In view of the construction of the self-adjoint operator $\mathcal{L}$ on $L^2(\Omega, \Lambda^{0,q})$ discussed above, we have

$$D_N^2 = \mathcal{L} \quad \text{on each factor } \ L^2(\Omega, \Lambda^{0,q}), \tag{8.34}$$

for $q \geq 1$. In particular, we have the identity of the domains of these operators.

The operator D_N provides an example of the following structure. $D_N = D$ has the form

$$D = \begin{pmatrix} 0 & A^* \\ A & 0 \end{pmatrix}, \tag{8.35}$$

a self-adjoint operator on a Hilbert space $H = H_0 \oplus H_1$, where $A : H_0 \to H_1$ is a closed, densely defined operator. In the present case, $H_j = L^2(\Omega, E_j)$. Thus

$$D^2 = \begin{pmatrix} A^*A & 0 \\ 0 & AA^* \end{pmatrix}, \tag{8.36}$$

and our results on $\mathcal{L}$ imply that

$$(D^2 + 1)^{-1} \ \text{ is compact on } \ H_1. \tag{8.37}$$

Of course, $(D^2 + 1)^{-1}$ is not compact on H_0 in this case, since it coincides with the identity on $\mathcal{H}(\Omega) \subset L^2(\Omega, \Lambda^{0,0}) \subset H_0$, which is an infinite-dimensional space. There is another important property, namely that, for any $f \in C^\infty(\overline{\Omega})$, M_f preserves $\mathcal{D}(D_N)$ and

$$[M_f, D] \ \text{ extends to a bounded operator on } \ H_0 \oplus H_1. \tag{8.38}$$

Using these properties, we will establish the following, which, as we will see, complements Proposition 7.3. Set

$$F = D(D^2+1)^{-1/2} = \begin{pmatrix} 0 & T^* \\ T & 0 \end{pmatrix}, \quad T = A(A^*A+1)^{-1/2}. \tag{8.39}$$

Proposition 8.6. *The operator F has closed range, and for all $f \in C(\overline{\Omega})$, we have a compact commutator:*

$$[M_f, F] \in \mathcal{K}(H_0 \oplus H_1). \tag{8.40}$$

To establish (8.40), we may as well assume $f \in C^\infty(\overline{\Omega})$. Then we can write

$$[M_f, F] = [M_f, D](D^2+1)^{-1/2} + D[M_f, (D^2+1)^{-1/2}]. \tag{8.41}$$

It follows from (8.38) that the first term on the right is compact on H_1. Before looking at the last term on the right, we derive a result that gives some information on the behavior of the first term on H_0.

Lemma 8.7. *The operator $(D^2+1)^{-1/2}$ is compact on the orthogonal complement of ker D (in H_0).*

Proof. We are saying that $(A^*A+1)^{-1/2}$ is compact on the orthogonal complement of ker A in H_0. We will deduce this via the identity

$$Ag(A^*A) = g(AA^*)A \quad \text{on } \mathcal{D}(A), \tag{8.42}$$

which holds for any bounded, continuous function g on $[0,\infty)$. The identity (8.42) is a consequence of the identity

$$Dg(D^2) = g(D^2)D \quad \text{on } \mathcal{D}(D). \tag{8.43}$$

Another ingredient in the proof of the lemma is the following. Since AA^* has compact resolvent, H_1 has an orthonormal basis of eigenvectors for AA^*, and we have

$$\begin{aligned} A : \text{Eigen}(\lambda, A^*A) &\longrightarrow \text{Eigen}(\lambda, AA^*), \\ A^* : \text{Eigen}(\lambda, AA^*) &\longrightarrow \text{Eigen}(\lambda, A^*A). \end{aligned} \tag{8.44}$$

If $\lambda \neq 0$, these maps are inverses of each other, up to a factor λ, so they are isomorphisms.

To prove the lemma, we first show that

$$\varphi \in C_0^\infty(\mathbb{R}),\ \varphi(0)=0 \Longrightarrow \varphi(A^*A) \in \mathcal{K}(H_0). \tag{8.45}$$

To do this, write $\varphi(s) = s\varphi_1(s)\varphi_2(s)\varphi_3(s),\ \varphi_j \in C_0^\infty(\mathbb{R})$. Then, applying (8.42) with $g = \varphi_1\varphi_2$, we have

$$\begin{aligned} \varphi(A^*A) &= A^*A(\varphi_1\varphi_2)(A^*A)\varphi_2(A^*A) \\ &= A^*\varphi_1(AA^*)\varphi_2(AA^*)A\varphi_3(A^*A). \end{aligned} \tag{8.46}$$

Here, $A\varphi_3(A^*A) \in \mathcal{L}(H_0, H_1)$, $\varphi_2(AA^*) \in \mathcal{K}(H_1)$, and $A^*\varphi_1(AA^*) \in \mathcal{L}(H_1, H_0)$, so (8.45) follows. Consequently, the spectrum of A^*A, which is contained in $[0,\infty)$, is discrete on any compact interval in $(0,\infty)$, of finite multiplicity. It remains to show that this spectrum cannot accumulate at 0. Indeed, the argument involving (8.44) shows that

$$\{0\} \cup \text{ spec } A^*A = \{0\} \cup \text{ spec } AA^*, \tag{8.47}$$

and since AA^* has compact resolvent, its spectrum does not accumulate at 0, so Lemma 8.7 is proved.

To proceed with the proof of Proposition 8.6, we next show that the last term in (8.41) is compact on H_1 and on the orthogonal complement of ker D in H_0. One tool will be the integral representation

$$(8.48)\qquad \begin{aligned}(D^2+1)^{-1/2} &= \frac{1}{\pi}\int_0^\infty \lambda^{-1/2}(D^2+1+\lambda)^{-1}\,d\lambda \\ &= \frac{2}{\pi}\int_0^\infty (D^2+1+s^2)^{-1}\,ds.\end{aligned}$$

In order to get a convenient formula for $[M_f, (D^2+1+s^2)^{-1}]$, set $t = \sqrt{1+s^2}$ and write

$$(8.49)\qquad (D^2+t^2)^{-1} = (D+it)^{-1}(D-it)^{-1},$$

so

$$(8.50)\qquad \begin{aligned}[M_f, (D^2+t^2)^{-1}] = [M_f, (D+it)^{-1}](D-it)^{-1} \\ + (D+it)^{-1}[M_f, (D-it)^{-1}].\end{aligned}$$

Since, for $f \in C^\infty(\overline{\Omega})$, M_f preserves $\mathcal{D}(D)$, we have

$$(8.51)\qquad [M_f, (D+it)^{-1}] = -(D+it)^{-1}[M_f, D](D+it)^{-1}.$$

Hence

$$(8.52)\qquad \begin{aligned}[M_f, (D^2+t^2)^{-1}] = &-(D+it)[M_f, D](D^2+t^2)^{-1} \\ &-(D^2+t^2)[M_f, D](D-it)^{-1}.\end{aligned}$$

Therefore the last term in (8.41) is equal to

$$(8.53)\qquad \begin{aligned}&-\frac{2}{\pi}\int_0^\infty D\left(D+i\sqrt{1+s^2}\right)^{-1}[M_f, D](D^2+1+s^2)^{-1}\,ds \\ &-\frac{2}{\pi}\int_0^\infty D(D^2+1+s^2)^{-1}[M_f, D]\left(D-i\sqrt{1+s^2}\right)^{-1}\,ds \\ &= T_1 + T_2.\end{aligned}$$

In view of the operator norm estimates

$$\|D(D\pm it)^{-1}\| \le 1, \quad \|(D\pm it)^{-1}\| \le t^{-1},$$

for $t \geq 0$, it follows that both integrals in (8.53) are convergent in the operator norm. Thus $T_1 + T_2$ is compact on any closed subspace of $H_0 \oplus H_1$ on which the integrands are compact, for all $s \in [0, \infty)$.

The integrand for T_1 is a product of bounded operators with the factor $(D^2 + 1 + s^2)^{-1}$, which we know to be compact on H_1 and on $(\ker D)^{\perp}$, so T_1 is compact there. The integrand for T_2 is a product of bounded operators and $D(D^2 + 1 + s^2)^{-1}$. It follows from Lemma 8.7 that this factor is compact on all of $H_0 \oplus H_1$, so T_2 is compact on $H_0 \oplus H_1$.

To complete the proof of compactness of the commutator (8.40), it remains to show that this commutator is compact on $\ker D$, for $f \in C^\infty(\overline{\Omega})$. In such a case we can write, in place of (8.41),

(8.54)
$$\begin{aligned}[M_f, D(D^2+1)^{-1/2}] &= [M_f, (D^2+1)^{-1/2} D] \\ &= [M_f, (D^2+1)^{-1/2}]D + (D^2+1)^{-1/2}[M_f, D],\end{aligned}$$

on $\mathcal{D}(D)$. On $\ker D$, this is equal to

$$(D^2+1)^{-1/2}[M_f, D]. \tag{8.55}$$

Now $[M_f, D]$ maps H_0 to H_1, and $(D^2+1)^{-1/2}$ is compact on H_1, so (8.55) is compact on H_0. This completes the proof of the compactness assertion (8.40).

Finally, the proof of Lemma 8.7 shows that zero is an isolated point of spec D^2, hence of spec D, so D has closed range, and hence F has closed range. This completes the proof of Proposition 8.6.

Another consequence of Lemma 8.7 is that

$$F - F_0 \in \mathcal{K}, \quad F_0 = P^+ - P^-, \tag{8.56}$$

where P^+ is the orthogonal projection onto the closed linear span of the positive eigenspaces of D, and P^- is the orthogonal projection onto the closed linear span of the negative eigenspaces of D. Thus (8.40) is equivalent to

$$[M_f, F_0] \in \mathcal{K}, \tag{8.57}$$

for all $f \in C(\overline{\Omega})$. Note that

$$F_0^2 = P^+ + P^- = I - P^0, \tag{8.58}$$

where P^0 is the orthogonal projection onto $\ker D$. Since

$$[M_f, F_0^2] = [M_f, F_0]F_0 + F_0[M_f, F_0],$$

we have the following variant of (8.40):

Proposition 8.8. *For all $f \in C(\overline{\Omega})$, we have compact commutators:*

$$[M_f, P^+],\ [M_f, P^-],\ [M_f, P^0] \in \mathcal{K}(H_0 \oplus H_1). \tag{8.59}$$

In our present situation, P^0 preserves each factor $L^2(\Omega, \Lambda^{0,q})$ in $H_0 \oplus H_1$. In fact, P^0 is zero on all these spaces except $L^2(\Omega, \Lambda^{0,0})$, on which it is the Bergman projection. Thus the compactness of $[M_f, P^0]$ is equivalent to the compactness of $[M_f, B]$, established in Proposition 7.3.

The value of Propositions 8.6–8.8 as a complement to Proposition 7.3 is particularly revealed in its relevance to the index identity (7.34). We will give only a brief description of this connection here, referring to [BDT] for details. As shown in [BDT], the results in Propositions 8.6–8.8 imply that D_N determines a relative K-homology class, $[D_N] \in K_0(\overline{\Omega}, \partial\Omega)$, and that the K-homology class $[\tau] \in K_1(\partial\Omega)$ described in §7 is obtained from $[D_N]$ by applying a natural boundary map:

$$\partial : K_0(\overline{\Omega}, \partial\Omega) \longrightarrow K_1(\partial\Omega).$$

It is then shown in [BDT] that a certain identity in $K_0(\overline{\Omega}, \partial\Omega)$ leads, via the application of this boundary map, to the identity $[\tau] = [\pi] \in K_1(\partial\Omega)$ mentioned in §7, an identity that in turn implies the index identity (7.34).

Exercises

1. Extend the results of this section to

$$\overline{\partial} : C^\infty(\overline{\Omega}, V \otimes \Lambda^{0,q}) \longrightarrow C^\infty(\overline{\Omega}, V \otimes \Lambda^{0,q+1}),$$

for any finite-dimensional, complex vector space V. Deduce results for

$$\overline{\partial} : \Lambda^{p,q}(\overline{\Omega}) \longrightarrow \Lambda^{p,q+1}(\overline{\Omega}).$$

2. Establish an analogue of Proposition A.4 for the $\overline{\partial}$-complex.

9. Reduction to pseudodifferential equations on the boundary

In this section we reduce the $\overline{\partial}$-Neumann problem on $\overline{\Omega}$ to a system of equations on $\partial\Omega$. This method provides an alternative to the sort of analysis carried out in §§2–5. We consider the boundary problem

$$\begin{gathered} \Box u = 0 \ \text{ on } \ \Omega, \\ \sigma_{\overline{\partial}^*}(x,\nu)u = 0, \quad \sigma_{\overline{\partial}^*}(x,\nu)\overline{\partial}u = f \ \text{ on } \ \partial\Omega. \end{gathered} \tag{9.1}$$

We write u in terms of a solution to the Dirichlet problem:

$$\Box u = 0 \ \text{ on } \ \Omega, \quad u\Big|_{\partial\Omega} = g, \tag{9.2}$$

that is, $u = PI\ g$. Thus, g satisfies the equation

$$Ag = f, \tag{9.3}$$

where

$$Ag = \sigma_{\overline{\partial}^*}(x, \nu)\, \overline{\partial}(\mathrm{PI}\, g)\Big|_{\partial\Omega}, \tag{9.4}$$

and we require

$$\sigma_{\overline{\partial}^*}(x, \nu) g = 0. \tag{9.5}$$

Thus, we can regard A as a linear operator

$$A : C^\infty(\partial\Omega, E) \longrightarrow C^\infty(\partial\Omega, E), \tag{9.6}$$

where $E \to \partial\Omega$ is the complex vector bundle

$$E = \operatorname{Ker} \sigma_{\overline{\partial}^*}(x, \nu) = \operatorname{Im} \sigma_{\overline{\partial}^*}(x, \nu). \tag{9.7}$$

Here, if we are looking at $\Box$ on $(0, q)$-forms, the first $\sigma_{\overline{\partial}^*}(x, \nu)$ in (9.7) acts on $\Lambda^{0,q}\Big|_{\partial\Omega}$ and the second acts on $\Lambda^{0,q+1}\Big|_{\partial\Omega}$. The second identity in (9.7) follows from the exactness of the symbol sequence for the $\overline{\partial}$-complex.

For simplicity, we confine attention to the case $q = 1$, which was studied in §§2–6. Thus $u = \mathrm{PI}\, g$ has the form

$$u = \sum u_j\, d\overline{z}_j, \quad u_j = g_j \text{ on } \partial\Omega. \tag{9.8}$$

Say $\nu = \sum(\nu_j\, dx_j + \mu_j\, dy_j)$, so

$$\nu + iJ^t\nu = \sum \varphi_j\, d\overline{z}_j, \quad \varphi_j = \nu_j - i\mu_j. \tag{9.9}$$

The condition that $\sigma_{\overline{\partial}^*}(x, \nu) u = 0$ on $\partial\Omega$ is equivalent to

$$\sum \varphi_j g_j = 0. \tag{9.10}$$

A computation gives

$$\sigma_{\overline{\partial}^*}(x, \nu)\, \overline{\partial} u = \sum f_j\, d\overline{z}_j, \quad f_j = \frac{1}{4} \sum_k \varphi_k \left(\frac{\partial u_j}{\partial \overline{z}_k} - \frac{\partial u_k}{\partial \overline{z}_j} \right). \tag{9.11}$$

Note that $\sum \varphi_j f_j = 0$, a fact consistent with the second identity in (9.7).

To express (9.11) in the form (9.3), write

$$\frac{\partial u}{\partial \bar{z}_k} = \frac{1}{2}(Y_k + a_k \mathcal{N})g, \tag{9.12}$$

for $u = PI\ g$, where Y_k is a (complex) vector field tangent to $\partial\Omega$, $a_k \in C^\infty(\partial\Omega)$, and $\mathcal{N}$ is the Neumann operator for the Dirichlet problem:

$$\mathcal{N}g = \frac{\partial u}{\partial \nu}, \quad u = \mathrm{PI}\, g. \tag{9.13}$$

Thus we get $Ag = f$, that is, $f_j = A_j{}^k g_k$, with

$$8 f_j = \sum_k \varphi_k (Y_k + a_k \mathcal{N}) g_j - \sum_k \varphi_k (Y_j + a_j \mathcal{N}) g_k. \tag{9.14}$$

Note that $\sum_k \varphi_k g_k = 0 \Rightarrow \sum_k \varphi_k Y_j g_k = -\sum_k (Y_j \varphi_k) g_k$, for each j. Thus we can write the system as

$$8 f_j = (iY + a\mathcal{N}) g_j - a_j \sum_k \varphi_k \mathcal{N} g_k + \sum_k (Y_j \varphi_k) g_k, \tag{9.15}$$

with

$$iY = \sum_k \varphi_k Y_k, \quad a = \sum_k a_k \varphi_k. \tag{9.16}$$

Now, (9.12) implies that $\operatorname{Re} a_k = \nu_k$ and $\operatorname{Im} a_k = \mu_k$, or

$$a_k = \overline{\varphi}_k. \tag{9.17}$$

Also, of course, $1 = \sum (\nu_j^2 + \mu_j^2) = \sum |\varphi_j|^2$, so we have $a = 1$ in (9.16). Furthermore, if ρ is a defining function for $\overline{\Omega}$, as in §2 (so $\rho = 0$ and $|\nabla \rho| = 1$ on $\partial\Omega$, $\rho < 0$ on Ω), then $\varphi_k = -2\partial\rho/\partial z_k$ and, for all $v \in C^\infty(\overline{\Omega})$,

$$\sum \frac{\partial \rho}{\partial z_k} \frac{\partial v}{\partial \bar{z}_k} = \frac{1}{2} \langle \overline{\partial} v, \overline{\partial} \rho \rangle; \tag{9.18}$$

so, if $u_j = PI\ g_j$, we have

$$(iY + a\mathcal{N}) g_j = \langle \overline{\partial} u_j, \overline{\partial} \rho \rangle \Big|_{\partial\Omega}. \tag{9.19}$$

Fix $p \in \partial\Omega$, and rotate coordinates so $\partial\Omega$ is tangent to $\{y_n = c\}$ at p, and $\nabla \rho = -\partial/\partial y_n$ at p. Then $\overline{\partial}\rho = -(i/2)\, d\bar{z}_n$ at p, so we have

$$(iY + a\mathcal{N}) g_j = -i \frac{\partial g_j}{\partial x_n} + \mathcal{N} g_j. \tag{9.20}$$

We recapture the identity $a = 1$ and see that, in (9.15), Y is a real vector field tangent to $\partial\Omega$, namely

$$Y = -J(\nabla\rho). \tag{9.21}$$

Note that we can write (9.10), and the analogous result for f, as

$$\sum g_j \overline{a}_j = 0 = \sum f_j \overline{a}_j. \tag{9.22}$$

If we define $Q \in C^\infty(\partial\Omega, \mathrm{End}(\mathbb{C}^n))$ to be the orthogonal projection that annihilates $(a_1, \dots, a_n)$, then $Qg = g$ and $Qf = f$, so we can apply Q to (9.15), and write

$$8f = (\mathcal{N} + iY)g + Cg, \tag{9.23}$$

where $C = C_1 + C_2$, with

$$C_1 g = [Q, \mathcal{N} + iY]g, \quad (C_2 g)_j = \sum_{k,\ell} Q_j{}^\ell (Y_\ell \varphi_k)(Qg)_k. \tag{9.24}$$

Note that, for each $x \in \partial\Omega$, $\mathcal{R}(Q(x)) = E_x = \mathfrak{H}_x(\partial\Omega)$ (defined by (2.33)) in this case.

Thus we need to analyze the pseudodifferential operator

$$\Box^+ = \mathcal{N} + iY + C \in OPS^1(\partial\Omega), \tag{9.25}$$

which we claim to be hypoelliptic. The principal symbol is given by

$$\sigma_{\Box^+}(x, \xi) = -|\xi| + \tau(x, \xi), \quad \tau(x, \xi) = \langle Y, \xi \rangle, \tag{9.26}$$

which is ≤ 0 everywhere and vanishes to second order on the ray bundle generated by $J^t(d\rho)$, which we will denote as $\Sigma^+ \subset T^*(\partial\Omega) \setminus 0$. Thus Σ^+ is the characteristic set of $\Box^+$.

Since $\sigma_{\Box^+} = \sigma_{\mathcal{N}+iY}$ vanishes to second order on Σ^+, it follows that $C_1 = [Q, \mathcal{N} + iY] \in OPS^0(\partial\Omega)$ satisfies

$$\sigma_{C_1}(x, \xi) = 0 \quad \text{on } \Sigma^+. \tag{9.27}$$

It will turn out that this implies that the presence of C_1 does not affect the hypoellipticity of $\Box^+$.

The operator C_2 (also of order zero) requires further study. If we fix $p \in \partial\Omega$ and rotate coordinates so that $T_p(\partial\Omega)$ is given by $\{y_n = c\}$, then $Q(p)$ annihilates the last component of a vector in $\mathbb{C}^n$, and we have

$$C_2 g(p) = 2\left(\sum_{k\le n-1} \frac{\partial\varphi_k}{\partial\bar z_1} g_k, \dots, \sum_{k\le n-1} \frac{\partial\varphi_k}{\partial\bar z_{n-1}} g_k, 0\right). \tag{9.28}$$

Note that the Levi form arises here:

$$2\frac{\partial\varphi_k}{\partial\bar z_j} = -4\frac{\partial^2\rho}{\partial\bar z_j\,\partial z_k} = -4\mathcal{L}_{kj} \tag{9.29}$$

on $\partial\Omega$. Thus, for any $g, h \in L^2(\partial\Omega, \mathbb{C}^n)$, we have

$$(C_2 g, h)_{L^2} = -4\int_{\partial\Omega} \mathcal{L}(Qg, Qh)\, dS. \tag{9.30}$$

If g and h are sections of $E \to \partial\Omega$, we can omit the Qs in (9.30).

We have used the fact that $\mathcal{N} = -\sqrt{-\Delta_X}$ mod $OPS^0(\partial\Omega)$ in the symbol calculation (9.26), where Δ_X is the Laplace operator on $X = \partial\Omega$. We next make use of a finer analysis of $\mathcal{N}$, given in §C at the end of this chapter, which says

$$\mathcal{N} = -\sqrt{-\Delta_X} + B, \quad B \in OPS^0(\partial\Omega), \quad \sigma_B(x,\xi) = \frac{1}{2}\operatorname{Tr}(A_N P_\xi^0), \tag{9.31}$$

where A_N is the Weingarten map (arising from the second fundamental form of $\partial\Omega \subset \mathbb{R}^{2n}$) and P_ξ^0 is the orthogonal projection of $T_x(\partial\Omega)$ onto the linear subspace annihilated by ξ. Thus

$$-\Box^+ = \sqrt{-\Delta_X} - iY + B_1, \quad B_1 = -B - C \in OPS^0(\partial\Omega). \tag{9.32}$$

In (9.29) we have related the principal symbol of the most important part C_2 of C to the Levi form. If we use (B.20), we can write the principal symbol of B on Σ^+ as

$$\sigma_B(x,\eta) = 2\Big(\operatorname{Tr}\widehat{\mathcal{L}}\Big)I, \quad \eta = J^t(d\rho), \tag{9.33}$$

where we use the Hermitian metric on $\mathfrak{H}(\partial\Omega)$ to produce the section

$$\widehat{\mathcal{L}} \in C^\infty\Big(\partial\Omega, \operatorname{End}\mathfrak{H}\Big)$$

from the Levi form. Hence

$$\sigma_{B_1}(x,\xi) = -2\Big[(\operatorname{Tr}\widehat{\mathcal{L}})I - 2\widehat{\mathcal{L}}Q\Big] + \beta(x,\xi), \tag{9.34}$$

where $\beta(x,\xi)$ vanishes on Σ^+.

The action of $\square^+$ on sections of E is of major interest, but it is convenient to analyze $\square^+$ on general functions $u \in L^2(\partial\Omega, \mathbb{C}^n)$. To do this, it is convenient to replace $\widehat{\mathcal{L}}Q$ by

$$\widehat{\mathcal{L}}_Q = \widehat{\mathcal{L}}Q + \gamma(x)(I - Q),$$

where $\gamma(x)$ is a positive function to be specified later. Thus, we replace (9.34) by

$$\sigma_{B_1}(x,\xi) = -2\Big[(\mathrm{Tr}\,\widehat{\mathcal{L}})I - 2\widehat{\mathcal{L}}_Q\Big] + \beta(x,\xi). \tag{9.35}$$

From now on, we work with this modified $\square^+$.

The structure of $\square^+$ is to some degree simplified by composing on the left by $\square^-$, defined by

$$-\square^- = \sqrt{-\Delta_X} + iY + B_2, \tag{9.36}$$

where $B_2 \in OPS^0(\partial\Omega)$ will be specified shortly. Note the different sign in front of iY. Thus $\square^-$ is elliptic on Σ^+; its characteristic set is Σ^-, the ray bundle in $T^*(\partial\Omega) \setminus 0$ generated by $-J^t(d\rho)$. We have

$$\square^-\square^+ = -\Delta_X + Y^2 + F, \tag{9.37}$$

where $F \in OPS^1(\partial\Omega)$ is given by

$$F = \left(\sqrt{-\Delta_X} + iY\right)B_1 - i\left[\sqrt{-\Delta_X}, Y\right] + B_2\left(\sqrt{-\Delta_X} - iY\right) + B_2B_1. \tag{9.38}$$

Since

$$\pm i\left[\sqrt{-\Delta_X}, Y\right] = \left[\sqrt{-\Delta_X}, \sqrt{-\Delta_X} \pm iY\right]$$

and $\sqrt{-\Delta_X} \pm iY$ is *doubly* characteristic on $\Sigma^{\mp}$, we see that

$$\sigma_{i[\sqrt{-\Delta_X}, Y]}(x,\xi) = 0 \quad \text{on } \Sigma^+ \cup \Sigma^-. \tag{9.39}$$

Now $\tau(x,\xi) = \pm|\xi|$ on $\Sigma^{\pm}$. Consequently, the principal symbol of F satisfies

$$\sigma_F(x,\xi) = \begin{Bmatrix} 2|\xi|\sigma_{B_1}(x,\xi) & \text{on } \Sigma^+ \\ 2|\xi|\sigma_{B_2}(x,\xi) & \text{on } \Sigma^- \end{Bmatrix}. \tag{9.40}$$

Thus, if B_2 is chosen so that

$$\sigma_{B_2}(x,\xi) = 2\Big[(\mathrm{Tr}\,\widehat{\mathcal{L}})I - 2\widehat{\mathcal{L}}_Q\Big] \quad \text{on } \Sigma^-, \tag{9.41}$$

then

$$\sigma_F(x,\xi) = -4\Big[(\mathrm{Tr}\,\widehat{\mathcal{L}})I - 2\widehat{\mathcal{L}}_Q\Big]\tau(x,\xi) \quad \text{on } \Sigma = \Sigma^+ \cup \Sigma^-. \tag{9.42}$$

If $\square^-$ is constructed in this fashion, we have

$$\square^-\square^+ = -\Delta_X + Y^2 - i\alpha(x)Y + R = \square_b + R, \tag{9.43}$$

with

$$\alpha(x) = 4\Big[(\mathrm{Tr}\,\widehat{\mathcal{L}})I - 2\widehat{\mathcal{L}}_Q\Big], \quad R \in OPS^1(\partial\Omega),\ \sigma_R(x,\xi) = 0 \text{ on } \Sigma. \tag{9.44}$$

As we have suggested, R will play a minor role in the analysis. Now the operator $\square_b = -\Delta_X + Y^2 + i\alpha(x)Y$ is a second-order *differential* operator, doubly characteristic on $\Sigma = \Sigma^+ \cup \Sigma^-$. It is essentially the "Kohn Laplacian" on $\partial\Omega$.

We now derive an analogue of the "$1/2$-estimate" for $\square_b + R$. In the analysis, R (or perhaps R' or R_j) will denote an arbitrary element of $OPS^1(\partial\Omega)$ (sometimes a differential operator) whose principal symbol vanishes on Σ; it might vary from line to line. We begin with an estimate on $\Big((\square_b + R)u, u\Big)_{L^2}$ when u is supported on an open set $\mathcal{O} \subset \partial\Omega$ diffeomorphic to a ball in $\mathbb{R}^{2n-1}$. Let $\{X_j : 1 \le j \le 2n\}$ be a smooth, orthonormal frame field for $\mathfrak{H}(\partial\Omega)$ over $\mathcal{O}$, such that $X_{n+j} = JX_j$.

Lemma 9.1. *Assume $\widehat{\mathcal{L}}$ is positive-definite on $\mathfrak{H}_x(\partial\Omega)$, for all $x \in \partial\Omega$. Also assume $n \ge 3$. If $u \in C_0^\infty(\mathcal{O}, \mathbb{C}^n)$, then, for some $C_j > 0$,*

$$\mathit{Re}\Big((\square_b + R)u, u\Big)_{L^2} \ge C_0 \sum \|X_j u\|^2_{L^2} + C_0\|u\|^2_{H^{1/2}} - C_1\|u\|^2_{L^2}. \tag{9.45}$$

Proof. Note that

$$-\Delta_X + Y^2 = -\sum_{j=1}^{2n} X_j^2 + R. \tag{9.46}$$

Now set

$$Z_j = X_j - iX_{j+n}, \quad \overline{Z}_j = X_j + iX_{j+n}. \tag{9.47}$$

We have

$$\begin{aligned} Z_j\overline{Z}_j &= X_j^2 + X_{j+n}^2 - \frac{1}{2}[Z_j, \overline{Z}_j], \\ \overline{Z}_j Z_j &= X_j^2 + X_{j+n}^2 + \frac{1}{2}[Z_j, \overline{Z}_j]. \end{aligned} \tag{9.48}$$

If we use (B.7), and recall that $Y = -J(\nabla\rho)$, we have

$$\frac{1}{2}[Z_j, \overline{Z}_j] = -i\beta_j(x)Y + R_j, \quad \beta_j(x) = 4\mathcal{L}(X_j, X_j) = 4\langle\widehat{\mathcal{L}}X_j, X_j\rangle. \tag{9.49}$$

(The factor of 4 arises from a slightly different definition of Z_j and $\overline{Z}_j$ in (B.10).) Note also that

$$Z_j \overline{Z}_j = -Z_j Z_j^* + R_j, \quad \overline{Z}_j Z_j = -Z_j^* Z_j + R_j'. \tag{9.50}$$

Hence, since $\sum \beta_j = 4 \operatorname{Tr} \widehat{\mathcal{L}}$,

$$\begin{aligned} -\Delta_X + Y^2 &= \sum Z_j Z_j^* + 4i(\operatorname{Tr} \widehat{\mathcal{L}})Y + R_1 \\ &= \sum Z_j^* Z_j - 4i(\operatorname{Tr} \widehat{\mathcal{L}})Y + R_2. \end{aligned} \tag{9.51}$$

Thus,

$$\begin{aligned} \Box_b + R &= \sum Z_j Z_j^* + iA_1(x)Y + R_1 \\ &= \sum Z_j^* Z_j - iA_2(x)Y + R_2, \end{aligned} \tag{9.52}$$

where

$$A_1(x) = 8\widehat{\mathcal{L}}_Q, \quad A_2(x) = 8\Big((\operatorname{Tr} \widehat{\mathcal{L}})I - \widehat{\mathcal{L}}_Q\Big). \tag{9.53}$$

Recall that $\widehat{\mathcal{L}}_Q = \widehat{\mathcal{L}}Q + \gamma(x)(I - Q)$. As long as $\gamma(x) > 0$, $A_1(x)$ is a positive-definite matrix function. Also, as long as $n \geq 3$, $\operatorname{Tr} \widehat{\mathcal{L}}$ exceeds any single eigenvalue of $\widehat{\mathcal{L}}$, so we can pick $\gamma(x) > 0$ small enough that $A_2(x)$ is also a positive-definite matrix function.

Given that $A_1(x)$ and $A_2(x)$ are positive-definite, we want to take a "convex combination" of the two expressions on the right side of (9.52) and obtain an expression for which the estimate (9.45) is obvious. Let $\varphi_j(x, \xi) \in S^0$ be real-valued and satisfy

$$\varphi_j(x, \xi) \geq \delta > 0, \quad \varphi_1(x, \xi)^2 + \varphi_2(x, \xi)^2 = 1. \tag{9.54}$$

Then the operators $\varphi_j(x, D)$ are elliptic, and we have

$$\begin{aligned} \Box_b + R = \varphi_1(x, D)^* &\sum Z_j Z_j^* \varphi_1(x, D) \\ &+ \varphi_2(x, D)^* \sum Z_j^* Z_j \varphi_2(x, D) \\ &+ i\Big[A_1(x)\varphi_1(x, D) - A_2(x)\varphi_2(x, D)\Big]Y + R'. \end{aligned} \tag{9.55}$$

Now the operator

$$V = i\Big[A_1(x)\varphi_1(x, D) - A_2(x)\varphi_2(x, D)\Big]Y \tag{9.56}$$

has symbol on Σ given by

$$\sigma_V(x,\xi) = \pm\Big[A_1(x)\varphi(x,\xi) - A_2(x)\varphi_2(x,\xi)\Big]|\xi| \quad \text{on } \Sigma^{\pm}. \tag{9.57}$$

Given that $A_1(x)$ is positive-definite, we see that $\sigma_V(x,\xi)$ is positive-definite on Σ^+ as long as $\varphi_2(x,\xi)$ is sufficiently small on Σ^+; similarly, if $A_2(x)$ is positive-definite, then $\sigma_V(x,\xi)$ is positive-definite on Σ^- as long as $\varphi_1(x,\xi)$ is sufficiently small on Σ^-.

Thus, under the hypotheses of Lemma 9.1, we can arrange

$$\sigma_V(x,\xi) \geq c_0|\xi|I \quad \text{on } \Sigma, \quad c_0 > 0. \tag{9.58}$$

Now we can write

$$V = W + R, \quad \sigma_W(x,\xi) \geq c_0|\xi|I \quad \text{on } T^*\mathcal{O}\setminus 0, \tag{9.59}$$

and deduce from (9.55) that

$$\begin{aligned}\operatorname{Re}\Big((\Box_b + R)u, u\Big)_{L^2} &\\ = \sum\Big\{\|Z_j^*\varphi_1(x,D)u\|_{L^2}^2 &+ \|Z_j\varphi_2(x,D)u\|_{L^2}^2\Big\} \\ &+ \operatorname{Re}(Wu,v)_{L^2} + \operatorname{Re}(Ru,u)_{L^2},\end{aligned} \tag{9.60}$$

Gårding's inequality implies

$$\operatorname{Re}(Wu,u)_{L^2} \geq C_0\|u\|_{H^{1/2}}^2 - C_1\|u\|_{L^2}^2. \tag{9.61}$$

If we note that $[Z_j,\varphi_1(x,D)]$ and $[Z_j^*,\varphi_2(x,D)]$ belong to $OPS^0(\partial\Omega)$ and use elliptic estimates, we see that the sum over j on the right side of (9.60) is

$$\geq C_0\sum_{j=1}^{2n}\|X_ju\|_{L^2}^2 - C_1\|u\|_{L^2}^2. \tag{9.62}$$

Finally, given $R \in OPS^1(\partial\Omega)$, $\sigma_R(x,\xi) = 0$ on Σ, we can write

$$R = \sum S_jX_j + S_0, \quad S_j \in OPS^0(\partial\Omega) \tag{9.63}$$

and obtain

$$|(Ru,u)| \leq C\sum\|X_ju\|_{L^2}\|u\|_{L^2} + C\|u\|_{L^2}^2. \tag{9.64}$$

From these estimates, we have (9.45).

Note that since $\operatorname{Re}\Big((\Box_b + R)u, u\Big)_{L^2} \le \|u\|_{H^{1/2}}\|(\Box_b + R)u\|_{H^{-1/2}}$, the inequality (9.45) implies the estimate

$$\sum \|X_j u\|_{L^2}^2 + \|u\|_{H^{1/2}}^2 \le C\|(\Box_b + R)u\|_{H^{-1/2}}^2 + C\|u\|_{L^2}^2, \tag{9.65}$$

for $u \in C_0^\infty(\mathcal{O}, \mathbb{C}^n)$.

We can localize the estimate (9.65) as follows. Given $\psi_0 \in C_0^\infty(\mathcal{O})$, we see that

$$\psi_0(x)(\Box_b + R) - (\Box_b + R)\psi_0(x) = R'. \tag{9.66}$$

Assuming $\psi_j \in C_0^\infty(\mathcal{O}),\ \psi_{j+1} = 1$ on supp ψ_j, we have

$$\psi_0(\Box_b + R) - \psi_1(\Box_b + R)\psi_0 = \psi_1 R'\psi_2 + \psi_1[\psi_2, R']. \tag{9.67}$$

Applying (9.65) with u replaced by $\psi_0 u$, we then have

$$\begin{aligned} &\sum \|X_j(\psi_0 u)\|_{L^2}^2 + \|\psi_0 u\|_{H^{1/2}}^2 \\ &\quad \le C\|\psi_0(\Box_b + R)u\|_{H^{-1/2}}^2 + C\sum \|X_j(\psi_2 u)\|_{H^{-1/2}}^2 \\ &\qquad + C\|\psi_0 u\|_{L^2}^2 + C\|u\|_{H^{-1}}^2, \end{aligned} \tag{9.68}$$

for $u \in C^\infty(\partial\Omega, \mathbb{C}^n)$.

If we cover $\partial\Omega$ by a finite collection of open sets $\mathcal{O}_\nu$, diffeomorphic to balls, and sum the resulting estimates, we obtain a global estimate of the form

$$\|\nabla_{\mathfrak{H}} u\|_{L^2}^2 + \|u\|_{H^{1/2}}^2 \le C\|(\Box_b + R)u\|_{H^{-1/2}}^2 + C\|\nabla_{\mathfrak{H}} u\|_{H^{-1/2}}^2 + C\|u\|_{L^2}^2, \tag{9.69}$$

for all $u \in C^\infty(\partial\Omega, \mathbb{C}^n)$, where, for each component u_j of u, $\nabla_{\mathfrak{H}} u_j(x)$ is the orthogonal projection of $\nabla u_j(x) \in \mathbb{C} \otimes T_x(\partial\Omega)$ onto $\mathbb{C} \otimes_{\mathbb{R}} \mathfrak{H}_x(\partial\Omega)$, so $\nabla_{\mathfrak{H}} u(x) \in \mathbb{C}^n \otimes_{\mathbb{R}} \mathfrak{H}_x(\partial\Omega)$. We can write

$$\|\nabla_{\mathfrak{H}} u\|_{H^{-1/2}}^2 \le \varepsilon\|\nabla_{\mathfrak{H}} u\|_{L^2}^2 + C(\varepsilon)\|\nabla_{\mathfrak{H}} u\|_{H^{-1}}^2 \tag{9.70}$$

and absorb the term $\varepsilon\|\nabla_{\mathfrak{H}} u\|_{L^2}^2$, obtaining

$$\|\nabla_{\mathfrak{H}} u\|_{L^2}^2 + \|u\|_{H^{1/2}}^2 \le C\|(\Box_b + R)u\|_{H^{-1/2}}^2 + C\|u\|_{L^2}^2. \tag{9.71}$$

We can obtain higher-order estimates as follows. With $\Lambda = \sqrt{-\Delta_X + I}$, we have

$$\Lambda^k(\Box_b + R) - (\Box_b + R)\Lambda^k = R'\Lambda^k, \tag{9.72}$$

so applying (9.71) with u replaced by $\Lambda^k u$ yields

$$\begin{aligned}(9.73)\qquad &\|\nabla_{\mathfrak{H}}\Lambda^k u\|^2_{L^2} + \|\Lambda^k u\|^2_{H^{1/2}} \\ &\le C\|\Lambda^k(\Box_b + R)u\|^2_{H^{-1/2}} + C\|R'\Lambda^k u\|^2_{H^{-1/2}} + C\|\Lambda^k u\|^2.\end{aligned}$$

Now we have $\|R'\Lambda^k u\|^2_{H^{-1/2}} \le C\|\nabla_{\mathfrak{H}}\Lambda^k u\|^2_{H^{-1/2}} + C\|\Lambda^k u\|^2_{H^{-1/2}}$, so standard methods yield the a priori estimate

$$(9.74)\qquad \|\nabla_{\mathfrak{H}} u\|^2_{H^k} + \|u\|^2_{H^{k+1/2}} \le C\|(\Box_b + R)u\|^2_{H^{k-1/2}} + C\|u\|^2_{L^2},$$

for all $u \in C^\infty(\partial\Omega, \mathbb{C}^n)$.

We can go from the estimate (9.74) to the associated regularity theorem:

Theorem 9.2. *If Ω is a bounded, strongly pseudoconvex domain in $\mathbb{C}^n$ and $n \ge 3$, then the operator $\Box_b + R$ given by (9.43)–(9.44) is hypoelliptic. If $u \in L^2(\partial\Omega, \mathbb{C}^n)$, then, for any $s \in [-1/2, \infty)$,*

$$(9.75)\qquad (\Box_b + R)u \in H^s \Longrightarrow u \in H^{s+1}.$$

Proof. We first establish the local version of (9.75), for $s = k - 1/2,\ k = 0, 1, 2,$ and so on. Let $\mathcal{O}$ be a coordinate patch on $\partial\Omega$, identified with a ball in $\mathbb{R}^{2n+1}$. Take $\psi_j \in C_0^\infty(\mathcal{O})$ such that $\psi_{j+1} = 1$ on supp ψ_j. Let $\varphi \in C_0^\infty(\mathbb{R}^{2n+1})$ satisfy $\varphi(\xi) = 1$ for $|\xi| \le 1,\ 0$ for $|\xi| \ge 3/2$. Consider the following families of operators, for $\varepsilon \in (0, 1]$:

$$(9.76)\qquad J_\varepsilon u = \psi_1(x)\varphi(\varepsilon D)\psi_0(x)u, \quad K_\varepsilon u = \psi_3(x)\varphi(2^{-1}\varepsilon D)\psi_2(x)u.$$

We have

$$(9.77)\qquad J_\varepsilon,\ K_\varepsilon \ \text{bounded in}\ OPS^0_{1,0}, \quad O(\varepsilon^{-k})\ \text{in}\ OPS^{-k}_{1,0},$$

and

$$(9.78)\qquad K_\varepsilon J_\varepsilon - J_\varepsilon \ \text{bounded in}\ OPS^{-\infty}_{1,0}.$$

The formula (9.46) yields

$$(9.79)\quad K_\varepsilon(\Box_b + R) - (\Box_b + R)K_\varepsilon = -\sum [K_\varepsilon, X_j^2] - i[K_\varepsilon, \alpha Y] + [K_\varepsilon, R'],$$

which is equal to

$$(9.80)\qquad \sum B_{j\varepsilon} R_j, \quad B_{j\varepsilon}\ \text{bounded in}\ OPS^0_{1,0},\ O(\varepsilon^{-k})\ \text{in}\ OPS^{-k}_{1,0}.$$

Also we have

$$B_{j\varepsilon}R_j J_\varepsilon \text{ bounded in } OPS^{-\infty}_{1,0}. \tag{9.81}$$

Now, we apply (9.74) with u replaced by $J_\varepsilon u$, to get

$$\begin{aligned}\|\nabla_{\mathfrak{H}} J_\varepsilon u\|^2_{H^k} &+ \|J_\varepsilon u\|^2_{H^{k+1/2}} \\ &\le C\|K_\varepsilon(\Box_b + R)u\|^2_{H^{k-1/2}} \\ &\quad + C\sum \|B_{j\varepsilon}R_j J_\varepsilon u\|^2_{H^{k-1/2}} + C\|J_\varepsilon u\|^2_{L^2} \\ &\le C\|(\Box_b + R)u\|^2_{H^{k-1/2}} + C\|u\|^2_{L^2}.\end{aligned} \tag{9.82}$$

Passing to the limit $\varepsilon \to 0$ yields the local version of the regularity result (9.75), for $s = k - 1/2,\ k = 0, 1, 2, \dots$, and the result (9.75) for general $s \in [-1/2, \infty)$ can be deduced via an interpolation argument.

We are now in a position to prove the main result of this section.

Theorem 9.3. *If Ω is a bounded, strongly pseudoconvex domain in $\mathbb{C}^n$, and $n \ge 2$, then the operator $\Box^+$ given by (9.25) is hypoelliptic. If $u \in L^2(\partial\Omega, \mathbb{C}^n)$, then, for any $s \in [1/2, \infty)$,*

$$\Box^+ u \in H^s \Longrightarrow u \in H^s. \tag{9.83}$$

Proof. If $n \ge 3$, this is immediate from Theorem 9.2. It remains to deal with the case $n = 2$.

What happens to the argument involving $\Box^-\Box^+ = \Box_b + R$ when $n = 2$ is that we cannot pick $\varphi_j(x, \xi)$ to satisfy (9.54) and arrange that $\sigma_V(x, \xi)\big|_\Sigma$, given by (9.57), be $\ge c_0|\xi|$, with $c_0 > 0$. The reason is that A_2 is not positive definite. Recall from (9.53) that $A_2(x) = 8\Big((\mathrm{Tr}\,\widehat{\mathcal{L}})I - \widehat{\mathcal{L}}_Q\Big)$, so if $n = 2$, zero must be an eigenvalue of $A_2(x),\ \forall\ x \in \partial\Omega$. This makes it impossible to make $\sigma_V(x, \xi) \ge c_0|\xi|$ on Σ^-. However, we can still arrange that $\sigma_V(x, \xi) \ge c_0|\xi|$ on Σ^+. In fact, for this we can just take $\varphi_1(x, \xi) = 1 - \delta,\ \varphi_2(x, \xi) = \delta$, for small, positive δ.

To fix $\Box_b + R$, we merely alter it on a small conic neighborhood of Σ^-. Set

$$\widetilde{\Box}_b = \Box_b + R + S, \tag{9.84}$$

where $S \in OPS^1,\ S(x, \xi)$ is supported on a small conic neighborhood of Σ^-, and furthermore $S(x, \xi)\big|_{\Sigma^-} \ge c_1|\xi|$, for sufficiently large, positive c_1. Then the arguments used to prove Lemma 9.1 and Theorem 9.2, starting with (9.58)–(9.60), show that $\widetilde{\Box}_b$ is hypoelliptic and

$$\widetilde{\Box}_b u \in H^s \Longrightarrow u \in H^{s+1}. \tag{9.85}$$

Since $\widetilde{\Box}_b$ is equal to $\Box^-\Box^+$ microlocally near Σ^+ and $\Box^+$ is elliptic away from Σ^+, this is enough to complete the proof of Theorem 9 .3.

Though we will not pursue details, we mention that there are constructions of parametrices for various classes of hypoelliptic operators with double characteristics which include the operators studied above. Constructions making use of Fourier integral operators are given in [B1, Sj], and [Tr2]. Another approach makes use of analysis on the Heisenberg group; this was initiated in [FS] and pursued in a number of papers, including [BG, BGS, BS, D, Gel, GS, RS], and [Tay]. The Heisenberg group approach provides a convenient way to obtain estimates in other function spaces, such as L^p-Sobolev spaces and Hölder spaces, on which results were first obtained, by other methods, in [Ker].

The reduction of the $\overline{\partial}$-Neumann problem to the study of $\Box^+$ was made in [GS], for (0,1)-forms, on a strongly pseudoconvex manifold, with a special metric, typically different from the Euclidean metric, called a Levi metric. A special property of a Levi metric is that $\widehat{\mathcal{L}}$, arising as in (9.33) and (9.34), is a multiple of the identity. A reduction of the $\overline{\partial}$-Neumann problem for (0,1)-forms, with a general metric, was made in [Cha]. The analogous study for $(0, q)$-forms, $q \geq 1$, is made in [LR], for a Levi metric, and in [BS] for a general metric. In these analyses, $\overline{\Omega}$ can be a general strongly pseudoconvex complex manifold, not necessarily a domain in $\mathbb{C}^n$. In §10 we will derive estimates for the $\overline{\partial}$-Neumann problem on such manifolds, via the sort of energy-estimate approach used in §§2–5 and 8. The details of the reduction to the boundary made in this section would have to be modified to treat the more general situation, since we made use of the fact that $\Box$ is (a constant multiple of) the Laplace operator, acting componentwise, on forms on a domain in $\mathbb{C}^n$, with its standard flat metric.

While we have emphasized C^∞ regularity, there are also results on the analytic regularity of solutions to the $\overline{\partial}$-Neumann problem when $\partial\Omega$ is real analytic and strongly pseudoconvex, in [Tar, Tr1], and [Gel], the latter two making use of analytic pseudodifferential operator calculi on $\partial\Omega$.

Exercises

1. Work out the formula for $\Box^+$ when Ω is the unit ball in $\mathbb{C}^n$, using (C.29), with $m = 2n$.
2. Extend the results of Theorems 9.2 and 9.3 to all s. (*Hint*: For any invertible, elliptic $\Lambda^\sigma \in OPS^\sigma(\partial\Omega)$ with scalar principal symbol, write
$$\Lambda^{-\sigma}(\Box_b + R)\Lambda^\sigma = \Box_b + R'.)$$
3. Show that the regularity results of §5 follow from Theorem 9.3. (*Hint*: If U solves $\Box U = F$, with homogeneous $\overline{\partial}$-Neumann boundary conditions of the form (1.12), write $U = U_0 + u$, where U_0 solves the Dirichlet problem
$$\Box U_0 = F \text{ on } \Omega, \quad U_0\Big|_{\partial\Omega} = 0,$$
and u solves (9.1), with
$$f = -\sigma_{\overline{\partial}^*}(x, \nu)\overline{\partial} U_0\Big|_{\partial\Omega}.)$$
4. Show that under the hypotheses of Theorem 9.3 (but with no restriction on $s \in \mathbb{R}$),

$$\Box^+ u \in H^s(\partial\Omega) \Longrightarrow u \in H^s(\partial\Omega) \text{ and } \nabla_{\mathfrak{H}} u \in H^{s-\frac{1}{2}}(\partial\Omega). \tag{9.86}$$

Let $v = \mathrm{PI}\, u$. If (9.86) holds, for some $s \geq -1/2$, then

$$v \in H^{s+1/2}(\Omega).$$

5. Let P be a first-order differential operator with constant coefficients, acting on $\Lambda^{0,1}(\overline{\Omega})$. Suppose

$$P v\Big|_{\partial\Omega} = A(x)\frac{\partial v}{\partial \nu} + P_{\tan} v, \tag{9.87}$$

where ν is the inward normal, and suppose

$$-A(x)|\xi| + \sigma_{P_{\tan}}(x,\xi) \ \text{ vanishes on } \ \Sigma^+. \tag{9.88}$$

Show that if v is as in Exercise 4, then

$$P v \in H^s(\Omega). \tag{9.89}$$

(*Hint*: Pv is harmonic on Ω. Write $P v\Big|_{\partial\Omega} = w = A(x)\mathcal{N}u + P_{\tan} u$, and show that $w \in H^{s-1/2}(\partial\Omega)$.)

6. Suppose v satisfies the hypotheses of Exercise 4 and, in addition, that $u = v\Big|_{\partial\Omega}$ is a section of E, so that $Qu = u$. Show that the conclusion (9.89) of Exercise 5 still holds, when the hypothesis (9.88) is weakened to

$$\Big(-A(x)|\xi| + \sigma_{P_{\tan}}(x,\xi)\Big) Q(x) \ \text{vanishes on } \ \Sigma^+. \tag{9.90}$$

7. Show that Exercise 6 applies to $\overline{\partial}^* v$. Using this, establish (6.19).

10. The $\overline{\partial}$-equation on complex manifolds and almost complex manifolds

Let $\overline{M}$ be a compact C^∞-manifold with boundary. We can assume $\overline{M}$ is contained in a smooth manifold $\mathcal{O}$ without boundary, such that the interior M is open in $\mathcal{O}$. An almost complex structure on $\overline{M}$ is a smooth section J of $\mathrm{End}(T\overline{M})$ such that $J^2 = -I$. If there is such a structure, the real dimension of M must be even, say $k = 2n$. Thus, for $p \in \overline{M}$, $T_p\overline{M}$, regarded as a complex vector space, has complex dimension n.

A (0, 1)-form on $\overline{M}$ is a section of the complexified cotangent bundle $\mathbb{C}T^*\overline{M}$ of the form

$$\alpha = \beta - iJ^t\beta, \tag{10.1}$$

where β is a section of $T^*\overline{M}$, and $J^t : T_p^*\overline{M} \to T_p^*\overline{M}$ is the adjoint of J. Similarly, a $(1,0)$-form on $\overline{M}$ has the form $\alpha' = \beta + iJ^t\beta$. We have vector bundles $\Lambda^{0,1}\overline{M}$ and $\Lambda^{1,0}\overline{M}$, and clearly $\mathbb{C}T^*\overline{M} = \Lambda^{0,1}\overline{M} \oplus \Lambda^{1,0}\overline{M}$. An obvious algebraic procedure yields subbundles $\Lambda^{p,q}\overline{M}$ of $\mathbb{C}\Lambda^r T^*\overline{M}$, $r = p+q$, and

$$\mathbb{C}\Lambda^r T^*\overline{M} = \bigoplus_{p+q=r} \Lambda^{p,q}\overline{M}. \tag{10.2}$$

We also use $\Lambda^{p,q}(\overline{M})$ to denote the spaces of C^∞-sections of these bundles.

Let Π_{pq} denote the natural projection of $\mathbb{C}\Lambda^r T^*\overline{M}$ onto $\Lambda^{p,q}\overline{M}$. We define the operators

$$\overline{\partial} : \Lambda^{p,q}(\overline{M}) \longrightarrow \Lambda^{p,q+1}(\overline{M}), \quad \partial : \Lambda^{p,q}(\overline{M}) \longrightarrow \Lambda^{p+1,q}(\overline{M}) \tag{10.3}$$

by

$$\overline{\partial} u = \Pi_{p,q+1}\, du, \quad \partial u = \Pi_{p+1,q}\, du. \tag{10.4}$$

The basic case of an almost complex manifold is a complex manifold. For example, we say $\mathcal{O}$ is a complex manifold if it has a covering by coordinate charts (into open subsets of $\mathbb{C}^n$) whose transition maps are holomorphic. Then $\overline{M} \subset \mathcal{O}$ is a complex manifold with boundary. In such a case, $\overline{\partial}$ and ∂ are as defined in §1 (in any local holomorphic coordinate patch), and the following relations hold:

$$\overline{\partial}^2 = 0, \quad \partial^2 = 0, \quad d = \partial + \overline{\partial}. \tag{10.5}$$

These relations need not hold for an arbitrary, almost complex manifold, but it is readily verified that if any one of them holds, so do the other two. In such a case, the almost complex structure is said to be *integrable*. Thus, for a complex manifold, the almost complex structure is integrable. The converse to this is known as the Newlander–Nirenberg theorem; any manifold with an integrable, almost complex structure has a holomorphic coordinate chart. We will say more about this later in this section. There is also a direct characterization of the integrability condition in terms of J, which we will not need for our analysis, but we will mention it in the exercises.

In the rest of this section, we assume $\overline{M}$ has an integrable, almost complex structure. As in earlier sections, we are interested in the equation

$$\overline{\partial} u = f \quad \text{on } \overline{M}, \tag{10.6}$$

for $u \in \Lambda^{0,q}(\overline{M})$, given $f \in \Lambda^{0,q+1}(\overline{M})$, satisfying $\overline{\partial} f = 0$. For this to lead to a $\overline{\partial}$-Neumann problem, we need an operator

$$\overline{\partial}^* : \Lambda^{p,q+1}(\overline{M}) \longrightarrow \Lambda^{p,q}(\overline{M}), \tag{10.7}$$

a formal adjoint of $\overline{\partial}$. We assume $\overline{M}$ has a Riemannian metric with the property that $J : T_p\overline{M} \to T_p\overline{M}$ is an isometry, which can be obtained from an arbitrary Riemannian metric by averaging over the action of $\{I, J, J^2, J^3\}$.

The Riemannian metric yields both a volume element on $\overline{M}$ and a Hermitian metric on $\mathbb{C}\Lambda^* T^*\overline{M}$, hence inner products $(u, v)_{L^2}$ for $u, v \in \Lambda^{p,q}(\overline{M})$, and

Hilbert spaces $L^2(\Omega, \Lambda^{p,q})$, as well as the operator $\overline{\partial}^*$. As in (2.2), we can define

$$\mathcal{D}^{p,q} = \{u \in \Lambda^{p,q}(\overline{M}) : \sigma_{\overline{\partial}^*}(x, \nu)u = 0 \text{ on } \partial M\} \tag{10.8}$$

and then consider

$$Q(u, v) = (\overline{\partial}u, \overline{\partial}v)_{L^2} + (\overline{\partial}^* u, \overline{\partial}^* v)_{L^2}, \quad u, v \in \mathcal{D}^{0,q}. \tag{10.9}$$

We can define the subbundle $\mathfrak{H}(\partial M)$ of $T(\partial M)$ in a fashion similar to that done in §2; given $p \in \partial M$,

$$\mathfrak{H}_p(\partial M) = \{X \in T_p(\partial M) : JX \in T_p(\partial M)\}. \tag{10.10}$$

We can define the notion of a strongly pseudoconvex manifold, though it is more convenient to use an approach of §B than that of §2. Suppose $\rho \in C^\infty(\overline{M}),\ \rho < 0$ on $M,\ \rho = 0$ on $\partial M,\ |\nabla\rho| = 1$ on ∂M. Then, as in (B.14)–(B.15), define the Levi form as a quadratic form on $\mathfrak{H}(\partial M)$ by

$$\mathcal{L}(X, X) = \Big\langle [JX, X], \alpha \Big\rangle = (d\alpha)(X, JX), \quad \alpha = J^t(d\rho). \tag{10.11}$$

If $\mathcal{L}(X, X) > 0$ for all nonzero $X \in \mathfrak{H}_p(\partial M)$, we say $\overline{M}$ is strongly pseudoconvex at p.

The version of Morrey's inequality available in this setting is a little weaker than (2.1). We will prove the following.

Proposition 10.1. *If $\overline{M}$ is strongly pseudoconvex, then, for some $C > 0$, all $q \geq 1$,*

$$\|\overline{\partial}u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2 + \|u\|_{L^2}^2 \geq C \int_{\partial M} |u|^2 \, dS, \quad \forall\, u \in \mathcal{D}^{0,q}. \tag{10.12}$$

It will suffice to establish (10.12) when $u \in \mathcal{D}^{0,q}$ has support in some coordinate patch U intersecting ∂M. We can assume that, over U, there is a smooth orthonormal frame field $\{\omega_j : 1 \leq j \leq n\}$ for $\Lambda^{1,0}$, with $\omega_n = \sqrt{2}\partial\rho$. Let $\{L_j : 1 \leq j \leq n\}$ be the dual basis, consisting of (complex) vector fields. Set

$$C_{jk} = \widetilde{\mathcal{L}}(L_j, L_k) = \Big\langle [L_j, \overline{L}_k], \omega_n \Big\rangle, \quad 1 \leq j, k \leq n - 1. \tag{10.13}$$

One can verify that this is essentially equivalent to the Levi form. In particular, $\overline{M}$ is strongly pseudoconvex if and only if (C_{jk}) is a positive-definite matrix, at each $p \in \partial M$. Then, for $u \in \Lambda^{0,q}(U)$, we can write

$$u = \sum u_\alpha \overline{\omega}^\alpha. \tag{10.14}$$

Here and below, all multi-indices will be increasing (see the conventions in §8). Note that

$$u \in \mathcal{D}^{0,q} \Longleftrightarrow u_\alpha = 0 \quad \text{on } \partial M \text{ whenever } n \in \alpha. \tag{10.15}$$

Using these frames, we have the following formulas:

$$\begin{aligned} \overline{\partial} u &= \sum_{k,\alpha,\beta} \operatorname{sgn} \binom{\beta}{k\alpha} (\overline{L}_k u_\alpha) \overline{\omega}^\beta + Au, \\ \overline{\partial}^* u &= -\sum_{k,\alpha,\gamma} \operatorname{sgn} \binom{\alpha}{k\gamma} (L_k u_\alpha) \overline{\omega}^\gamma + Bu, \end{aligned} \tag{10.16}$$

where A and B are operators of order zero.

Let us set

$$E(u)^2 = \sum_{\alpha,k} \|\overline{L}_k u_\alpha\|_{L^2}^2 + \|u\|_{L^2}^2 + \int_{\partial M} |u|^2 \, dS. \tag{10.17}$$

The following result will suffice to prove Proposition 10.1.

Lemma 10.2. *Assume that (C_{jk}) is diagonal at $p \in \partial M$, with eigenvalues $\lambda_1, \dots, \lambda_{n-1}$. Let $\delta > 0$ be given. There is a neighborhood U of p such that if $q \geq 1$ and $u \in \mathcal{D}^{0,q}$ is supported in U, then*

$$\|\overline{\partial} u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2 = \sum_{k,\alpha} \|\overline{L}_k u_\alpha\|_{L^2}^2 + \sum_\alpha \sum_{k\in\alpha} \lambda_k \int_{\partial M} |u_\alpha|^2 \, dS + R(u), \tag{10.18}$$

where

$$|R(u)| \leq \delta E(u)^2 + C\|u\|_{L^2} E(u). \tag{10.19}$$

To begin the estimates, we use (10.16) to write

$$\begin{aligned} \|\overline{\partial} u\|_{L^2}^2 &= \sum_{k\notin\alpha} \|\overline{L}_k u_\alpha\|_{L^2}^2 \\ &\quad + \sum \operatorname{sgn} \binom{\gamma}{k\alpha} \operatorname{sgn} \binom{\sigma}{j\beta} (\overline{L}_k u_\alpha, \overline{L}_j u_\beta)_{L^2} + R_2, \end{aligned} \tag{10.20}$$

where $|R_2| \leq C \sum \|\overline{L}_k u_\alpha\|_{L^2} \|u\|_{L^2} + C\|u\|_{L^2}^2$. Here and below, the quantities R_j will all satisfy estimates (either stronger than or) of the form (10.19). From this we can deduce

$$\begin{aligned}\|\overline{\partial}u\|_{L^2}^2 &= \sum_{k,\alpha}\|\overline{L}_k u_\alpha\|_{L^2}^2 \\ &\quad - \sum \operatorname{sgn}\begin{pmatrix}\alpha\\ j\sigma\end{pmatrix}\operatorname{sgn}\begin{pmatrix}\beta\\ k\gamma\end{pmatrix}(\overline{L}_k u_\alpha, \overline{L}_j u_\beta)_{L^2} + R_3.\end{aligned} \tag{10.21}$$

Now

$$\begin{aligned}-(\overline{L}_k u_\alpha, \overline{L}_j u_\beta)_{L^2} &= (L_j \overline{L}_k u_\alpha, u_\beta)_{L^2} + R_4 \\ &= -(L_j u_\alpha, L_k u_\beta)_{L^2} + ([L_j, \overline{L}_k]u_\alpha, u_\beta)_{L^2} + R_5,\end{aligned} \tag{10.22}$$

so the second term on the right side of (10.21) is equal to

$$\begin{aligned}&-\sum \operatorname{sgn}\begin{pmatrix}\alpha\\ j\gamma\end{pmatrix}\operatorname{sgn}\begin{pmatrix}\beta\\ k\sigma\end{pmatrix}(L_j u_\alpha, L_k u_\beta)_{L^2} \\ &\quad + \sum \operatorname{sgn}\begin{pmatrix}\alpha\\ j\gamma\end{pmatrix}\operatorname{sgn}\begin{pmatrix}\beta\\ k\sigma\end{pmatrix}\Big([L_j, \overline{L}_k]u_\alpha, u_\beta)_{L^2} + R_6.\end{aligned} \tag{10.23}$$

Meanwhile, the formula for $\overline{\partial}^* u$ in (10.16) implies that the first sum in (10.23) is equal to

$$-\|\overline{\partial}^* u\|_{L^2}^2 + R_7. \tag{10.24}$$

Putting together (10.21)–(10.24), we have

$$\begin{aligned}\|\overline{\partial}u\|_{L^2}^2 + \|\overline{\partial}^* u\|_{L^2}^2 &= \sum_{k,\alpha}\|\overline{L}_k u_\alpha\|_{L^2}^2 \\ &\quad + \sum \operatorname{sgn}\begin{pmatrix}\alpha\\ j\gamma\end{pmatrix}\operatorname{sgn}\begin{pmatrix}\beta\\ k\sigma\end{pmatrix}\Big([L_j, \overline{L}_k]u_\alpha, u_\beta\Big)_{L^2} + R_8.\end{aligned} \tag{10.25}$$

To pass from here to (10.20), it remains to consider the second sum on the right side of (10.25). Making use of (10.15), one can show that if $j \in \alpha$ and $k \in \beta$, then

(10.26)
$$\Big([L_j, \overline{L}_k]u_\alpha, u_\beta\Big)_{L^2} = \begin{cases}\sqrt{2}(C_{jk} L_n u_\alpha, u_\beta)_{L^2} + R_9 & \text{if } j, k \le n-1, \\ R_{10} & \text{if } j = n \text{ or } k = n.\end{cases}$$

Now write

$$C_{jk} = \lambda_k \delta_{jk} + b_{jk}, \quad b_{jk}(p) = 0. \tag{10.27}$$

Note that $\alpha = \beta$ for nonzero terms for which $j = k$, in the last sum in (10.25). Also

$$(10.28)\quad \begin{aligned}(\lambda_k \delta_{jk} L_n u_\alpha, u_\alpha)_{L^2} &= \lambda_k \delta_{jk} (u_\alpha, \overline{L}_n u_\alpha)_{L^2} \\ &\quad + \lambda_k \delta_{jk} \int\limits_{\partial M} \frac{1}{i} \Big\langle \sigma_{L_n}(x,\nu) u_\alpha, u_\alpha \Big\rangle \, dS + R_{11} \\ &= \frac{1}{\sqrt{2}} \lambda_k \delta_{jk} \int\limits_{\partial M} |u_\alpha|^2 \, dS + R_{12}.\end{aligned}$$

Similarly,

$$(10.29)\quad (b_{jk} L_n u_\alpha, u_\beta)_{L^2} = \frac{1}{\sqrt{2}} \int\limits_{\partial M} b_{jk} u_\alpha \overline{u}_\beta \, dS + R_{13},$$

and if U is such a small neighborhood of p that $\sup |b_{jk}|$ is small compared to δ, this can also be denoted as R_{14}.

Combining (10.26)–(10.29) and summing, we have

$$(10.30)\quad \begin{aligned}&\sum \operatorname{sgn} \begin{pmatrix} \alpha \\ j\gamma \end{pmatrix} \operatorname{sgn} \begin{pmatrix} \beta \\ k\sigma \end{pmatrix} \Big([L_j, \overline{L}_k] u_\alpha, u_\beta \Big)_{L^2} \\ &\qquad = \sum_{k \in \alpha} \lambda_k \int\limits_{\partial M} |u_\alpha|^2 \, dS + R_{15}.\end{aligned}$$

Using this in (10.25), we have Lemma 10.2.

We can now use the Friedrichs method to define an unbounded, self-adjoint operator $\mathcal{L}$ on $L^2(M, \Lambda^{0,q})$, for any $q \geq 1$, such that $\mathcal{D}(\mathcal{L}^{1/2})$ is the completion of $\mathcal{D}^{0,q}$ with respect to the square norm $Q(u,u) + \|u\|_{L^2}^2$, where $Q(u,v)$ is given by (10.9), and $(\mathcal{L}u, v)_{L^2} = Q(u,v)$ for $u \in \mathcal{D}(\mathcal{L}),\ v \in \mathcal{D}^{0,q}$. One difference between this situation and those that arose in §§3 and 8 is that, while $\mathcal{L} \geq 0$, we might possibly have $0 \in \operatorname{spec} \mathcal{L}$. The estimates of §§3 and 4 and the regularity result of §5 extend without difficulty if $\overline{M}$ is strongly pseudoconvex, which we will assume in the rest of this section.

We have

$$(10.31)\quad u \in \mathcal{D}(\mathcal{L}),\ \mathcal{L}u = f \in H^j(M, \Lambda^{0,q}) \Longrightarrow u \in H^{j+1}(M, \Lambda^{0,q});$$

for $q \geq 1$; in particular,

$$(10.32)\quad \operatorname{Ker} \mathcal{L} \subset \Lambda^{0,q}(\overline{M}).$$

Denote this space by $\mathcal{H}^{0,q}(M)$. In fact, we have

$$(10.33)\quad \mathcal{H}^{0,q}(M) = \{u \in \mathcal{D}^{0,q} : \overline{\partial} u = 0 = \overline{\partial}^* u\}.$$

Let P_h denote the orthogonal projection of $L^2(M, \Lambda^{0,q})$ onto Ker $\mathcal{L}$. Given $f \in L^2(M, \Lambda^{0,q})$, let Gf denote the unique element u of $(\text{Ker}\,\mathcal{L})^\perp$ such that $\mathcal{L}u = (I - P_h)f$. Clearly,

$$G : H^j(M, \Lambda^{0,q}) \longrightarrow H^{j+1}(M, \Lambda^{0,q}), \tag{10.34}$$

for $q \geq 1$. We have the following Hodge decomposition for $u \in \Lambda^{0,q}(\overline{M})$:

$$u = \overline{\partial}\overline{\partial}^* Gu + \overline{\partial}^*\overline{\partial}Gu + P_h u = P_{\overline{\partial}}u + P_{\overline{\partial}^*}u + P_h u. \tag{10.35}$$

Arguments used in §§6 and 8 extend to show that these three terms are mutually orthogonal in $L^2(M, \Lambda^{0,q})$, so $P_{\overline{\partial}}$ and $P_{\overline{\partial}^*}$ (as well as P_h) extend to bounded operators on $L^2(M, \Lambda^{0,q})$, which in fact are orthogonal projections. Such arguments as used before also yield

$$P_{\overline{\partial}},\ P_{\overline{\partial}^*} : H^j(M) \longrightarrow H^{j-\varepsilon}(M), \quad \forall\, \varepsilon > 0. \tag{10.36}$$

In connection with the Hodge decomposition, note that just as in (6.4), we have

$$w \in \mathcal{D}^{0,q+1} \Longrightarrow \overline{\partial}^* w \perp \ \ker \overline{\partial} \cap \Lambda^{0,q}(\overline{M}). \tag{10.37}$$

We hence have the following extension of Theorem 6.2 and Corollary 8.5:

Proposition 10.3. *If $q \geq 1$ and $f \in \Lambda^{0,q}(\overline{M})$ satisfies*

$$\overline{\partial} f = 0 \quad \text{and} \quad f \perp \mathcal{H}^{0,q}(M), \tag{10.38}$$

then there exists $u \in \Lambda^{0,q-1}(\overline{M})$ satisfying $\overline{\partial}u = f$.

Proof. With $g = Gf \in \Lambda^{0,q}(\overline{M})$, we have the decomposition into orthogonal pieces

$$f = \overline{\partial}\overline{\partial}^* g + \overline{\partial}^*\overline{\partial}g + P_h f. \tag{10.39}$$

Now (10.37), applied to $w = \overline{\partial}g$, implies $\overline{\partial}^*\overline{\partial}g \perp f$, while the second hypothesis in (10.38) implies $P_h f = 0$. This gives

$$f = \overline{\partial}(\overline{\partial}^* g), \tag{10.40}$$

so we have the desired result, with $u = \overline{\partial}^* g$.

Note also by (6.3) that if $u \in \Lambda^{0,q-1}(\overline{M})$ and $f = \overline{\partial}u$, then $f \perp w$ whenever $w \in \mathcal{D}^{0,q}$ and $\overline{\partial}^* w = 0$, so the condition $f \perp \mathcal{H}^{0,q}(M)$ is necessary for solvability.

Since solving $\overline{\partial} u = f$ is of primary importance, the last result suggests two objects of study: Determine when $\mathcal{H}^{0,q}(M) = 0$, and work out how to deal with the requirement that $f \perp \mathcal{H}^{0,q}(M)$ if you cannot show that $\mathcal{H}^{0,q}(M) = 0$.

Here is an example of the first sort. Suppose J_s is a smooth family of integrable, almost complex structures on a compact Riemannian manifold with boundary, parameterized by $s \in [0, a]$. We can adjust the metric to depend smoothly on s and make each J_s an isometry. Denote the resulting object by $\overline{M}_s$. We have operators $\overline{\partial}_s$, $\overline{\partial}_s^*$, $\mathcal{L}_s$, etc.; we often drop the subscript when $s = 0$. Assume $\overline{M}_0$ is strongly pseudoconvex; then $\overline{M}_s$ is strongly pseudoconvex if $|s|$ is sufficiently small.

Proposition 10.4. *If $q \geq 1$ and $\mathcal{H}^{0,q}(M_0) = 0$, then $\mathcal{H}^{0,q}(M_s) = 0$ for $|s|$ sufficiently small.*

Proof. The proof of the Morrey-type estimates and consequent derivation of the $1/2$-estimate yields

$$(I + \mathcal{L}_s)^{-1} \text{ bounded in } \mathcal{L}\Big(L^2(M_s, \Lambda^{0,q}), H^{1/2}(M_s, \Lambda^{0,q})\Big), \tag{10.41}$$

for $|s|$ small. Now, suppose $s_j \to 0$, $\mathcal{H}^{0,q}(M_{s_j}) \neq 0$. Pick

$$u_{s_j} \in \mathcal{H}^{0,q}(M_{s_j}), \quad \|u_{s_j}\|_{L^2} = 1. \tag{10.42}$$

Then (10.41) implies $\|u_{s_j}\|_{H^{1/2}} \leq K$. Hence, passing to a subsequence, we have

$$u_{s_j} \to u_0 \text{ strongly in } L^2(M), \text{ weakly in } H^{1/2}(M). \tag{10.43}$$

In particular, $\|u_0\|_{L^2} = 1$. Now, via (10.37), we can say

$$u_{s_j} \perp \overline{\partial}_{s_j}^*(\mathcal{D}_{s_j}^{0,q+1}) \Longrightarrow u_0 \perp \overline{\partial}^*(\mathcal{D}^{0,q+1}), \tag{10.44}$$

while, by the remark after (10.40), we can say

$$u_{s_j} \perp \overline{\partial}_{s_j}\Big(\Lambda^{0,q-1}(\overline{M}_{s_j})\Big) \Longrightarrow u_0 \perp \overline{\partial}\Big(\Lambda^{0,q-1}(\overline{M})\Big). \tag{10.45}$$

The conclusions of (10.44) and (10.45) imply that $u_0 \in \mathcal{H}^{0,q}(M_0)$. Since $\|u_0\|_{L^2} = 1$, this means $\mathcal{H}^{0,q}(M_0) \neq 0$, so we have a contradiction to the hypothesis that $\mathcal{H}^{0,q}(M_{s_j}) \neq 0$.

Corollary 10.5. *Under the hypotheses of Proposition 10.4, we have, for each $j \geq 0$,*

$$\mathcal{L}_s^{-1} \text{ bounded in } \mathcal{L}\Big(H^j(M_s, \Lambda^{0,q}), H^{j+1}(M_s, \Lambda^{0,q})\Big), \tag{10.46}$$

for $|s|$ sufficiently small.

Proof. A check of the sorts of estimates arising in §4 shows that

$$(I+\mathcal{L}_s)^{-1} \text{ bounded in } \mathcal{L}\Big(H^j(M_s,\Lambda^{0,q}), H^{j+1}(M_s,\Lambda^{0,q})\Big), \tag{10.47}$$

for $|s|$ sufficiently small. Passing from this to (10.46) can be done by arguments similar to those used in the proof of Proposition 10.4.

Let us give an example of a situation where Proposition 10.4 and Corollary 10.5 apply. Let $\mathcal{O}$ be a manifold of real dimension $2n$, with an integrable, almost complex structure J, and fix $p \in \mathcal{O}$. Without loss of generality, we identify a neighborhood U of p with an open set in $\mathbb{C}^n$, and suppose $J\big|_{T_pU}$ coincides with the "standard" complex structure on $\mathbb{C}^n$, which for now we denote J_0. We may as well suppose $p = 0$. Use the standard metric on $\mathbb{C}^n$, and consider $B_\varepsilon = \{z \in \mathbb{C}^n : |z| < \varepsilon\}$. For small ε, B_ε is strongly pseudoconvex both for J_0 and for J. Now we produce a family $\mathcal{M}_s$ of almost complex manifolds as follows. As a set, $\mathcal{M}_s = \mathcal{M} = B_1 = \{z \in \mathbb{C}^n : |z| < 1\}$. We have $\varphi_s : \mathcal{M}_s \to B_s$, given by $\varphi_s(z) = sz$, and we pull back J (restricted to B_s) to get integrable, almost complex structures J_s on $\mathcal{M}_s$, for $0 < s \le a$ (for some $a > 0$). Clearly, such J_s and J_0 fit together smoothly. Also, since $\overline{\mathcal{M}}_0$ is a strongly pseudoconvex domain in $\mathbb{C}^n$, the results of §§2 and 8 imply $\mathcal{H}^{0,q}(\mathcal{M}_0) = 0$ for $q \ge 1$. Using this family, we will now prove the Newlander–Nirenberg theorem. This proof was given by Kohn in [K1], following a suggestion of D. Spencer.

Theorem 10.6. *If $\mathcal{O}$ has an integrable, almost complex structure, then $\mathcal{O}$ has holomorphic coordinate charts, so $\mathcal{O}$ is a complex manifold.*

Proof. It suffices to show that any point $p \in \mathcal{O}$ has a neighborhood $\mathcal{B}$ such that there are smooth complex-valued functions u_j on $\mathcal{B}$ (i.e., $u_j \in \Lambda^{0,0}(\mathcal{B})$) that are holomorphic (i.e., $\overline{\partial} u_j = 0$), and such that $du_1, \dots, du_n$ are linearly independent at p.

Let us bring in the structure described in the preceding paragraph. Thus we have a family of small neighborhoods B_s of p, blown up to $\mathcal{M}_s$, with integrable, almost complex structures, and Proposition 10.4 and Corollary 10.5 apply to $\mathcal{M}_s$ for $|s|$ small. As a set, $\mathcal{M}_s = \mathcal{M}$ is the unit ball in $\mathbb{C}^n$. We will be done if we produce some $s_0 > 0$ and $u_1, \dots, u_n \in \Lambda^{0,0}(\mathcal{M}_{s_0})$ such that $\overline{\partial}_{s_0} u_j = 0$ and $du_j(0)$ are linearly independent.

We write $u_{js} = z_j + v_{js}$, where the functions z_j are the standard coordinate functions on $\mathbb{C}^n$, and we pick v_{js} to be convenient solutions to

$$\overline{\partial}_s v_{js} = r_{js} \ \text{ on } \ \mathcal{M}_s, \quad r_{js} = -\overline{\partial}_s z_j. \tag{10.48}$$

Namely, we take

$$v_{js} = \overline{\partial}_s^* \mathcal{L}_s^{-1} r_{js}. \tag{10.49}$$

It follows from Proposition 10.4 and Corollary 10.5 that, for $|s|$ sufficiently small, (10.49) is well defined, and, for each $\ell \in \mathbb{Z}^+$,

$$\|v_{js}\|_{H^\ell(\mathcal{M})} \le K_\ell \|r_{js}\|_{H^\ell(\mathcal{M})}, \tag{10.50}$$

with K_ℓ independent of s.

On the other hand, since J_s approaches J_0 as $s \to 0$, we have

$$\|r_{js}\|_{H^\ell(\mathcal{M})} \le C_\ell s. \tag{10.51}$$

Now if we pick $\ell > n + 1$, we deduce that

$$\overline{\partial}_s u_{js} = 0, \quad u_{js} = z_j + v_{js}, \quad \|v_{js}\|_{C^1} \le Cs. \tag{10.52}$$

It is thus clear that, for $s = s_0$ sufficiently small, a desired coordinate system is produced.

The reader can compare this argument with the proof of the existence of isothermal coordinates on a two-dimensional Riemannian manifold, given in §10 of Chap. 5.

An important class of strongly pseudoconvex complex manifolds arises as follows. Let X be a compact, real analytic manifold; we can regard $X \subset TX$ as the zero section. Then there is a neighborhood U of X in TX that has the structure of a complex manifold, and U contains a strongly pseudoconvex neighborhood $\overline{M}$ of X, diffeomorphic to the unit ball bundle of X (given some Riemannian metric), so ∂M is diffeomorphic to the unit sphere bundle of X. The solution to the $\overline{\partial}$-Neumann problem on such $\overline{M}$ yields the result that there is a real analytic imbedding of X into Euclidean space $\mathbb{R}^N$. See Chap. 8 of [Mor] for an account of this. It was in the process of tackling this problem that the "Morrey inequality" was derived.

In this section we have continued to restrict our attention to the case of strongly pseudoconvex manifolds. However, as discovered in [Ho1], the basic estimate (10.12) holds for $u \in \mathcal{D}^{p,q}$ under a more general condition, called "condition $Z(q)$." This condition is that (if $\dim \partial M = 2n - 1$) the Levi form has either at least $n - q$ positive eigenvalues or at least $q + 1$ negative eigenvalues. A strongly pseudoconvex manifold satisfies condition $Z(q)$ for all $q \ge 1$, and for a bounded domain in $\mathbb{C}^n$ this is the only way condition $Z(q)$ can be satisfied, at least over all of ∂M. But there are open domains M with smooth boundary in compact, complex manifolds (such as complex projective space) which can satisfy condition $Z(q)$ for some but not all q, by virtue of the Levi form on ∂M having some negative eigenvalues. A proof of the estimate (10.12) under condition $Z(q)$ is given in [FK]. Also, [BS] analyzes the $\overline{\partial}$-Neumann problem via reduction to pseudodifferential operators on ∂M, under condition $Z(q)$.

Exercises

In Exercises 1–7, M is a Riemannian manifold with an almost complex structure J satisfying $\langle X, Y\rangle = \langle JX, JY\rangle$, where $\langle\ ,\ \rangle$ is the Riemannian inner product.

1. For $X, Y \in TM$, set
$$(X, Y) = \langle X, Y\rangle + i\langle X, JY\rangle.$$
Show that
$$(X, X) = \langle X, X\rangle, \quad (JX, Y) = i(X, Y), \quad (X, Y) = \overline{(Y, X)},$$
so we have a Hermitian metric. Thus we will call such M a Hermitian manifold.
2. Show that if M is actually a complex manifold, the Lie bracket of two vector fields $X + iJX$ and $Y + iJY$ has the same form, that is,
$$\mathcal{N}(X, Y) = J\Big([X, Y] - [JX, JY]\Big) - \Big([JX, Y] + [X, JY]\Big)$$
vanishes. Show that $\mathcal{N}(fX, gY) = fg\mathcal{N}(X, Y)$ for $f, g \in C^\infty(M)$, so $\mathcal{N}$ defines a tensor field of type (1,2). A related tensor N, defined by $N(X, Y) = 2J\mathcal{N}(X, Y)$, is called the Nijenhuis tensor.
3. Show that on any almost complex manifold, the vanishing of $\mathcal{N}$ is equivalent to the integrability condition (10.5).
4. Let ∇ be the Levi–Civita connection on M, and set
$$\omega(X, Y) = \langle X, JY\rangle,$$
a 2-form on M. Show that
$$2\Big\langle(\nabla_X J)Y, Z\Big\rangle = (d\omega)(X, JY, JZ) - (d\omega)(X, Y, Z) + \Big\langle\mathcal{N}(Y, Z), X\Big\rangle.$$

A Riemannian manifold M is called a symmetric space if, for each $p \in M$, there is an isometry $\iota_p : M \to M$ such that $\iota_p(p) = p$ and $D\iota_p(p) = -I$ on T_pM. If M is an almost complex manifold with metric as above (i.e., a Hermitian manifold), and is also a symmetric space, and if each isometry ι_p preserves the almost complex structure J, then M is called a Hermitian symmetric space.

5. Show that if M is a Hermitian symmetric space, then, for all vector fields X,
$$\nabla_X J = 0.$$
(*Hint*: Consider the tensor field $F = \nabla J$, of type (1,2). Show that $\iota_p^* F = F$ on M and that $\iota_p^* F = -F$ at p, so that $F = -F$ at p, for all $p \in M$.)
6. Show that the almost complex structure of any Hermitian symmetric space is integrable. (*Hint*: Show that $\nabla J = 0 \Rightarrow \nabla\omega = 0 \Rightarrow d\omega = 0$, and then use Exercise 4.)
7. More generally, a Hermitian manifold M is said to be a Kähler manifold if $\nabla J = 0$. Show that M is Kähler if and only if the almost complex structure J is integrable and $d\omega = 0$.
8. Show that the $\overline{\partial}$-operator is well defined:
$$\overline{\partial} : C^\infty(\overline{M}, E \otimes \Lambda^{0,q}) \longrightarrow C^\infty(\overline{M}, E \otimes \Lambda^{0,q+1}),$$
for any holomorphic vector bundle E over the complex manifold with boundary $\overline{M}$. Extend the results of this section to this case.

B. Complements on the Levi form

In this appendix we will give further formulas and other results for the Levi form on a hypersurface in $\mathbb{C}^n$. As a preliminary, we reexamine the formulas (2.7) and (2.8) in terms of the complex vector fields

$$Z = \sum u_k \frac{\partial}{\partial z_k}, \quad \overline{Z} = \sum \overline{u}_k \frac{\partial}{\partial \overline{z}_k}. \tag{B.1}$$

We assume that at each $p \in \partial\Omega$, $u(p) = (u_1, \dots, u_n)$ belongs to $\mathfrak{H}_p(\partial\Omega)$, defined by (2.13). As noted in §2, this hypothesis is equivalent both to $Z\rho = 0$ and to $\overline{Z}\rho = 0$ on $\partial\Omega$. Then (2.7) simply says $\overline{Z}Z\rho = 0$ on $\partial\Omega$. Also, of course, $Z\overline{Z}\rho = 0$ on $\partial\Omega$, and hence

$$[Z, \overline{Z}]\rho = 0 \quad \text{on } \partial\Omega,$$

but this is not the content of (2.8). To restate (2.8), note that

$$[Z, \overline{Z}] = \sum_{j,k} \left(u_k \frac{\partial \overline{u}_j}{\partial z_k} \frac{\partial}{\partial \overline{z}_j} - \overline{u}_j \frac{\partial u_k}{\partial \overline{z}_j} \frac{\partial}{\partial z_k} \right). \tag{B.2}$$

Now let us apply the operator J that gives the complex structure of $\mathbb{C}^n$, so

$$J \frac{\partial}{\partial x_j} = \frac{\partial}{\partial y_j}, \quad J \frac{\partial}{\partial y_j} = -\frac{\partial}{\partial x_j}, \tag{B.3}$$

and hence

$$J \frac{\partial}{\partial z_j} = i \frac{\partial}{\partial z_j}, \quad J \frac{\partial}{\partial \overline{z}_j} = -i \frac{\partial}{\partial \overline{z}_j}. \tag{B.4}$$

We have

$$W = J[Z, \overline{Z}] = -i \sum_{j,k} \left(u_k \frac{\partial \overline{u}_j}{\partial z_k} \frac{\partial}{\partial \overline{z}_j} + \overline{u}_k \frac{\partial u_j}{\partial \overline{z}_k} \frac{\partial}{\partial z_j} \right), \tag{B.5}$$

where to get the last term in parentheses from J applied to (B.2), we have interchanged the roles of j and k. Hence

$$W\rho = -2i \ \mathrm{Re} \left(\sum_{j,k} \overline{u}_k \frac{\partial u_j}{\partial \overline{z}_k} \frac{\partial \rho}{\partial z_j} \right). \tag{B.6}$$

Now the quantity in parentheses here is precisely the left side of (2.8). Since the right side of (2.8) is clearly real-valued, we have

(B.7)
$$\frac{1}{2i}\Big\langle J[Z,\overline{Z}], d\rho\Big\rangle = \sum_{j,k} \mathcal{L}_{jk} u_j \overline{u}_k \quad \text{on } \partial\Omega.$$

Let $\alpha = J^t d\rho$, so the left side of (B.7) is $1/2i$ times $\alpha([Z,\overline{Z}])$. Since

$$(d\alpha)(Z,\overline{Z}) = Z\cdot\alpha(\overline{Z}) - \overline{Z}\cdot\alpha(z) - \alpha([Z,\overline{Z}])$$

and $\alpha(\overline{Z}) = d\rho(J\overline{Z}) = -i\overline{Z}\rho = 0$, we have $(d\alpha)(Z,\overline{Z}) = -\alpha([Z,\overline{Z}])$, so (B.7) implies

(B.8)
$$\sum_{j,k} \mathcal{L}_{jk} u_j \overline{u}_k = -\frac{1}{2i}(d\alpha)(Z,\overline{Z}),$$

another useful formula for the Levi form.

It is also useful to write these formulas in terms of

(B.9)
$$X = \sum_k \left(f_k \frac{\partial}{\partial x_k} + g_k \frac{\partial}{\partial y_k} \right), \quad u_k = f_k + i g_k,$$

where $f_k = \operatorname{Re} u_k,\ g_k = \operatorname{Im} u_k$. The hypothesis $Z\rho = 0$ is still in effect, so, for $p \in \partial\Omega,\ X(p) \in \mathfrak{H}_p(\partial\Omega) \subset T_p\partial\Omega \subset \mathbb{R}^{2n}$. Note that

(B.10)
$$2Z = X - iJX, \quad 2\overline{Z} = X + iJX.$$

Thus (B.8) implies

(B.11)
$$4\sum_{j,k} \mathcal{L}_{jk} u_j \overline{u}_k = -(d\alpha)(X, JX).$$

Following the trail (B.7)⇒(B.8) backward, we note that $(d\alpha)(X,JX) = X\cdot\alpha(JX) - (JX)\cdot\alpha(X) - \alpha([X,JX])$ and $\alpha(X) = 0 = \alpha(JX)$, so $(d\alpha)(X,JX) = -\alpha([X,JX])$, and hence

(B.12)
$$4\sum_{j,k} \mathcal{L}_{jk} u_j \overline{u}_k = \alpha([X,JX]) = \Big\langle J[X,JX], d\rho\Big\rangle.$$

Note also that, by direct calculation,

(B.13)
$$4\sum_{j,k} \mathcal{L}_{jk} u_j \overline{u}_k = H(X,X) + H(JX,JX),$$

where H is the $(2n)\times(2n)$ real Hessian matrix of second-order partial derivatives of ρ with respect to $(x_1,\dots,x_n,y_1,\dots,y_n)$.

We can recast the Levi form in the following more invariant way, as done in [HN]. For a local section X of $\mathfrak{H}(\partial\Omega)$, we will define

$$\mathcal{L}_p(X, X) \in \mathfrak{H}_p^0(\partial\Omega), \tag{B.14}$$

where $\mathfrak{H}_p^0(\partial\Omega) \subset T_p^*(\partial\Omega)$ is the annihilator of $\mathfrak{H}_p(\partial\Omega) \subset T_p(\partial\Omega)$. To do this, we set

$$\begin{aligned} 4\mathcal{L}(X, X)(\alpha) &= \langle [X, JX], \alpha \rangle \\ &= -(d\alpha)(X, JX). \end{aligned} \tag{B.15}$$

When $\alpha = J^t d\rho$, this coincides with (B.11)–(B.12). This object is clearly invariant under conjugation by biholomorphic maps (i.e., under biholomorphic changes of coordinates). The property of positivity of $\mathcal{L}$ is invariantly defined, since the real line bundle $\mathfrak{H}_p^0(\partial\Omega)$ has a natural orientation, defined by declaring that $J^t d\rho > 0$. Thus we have the following:

Proposition B.1. *If $\overline{\Omega}$ is strongly pseudoconvex at $p \in \partial\Omega$ and if $F : \mathcal{O} \to U \subset \mathbb{C}^n$ is a biholomorphic map defined on a neighborhood of p, then $F(\mathcal{O} \cap \overline{\Omega}) = \widetilde{\Omega}$ is strongly pseudoconvex at $\tilde{p} = F(p)$.*

It follows readily from (B.13) that $\overline{\Omega}$ is strongly pseudoconvex at any $p \in \partial\Omega$ at which $\overline{\Omega}$ is strongly *convex*. By Proposition B.1 we see then that any (local) biholomorphic image of a strongly convex $\overline{\Omega} \subset \mathbb{C}^n$ is (locally) strongly pseudoconvex.

We can also relate the Levi form to the second fundamental form of $\partial\Omega$ as a hypersurface of $\mathbb{R}^{2n}$, using the following:

Lemma B.2. *If II is the second fundamental form of $\partial\Omega \subset \mathbb{R}^{2n}$, and if X is a section of $\mathfrak{H}(\partial\Omega)$, then*

$$II(X, X) = -P_N J \, \nabla_X(JX) = -J P_{JN} \, \nabla_X(JX). \tag{B.16}$$

Here, ∇ is the Levi–Civita connection on $\partial\Omega$, P_N is the orthogonal projection of $\mathbb{R}^{2n}$ onto the span of $N = -\nabla\rho$ (the sign chosen so N points inward), and P_{JN} is the orthogonal projection of $\mathbb{R}^{2n}$ onto the span of JN. We denote the span of JN by $\mathfrak{H}^\perp(\partial\Omega)$, which is isomorphic to $\mathfrak{H}^0(\partial\Omega)$, via the Riemannian metric on $\partial\Omega$.

To prove the lemma, recall from §4 of Appendix C (Connections and Curvature) that if X and Y are tangent to $\partial\Omega$, then $II(X, Y) = P_N D_X Y$, where D_X denotes the standard flat connection on $\mathbb{R}^{2n}$. Of course, also $II(X, Y) = D_X Y - \nabla_X Y$. Note that $D_X(JY) = J D_X Y$, so $II(JX, X) = II(X, JX) = P_N D_X(JX) = P_N J(D_X X)$. Hence

$$II(JX, X) = P_N J \, II(X, X) + P_N J \, \nabla_X X. \tag{B.17}$$

Similarly, $II(JX, JX) = P_N J D_{JX} X = P_N J\, II(JX, X) + P_N J\, \nabla_{JX} X$, and substituting (B.17) yields

$$II(JX, JX) = P_N J P_N J\, II(X, X) + P_N J P_N J\, \nabla_X X + P_N J\, \nabla_{JX} X.$$

Now, $JP_N J = -P_{JN}$, which is orthogonal to P_N, so $P_N J P_N J = 0$, and we have

$$II(JX, JX) = P_N J\, \nabla_{JX} X = J P_{JN}\, \nabla_{JX} X. \tag{B.18}$$

Replacing X by JX hence yields (B.16) and proves the lemma.

We can add (B.16) and (B.18), obtaining

$$\begin{aligned} &II(X, X) + II(JX, JX) \\ &\quad = P_N J\Big[\nabla_{JX} X - \nabla_X (JX)\Big] = P_N J\, [JX, X]. \end{aligned} \tag{B.19}$$

Comparing this with (B.12) and using the notation $II(X, Y) = \widetilde{II}(X, Y)N$, as in (4.15) of Appendix C, we see that

$$4 \sum_{j,k} \mathcal{L}_{jk} u_j \overline{u}_k = \widetilde{II}(X, X) + \widetilde{II}(JX, JX). \tag{B.20}$$

This can also be obtained from (B.13), plus formula (4.25) of Appendix C.

We will consider one more formula for the Levi form, in terms of the geometry of $\mathfrak{H}(\partial\Omega)$ as a subbundle of the trivial bundle $\partial\Omega \times \mathbb{R}^{2n} \approx \partial\Omega \times \mathbb{C}^n$. Associated to this subbundle there is a second fundamental form $II_{\mathfrak{H}}$, defined as in (4.40) of Appendix C. A formula for $II_{\mathfrak{H}}$ can be given as follows. Let $\mathfrak{K}(\partial\Omega)$ denote the orthogonal complement of $\mathfrak{H}(\partial\Omega)$; this can be viewed as a real vector bundle of rank 2, generated by N and JN, or as a complex line bundle generated by N. If $P_{\mathfrak{K}}$ denotes the orthogonal projection of $\mathbb{R}^{2n}$ onto $\mathfrak{K}$, then we have

$$II_{\mathfrak{H}}(X, Y) = P_{\mathfrak{K}} D_X Y, \tag{B.21}$$

when X and Y are sections of $\mathfrak{H}(\partial\Omega)$.

We want to relate $II_{\mathfrak{H}}$ to the Levi form. It is convenient to use the previous analysis of II. Since $P_{\mathfrak{K}} = P_N + P_{JN}$, we have

$$II_{\mathfrak{H}}(X, X) = II(X, X) + P_{JN} D_X X.$$

As noted in the proof of (B.16), $II(JX, X) = P_N J\; D_X X$, which is equal to $JP_{JN}\; D_X X$, so we have

$$II_{\mathfrak{H}}(X, X) = II(X, X) - J\; II(JX, X). \tag{B.22}$$

Substituting JX for X, we have

$$II_{\mathfrak{H}}(JX, JX) = II(JX, JX) + J\, II(JX, X), \tag{B.23}$$

and adding this to (B.22) and using (B.20), we obtain

$$II_{\mathfrak{H}}(X, X) + II_{\mathfrak{H}}(JX, JX) = 4\Big(\sum_{j,k} \mathcal{L}_{jk} u_j \overline{u}_k\Big) N. \tag{B.24}$$

C. The Neumann operator for the Dirichlet problem

Let $\overline{M}$ be a compact Riemannian manifold with boundary $\partial M = X$. Then X has an induced Riemannian metric, and $X \hookrightarrow \overline{M}$ has a second fundamental form, with associated Weingarten map

$$A_N : T_x X \longrightarrow T_x X, \tag{C.1}$$

defined as in §4 of Appendix C, Connections and Curvature. We take N to be the unit normal to X, pointing into M.

Both $\overline{M}$ and X have Laplace operators, which we denote Δ and Δ_X, respectively. The *Neumann operator* $\mathcal{N}$ is an operator on $\mathcal{D}'(X)$ defined as follows:

$$\mathcal{N} f = \frac{\partial u}{\partial N}, \quad u = \mathrm{PI}\, f, \tag{C.2}$$

where to say $u = \mathrm{PI}\, f$ is to say

$$\Delta u = 0 \ \text{ on } \ M, \quad u\Big|_{\partial M} = f. \tag{C.3}$$

As shown in §§11 and 12 of Chap. 7, $\mathcal{N}$ is a negative-semidefinite, self-adjoint operator, and also an elliptic operator in $OPS^1(X)$. It is fairly easy to see that

$$\mathcal{N} = -\sqrt{-\Delta_X} \mod OPS^0(X). \tag{C.4}$$

Our main purpose here is to capture the principal part of the difference.

Proposition C.1. *The Neumann operator $\mathcal{N}$ is given by*

$$\mathcal{N} = -\sqrt{-\Delta_X} + B \quad \mathit{mod}\ OPS^{-1}(X), \tag{C.5}$$

where $B \in OPS^0(X)$ has principal symbol

$$\sigma_B(x,\xi) = \frac{1}{2}\left(Tr\, A_N - \frac{\langle A_N^* \xi, \xi\rangle}{\langle \xi, \xi\rangle}\right). \tag{C.6}$$

Here, $A_N^* : T_x^*X \to T_x^*X$ is the adjoint of (C.1), and $\langle\ ,\ \rangle$ is the inner product on T_x^*X arising from the given Riemannian metric.

To prove this, we choose coordinates $x = (x_1, \dots, x_{m-1})$ on an open set in X (if dim $M = m$) and then coordinates (x, y) on a neighborhood in $\overline{M}$ such that $y = 0$ on X and $|\nabla y| = 1$ near X while $y > 0$ on M and such that x is constant on each geodesic segment in $\overline{M}$ normal to X. Then the metric tensor on $\overline{M}$ has the form

$$(\text{C.7}) \qquad \Big(g_{jk}(x, y)\Big) = \begin{pmatrix} h_{jk}(x, y) & 0 \\ 0 & 1 \end{pmatrix},$$

where, in the first matrix, $1 \le j, k \le m$, and in the second, $1 \le j, k \le m - 1$. Thus the Laplace operator Δ on $\overline{M}$ is given in local coordinates by

$$(\text{C.8}) \qquad \begin{aligned} \Delta u &= g^{-1/2}\partial_j \Big(g^{1/2} g^{jk} \partial_k u\Big) \\ &= h^{-1/2}\partial_y(h^{1/2}\partial_y u) + h^{-1/2}\partial_j\Big(h^{1/2}h^{jk}\partial_k u\Big) \\ &= \partial_y^2 u + \frac{1}{2}\frac{h_y}{h}\partial_y u + L(y, x, D_x)u, \end{aligned}$$

where, as usual,

$$(\text{C.9}) \qquad g = \det(g_{jk}), \quad h = \det(h_{jk});$$

we set $h_y = \partial h/\partial y$, and $L(y) = L(y, x, D_x)$ is a family of Laplace operators on X, associated to the family of metrics $\Big(h_{jk}(y)\Big)$ on X, so $L(0) = \Delta_X$. In other words,

$$(\text{C.10}) \qquad \Delta u = \partial_y^2 u + a(y)\partial_y u + L(y)u, \quad a(y) = \frac{1}{2}\frac{h_y}{h}.$$

We will construct smooth families of operators $A_j(y) \in OPS^1(X)$ such that

$$(\text{C.11}) \qquad \partial_y^2 + a(y)\partial_y + L(y) = \Big(\partial_y - A_1(y)\Big)\Big(\partial_y + A_2(y)\Big),$$

modulo a smoothing operator. The principal parts of $A_1(y)$ and $A_2(y)$ will be $\sqrt{-L(y)}$. It will follow that

$$(\text{C.12}) \qquad \mathcal{N} = -A_2(0) \mod OPS^{-\infty}(X),$$

and we can then read off (C.5)–(C.6).

To construct $A_j(y)$, we compute that the right side of (C.11) is equal to

$$(\text{C.13}) \qquad \partial_y^2 - A_1(y)\partial_y + A_2(y)\partial_y + A_2'(y) - A_1(y)A_2(y),$$

so we need

$$A_2(y) - A_1(y) = a(y), \quad -A_1(y)A_2(y) + A_2'(y) = L(y). \tag{C.14}$$

Substituting $A_2(y) = A_1(y) + a(y)$ into the second identity, we get an equation for $A_1(y)$:

$$A_1(y)^2 + A_1(y)a(y) - A_1'(y) = -L(y) + a'(y). \tag{C.15}$$

Now set

$$A_1(y) = \Lambda(y) + B(y), \quad \Lambda(y) = \sqrt{-L(y)}. \tag{C.16}$$

We get an equation for $B(y)$:

$$2B(y)\Lambda(y) + [\Lambda(y), B(y)] + B(y)^2 - B'(y) + B(y)a(y) = \Lambda'(y) - \Lambda(y)a(y) + a'(y). \tag{C.17}$$

Granted that $B(y)$ is a smooth family in $OPS^0(X)$, the principal part $B_0(y)$ must satisfy $2B_0(y)\Lambda(y) = \Lambda'(y) - a(y)\Lambda(y)$, or

$$B_0(y) = \frac{1}{2}\Lambda'(y)\Lambda(y)^{-1} - \frac{1}{2}a(y) \mod OPS^{-1}(X). \tag{C.18}$$

We can inductively obtain further terms $B_j(y) \in OPS^{-j}(X)$ and establish that, with $B(y) \sim \sum_{j\ge 0} B_j(y)$, the operators

$$A_1(y) = \sqrt{-L(y)} + B(y), \quad A_2(y) = \sqrt{-L(y)} + B(y) + a(y)$$

do yield (C.11) modulo a smoothing operator. Details are similar to those arising in the decoupling procedure in §12 of Chap. 7.

Given this, we have (C.5) with

$$-B = B_0(0) + a(0) = \frac{1}{2}\Lambda'(0)\Lambda(0)^{-1} + \frac{1}{2}a(0) \mod OPS^{-1}(X). \tag{C.19}$$

In turn, since $\Lambda(y) = \sqrt{-L(y)}$, we have

$$\Lambda'(0)\Lambda(0)^{-1} = \frac{1}{2}L'(0)L(0)^{-1} \mod OPS^{-1}(X). \tag{C.20}$$

Hence

$$-B = \frac{1}{4}\Big(L'(0)L(0)^{-1} + \frac{h_y}{h}(0,x)\Big). \tag{C.21}$$

To compute the symbol of B, note that

$$\sigma_{L'(0)}(x,\xi) = -\sum \partial_y h^{jk}(0,x)\xi_j\xi_k, \tag{C.22}$$

while, of course, $\sigma_{L(0)}(x,\xi) = -\sum h^{jk}(0,x)\xi_j\xi_k = -\langle\xi,\xi\rangle$. Now, (4.68)–(4.70) of Appendix C, Connections and Curvature, we have

$$\sum \partial_y h_{jk}(0,x)U_jV_k = -2\langle A_N U, V\rangle, \tag{C.23}$$

so

$$\sum \partial_y h^{jk}(0,x)\xi_j\xi_k = 2\langle A_N^*\xi,\xi\rangle. \tag{C.24}$$

Thus,

$$\sigma_{L'(0)L(0)^{-1}}(x,\xi) = 2\frac{\langle A_N^*\xi,\xi\rangle}{\langle\xi,\xi\rangle}. \tag{C.25}$$

Next, for $h = \operatorname{Det}(h_{jk}) = \operatorname{Det} H$, we have $h_y = h\operatorname{Tr}(H^{-1}H_y)$. Looking in a normal coordinate system on X, centered at x_0, we have

$$\frac{h_y}{h}(0,x_0) = \sum_j \partial_y h_{jj}(0,x_0) = -2\operatorname{Tr} A_N, \tag{C.26}$$

the last identity by (C.23). Combining (C.25) and (C.26) yields the desired formula (C.6).

The following alternative way of writing (C.6) is useful. We have

$$\sigma_B(x,\xi) = \frac{1}{2}\operatorname{Tr}(A_N^* P_\xi^\perp), \tag{C.27}$$

where, for nonzero $\xi \in T_x^*X$, $P_\xi^\perp$ is the orthogonal projection of T_x^*X onto the orthogonal complement of the linear span of ξ. Another equivalent formula is

$$\sigma_B(x,\xi) = \frac{1}{2}\operatorname{Tr}(A_N P_\xi^0), \tag{C.28}$$

where P_ξ^0 is the orthogonal projection of T_xX onto the subspace annihilated by ξ.

To close, we mention the special case where $\overline{M}$ is the closed unit ball in $\mathbb{R}^m$, so $\partial M = S^{m-1}$. It follows from (4.5)–(4.6) of Chap. 8 that

$$\mathcal{N} = -\sqrt{-\Delta_X + c_m^2} + c_m, \quad c_m = \frac{m-2}{2}, \tag{C.29}$$

in this case. Note that, in this case, $A_N = I$, so this formula is consistent with (C.5)–(C.6).

We mention that calculations of the symbol of $\mathcal{N}$ in a similar spirit (but for a different purpose) were done in [LU]. Another approach was taken in [CNS].

References

[AHLM] A. Andreotti, C. D. Hill, S. Lojasiewicz, and B. MacKichan, Complexes of differential operators, The Mayer–Vietoris sequence, *Invent. Math*. 35(1976), 43–86.

[AV] A. Andreotti and E. Vesentini, Carleman estimates for the Laplace–Beltrami equation on complex manifolds, *Publ. Math. Inst. Hautes Etudes Sci*. 25(1965), 81–130.

[BDT] P. Baum, R. Douglas, and M. Taylor, Cycles and relative cycles in analytic K-homology, *J. Differential Geom*. 30(1989), 761–804.

[BFG] M. Beals, C. Fefferman, and R. Grossman, Strictly pseudoconvex domains in $\mathbb{C}^n$, *Bull. AMS* 8(1983), 125–322.

[BG] R. Beals and P. Greiner, *Calculus on Heisenberg Manifolds*, Princeton University Press, Princeton, NJ, 1988.

[BGS] R. Beals, P. Greiner, and N. Stanton, L^p and Lipschitz estimates for the $\overline{\partial}$-equation and the $\overline{\partial}$-Neumann problem, *Math. Ann*. 277(1987), 185–196.

[BS] R. Beals and N. Stanton, Estimates on kernels for the $\overline{\partial}$-equation and the $\overline{\partial}$-Neumann problem, *Math. Ann*. 289(1991), 73–83.

[BL] S. Bell and E. Ligocka, A simplification and extension of Fefferman's theorem on holomorphic mappings, *Invent. Math*. 57(1980), 283–289.

[Bog] A. Boggess, *CR Manifolds and the Tangential Cauchy–Riemann Complex*, CRC, Boca Raton, FL, 1991.

[B1] L. Boutet de Monvel, Hypoelliptic operators with double characteristics and related pseudodifferential operators, *CPAM* 27(1974), 585–639.

[B2] L. Boutet de Monvel, On the index of Toeplitz operators in several complex variables, *Invent. Math*. 50(1979), 249–272.

[BGu] L. Boutet de Monvel and V. Guillemin, *The Spectral Theory of Toeplitz Operators*, Princeton University Press, Princeton, NJ, 1981.

[BSj] L. Boutet de Monvel and J. Sjöstrand, Sur la singularite des noyaux de Bergman et du Szegö, *Asterisque* 34–36(1976), 123–164.

[BT] L. Boutet de Monvel and F. Treves, On a class of pseudodifferential operators with double characteristics, *CPAM* 27(1974), 59–89.

[BDF] L. Brown, R. Douglas, and P. Fillmore, Extensions of C^*-algebras and K-homology, *Ann. Math*. 105(1977), 265–324.

[Cat] D. Catlin, Subelliptic estimates for the $\overline{\partial}$-Neumann problem on pseudoconvex domains, *Ann. Math*. 126(1987), 131–191.

[Cha] D. Chang, On L^p and Hölder estimates for the $\overline{\partial}$-Neumann problem on strongly pseudoconvex domains, *Math. Ann*. 282(1988), 267–297.

[CNS] D. Chang, A. Nagel, and E. Stein, Estimates for the $\overline{\partial}$-Neumann problem in pseudoconvex domains of finite type in $\mathbb{C}^2$, *Acta Math*. 169(1992), 153–228.

[Chr] M. Christ, Regularity properties of the $\overline{\partial}_b$-equation on weakly pseudoconvex CR manifolds of dimension three, *J. AMS* 1(1988), 587–646.

[D] A. Dynin, An algebra of pseudodifferential operators on the Heisenberg group; symbolic calculus, *Soviet Math. Dokl.* 17(1976), 508–512.

[EMM] C. Epstein, R. Melrose, and G. Mendoza, Resolvent of the Laplacian on strictly pseudoconvex domains, *Acta Math.* 167(1991), 1–106.

[F] C. Fefferman, The Bergman kernel and biholomorphic mappings of pseudoconvex domains, *Invent. Math.* 26(1974), 1–65.

[FeK] C. Fefferman and J. J. Kohn, Hölder estimates on domains of complex dimension 2 and on 3 dimensional CR manifolds, *Advances Math.* 69(1988), 223–303.

[FK] G. Folland and J. J. Kohn, *The Neumann Problem for the Cauchy–Riemann Complex*, Princeton University Press, Princeton, NJ, 1972.

[FS] G. Folland and E. Stein, Estimates for the $\overline{\partial}_b$ complex and analysis on the Heisenberg group, *CPAM* 27(1974), 429–522.

[Gel] D. Geller, *Analytic Pseudodifferential Operators on the Heisenberg Group and Local Solvability*, Princeton University Press, Princeton, NJ, 1990.

[GS] P. Greiner and E. Stein, *Estimates for the $\overline{\partial}$-Neumann Problem*, Princeton University Press, Princeton, NJ, 1977.

[GR] R. Gunning and H. Rossi, *Analytic Functions of Several Complex Variables*, Prentice-Hall, Englewood Cliffs, NJ, 1965.

[HP1] R. Harvey and J. Polking, Fundamental solutions in complex analysis, *Duke Math. J.* 46(1979), 253–340.

[HP2] R. Harvey and J. Polking, The $\overline{\partial}$-Neumann kernel in the ball in $\mathbb{C}^n$, *Proc. Symp. Pure Math.* 41(1984), 117–136.

[HN] C. D. Hill and M. Nacinovich, Pseudoconcave CR manifolds, Preprint, 1993.

[Ho1] L. Hörmander, L^2 estimates and existence theorems for the $\overline{\partial}$ operator, *Acta Math.* 113(1965), 89–152.

[Ho2] L. Hörmander, The Frobenius–Nirenberg theorem, *Arkiv. Mat.* 5(1965), 425–432.

[Ho3] L. Hörmander, *Introduction to Complex Analysis in Several Variables*, Van Nostrand, Princeton, NJ, 1966.

[Ho4] L. Hörmander, Pseudo-differential operators and non-elliptic boundary problems, *Ann. Math.* 83(1966), 129–209.

[Ker] N. Kerzman, Hölder and L^p estimates for solutions to $\overline{\partial}u = f$ in strongly pseudoconvex domains, *CPAM* 24(1971), 301–379.

[Ker2] N. Kerzman, The Bergman kernel function: Differentiability at the boundary, *Math. Annalen* 195(1972), 149–158.

[KS] N. Kerzman and E. Stein, The Szegö kernel in terms of Cauchy–Fantappie kernels, *Duke Math. J.* 45(1978), 197–224.

[Kh] G. Khenkin, The method of integral representations in complex analysis, pp. 19–116 in [Vit].

[K1] J. J. Kohn, Harmonic integrals on strictly pseudoconvex manifolds I, *Ann. Math.* 78(1963), 112–148; II, *Ann. Math.* 79(1964), 450–472.

[K2] J. J. Kohn, Subellipticity of the $\overline{\partial}$-Neumann problem on pseudoconvex domains; sufficient conditions, *Acta Math.* 142(1979), 79–122.

[K3] J. J. Kohn, A survey of the $\overline{\partial}$-Neumann problem, *Proc. Symp. Pure Math.* 41(1984), 137–145.

[KN] J. J. Kohn and L. Nirenberg, Non-coercive boundary problems, *CPAM* 18(1965), 443–492.

[KN2] J. J. Kohn and L. Nirenberg, A pseudo-convex domain not admitting a holomorphic support function, *Math. Ann.* 201(1973), 265–268.

[KR] J. J. Kohn and H. Rossi, On the extension of holomorphic functions from the boundary of a complex manifold, *Ann. Math.* 81(1965), 451–472.

[Kr1] S. Krantz, *Function Theory of Several Complex Variables*, Wadsworth, Belmont, CA, 1992.

[Kr2] S. Krantz, *Partial Differential Equations and Complex Analysis*, CRC, Boca Raton, FL, 1992.

[LU] J. Lee and G. Uhlmann, Determining anisotropic real-analytic conductivities by boundary measurements, *CPAM* 42(1989), 1097–1112.

[Lel] P. Lelong, *Plurisubharmonic Functions and Positive Differential Forms*, Gordon and Breach, New York, 1969.

[LR] I. Lieb and R. Range, Integral representations and estimates in the theory of the $\bar{\partial}$-Neumann problem, *Ann. Math.* 123(1986), 265–301

[Mor] C. B. Morrey, *Multiple Integrals in the Calculus of Variations*, Springer, New York, 1966.

[NRSW] A. Nagel, J. Rosay, E. Stein, and S. Wainger, Estimates for the Bergman and Szegö kernels in $\mathbb{C}^2$, *Ann. Math.* 129(1989), 113–149.

[NeNi] A. Newlander and L. Nirenberg, Complex analytic coordinates in almost-complex manifolds, *Ann. Math.* 65(1957), 391–404.

[Ni] L. Nirenberg, *Lectures on Linear Partial Differential Equations*, Reg. Conf. Ser. in Math., no. 17, AMS, 1972.

[PS] D. Phong and E. Stein, Hilbert integrals, singular integrals, and Radon transforms, II, *Invent. Math.* 86(1986), 75–113.

[RS] L. Rothschild and E. Stein, Hypoelliptic differential operators and nilpotent groups, *Acta Math.* 137(1976), 247–320.

[Sj] J. Sjöstrand, Parametrices for pseudodifferential operators with multiple characteristics, *Ark. Math.* 2(1974), 85–130.

[Sm] H. Smith, A calculus for three-dimensional CR manifolds of finite type, *J. Func. Anal.* 120(1994), 135–162.

[Sp] D. Spencer, Overdetermined systems of linear partial differential equations, *Bull. AMS* 75(1969), 179–239.

[St] E. Stein, *Harmonic Analysis*, Princeton University Press, Princeton, NJ, 1993.

[Sw] W. Sweeney, The D-Neumann problem, *Acta Math.* 120(1968), 224–277.

[Tai] G. Taiani, *Cauchy–Riemann (CR) Manifolds*, Pace University, New York, 1989.

[Tar] D. Tartakoff, The local real analyticity of solutions to $\Box_b$ and the $\bar{\partial}$-Neumann problem, *Acta Math.* 145(1980), 177–204.

[Tay] M. Taylor, *Noncommutative Microlocal Analysis I*, Mem. AMS #313 (1984).

[Tr1] F. Treves, Analytic hypo-ellipticity of a class of pseudodifferential operators with double characteristics and applications to the $\bar{\partial}$-Neumann problem, *Comm. PDE* 3(1978), 475–642.

[Tr2] F. Treves, *Introduction to Pseudodifferential and Fourier Integral Operators*, Vols. 1, 2, Plenum, New York, 1980.

[Ven] U. Venugopalkrishna, Fredholm operators associated with strongly pseudoconvex domains, *J. Func. Anal.* 9(1972), 349–373.

[Vit] A. Vitushkin (ed.), *Several Complex Variables I, Encycl. of Math. Sci.*, Vol.7, Springer, New York, 1990.

C

Connections and Curvature

Introduction

In this appendix we present results in differential geometry that serve as a useful background for material in the main body of the book. Material in §1 on connections is somewhat parallel to the study of the natural connection on a Riemannian manifold made in §11 of Chap. 1, but here we also study the curvature of a connection. Material in §2 on second covariant derivatives is connected with material in Chap. 2 on the Laplace operator. Ideas developed in §§3 and 4, on the curvature of Riemannian manifolds and submanifolds, make contact with such material as the existence of complex structures on two-dimensional Riemannian manifolds, established in Chap. 5, and the uniformization theorem for compact Riemann surfaces and other problems involving nonlinear, elliptic PDE, arising from studies of curvature, treated in Chap. 14. Section 5 on the Gauss–Bonnet theorem is useful both for estimates related to the proof of the uniformization theorem and for applications to the Riemann–Roch theorem in Chap. 10. Furthermore, it serves as a transition to more advanced material presented in §§6–8.

In §6 we discuss how constructions involving vector bundles can be derived from constructions on a principal bundle. In the case of ordinary vector fields, tensor fields, and differential forms, one can largely avoid this, but it is a very convenient tool for understanding spinors. The principal bundle picture is used to construct characteristic classes in §7. The material in these two sections is needed in Chap. 10, on the index theory for elliptic operators of Dirac type. In §8 we show how one particular characteristic class, arising from the Pfaffian, figures into the higher-dimensional version of the Gauss–Bonnet formula. The proof given here is geometrical and uses the elements of Morse theory. In Chap. 10 this result is derived as a special case of the Atiyah–Singer index formula.

M.E. Taylor, *Partial Differential Equations II: Qualitative Studies of Linear Equations*, Applied Mathematical Sciences 116, DOI 10.1007/978-1-4419-7052-7_7,

1. Covariant derivatives and curvature on general vector bundles

Let $E \to M$ be a vector bundle, either real or complex. A *covariant derivative*, or *connection*, on E is a map

$$\nabla_X : C^\infty(M, E) \longrightarrow C^\infty(M, E) \tag{1.1}$$

assigned to each vector field X on M, satisfying the following three conditions:

$$\nabla_X(u + v) = \nabla_X u + \nabla_X v, \tag{1.2}$$

$$\nabla_{(fX+Y)} u = f \nabla_X u + \nabla_Y u, \tag{1.3}$$

$$\nabla_X(fu) = f \nabla_X u + (Xf)u, \tag{1.4}$$

where u, v are sections of E, and f is a smooth scalar function. The examples contained in Chaps. 1 and 2 are the Levi–Civita connection on a Riemannian manifold, in which case E is the tangent bundle, and associated connections on tensor bundles, discussed in §2.2.

One general construction of connections is the following. Let F be a vector space, with an inner product; we have the trivial bundle $M \times F$. Let E be a subbundle of this trivial bundle; for each $x \in M$, let P_x be the orthogonal projection of F on $E_x \subset F$. Any $u \in C^\infty(M, E)$ can be regarded as a function from M to F, and for a vector field X, we can apply X componentwise to any function on M with values in F; call this action $u \mapsto D_X u$. Then a connection on M is given by

$$\nabla_X u(x) = P_x D_X u(x). \tag{1.5}$$

If M is imbedded in a Euclidean space $\mathbb{R}^N$, then $T_x M$ is naturally identified with a linear subspace of $\mathbb{R}^N$ for each $x \in M$. In this case it is easy to verify that the connection defined by (1.5) coincides with the Levi–Civita connection, where M is given the metric induced from its imbedding in $\mathbb{R}^N$. Compare with the discussion of submanifolds in §4 below.

Generally, a connection defines the notion of "parallel transport" along a curve γ in M. A section u of E over γ is obtained from $u(\gamma(t_0))$ by parallel transport if it satisfies $\nabla_T u = 0$ on γ, where $T = \dot{\gamma}(t)$.

Formulas for covariant derivatives, involving indices, are produced in terms of a choice of "local frame" for E, that is, a set e_α, $1 \le \alpha \le K$, of sections of E over an open set U which forms a basis of E_x for each $x \in U$; $K = \dim E_x$. Given such a local frame, a smooth section u of E over U is specified by

$$u = u^\alpha e_\alpha \quad \text{(summation convention)}. \tag{1.6}$$

If $D_j = \partial/\partial x_j$ in a coordinate system on U, we set

$$\nabla_{D_j} u = u^\alpha{}_{;j} e_\alpha = (\partial_j u^\alpha + u^\beta \Gamma^\alpha{}_{\beta j}) e_\alpha, \tag{1.7}$$

the connection coefficients $\Gamma^\alpha{}_{\beta j}$ being defined by

$$\nabla_{D_j} e_\beta = \Gamma^\alpha{}_{\beta j} e_\alpha. \tag{1.8}$$

A vector bundle $E \to M$ may have an inner product on its fibers. In that case, a connection on E is called a *metric connection* provided that

$$X\langle u, v\rangle = \langle \nabla_X u, v\rangle + \langle u, \nabla_X v\rangle, \tag{1.9}$$

for any vector field X and smooth sections u, v of E.

The curvature of a connection is defined by

$$R(X, Y)u = [\nabla_X, \nabla_Y]u - \nabla_{[X,Y]}u, \tag{1.10}$$

where X and Y are vector fields and u is a section of E. It is easy to verify that (1.10) is linear in X, Y, and u, over $C^\infty(M)$. With respect to local coordinates, giving $D_j = \partial/\partial x_j$, and a local frame $\{e_\alpha\}$ on E, as in (1.6), we define the components $R^\alpha{}_{\beta jk}$ of the curvature by

$$R(D_j, D_k)e_\beta = R^\alpha{}_{\beta jk} e_\alpha, \tag{1.11}$$

as usual, using the summation convention. Since D_j and D_k commute, $R(D_j, D_k)e_\beta = [\nabla_{D_j}, \nabla_{D_k}]e_\beta$. Applying the formulas (1.7) and (1.8), we can express the components of R in terms of the connection coefficients. The formula is seen to be

$$R^\alpha{}_{\beta jk} = \partial_j \Gamma^\alpha{}_{\beta k} - \partial_k \Gamma^\alpha{}_{\beta j} + \Gamma^\alpha{}_{\gamma j}\Gamma^\gamma{}_{\beta k} - \Gamma^\alpha{}_{\gamma k}\Gamma^\gamma{}_{\beta j}. \tag{1.12}$$

The formula (1.12) can be written in a shorter form, as follows. Given a choice of local frame $\{e_\alpha : 1 \le \alpha \le K\}$, we can define $K \times K$ matrices $\Gamma_j = (\Gamma^\alpha{}_{\beta j})$ and also $\mathfrak{R}_{jk} = (R^\alpha{}_{\beta jk})$. Then (1.12) is equivalent to

$$\mathfrak{R}_{jk} = \partial_j \Gamma_k - \partial_k \Gamma_j + [\Gamma_j, \Gamma_k]. \tag{1.13}$$

Note that $\mathfrak{R}_{jk}$ is antisymmetric in j and k. Now we can define a "connection 1-form" Γ and a "curvature 2-form" Ω by

$$\Gamma = \sum_j \Gamma_j \, dx_j, \quad \Omega = \frac{1}{2}\sum_{j,k} \mathfrak{R}_{jk} \, dx_j \wedge dx_k. \tag{1.14}$$

Then the formula (1.12) is equivalent to

$$\Omega = d\Gamma + \Gamma \wedge \Gamma. \tag{1.15}$$

The curvature has symmetries, which we record here, for the case of general vector bundles. The Riemann curvature tensor, associated with the Levi–Civita connection, has additional symmetries, which will be described in §3.

Proposition 1.1. *For any connection ∇ on $E \to M$, we have*

$$R(X,Y)u = -R(Y,X)u. \tag{1.16}$$

If ∇ is a metric connection, then

$$\langle R(X,Y)u, v\rangle = -\langle u, R(X,Y)v\rangle. \tag{1.17}$$

Proof. Equation (1.16) is obvious from the definition (1.10); this is equivalent to the antisymmetry of $R^\alpha{}_{\beta jk}$ in j and k noted above. If ∇ is a metric connection, we can use (1.9) to deduce

$$\begin{aligned} 0 &= (XY - YX - [X,Y])\langle u, v\rangle \\ &= \langle R(X,Y)u, v\rangle + \langle u, R(X,Y)v\rangle, \end{aligned}$$

which gives (1.17).

Next we record the following implication of a connection having *zero* curvature. A section u of E is said to be "parallel" if $\nabla_X u = 0$ for all vector fields X.

Proposition 1.2. *If $E \to M$ has a connection ∇ whose curvature is zero, then any $p \in M$ has a neighborhood U on which there is a frame $\{e_\alpha\}$ for E consisting of parallel sections: $\nabla_X e_\alpha = 0$ for all X.*

Proof. If U is a coordinate neighborhood, then e_α is parallel provided $\nabla_j e_\alpha = 0$ for $j = 1, \dots, n = \dim M$. The condition that $R = 0$ is equivalent to the condition that the operators ∇_{D_j} all commute with each other, for $1 \le j \le n$. Consequently, Frobenius's theorem (as expanded in Exercise 5 in §9 of Chap. 1) allows us to solve the system of equations

$$\nabla_{D_j} e_\alpha = 0, \quad j = 1, \dots, n, \tag{1.18}$$

on a neighborhood of p, with e_α prescribed at the point p. If we pick $e_\alpha(p)$, $1 \le \alpha \le K$, to be a basis of E_p, then $e_\alpha(x)$, $1 \le \alpha \le K$, will be linearly independent in E_x for x close to p, so the local frame of parallel sections is constructed.

It is useful to note, in general, several formulas that result from choosing a local frame $\{e_\alpha\}$ by parallel translation along rays through a point $p \in M$, the origin in some coordinate system $(x_1, \dots, x_n)$, so

$$\nabla_{r\partial/\partial r} e_\alpha = 0, \quad 1 \le \alpha \le K. \tag{1.19}$$

This means $\sum x_j \nabla_{D_j} e_\alpha = 0$. Consequently, the connection coefficients (1.8) satisfy

$$x_1 \Gamma^\alpha{}_{\beta 1} + \cdots + x_n \Gamma^\alpha{}_{\beta n} = 0. \tag{1.20}$$

Differentiation with respect to x_j gives

$$\Gamma^\alpha{}_{\beta j} = -x_1 \partial_j \Gamma^\alpha{}_{\beta 1} - \cdots - x_n \partial_j \Gamma^\alpha{}_{\beta n}. \tag{1.21}$$

In particular,

$$\Gamma^\alpha{}_{\beta j}(p) = 0. \tag{1.22}$$

Comparison of (1.21) with

$$\Gamma^\alpha{}_{\beta j} = x_1 \partial_1 \Gamma^\alpha{}_{\beta j}(p) + \cdots + x_n \partial_n \Gamma^\alpha{}_{\beta j}(p) + O(|x|^2) \tag{1.23}$$

gives

$$\partial_k \Gamma^\alpha{}_{\beta j} = -\partial_j \Gamma^\alpha{}_{\beta k}, \ \text{ at } p. \tag{1.24}$$

Consequently, the formula (1.12) for curvature becomes

$$R^\alpha{}_{\beta jk} = 2\, \partial_j \Gamma^\alpha{}_{\beta k}, \ \text{ at } p, \tag{1.25}$$

with respect to such a local frame. Note that, near p,

$$R^\alpha{}_{\beta jk} = \partial_j \Gamma^\alpha{}_{\beta k} - \partial_k \Gamma^\alpha{}_{\beta j} + O(|x|^2). \tag{1.26}$$

Given vector bundles $E_j \to M$ with connections ∇^j, there is a natural covariant derivative on the tensor-product bundle $E_1 \otimes E_2 \to M$, defined by the derivation property

$$\nabla_X(u \otimes v) = (\nabla^1_X u) \otimes v + u \otimes (\nabla^2_X v). \tag{1.27}$$

Also, if A is a section of $\mathrm{Hom}(E_1, E_2)$, the formula

$$(\nabla^\#_X A)u = \nabla^2_X(Au) - A(\nabla^1_X v) \tag{1.28}$$

defines a connection on $\mathrm{Hom}(E_1, E_2)$.

Regarding the curvature tensor R as a section of $(\otimes^2 T^*) \otimes \mathrm{End}(E)$ is natural in view of the linearity properties of R given after (1.10). Thus if $E \to M$ has a connection with curvature R, and if M also has a Riemannian metric, yielding a connection on T^*M, then we can consider $\nabla_X R$. The following, known as *Bianchi's identity*, is an important result involving the covariant derivative of R.

Proposition 1.3. *For any connection on $E \to M$, the curvature satisfies*

$$(\nabla_Z R)(X,Y) + (\nabla_X R)(Y,Z) + (\nabla_Y R)(Z,X) = 0, \tag{1.29}$$

or equivalently

$$R^\alpha{}_{\beta ij;k} + R^\alpha{}_{\beta jk;i} + R^\alpha{}_{\beta ki;j} = 0. \tag{1.30}$$

Proof. Pick any $p \in M$. Choose normal coordinates centered at p, and choose a local frame field for E by radial parallel translation, as above. Then, by (1.22) and (1.26),

$$R^\alpha{}_{\beta ij;k} = \partial_k \partial_i \Gamma^\alpha{}_{\beta j} - \partial_k \partial_j \Gamma^\alpha{}_{\beta i}, \ \text{ at } p. \tag{1.31}$$

Cyclically permuting (i,j,k) here and summing clearly give 0, proving the proposition.

Note that we can regard a connection on E as defining an operator

$$\nabla : C^\infty(M,E) \longrightarrow C^\infty(M, T^* \otimes E), \tag{1.32}$$

in view of the linear dependence of ∇_X on X. If M has a Riemannian metric and E a Hermitian metric, it is natural to study the adjoint operator

$$\nabla^* : C^\infty(M, T^* \otimes E) \longrightarrow C^\infty(M,E). \tag{1.33}$$

If u and v are sections of E, ξ a section of T^*, we have

$$\begin{aligned} \big(v, \nabla^*(\xi \otimes u)\big) &= (\nabla v, \xi \otimes u) \\ &= (\nabla_X v, u) \\ &= (v, \nabla_X^* u), \end{aligned} \tag{1.34}$$

where X is the vector field corresponding to ξ via the Riemannian metric. Using the divergence theorem we can establish:

Proposition 1.4. *If E has a metric connection, then*

$$\nabla^*(\xi \otimes u) = \nabla_X^* u = -\nabla_X u - (\mathit{div}\, X)u. \tag{1.35}$$

Proof. The first identity follows from (1.34) and does not require E to have a metric connection. If E does have a metric connection, integrating

$$\langle \nabla_X v, u \rangle = -\langle v, \nabla_X u \rangle + X\langle v, u \rangle$$

and using the identity

$$\int_M Xf\,dV = -\int_M (\operatorname{div} X) f\,dV, \quad f \in C_0^\infty(M), \tag{1.36}$$

give the second identity in (1.35) and complete the proof.

Exercises

1. If ∇ and $\widetilde{\nabla}$ are two connections on a vector bundle $E \to M$, show that

$$\nabla_X u = \widetilde{\nabla}_X u + C(X, u), \tag{1.37}$$

   where C is a smooth section of $\operatorname{Hom}(T \otimes E, E) \approx T^* \otimes \operatorname{End}(E)$. Show that conversely, if C is such a section and ∇ a connection, then (1.37) defines $\widetilde{\nabla}$ as a connection.
2. If ∇ and $\widetilde{\nabla}$ are related as in Exercise 1, show that their curvatures R and $\widetilde{R}$ are related by

$$(R - \widetilde{R})(X, Y)u = [C_X, \widetilde{\nabla}_Y]u - [C_Y, \widetilde{\nabla}_X]u - C_{[X,Y]}u + [C_X, C_Y]u, \tag{1.38}$$

   where C_X is the section of $\operatorname{End}(E)$ defined by $C_X u = C(X, u)$.
   In Exercises 3–5, let $P(x)$, $x \in M$, be a smooth family of projections on a vector space F, with range E_x, forming a vector bundle $E \to M$; E gets a natural connection via (1.5).
3. Let $\gamma : I \to M$ be a smooth curve through $x_0 \in M$. Show that parallel transport of $u(x_0) \in E_{x_0}$ along I is characterized by the following (with $P'(t) = dP(\gamma(t))/dt$):

$$\frac{du}{dt} = P'(t)u.$$

4. If each $P(x)$ is an orthogonal projection of the inner-product space F onto E_x, show that you get a metric connection. (*Hint*: Show that $du/dt \perp u(\gamma(t))$ via $P'P = (I - P)P'$.)
5. In what sense can $\Gamma = -dP\ P = -(I - P)\,dP$ be considered the connection 1-form, as in (1.13)? Show that the curvature form (1.15) is given by

$$\Omega = P\ dP \wedge dP\ P. \tag{1.39}$$

   For more on this, see (4.50)–(4.53).
6. Show that the formula

$$d\Omega = \Omega \wedge \Gamma - \Gamma \wedge \Omega \tag{1.40}$$

   follows from (1.15). Relate this to the Bianchi identity. Compare with (2.13) in the next section.
7. Let $E \to M$ be a vector bundle with connection ∇, with two local frame fields $\{e_\alpha\}$ and $\{f_\alpha\}$, defined over $U \subset M$. Suppose

$$f_\alpha(x) = g^\beta{}_\alpha(x) e_\beta(x), \quad e_\alpha(x) = h^\beta{}_\alpha(x) f_\beta(x);$$

   note that $g^\beta{}_\gamma(x) h^\gamma{}_\alpha(x) = \delta^\beta{}_\alpha$. Let $\Gamma^\alpha{}_{\beta j}$ be the connection coefficients for the frame field $\{e_\alpha\}$, as in (1.7) and (1.8), and let $\widetilde{\Gamma}^\alpha{}_{\beta j}$ be the connection coefficients for the frame field $\{f_\alpha\}$. Show that

$$\widetilde{\Gamma}^{\alpha}{}_{\beta j} = h^{\alpha}{}_{\mu}\Gamma^{\mu}{}_{\gamma j}g^{\gamma}{}_{\beta} + h^{\alpha}{}_{\gamma}(\partial_j g^{\gamma}{}_{\beta}). \tag{1.41}$$

2. Second covariant derivatives and covariant-exterior derivatives

Let M be a Riemannian manifold, with Levi–Civita connection, and let $E \to M$ be a vector bundle with connection. In §1 we saw that the covariant derivative acting on sections of E yields an operator

$$\nabla : C^\infty(M, E) \longrightarrow C^\infty(M, T^* \otimes E). \tag{2.1}$$

Now on $T^* \otimes E$ we have the product connection, defined by (1.27), yielding

$$\nabla : C^\infty(T^* \otimes E) \longrightarrow C^\infty(M, T^* \otimes T^* \otimes E). \tag{2.2}$$

If we compose (2.1) and (2.2), we get a second-order differential operator called the *Hessian*:

$$\nabla^2 : C^\infty(M, E) \longrightarrow C^\infty(T^* \otimes T^* \otimes E). \tag{2.3}$$

If u is a section of E and X and Y are vector fields, (2.3) defines $\nabla^2_{X,Y}$ as a section of E; using the derivation properties, we have the formula

$$\nabla^2_{X,Y} u = \nabla_X \nabla_Y u - \nabla_{(\nabla_X Y)} u. \tag{2.4}$$

Note that the antisymmetric part is given by the curvature of the connection on E:

$$\nabla^2_{X,Y} u - \nabla^2_{Y,X} u = R(X, Y)u. \tag{2.5}$$

Now the metric tensor on M gives a linear map $T^* \otimes T^* \to \mathbb{R}$, hence a linear bundle map $\gamma : T^* \otimes T^* \otimes E \to E$. We can consider the composition of this with ∇^2 in (2.3):

$$\gamma \circ \nabla^2 : C^\infty(M, E) \longrightarrow C^\infty(M, E). \tag{2.6}$$

We want to compare $\gamma \circ \nabla^2$ and $\nabla^*\nabla$, in the case when E has a Hermitian metric and a metric connection.

Proposition 2.1. *If ∇ is a metric connection on E, then*

$$\nabla^*\nabla = -\gamma \circ \nabla^2 \ \ \textit{on } C^\infty(M, E). \tag{2.7}$$

Proof. Pick a local orthonormal frame of vector fields $\{e_j\}$, with dual frame $\{v_j\}$. Then, for $u \in C^\infty(M, E)$, $\nabla u = \sum v_j \otimes \nabla_{e_j} u$, so (1.35) implies

$$\nabla^* \nabla u = \sum \bigl[-\nabla_{e_j} \nabla_{e_j} u - (\operatorname{div} e_j) u\bigr]. \tag{2.8}$$

Using (2.4), we have

$$\nabla^* \nabla u = -\sum \nabla^2_{e_j, e_j} u - \sum \bigl[\nabla_{\nabla_{e_j} e_j} u + (\operatorname{div} e_j) \nabla_{e_j} u\bigr]. \tag{2.9}$$

The first term on the right is equal to $-\gamma \circ \nabla^2 u$. Now, given $p \in M$, if we choose the local frame $\{e_j\}$ such that $\nabla_{e_j} e_k = 0$ at p, the rest of the right side vanishes at p. This establishes the identity (2.7).

We next define a "covariant-exterior derivative" operator

$$d^\nabla : C^\infty(M, \Lambda^k T^* \otimes E) \longrightarrow C^\infty(M, \Lambda^{k+1} T^* \otimes E) \tag{2.10}$$

as follows. For $k = 0$, $d^\nabla = \nabla$, given by (2.1), and we require

$$d^\nabla(\beta \wedge u) = (d\beta) \wedge u - \beta \wedge d^\nabla u \tag{2.11}$$

whenever β is a 1-form and u is a section of $\Lambda^k T^* \otimes E$. The operator d^∇ is also called the "gauge exterior derivative." Unlike the case of the ordinary exterior derivative,

$$d^\nabla \circ d^\nabla : C^\infty \Bigl(M, \Lambda^k T^* \otimes E\Bigr) \longrightarrow C^\infty \Bigl(M, \Lambda^{k+2} T^* \otimes E\Bigr)$$

is not necessarily zero, but rather

$$d^\nabla d^\nabla u = \Omega \wedge u, \tag{2.12}$$

where Ω is the curvature, and we use the antisymmetry (1.16) to regard Ω as a section of $\Lambda^2 T^* \otimes \operatorname{End}(E)$, as in (1.15). The verification of (2.12) is a straightforward calculation; (2.5) is in fact the special case of this, for $k = 0$.

The following is an alternative form of Bianchi's identity (1.29):

$$d^\nabla \Omega = 0, \tag{2.13}$$

where the left side is a priori a section of $\Lambda^3 T^* \otimes \operatorname{End}(E)$. This can also be deduced from (2.12), the associative law $d^\nabla(d^\nabla d^\nabla) = (d^\nabla d^\nabla) d^\nabla$, and the natural derivation property generalizing (2.11):

$$d^\nabla (A \wedge u) = \bigl(d^\nabla A\bigr) \wedge u + (-1)^j A \wedge d^\nabla u, \tag{2.14}$$

where u is a section of $\Lambda^k T^* \otimes E$ and A is a section of $\Lambda^j T^* \otimes \operatorname{End}(E)$.

Exercises

1. Let $E \to M$ be a vector bundle with connection ∇, $u \in C^\infty(M, E)$. Fix $p \in M$. Show that if $\nabla u(p) = 0$, then $\nabla^2_{X,Y} u(p)$ is independent of the choice of connection on M.
2. In particular, Exercise 1 applies to the trivial bundle $\mathbb{R} \times M$, with trivial flat connection, for which $\nabla_X u = \langle X, du \rangle = Xu$. Thus, if $u \in C^\infty(M)$ is real-valued and $du(p) = 0$, then $D^2 u(p)$ is well defined as a symmetric bilinear form on $T_p M$. If, in a coordinate system, $X = \sum X_j\, \partial/\partial x_j$, $Y = \sum Y_j\, \partial/\partial x_j$, show that

$$D^2_{X,Y} u(p) = \sum \frac{\partial^2 u}{\partial x_j \partial x_k}(p)\, X_j Y_k. \tag{2.15}$$

   Show that this invariance fails if $du(p) \neq 0$.
3. If u is a smooth section of $\Lambda^k T^* \otimes E$, show that

$$\begin{aligned} d^\nabla u(X_0, \dots, X_k) = &\sum_j (-1)^j \nabla_{X_j} u\left(X_0, \dots, \widehat{X}_j, \dots, X_k\right) \\ &+ \sum_{j<\ell} (-1)^{j+\ell} u\left([X_j, X_\ell], X_0, \dots, \widehat{X}_j, \dots, \widehat{X}_\ell, \dots, X_k\right). \end{aligned} \tag{2.16}$$

   Compare with formula (13.56) of Chap. 1 and Exercises 2 and 3 in §3 of Chap. 2.
4. Verify the identity (2.12), namely, $d^\nabla d^\nabla u = \Omega \wedge u$.
5. If ∇ and $\widetilde{\nabla}$ are connections on $E \to M$, related by $\nabla_X u = \widetilde{\nabla}_X u + C(X, u)$, $C \in C^\infty(M, T^* \otimes \mathrm{End}(E))$, with curvatures R and $\widetilde{R}$, and curvature forms Ω and $\widetilde{\Omega}$, show that

$$\Omega - \widetilde{\Omega} = d^\nabla C + C \wedge C. \tag{2.17}$$

   Here the wedge product of two sections of $T^* \otimes \mathrm{End}(E)$ is a section of the bundle $\Lambda^2 T^* \otimes \mathrm{End}(E)$, produced in a natural fashion, as in (1.15). Show that (2.17) is equivalent to (1.38).

3. The curvature tensor of a Riemannian manifold

The Levi–Civita connection, which was introduced in §11 of Chap. 1, is a metric connection on the tangent bundle TM of a manifold M with a Riemannian metric, uniquely specified among all such connections by the zero-torsion condition

$$\nabla_Y X - \nabla_X Y = [Y, X]. \tag{3.1}$$

We recall the defining formula

$$\begin{aligned} 2\langle \nabla_X Y, Z \rangle = &X\langle Y, Z \rangle + Y\langle X, Z \rangle - Z\langle X, Y \rangle \\ &+ \langle [X, Y], Z \rangle - \langle [X, Z], Y \rangle - \langle [Y, Z], X \rangle, \end{aligned} \tag{3.2}$$

derived in (11.22) of Chap. 1. Thus, in a local coordinate system with the naturally associated frame field on the tangent bundle, the connection coefficients (1.8) are given by

$$\Gamma^{\ell}{}_{jk} = \frac{1}{2} g^{\ell\mu} \left[\frac{\partial g_{j\mu}}{\partial x_k} + \frac{\partial g_{k\mu}}{\partial x_j} - \frac{\partial g_{jk}}{\partial x_\mu} \right]. \tag{3.3}$$

The associated curvature tensor is the Riemann curvature tensor:

$$R(X,Y)Z = [\nabla_X, \nabla_Y]Z - \nabla_{[X,Y]}Z. \tag{3.4}$$

In a local coordinate system such as that discussed above, the expression for the Riemann curvature is a special case of (1.12), namely,

$$R^j{}_{k\ell m} = \partial_\ell \Gamma^j{}_{km} - \partial_m \Gamma^j{}_{k\ell} + \Gamma^j{}_{\nu\ell}\Gamma^\nu{}_{km} - \Gamma^j{}_{\nu m}\Gamma^\nu{}_{k\ell}. \tag{3.5}$$

Consequently, we have an expression of the form

$$R^j{}_{k\ell m} = L\left(g_{\alpha\beta}, \partial_\mu\partial_\nu g_{\gamma\delta}\right) + Q(g_{\alpha\beta}, \partial_\mu g_{\gamma\delta}), \tag{3.6}$$

where L is linear in the second-order derivatives of $g_{\alpha\beta}(x)$ and Q is quadratic in the first-order derivatives of $g_{\alpha\beta}(x)$, each with coefficients depending on $g_{\alpha\beta}(x)$.

Building on Proposition 1.2, we have the following result on metrics whose Riemannian curvature is zero.

Proposition 3.1. *If (M, g) is a Riemannian manifold whose curvature tensor vanishes, then the metric g is flat; that is, there is a coordinate system about each $p \in M$ in which $g_{jk}(x)$ is constant.*

Proof. It follows from Proposition 1.2 that on a neighborhood U of p there are parallel vector fields $V_{(j)},\ j = 1, \dots, n = \dim M$, namely, in a given coordinate system

$$\nabla_{D_k} V_{(j)} = 0, \quad 1 \le j,\ k \le n, \tag{3.7}$$

such that $V_{(j)}(p)$ form a basis of T_pM. Let $v_{(j)}$ be the 1-forms associated to $V_{(j)}$ by the metric g, so

$$v_{(j)}(X) = g(X, V_{(j)}), \tag{3.8}$$

for all vector fields X. Hence

$$\nabla_{D_k} v_{(j)} = 0, \quad 1 \le j,\ k \le n. \tag{3.9}$$

We have $v_{(j)} = \sum v^k_{(j)}\, dx_k$, with $v^k_{(j)} = v_{(j)}(D_k) = \langle D_k, v_{(j)}\rangle$. The zero-torsion condition (3.1), in concert with (3.8), gives

$$\partial_\ell \langle v_{(j)}, D_k\rangle - \partial_k \langle v_{(j)}, D_\ell\rangle = \langle v_{(j)}, \nabla_{D_\ell} D_k\rangle - \langle v_{(j)}, \nabla_{D_k} D_\ell\rangle = 0, \tag{3.10}$$

which is equivalent to

$$d\,v_{(j)} = 0, \quad j = 1, \dots, n. \tag{3.11}$$

Hence, locally, there exist functions x_j, $j = 1, \dots, n$, such that

$$v_{(j)} = dx_j. \tag{3.12}$$

The functions $(x_1, \dots, x_n)$ give a coordinate system near p. In this coordinate system the inverse of the matrix $\big(g_{jk}(x)\big)$ has entries $g^{jk}(x) = \langle dx_j, dx_k\rangle$. Now, by (1.9),

$$\partial_\ell g^{jk}(x) = \langle \nabla_{D_\ell}\, dx_j, dx_k\rangle + \langle dx_j, \nabla_{D_\ell}\, dx_k\rangle = 0, \tag{3.13}$$

so the proof is complete.

We have seen in Proposition 1.1 that R has the following symmetries:

$$R(X,Y) = -R(Y,X), \tag{3.14}$$

$$\langle R(X,Y)Z, W\rangle = -\langle Z, R(X,Y)W\rangle. \tag{3.15}$$

In other words, in terms of

$$R_{jk\ell m} = \langle R(D_\ell, D_m)D_k, D_j\rangle, \tag{3.16}$$

we have

$$R_{jk\ell m} = -R_{jkm\ell} \tag{3.17}$$

and

$$R_{jk\ell m} = -R_{kj\ell m}. \tag{3.18}$$

The Riemann tensor has additional symmetries:

Proposition 3.2. *The Riemann tensor satisfies*

$$R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0 \tag{3.19}$$

and

$$\langle R(X,Y)Z, W\rangle = \langle R(Z,W)X, Y\rangle, \tag{3.20}$$

or, in index notation,

$$R_{ijk\ell} + R_{ik\ell j} + R_{i\ell jk} = 0 \tag{3.21}$$

and

$$R_{ijk\ell} = R_{k\ell ij}. \tag{3.22}$$

Proof. Plugging in the definition of each of the three terms of (3.19), one gets a sum that is seen to cancel out by virtue of the zero-torsion condition (3.1). This gives (3.19) and hence (3.21). The identity (3.22) is an automatic consequence of (3.17), (3.18), and (3.21), by elementary algebraic manipulations, which we leave as an exercise, to complete the proof. Also, (3.22) follows from (3.50) below.

The identity (3.19) is sometimes called *Bianchi's first identity*, with (1.29) then called *Bianchi's second identity*.

There are important contractions of the Riemann tensor. The *Ricci tensor* is defined by

$$\mathrm{Ric}_{jk} = R^i{}_{jik} = g^{\ell m} R_{\ell jmk}, \tag{3.23}$$

where the summation convention is understood. By (3.22), this is symmetric in j, k. We can also raise indices:

$$\mathrm{Ric}^j{}_k = g^{j\ell}\mathrm{Ric}_{\ell k}; \quad \mathrm{Ric}^{jk} = g^{k\ell}\mathrm{Ric}^j{}_\ell. \tag{3.24}$$

Contracting again defines the scalar curvature:

$$S = \mathrm{Ric}^j{}_j. \tag{3.25}$$

As we will see below, the special nature of $R_{ijk\ell}$ for $\dim M = 2$ implies

$$\mathrm{Ric}_{jk} = \frac{1}{2} S g_{jk} \ \text{ if } \dim M = 2. \tag{3.26}$$

The Bianchi identity (1.29) yields an important identity for the Ricci tensor. Specializing (1.30) to $\alpha = i,\ \beta = j$ and raising the second index give

$$R^{ij}{}_{ij;k} + R^{ij}{}_{jk;i} + R^{ij}{}_{ki;j} = 0, \tag{3.27}$$

hence, $S_{;k} - \mathrm{Ric}^i{}_{k;i} - \mathrm{Ric}^j{}_{k;j} = 0$, or

$$S_{;k} = 2\, \mathrm{Ric}^j{}_{k;j}. \tag{3.28}$$

This is called the *Ricci identity*. An equivalent form is

$$\mathrm{Ric}^{jk}{}_{;j} = \frac{1}{2}(S\ g^{jk})_{;j}. \tag{3.29}$$

The identity in this form leads us naturally to a tensor known as the *Einstein tensor*:

$$G^{jk} = \mathrm{Ric}^{jk} - \frac{1}{2} S\, g^{jk}. \tag{3.30}$$

The Ricci identity is equivalent to

$$G^{jk}{}_{;j} = 0. \tag{3.31}$$

As shown in Chap. 2, this means the Einstein tensor has zero divergence. This fact plays an important role in Einstein's equation for the gravitational field. Note that by (3.26) the Einstein tensor always vanishes when $\dim M = 2$. On the other hand, the identity (3.31) has the following implication when $\dim M > 2$.

Proposition 3.3. *If* $\dim M = n > 2$ *and the Ricci tensor is a scalar multiple of the metric tensor, the factor necessarily being* $1/n$ *times the scalar curvature:*

$$Ric_{jk} = \frac{1}{n} S g_{jk}, \tag{3.32}$$

then S *must be a constant.*

Proof. Equation (3.32) is equivalent to

$$G^{jk} = \left(\frac{1}{n} - \frac{1}{2}\right) S g^{jk}. \tag{3.33}$$

By (3.31) and the fact that the covariant derivative of the metric tensor is 0, we have

$$0 = \left(\frac{1}{n} - \frac{1}{2}\right) S_{;k} g^{jk},$$

or $S_{;k} = 0$, which proves the proposition.

We now make some comments on the curvature of Riemannian manifolds M of dimension 2. By (3.17) and (3.18), in this case each component $R_{jk\ell m}$ of the curvature tensor is either 0 or $\pm$ the quantity

$$R_{1212} = R_{2121} = gK, \quad g = \det(g_{jk}). \tag{3.34}$$

One calls K the *Gauss curvature* of M when $\dim M = 2$.

Suppose we pick normal coordinates centered at $p \in M$, so $g_{jk}(p) = \delta_{jk}$. We see that if $\dim M = 2$,

$$\mathrm{Ric}_{jk}(p) = R_{1j1k} + R_{2j2k}.$$

Now, the first term on the right is zero unless $j = k = 2$, and the second term is zero unless $j = k = 1$. Hence, $\mathrm{Ric}_{jk}(p) = K(p)\delta_{jk}$, in normal coordinates, so in arbitrary coordinates

$$\mathrm{Ric}_{jk} = K g_{jk}; \ \text{ hence } \ K = \frac{1}{2} S \ \text{ if } \dim M = 2. \tag{3.35}$$

Explicit formulas for K when M is a surface in $\mathbb{R}^3$ are given by (4.22) and (4.29), in the next section. (See also Exercises 2 and 5–7 below.) The following is a fundamental calculation of the Gauss curvature of a two-dimensional surface whose metric tensor is expressed in orthogonal coordinates:

$$ds^2 = E(x)\,dx_1^2 + G(x)\,dx_2^2. \tag{3.36}$$

Proposition 3.4. *Suppose* dim $M = 2$ *and the metric is given in coordinates by* (3.36)*. Then the Gauss curvature* $k(x)$ *is given by*

$$k(x) = -\frac{1}{2\sqrt{EG}}\left[\partial_1\left(\frac{\partial_1 G}{\sqrt{EG}}\right) + \partial_2\left(\frac{\partial_2 E}{\sqrt{EG}}\right)\right]. \tag{3.37}$$

To establish (3.37), one can first compute that

$$\begin{aligned}
\Gamma_1 = \left(\Gamma^j{}_{k1}\right) &= \frac{1}{2}\begin{pmatrix} E^{-1}\partial_1 E & E^{-1}\partial_2 E \\ -G^{-1}\partial_2 E & G^{-1}\partial_1 G \end{pmatrix}, \\
\Gamma_2 = \left(\Gamma^j{}_{k2}\right) &= \frac{1}{2}\begin{pmatrix} E^{-1}\partial_2 E & -E^{-1}\partial_1 G \\ G^{-1}\partial_1 G & G^{-1}\partial_2 G. \end{pmatrix}
\end{aligned}$$

Then, computing $\mathfrak{R}_{12} = (R^j{}_{k12}) = \partial_1\Gamma_2 - \partial_2\Gamma_1 + \Gamma_1\Gamma_2 - \Gamma_2\Gamma_1$, we have

$$\begin{aligned}
R^1{}_{212} = &-\frac{1}{2}\partial_1\left(\frac{\partial_1 G}{E}\right) - \frac{1}{2}\partial_2\left(\frac{\partial_2 E}{E}\right) \\
&+\frac{1}{4}\left(-\frac{\partial_1 E}{E}\frac{\partial_1 G}{E} + \frac{\partial_2 E}{E}\frac{\partial_2 G}{G}\right) - \frac{1}{4}\left(\frac{\partial_2 E}{E}\frac{\partial_2 E}{E} - \frac{\partial_1 G}{E}\frac{\partial_1 G}{G}\right).
\end{aligned} \tag{3.38}$$

Now $R_{1212} = E\ R^1{}_{212}$ in this case, and (3.34) yields

$$k(x) = \frac{1}{EG}\,R_{1212} = \frac{1}{G}\,R^1{}_{212}. \tag{3.39}$$

If we divide (3.38) by G and then in the resulting formula for $k(x)$ interchange E and G, and ∂_1 and ∂_2, and sum the two formulas for $k(x)$, we get

$$\begin{aligned}
k(x) = &-\frac{1}{4}\left[\frac{1}{G}\partial_1\left(\frac{\partial_1 G}{E}\right) + \frac{1}{E}\partial_1\left(\frac{\partial_1 G}{G}\right)\right] \\
&-\frac{1}{4}\left[\frac{1}{E}\partial_2\left(\frac{\partial_2 E}{G}\right) + \frac{1}{G}\partial_2\left(\frac{\partial_2 E}{E}\right)\right],
\end{aligned}$$

which is easily transformed into (3.37).

If $E = G = e^{2v}$, we obtain a formula for the Gauss curvature of a surface whose metric is a conformal multiple of the flat metric:

Corollary 3.5. *Suppose* dim $M = 2$ *and the metric is given in coordinates by*

$$g_{jk}(x) = e^{2v}\delta_{jk}, \tag{3.40}$$

for a smooth v. *Then the Gauss curvature* $k(x)$ *is given by*

$$k(x) = -(\Delta_0 v)e^{-2v}, \tag{3.41}$$

where Δ_0 *is the flat Laplacian in these coordinates:*

$$\Delta_0 v = \frac{\partial^2 v}{\partial x_1^2} + \frac{\partial^2 v}{\partial x_2^2}. \tag{3.42}$$

For an alternative formulation of (3.41), note that the Laplace operator for the metric g_{jk} is given by

$$\Delta f = g^{-1/2}\,\partial_j(g^{jk}g^{1/2}\,\partial_k f),$$

and in the case (3.40), $g^{jk} = e^{-2v}\delta^{jk}$ and $g^{1/2} = e^{2v}$, so we have

$$\Delta f = e^{-2v}\Delta_0 f, \tag{3.43}$$

and hence (3.41) is equivalent to

$$k(x) = -\Delta v. \tag{3.44}$$

The comparison of the Gauss curvature of two surfaces that are conformally equivalent is a source of a number of interesting results. The following generalization of Corollary 3.5 is useful.

Proposition 3.6. *Let* M *be a two-dimensional manifold with metric* g, *whose Gauss curvature is* $k(x)$. *Suppose there is a conformally related metric*

$$g' = e^{2u}g. \tag{3.45}$$

Then the Gauss curvature $K(x)$ *of* g' *is given by*

$$K(x) = \big(-\Delta u + k(x)\big)e^{-2u}, \tag{3.46}$$

where Δ *is the Laplace operator for the metric* g.

Proof. We will use Corollary 3.5 as a tool in this proof. It is shown in Chap. 5, §11, that (M, g) is locally conformally flat, so we can assume without loss of generality that (3.40) holds; hence $k(x)$ is given by (3.41). Then

$$(g')_{jk} = e^{2w}\delta_{jk}, \quad w = u + v, \tag{3.47}$$

and (3.41) gives

$$K(x) = -(\Delta_0 w)e^{-2w} = \big[-(\Delta_0 u)e^{-2v} - (\Delta_0 v)e^{-2v}\big]e^{-2u}. \tag{3.48}$$

By (3.43) we have $(\Delta_0 u)e^{-2v} = \Delta u$, and applying (3.41) for $k(x)$ gives (3.46).

We end this section with a study of $\partial_j\partial_k g_{\ell m}(p_0)$ when one uses a geodesic normal coordinate system centered at p_0. We know from §11 of Chap. 1 that in such a coordinate system, $\Gamma^\ell{}_{jk}(p_0) = 0$ and hence $\partial_j g_{k\ell}(p_0) = 0$. Thus, in such a coordinate system, we have

$$R^j{}_{k\ell m}(p_0) = \partial_\ell \Gamma^j{}_{km}(p_0) - \partial_m \Gamma^j{}_{k\ell}(p_0), \tag{3.49}$$

and hence (3.3) yields

$$R_{jk\ell m}(p_0) = \frac{1}{2}\left(\partial_j\partial_m g_{k\ell} + \partial_k\partial_\ell g_{jm} - \partial_j\partial_\ell g_{km} - \partial_k\partial_m g_{j\ell}\right). \tag{3.50}$$

In light of the complexity of this formula, the following may be somewhat surprising. Namely, as Riemann showed, one has

$$\partial_j\partial_k g_{\ell m}(p_0) = -\frac{1}{3}R_{\ell jmk} - \frac{1}{3}R_{\ell kmj}. \tag{3.51}$$

This is related to the existence of nonobvious symmetries at the center of a geodesic normal coordinate system, such as $\partial_j\partial_k g_{\ell m}(p_0) = \partial_\ell\partial_m g_{jk}(p_0)$. To prove (3.51), by polarization it suffices to establish

$$\partial_j^2 g_{\ell\ell}(p_0) = -\frac{2}{3}R_{\ell j\ell j}, \quad \forall\, j, \ell. \tag{3.52}$$

Proving this is a two-dimensional problem, since (by (3.50)) both sides of the asserted identity in (3.52) are unchanged if M is replaced by the image under Exp_p of the two-dimensional linear span of D_j and D_ℓ. All one needs to show is that if $\dim M = 2$,

$$\partial_1^2 g_{22}(p_0) = -\frac{2}{3}K(p_0) \ \text{ and } \ \partial_1^2 g_{11}(p_0) = 0, \tag{3.53}$$

where $K(p_0)$ is the Gauss curvature of M at p_0. Of these, the second part is trivial, since $g_{11}(x) = 1$ on the horizontal line through p_0. To establish the first part of (3.53), it is convenient to use geodesic polar coordinates, (r, θ), in which

$$ds^2 = dr^2 + G(r,\theta)\, d\theta^2. \tag{3.54}$$

It is not hard to show that $G(r,\theta) = r^2 H(r,\theta)$, with $H(r,\theta) = 1 + O(r^2)$. For the metric (3.54), the formula (3.37) implies that the Gauss curvature is

$$K = -\frac{1}{2G}\partial_r^2 G + \frac{1}{4G^2}(\partial_r G)^2 = -\frac{H_r}{rH} - \frac{H_{rr}}{2H} + \frac{H_r^2}{4H^2}, \tag{3.55}$$

so at the center

$$K(p_0) = -H_{rr} - \frac{1}{2}H_{rr} = -\frac{3}{2}H_{rr}. \tag{3.56}$$

On the other hand, in normal coordinates (x_1, x_2), along the x_1-axis, we have $g_{22}(s,0) = G(s,0)/s^2 = H(s,0)$, so the rest of the identity (3.53) is established.

Exercises

Exercises 1–3 concern the problem of producing two-dimensional surfaces with constant Gauss curvature.

1. For a two-dimensional Riemannian manifold M, take geodesic polar coordinates, so the metric is

$$ds^2 = dr^2 + G(r,\theta)\,d\theta^2.$$

Use the formula (3.55) for the Gauss curvature, to deduce that

$$K = -\frac{\partial_r^2\sqrt{G}}{\sqrt{G}}.$$

Hence, if $K = -1$, then

$$\partial_r^2\sqrt{G} = \sqrt{G}.$$

Show that

$$\sqrt{G}(0,\theta) = 0, \quad \partial_r\sqrt{G}(0,\theta) = 1,$$

and deduce that $\sqrt{G}(r,\theta) = \varphi(r)$ is the unique solution to

$$\varphi''(r) - \varphi(r) = 0, \quad \varphi(0) = 0,\ \varphi'(0) = 1.$$

Deduce that

$$G(r,\theta) = \sinh^2 r.$$

Use this computation to deduce that any two surfaces with Gauss curvature -1 are locally isometric.

2. Suppose M is a surface of revolution in $\mathbb{R}^3$, of the form

$$x^2 + y^2 = g(z)^2.$$

If it is parameterized by $x = g(u)\ \cos v,\ y = g(u)\ \sin v,\ z = u$, then

$$ds^2 = \bigl(1 + g'(u)^2\bigr)\,du^2 + g(u)^2\,dv^2.$$

Deduce from (3.37) that

$$K = -\frac{g''(u)}{g(u)\bigl(1 + g'(u)^2\bigr)^2}.$$

Hence, if $K = -1$,

$$g''(u) = g(u)\bigl(1 + g'(u)^2\bigr)^2.$$

Note that a sphere of radius R is given by such a formula with $g(u) = \sqrt{R^2 - u^2}$. Compute K in this case.

2A. Suppose instead that M is a surface of revolution, described in the form

$$z = f\left(\sqrt{x^2 + y^2}\right).$$

If it is parameterized by $x = u \cos v,\ y = u \sin v,\ z = f(u)$, then

$$ds^2 = \big(1 + f'(u)^2\big)\, du^2 + u^2\, dv^2.$$

Show that

$$K = -\frac{1}{u\sqrt{1 + f'(u)^2}} \frac{d}{du}\left(\frac{1}{\sqrt{1 + f'(u)^2}}\right) = -\frac{\varphi'(u)}{2u}, \quad \varphi(u) = \frac{1}{1 + f'(u)^2}.$$

Thus deduce that

$$K = -1 \Rightarrow \varphi(u) = u^2 + c \Rightarrow f(u) = \int \sqrt{\frac{1}{u^2 + c} - 1}\, du.$$

We note that this is an elliptic integral, for most values of c. Show that, for $c = 0$, you get

$$f(u) = \sqrt{1 - u^2} - \frac{1}{2}\log\left(1 + \sqrt{1 - u^2}\right) + \frac{1}{2}\log\left(1 - \sqrt{1 - u^2}\right).$$

3. Suppose M is a region in $\mathbb{R}^2$ whose metric tensor is a conformal multiple of the standard flat metric

$$g_{jk} = E(x)\delta_{jk} = e^{2v}\,\delta_{jk}.$$

Suppose $E = E(r),\ v = v(r)$. Deduce from (3.37) and (3.41) that

$$K = -\frac{1}{2E^2}\left(E''(r) + \frac{1}{r}E'(r)\right) + \frac{1}{2E^3}E'(r)^2 = -\left(v''(r) + \frac{1}{r}v'(r)\right)e^{-2v}.$$

Hence, if $K = -1$,

$$v''(r) + \frac{1}{r}v'(r) = e^{2v}.$$

Compute K when

$$g_{jk} = \frac{4}{(1 - r^2)^2}\,\delta_{jk}.$$

4. Show that whenever $g_{jk}(x)$ satisfies $g_{jk}(p_0) = \delta_{jk},\ \partial_\ell g_{jk}(p_0) = 0$, at some point p_0, then (3.50) holds at p_0. If $\dim M = 2$, deduce that

$$K(p_0) = -\frac{1}{2}\big(\partial_1^2 g_{22} + \partial_2^2 g_{11} - 2\partial_1\partial_2 g_{12}\big). \tag{3.57}$$

5. Suppose $M \subset \mathbb{R}^3$ is the graph of

$$x_3 = f(x_1, x_2),$$

so, using the natural (x_1, x_2)-coordinates on M,

$$ds^2 = (1 + f_1^2)\, dx_1^2 + 2 f_1 f_2\, dx_1\, dx_2 + (1 + f_2^2)\, dx_2^2,$$

where $f_j = \partial_j f$. Show that if $\nabla f(0) = 0$, then Exercise 4 applies, so

$$\nabla f(0) = 0 \Longrightarrow K(0) = f_{11} f_{22} - f_{12}^2. \tag{3.58}$$

Compare the derivation of (4.22) in the next section.

6. If $M \subset \mathbb{R}^3$ is the surface of Exercise 5, then the *Gauss map* $N : M \to S^2$ is given by
$$N\big(x, f(x)\big) = \frac{(-f_1, -f_2, 1)}{\sqrt{1 + f_1^2 + f_2^2}}.$$
Show that if $\nabla f(0) = 0$, then, at $p_0 = \big(0, f(0)\big)$, $DN(p_0) : \mathbb{R}^2 \to \mathbb{R}^2$ is given by
$$DN(p_0) = -\begin{pmatrix} \partial_1^2 f(0) & \partial_1 \partial_2 f(0) \\ \partial_2 \partial_1 f(0) & \partial_2^2 f(0). \end{pmatrix} \tag{3.59}$$
Here, $T_{p_0} M$ and $T_{(0,0,1)} S^2$ are both identified with the (x_1, x_2)-plane. Deduce from Exercise 5 that
$$K(p_0) = \det DN(p_0).$$

7. Deduce from Exercise 6 that whenever M is a smooth surface in $\mathbb{R}^3$, with Gauss map $N : M \to S^2$, then, with $DN(x) : T_x M \to T_{N(x)} S^2$,
$$K(x) = \det DN(x), \quad \forall\, x \in M. \tag{3.60}$$
(*Hint*: Given $x \in M$, rotate coordinates so that $T_x M$ is parallel to the (x_1, x_2)-plane.) This result is Gauss' *Theorema Egregium* for surfaces in $\mathbb{R}^3$. See Theorem 4.4 for a more general formulation; see also (4.35), and Exercises 5, 8, 9, and 14 of §4.

8. Recall from §11 of Chap. 1 that if $\gamma_s(t)$ is a family of curves $\gamma_s : [a, b] \to M$ satisfying $\gamma_s(a) = p$, $\gamma_s(b) = q$, and if $E(s) = \int_a^b \langle T, T\rangle dt$, $T = \gamma_s'(t)$, then, with $V = (\partial/\partial s)\gamma_s(t)|_{s=0}$, $E'(s) = -2\int_a^b \langle V, \nabla_T T\rangle\, dt$, leading to the stationary condition for E that $\nabla_T T = 0$, which is the geodesic equation. Now suppose $\gamma_{r,s}(t)$ is a two-parameter family of curves, $\gamma_{r,s}(a) = p$, $\gamma_{r,s}(b) = q$. Let $V = (\partial/\partial s)\gamma_{r,s}(t)|_{0,0}$, $W = (\partial/\partial r)\gamma_{r,s}(t)|_{0,0}$. Show that
$$\frac{\partial^2}{\partial s \partial r} E(0,0) = 2\int_a^b \big[\langle R(W,T)V, T\rangle + \langle \nabla_T V, \nabla_T W\rangle - \langle \nabla_W V, \nabla_T T\rangle\big]\, dt. \tag{3.61}$$
Note that the last term in the integral vanishes if $\gamma_{0,0}$ is a geodesic.

9. If Z is a Killing field, generating an isometry on M (as in Chap. 2, §3), show that
$$Z_{j;k;\ell} = R^m{}_{\ell k j}\, Z_m.$$
(*Hint*: From Killing's equation $Z_{j;k} + Z_{k;j} = 0$, derive $Z_{j;k;\ell} = -Z_{k;\ell;j} - R^m{}_{k\ell j}\, Z_m$. Iterate this process two more times, going through the cyclic permutations of (j, k, ℓ). Use Bianchi's first identity.) Note that the identity desired is equivalent to
$$\nabla^2_{(X,Y)} Z = R(Y, Z)X \quad \text{if } Z \text{ is a Killing field.}$$

10. Derive the following equation of Jacobi for a variation of geodesics. If $\gamma_s(t)$ is a one-parameter family of geodesics, $X = \gamma_s'(t)$, and $W = (\partial/\partial s)\gamma_s$, then
$$\nabla_X \nabla_X\, W = R(X, W)X.$$
(*Hint*: Start with $0 = \nabla_W \nabla_X X$, and use $[X, W] = 0$.)

11. Raising the second index of $R^j{}_{k\ell m}$, you obtain $R^{jk}{}_{\ell m}$, the coordinate expression for $\mathcal{R}$, which can be regarded as a section of $\mathrm{End}(\Lambda^2 T)$. Suppose $M = X \times Y$ with a product Riemannian metric and associated curvatures $\mathcal{R}, \mathcal{R}_X, \mathcal{R}_Y$. Using the splitting
$$\Lambda^2(V \oplus W) = \Lambda^2 V \oplus \big(\Lambda^1 V \otimes \Lambda^1 W\big) \oplus \Lambda^2 W,$$
write $\mathcal{R}$ as a 3×3 block matrix. Show that
$$\mathcal{R} = \begin{pmatrix} \mathcal{R}_X & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathcal{R}_Y \end{pmatrix}.$$

In Exercises 12–14, let X, Y, Z, and so forth, belong to the space $\mathfrak{g}$ of left-invariant vector fields on a Lie group G, assumed to have a bi-invariant Riemannian metric. (Compact Lie groups have these.)

12. Show that any (constant-speed) geodesic γ on G with $\gamma(0) = e$, the identity element, is a subgroup of G, that is, $\gamma(s+t) = \gamma(s)\gamma(t)$. Deduce that $\nabla_X X = 0$ for $X \in \mathfrak{g}$. (*Hint*: Given $p = \gamma(t_0)$, consider the "reflection" $R_p(g) = pg^{-1}p$, an isometry on G that fixes p and leaves γ invariant, though reversing its direction. From this, one can deduce that $p^2 = \gamma(2t_0)$.)
13. Show that $\nabla_X Y = (1/2)[X, Y]$ for $X, Y \in \mathfrak{g}$. (*Hint*: $0 = \nabla_X X = \nabla_Y Y = \nabla_{(X+Y)}(X+Y)$.)
This identity is called the *Maurer-Cartan* structure equation.
14. Show that
$$R(X,Y)Z = -\frac{1}{4}[[X,Y],Z], \quad \langle R(X,Y)Z, W\rangle = -\frac{1}{4}\big\langle [X,Y],[Z,W]\big\rangle.$$

15. If $E \to M$ is a vector bundle with connection $\widetilde{\nabla}$, and $\nabla = \widetilde{\nabla} + C$, as in Exercises 1 and 2 of §1, and M has Levi–Civita connection D, so that $\mathrm{Hom}(T \otimes E, E)$ acquires a connection from D and $\widetilde{\nabla}$, which we'll also denote as $\widetilde{\nabla}$, show that (1.38) is equivalent to
$$(R - \widetilde{R})(X,Y)u = (\widetilde{\nabla}_X C)(Y,u) - (\widetilde{\nabla}_Y C)(X,u) + [C_X, C_Y]u. \tag{3.62}$$
This is a general form of the "Palatini identity."
16. If g is a metric tensor and h a symmetric, second-order tensor field, consider the family of metric tensors $g_\tau = g + \tau h$, for τ close to zero, yielding the Levi–Civita connections
$$\nabla^\tau = \nabla + C(\tau),$$
where $\nabla = \nabla^0$. If $C' = C'(0)$, show that
$$\langle C'(X,Y), Z\rangle = \frac{1}{2}(\nabla_X h)(Y,Z) + \frac{1}{2}(\nabla_Y h)(X,Z) - \frac{1}{2}(\nabla_Z h)(X,Y). \tag{3.63}$$
(*Hint*: Use (3.2).)
17. Let $R(\tau)$ be the Riemann curvature tensor of g_τ, and set $R' = R'(0)$. Show that (3.62) yields
$$R'(X,Y)Z = (\nabla_X C')(Y,Z) - (\nabla_Y C')(X,Z). \tag{3.64}$$

Using (3.63), show that

$$2\langle R'(X,Y)Z,W\rangle = (\nabla^2_{Y,W}h)(X,Z) + (\nabla^2_{X,Z}h)(Y,W) - (\nabla^2_{X,W}h)(Y,Z) - (\nabla^2_{Y,Z}h)(X,W) + h\big(R(X,Y)Z,W\big) + h\big(R(X,Y)W,Z\big). \tag{3.65}$$

(*Hint*: Use the derivation property of the covariant derivative to obtain a formula for $\nabla_X C'$ from (3.63).)

18. Show that

$$\begin{aligned} 6\langle R(X,Y)Z,W\rangle = {} & \tilde{K}(X+W,Y+Z) - \tilde{K}(Y+W,X+Z) \\ & - \tilde{K}(X,Y+Z) - \tilde{K}(Y,X+W) - \tilde{K}(Z,X+W) \\ & - \tilde{K}(W,Y+Z) + \tilde{K}(X,Y+W) + \tilde{K}(Y,Z+W) \\ & + \tilde{K}(Z,Y+W) + \tilde{K}(W,X+Z) + \tilde{K}(X,Z) \\ & + \tilde{K}(Y,W) - \tilde{K}(X,Y) - \tilde{K}(Y,Z), \end{aligned} \tag{3.66}$$

where

$$\tilde{K}(X,Y) = \langle R(X,Y)Y,X\rangle. \tag{3.67}$$

See (4.34) for an interpretation of the right side of (3.67).

19. Using (3.51), show that, in exponential coordinates centered at p, the function $g = \det(g_{jk})$ satisfies, for $|x|$ small,

$$g(x) = 1 - \frac{1}{3}\sum_{\ell,m} \mathrm{Ric}_{\ell m}(p)x_\ell x_m + O\big(|x|^3\big). \tag{3.68}$$

Deduce that if A_{n-1} = area of $S^{n-1} \subset \mathbb{R}^n$ and V_n = volume of unit ball in $\mathbb{R}^n$, then, for r small,

$$V\big(B_r(p)\big) = \left(V_n - \frac{A_{n-1}}{6n(n+2)}S(p)r^2 + O(r^3)\right)r^n. \tag{3.69}$$

4. Geometry of submanifolds and subbundles

Let M be a Riemannian manifold, of dimension n, and let S be a submanifold, of dimension k, with the induced metric tensor. M has a Levi–Civita connection ∇ and Riemann tensor R. Denote by ∇^0 and R_S the connection and curvature of S, respectively. We aim to relate these objects. The *second fundamental form* is defined by

$$II(X,Y) = \nabla_X Y - \nabla^0_X Y, \tag{4.1}$$

for X and Y tangent to S. Note that II is linear in X and in Y over $C^\infty(S)$. Also, by the zero-torsion condition,

$$II(X,Y) = II(Y,X). \tag{4.2}$$

Proposition 4.1. *$II(X,Y)$ is normal to S at each point.*

Proof. If X, Y and Z are tangent to S, we have

$$\langle \nabla_X Y, Z\rangle - \langle \nabla^0_X Y, Z\rangle = -\langle Y, \nabla_X Z\rangle + X\langle Y, Z\rangle + \langle Y, \nabla^0_X Z\rangle - X\langle Y, Z\rangle,$$

and making the obvious cancelation, we obtain

$$\langle II(X,Y), Z\rangle = -\langle Y, II(X,Z)\rangle. \tag{4.3}$$

Using (4.2), we have

$$\langle II(X,Y), Z\rangle = -\langle Y, II(Z,X)\rangle; \tag{4.4}$$

that is, the trilinear form given by the left side changes sign under a cyclic permutation of its arguments. Since three such permutations produce the original form, the left side of (4.4) must equal its own negative, hence be 0. This proves the proposition.

Denote by $\nu(S)$ the bundle of normal vectors to S, called the normal bundle of S. It follows that II is a section of $\mathrm{Hom}(TS \otimes TS, \nu(S))$.

Corollary 4.2. *For X and Y tangent to S, $\nabla^0_X Y$ is the tangential projection on TS of $\nabla_X Y$.*

Let ξ be normal to S. We have a linear map, called the *Weingarten map*,

$$A_\xi : T_pS \longrightarrow T_pS \tag{4.5}$$

uniquely defined by

$$\langle A_\xi X, Y\rangle = \langle \xi, II(X,Y)\rangle. \tag{4.6}$$

We also define the section A of $\mathrm{Hom}(\nu(S)\otimes TS, TS)$ by

$$A(\xi, X) = A_\xi X. \tag{4.7}$$

We define a connection on $\nu(S)$ as follows; if ξ is a section of $\nu(S)$, set

$$\nabla^1_X \xi = P^\perp \nabla_X \xi,$$

where $P^\perp(x)$ is the orthogonal projection of T_xM onto $\nu_x(S)$. The following identity is called the *Weingarten formula*.

Proposition 4.3. *If ξ is a section of $\nu(S)$,*

$$\nabla^1_X \xi = \nabla_X \xi + A_\xi X. \tag{4.8}$$

Proof. It suffices to show that $\nabla_X \xi + A_\xi X$ is normal to S. In fact, if Y is tangent to S,

$$\begin{aligned}\langle \nabla_X \xi, Y\rangle + \langle A_\xi X, Y\rangle &= X\langle \xi, Y\rangle - \langle \xi, \nabla_X Y\rangle + \langle \xi, II(X,Y)\rangle \\ &= 0 - \langle \xi, \nabla^0_X Y\rangle - \langle \xi, II(X,Y)\rangle + \langle \xi, II(X,Y)\rangle \\ &= 0,\end{aligned}$$

which proves the proposition.

An equivalent statement is that, for X tangent to S, ξ normal to S,

$$\nabla_X \xi = \nabla^1_X \xi - A_\xi X \tag{4.9}$$

is an orthogonal decomposition, into components normal and tangent to S, respectively. Sometimes this is taken as the definition of A_ξ or, equivalently, by (4.6), of the second fundamental form.

In the special case that S is a hypersurface of M (i.e., dim M = dim $S + 1$), if $\xi = N$ is a smooth unit normal field to S, we see that, for X tangent to S,

$$\langle \nabla_X N, N\rangle = \frac{1}{2} X\langle N, N\rangle = 0,$$

so $\nabla^1_X N = 0$ in this case, and (4.9) takes the form

$$\nabla_X N = -A_N X, \tag{4.10}$$

the classical form of the Weingarten formula.

We now compare the tensors R and R_S. Let X, Y and Z be tangent to S. Then

$$\begin{aligned}\nabla_X \nabla_Y Z &= \nabla_X\big(\nabla^0_Y Z + II(Y,Z)\big) \\ &= \nabla^0_X \nabla^0_Y Z + II(X, \nabla^0_Y Z) - A_{II(Y,Z)} X + \nabla^1_X II(Y,Z).\end{aligned} \tag{4.11}$$

Reversing X and Y, we have

$$\nabla_Y \nabla_X Z = \nabla^0_Y \nabla^0_X Z + II(Y, \nabla^0_X Z) - A_{II(X,Z)} Y + \nabla^1_Y II(X,Z).$$

Also,

$$\nabla_{[X,Y]} Z = \nabla^0_{[X,Y]} Z + II([X,Y], Z). \tag{4.12}$$

From (4.11) and (4.12) we obtain the important identity

$$\begin{aligned}(R - R_S)(X,Y)Z = &\big\{II(X, \nabla^0_Y Z) - II(Y, \nabla^0_X Z) - II([X,Y], Z) \\ &+ \nabla^1_X II(Y,Z) - \nabla^1_Y II(X,Z)\big\} \\ &- \big\{A_{II(Y,Z)} X - A_{II(X,Z)} Y\big\}.\end{aligned} \tag{4.13}$$

Here, the quantity in the first set of braces $\{\quad\}$ is normal to S, and the quantity in the second pair of braces is tangent to S. The identity (4.13) is called the *Gauss–Codazzi equation*. A restatement of the identity for the tangential components is the following, known as Gauss' *Theorem Egregium*.

Theorem 4.4. *For X, Y, Z and W tangent to S,*

$$\begin{aligned}\langle (R - R_S)(X,Y)Z, W\rangle = \langle II(Y,W), II(X,Z)\rangle \\ - \langle II(X,W), II(Y,Z)\rangle.\end{aligned} \tag{4.14}$$

The normal component of the identity (4.13) is specifically Codazzi's equation. It takes a shorter form in case S has codimension 1 in M. In that case, choose a unit normal field N, and let

$$II(X,Y) = \widetilde{II}(X,Y)N; \tag{4.15}$$

$\widetilde{II}$ is a tensor field of type $(0,2)$ on S. Then Codazzi's equation is equivalent to

$$\langle R(X,Y)Z, N\rangle = (\nabla^0_X \widetilde{II})(Y,Z) - (\nabla^0_Y \widetilde{II})(X,Z), \tag{4.16}$$

for X, Y, Z tangent to S, since of course $R_S(X,Y)Z$ is tangent to S.

In the classical case, where S is a hypersurface in flat Euclidean space, $R = 0$, and Codazzi's equation becomes

$$(\nabla^0_Y \widetilde{II})(X,Z) - (\nabla^0_X \widetilde{II})(Y,Z) = 0, \tag{4.17}$$

that is, $\nabla^0 \widetilde{II}$ is a symmetric tensor field of type $(0,3)$. In this case, from the identity $\widetilde{II}_{jk;\ell} = \widetilde{II}_{\ell k;j}$, we deduce $A_j{}^k{}_{;k} = A_k{}^k{}_{;j} = (\mathrm{Tr}\, A)_{;j}$, where $A = A_N$ is the Weingarten map. Equivalently,

$$\operatorname{div} A = d(\mathrm{Tr}\, A). \tag{4.18}$$

An application of the Codazzi equation to minimal surfaces can be found in the exercises after §6 of Chap. 14.

It is useful to note the following characterization of the second fundamental form for a hypersurface M in $\mathbb{R}^n$. Translating and rotating coordinates, we can move a specific point $p \in M$ to the origin in $\mathbb{R}^n$ and suppose M is given locally by

$$x_n = f(x'), \quad \nabla f(0) = 0,$$

where $x' = (x_1, \dots, x_{n-1})$. We can then identify the tangent space of M at p with $\mathbb{R}^{n-1}$.

Proposition 4.5. *The second fundamental form of M at p is given by the Hessian of f:*

$$\widetilde{II}(X,Y)=\sum_{j,k=1}^{n-1}\frac{\partial^2 f}{\partial x_j\,\partial x_k}(0)\;X_jY_k. \tag{4.19}$$

Proof. From (4.9) we have, for any ξ normal to M,

$$\langle II(X,Y),\xi\rangle=-\langle\nabla_X\xi,Y\rangle, \tag{4.20}$$

where ∇ is the flat connection on $\mathbb{R}^n$. Taking

$$\xi=(-\partial_1 f,\dots,-\partial_{n-1}f,1) \tag{4.21}$$

gives the desired formula.

If S is a surface in $\mathbb{R}^3$, given locally by $x_3=f(x_1,x_2)$ with $\nabla f(0)=0$, then the Gauss curvature of S at the origin is seen by (4.14) and (4.19) to equal

$$\det\left(\frac{\partial^2 f(0)}{\partial x_j\,\partial x_k}\right). \tag{4.22}$$

Consider the example of the unit sphere in $\mathbb{R}^3$, centered at $(0,0,1)$. Then the "south pole" lies at the origin, near which S^2 is given by

$$x_3=1-(1-x_1^2-x_2^2)^{1/2}. \tag{4.23}$$

In this case (4.22) implies that the Gauss curvature K is equal to 1 at the south pole. Of course, by symmetry it follows that $K=1$ everywhere on the unit sphere S^2.

Besides providing a good conception of the second fundamental form of a hypersurface in $\mathbb{R}^n$, Proposition 4.5 leads to useful formulas for computation, one of which we will give in (4.29). First, we give a more invariant reformulation of Proposition 4.5. Suppose the hypersurface M in $\mathbb{R}^n$ is given by

$$u(x)=c, \tag{4.24}$$

with $\nabla u\neq 0$ on M. Then we can use the computation (4.20) with $\xi=$ grad u to obtain

$$\langle II(X,Y),\text{grad }u\rangle=-(D^2u)(X,Y), \tag{4.25}$$

where D^2u is the Hessian of u; we can think of $(D^2u)(X,Y)$ as $Y\cdot(D^2u)X$, where D^2u is the $n\times n$ matrix of second-order partial derivatives of u. In other words,

$$\widetilde{II}(X,Y)=-|\text{grad }u|^{-1}(D^2u)(X,Y), \tag{4.26}$$

for X,Y tangent to M.

In particular, if M is a two-dimensional surface in $\mathbb{R}^3$ given by (4.24), then the Gauss curvature at $p \in M$ is given by

$$K(p) = |\text{grad}\, u|^{-2} \det (D^2 u)|_{T_pM}, \tag{4.27}$$

where $D^2 u|_{T_pM}$ denotes the restriction of the quadratic form $D^2 u$ to the tangent space T_pM, producing a linear transformation on T_pM via the metric on T_pM. With this calculation we can derive the following formula, extending (4.22).

Proposition 4.6. *If $M \subset \mathbb{R}^3$ is given by*

$$x_3 = f(x_1, x_2), \tag{4.28}$$

then, at $p = (x', f(x')) \in M$, the Gauss curvature is given by

$$K(p) = \bigl(1 + |\nabla f(x')|^2\bigr)^{-2} \, det \left(\frac{\partial^2 f}{\partial x_j \partial x_k} \right). \tag{4.29}$$

Proof. We can apply (4.27) with $u(x) = f(x_1, x_2) - x_3$. Note that $|\nabla u|^2 = 1 + |\nabla f(x')|^2$ and

$$D^2 u = \begin{pmatrix} D^2 f & 0 \\ 0 & 0 \end{pmatrix}. \tag{4.30}$$

Noting that a basis of T_pM is given by $(1, 0, \partial_1 f) = v_1,\ (0, 1, \partial_2 f) = v_2$, we readily obtain

$$\det D^2 u|_{T_pM} = \frac{\det\bigl(v_j \cdot (D^2 u) v_k\bigr)}{\det(v_j \cdot v_k)} = \bigl(1 + |\nabla f(x')|^2\bigr)^{-1} \det D^2 f, \tag{4.31}$$

which yields (4.29).

If you apply Proposition 4.6 to the case (4.23) of a hemisphere of unit radius, the calculation that $K = 1$ everywhere is easily verified. The formula (4.29) gives rise to interesting problems in nonlinear PDE, some of which are studied in Chap. 14.

We now define the sectional curvature of a Riemannian manifold M. Given $p \in M$, let Π be a 2-plane in T_pM, $\Sigma = \text{Exp}_p(\Pi)$. The sectional curvature of M at p is

$$K_p(\Pi) = \text{Gauss curvature of } \Sigma \text{ at } p. \tag{4.32}$$

If U and V form an orthonormal basis of $T_p\Sigma = \Pi$, then by the definition of Gauss curvature,

$$K_p(\Pi) = \langle R_\Sigma(U, V)V, U \rangle. \tag{4.33}$$

We have the following more direct formula for the sectional curvature.

Proposition 4.7. *With U and V as above, R the Riemann tensor of M,*

$$K_p(\Pi) = \langle R(U,V)V,U\rangle. \tag{4.34}$$

Proof. It suffices to show that the second fundamental form of Σ vanishes at p. Since $II(X,Y)$ is symmetric, it suffices to show that $II(X,X) = 0$ for each $X \in T_pM$. So pick a geodesic γ in M such that $\gamma(0) = p$, $\gamma'(0) = X$. Then $\gamma \subset \Sigma$, and γ must also be a geodesic in S, so

$$\nabla_T T = \nabla^0_T T, \quad T = \gamma'(t),$$

which implies $II(X,X) = 0$. This proves (4.34).

Note that if $S \subset M$ has codimension 1, $p \in S$, and $\Pi \subset T_pS$, then, by (4.14),

$$K^S_p(\Pi) - K^M_p(\Pi) = \det\begin{pmatrix} \widetilde{II}(U,U) & \widetilde{II}(U,V) \\ \widetilde{II}(V,U) & \widetilde{II}(V,V) \end{pmatrix}. \tag{4.35}$$

Note how this is a direct generalization of (3.60).

The results above comparing connections and curvatures of a Riemannian manifold and a submanifold are special cases of more general results on subbundles, which arise in a number of interesting situations. Let E be a vector bundle over a manifold M, with an inner product and a metric connection ∇. Let $E_0 \to M$ be a subbundle. For each $x \in M$, let P_x be the orthogonal projection of E_x onto E_{0x}. Set

$$\nabla^0_X u(x) = P_x \nabla_X u(x), \tag{4.36}$$

when u is a section of E_0. Note that, for scalar f,

$$\begin{aligned} \nabla^0_X fu(x) &= P_x\big(f\nabla_X u(x) + (Xf)u\big) \\ &= fP_x\nabla_X u(x) + (Xf)u(x), \end{aligned}$$

provided u is a section of E_0, so $P_x u(x) = u(x)$. This shows that (4.36) defines a connection on E_0. Since $\langle \nabla^0_X u, v\rangle = \langle \nabla_X u, v\rangle$ for sections u, v of E_0, it is clear that ∇^0 is also a metric connection. Similarly, if E_1 is the orthogonal bundle, a subbundle of E, a metric connection on E_1 is given by

$$\nabla^1_X v(x) = (I - P_x)\nabla_X v(x), \tag{4.37}$$

for a section v of E_1.

It is useful to treat ∇^0 and ∇^1 on an equal footing, so we define a new connection $\widetilde{\nabla}$ on E, also a metric connection, by

$$\widetilde{\nabla} = \nabla^0 \oplus \nabla^1. \tag{4.38}$$

Then there is the relation

$$\nabla_X = \widetilde{\nabla}_X + C_X, \tag{4.39}$$

where

$$C_X = \begin{pmatrix} 0 & II_X^1 \\ II_X^0 & 0 \end{pmatrix}. \tag{4.40}$$

Here, $II_X^0 : E_0 \to E_1$ is the second fundamental form of $E_0 \subset E$, and $II_X^1 : E_1 \to E_0$ is the second fundamental form of $E_1 \subset E$. We also set $II^j(X, u) = II_X^j u$. In this context, the Weingarten formula has the form

$$C_X^t = -C_X, \ \text{ i.e., } II_X^1 = -\big(II_X^0\big)^t. \tag{4.41}$$

Indeed, for any two connections related by (4.39), with $C \in \operatorname{Hom}(TM \otimes E, E)$, if ∇ and $\widetilde{\nabla}$ are both metric connections, the first part of (4.41) holds.

We remark that when γ is a curve in a Riemannian manifold M, and for $p \in \gamma$, $E_p = T_pM$, $E_{0p} = T_p\gamma$, $E_{1p} = \nu(\gamma)$, the normal space, and if ∇ is the Levi–Civita connection on M, then $\widetilde{\nabla}$ is sometimes called the *Fermi-Walker connection* on γ. One also (especially) considers a timelike curve in a Lorentz manifold.

Let us also remark that if we start with metric connections ∇^j on E_j, then form $\widetilde{\nabla}$ on E by (4.38), and then define ∇ on E by (4.39), provided that (4.40) holds, it follows that ∇ is also a metric connection on E, and the connections ∇^j are recovered by (4.36) and (4.37).

In general, for any two connections ∇ and $\widetilde{\nabla}$, related by (4.39) for some $\operatorname{End}(E)$ valued 1-form C, we have the following relation between their curvature tensors R and $\widetilde{R}$, already anticipated in Exercise 2 of §1:

$$(R - \widetilde{R})(X, Y)u = \big\{[C_X, \widetilde{\nabla}_Y] - [C_Y, \widetilde{\nabla}_X] - C_{[X,Y]}\big\}u + [C_X, C_Y]u. \tag{4.42}$$

In case $\widetilde{\nabla} = \nabla^0 \oplus \nabla^1$ on $E = E_0 \oplus E_1$, and ∇ has the form (4.39), where C_X exchanges E_0 and E_1, it follows that the operator in brackets $\{\quad\}$ on the right side of (4.42) exchanges sections of E_0 and E_1, while the last operator $[C_X, C_Y]$ leaves invariant the sections of E_0 and E_1. In such a case these two components express respectively the Codazzi identity and Gauss' *Theorema Egregium*.

We will expand these formulas, writing $R(X, Y) \in \operatorname{End}(E_0 \oplus E_1)$ in the block matrix form

$$R = \begin{pmatrix} R_{00} & R_{01} \\ R_{10} & R_{11} \end{pmatrix}. \tag{4.43}$$

Then Gauss' equations become

$$(4.44)\qquad \begin{aligned}(R_{00}-R_0)(X,Y)u &= II_X^1 II_Y^0 u - II_Y^1 II_X^0 u,\\ (R_{11}-R_1)(X,Y)u &= II_X^0 II_Y^1 u - II_Y^0 II_X^1 u,\end{aligned}$$

for a section u of E_0 or E_1, respectively. Equivalently, if v is also a section of E_0 or E_1, respectively,

$$(4.45)\qquad \begin{aligned}\langle (R_{00}-R_0)(X,Y)u, v\rangle &= \langle II_X^0 u, II_Y^0 v\rangle - \langle II_Y^0 u, II_X^0 v\rangle,\\ \langle (R_{11}-R_1)(X,Y)u, v\rangle &= \langle II_X^1 u, II_Y^1 v\rangle - \langle II_Y^1 u, II_X^1 v\rangle.\end{aligned}$$

The second part of (4.45) is also called the Ricci equation.

Codazzi's equations become

(4.46)
$$\begin{aligned}R_{10}(X,Y)u &= II_X^0 \nabla_Y^0 u - II_Y^0 \nabla_X^0 u - II_{[X,Y]}^0 u + \nabla_X^1 II_Y^0 u - \nabla_Y^1 II_X^0 u,\\ R_{01}(X,Y)u &= II_X^1 \nabla_Y^1 u - II_Y^1 \nabla_X^1 u - II_{[X,Y]}^1 u + \nabla_X^0 II_Y^1 u - \nabla_Y^0 II_X^1 u,\end{aligned}$$

for sections u of E_0 and E_1, respectively. If we take the inner product of the first equation in (4.46) with a section v of E_1, we get

$$(4.47)\qquad \begin{aligned}&\langle R_{10}(X,Y)u, v\rangle\\ &\quad = -\langle \nabla_Y^0 u, II_X^1 v\rangle + \langle \nabla_X^0 u, II_Y^1 v\rangle - \langle II_{[X,Y]}^0 u, v\rangle\\ &\qquad + \langle II_X^0 u, \nabla_Y^1 v\rangle - \langle II_Y^0 u, \nabla_X^1 v\rangle + X\langle II_Y^0 u, v\rangle - Y\langle II_X^0 u, v\rangle,\end{aligned}$$

using the metric property of ∇^0 and ∇^1, and the antisymmetry of (4.40). If we perform a similar calculation for the second part of (4.46), in light of the fact that $R_{10}(X,Y)^t = -R_{01}(X,Y)$, we see that these two parts are equivalent, so we need retain only one of them. Furthermore, we can rewrite the first equation in (4.46) as follows. Form a connection on $\mathrm{Hom}(TM \otimes E_0, E_1)$ via the connections ∇^j on E_j and a Levi–Civita connection ∇^M on TM, via the natural derivation property, that is,

$$(4.48)\qquad (\widetilde{\nabla}_X II^0)(Y,u) = \nabla_X^1 II_Y^0 u - II_Y^0 \nabla_X^0 u - II^0(\nabla_X^M Y, u).$$

Then (4.46) is equivalent to

$$(4.49)\qquad R_{10}(X,Y)u = (\widetilde{\nabla}_X II^0)(Y,u) - (\widetilde{\nabla}_Y II^0)(X,u).$$

One case of interest is when E_1 is the trivial bundle $E_1 = M \times \mathbb{R}$, with one-dimensional fiber. For example, E_1 could be the normal bundle of a codimension-one surface in $\mathbb{R}^n$. In this case, it is clear that both sides of the last half of (4.45) are tautologically zero, so Ricci's equation has no content in this case.

As a parenthetical comment, suppose E is a trivial bundle $E = M \times \mathbb{R}^n$, with complementary subbundles E_j, having metric connections constructed as in (4.36) and (4.37), from the trivial connection D on E, defined by componentwise differentiation, so

(4.50) $$\nabla^0_X u = P D_X u, \quad \nabla^1_X u = (I - P) D_X u,$$

for sections of E_0 and E_1, respectively. There is the following alternative approach to curvature formulas. For $\widetilde{\nabla} = \nabla^0 \oplus \nabla^1$, we have

(4.51) $$\widetilde{\nabla}_X u = D_X u + (D_X P)(I - 2P)u.$$

Note that with respect to a choice of basis of $\mathbb{R}^n$ as a global frame field on $M \times \mathbb{R}^n$, we have the connection 1-form (1.13) given by

(4.52) $$\Gamma = dP(I - 2P).$$

Since $dP = dP\ P + P\ dP$, we have $dP\ P = (I - P)\,dP$. Thus, writing the connection 1-form as $\Gamma = P\ dP(I - P) - (I - P)\,dP\ P$ casts $\Gamma = -C$ in the form (4.40). We obtain directly from the formula $\Omega = d\Gamma + \Gamma \wedge \Gamma$, derived in (4.15), that the curvature of $\widetilde{\nabla}$ is given by

(4.53) $$\Omega = dP \wedge dP = P\ dP \wedge dP\ P + (I - P)\ dP \wedge dP\ (I - P),$$

the last identity showing the respective curvatures of E_0 and E_1. Compare with Exercise 5 of §1.

Our next goal is to invert the process above. That is, rather than starting with a flat bundle $E = M \times \mathbb{R}^n$ and obtaining connections on subbundles and second fundamental forms, we want to start with bundles $E_j \to M$, $j = 1, 2$, with metric connections ∇^j, and proposed second fundamental forms II^j, sections of $\mathrm{Hom}(TM \otimes E_j, E_{j'})$, and then obtain a flat connection ∇ on E via (4.38)–(4.40). Of course, we assume II^0 and II^1 are related by (4.41), so (4.39) makes ∇ a metric connection. Thus, according to equations (4.45) and (4.49), the connection ∇ is flat if and only if, for all sections u, v of E_0,

(4.54)
$$(\widetilde{\nabla}_X II^0)(Y, u) - (\widetilde{\nabla}_Y II^0)(X, u) = 0,$$
$$\langle R_0(X, Y)u, v\rangle = \langle II^0_Y u, II^0_X v\rangle - \langle II^0_X u, II^0_Y v\rangle,$$

and, for all sections u, v of E_1,

(4.55) $$\langle R_1(X, Y)u, v\rangle = \langle II^1_Y u, II^1_X v\rangle - \langle II^1_X u, II^1_Y v\rangle.$$

If these conditions are satisfied, then E will have a global frame field of sections $e_1, \dots, e_n$, such that $\nabla e_j = 0$, at least provided M is simply connected. Then, for each $p \in M$, we have an isometric isomorphism

(4.56) $$J(p) : E_p \longrightarrow \mathbb{R}^n$$

by expanding elements of E_p in terms of the basis $\{e_j(p)\}$. Thus $E_0 \subset E$ is carried by $J(p)$ to a family of linear subspaces $J(p)E_{0p} = V_p \subset \mathbb{R}^n$, with orthogonal complements $J(p)E_{1p} = N_p \subset \mathbb{R}^n$.

We now specialize to the case $E_0 = TM$, where M is an m-dimensional Riemannian manifold, with its Levi–Civita connection; E_1 is an auxiliary bundle over M, with metric connection ∇^1. We will assume M is simply connected. The following result is sometimes called the *fundamental theorem of surface theory*.

Theorem 4.8. *Let II^0 be a section of Hom$(TM \otimes TM, E_1)$, and set $II^1_X = -(II^0_X)^t$. Make the symmetry hypothesis*

$$II^0(X,Y) = II^0(Y,X). \tag{4.57}$$

Assume the equations (4.54) *and* (4.55) *are satisfied, producing a trivialization of $E = E_0 \oplus E_1$, described by* (4.56)*. Then there is an isometric immersion*

$$X : M \longrightarrow \mathbb{R}^n, \tag{4.58}$$

and a natural identification of E_1 with the normal bundle of $S = X(M) \subset \mathbb{R}^n$, such that the second fundamental form of S is given by II^0.

To get this, we will construct the map (4.58) so that

$$DX(p) = J(p)\big|_{TM}, \tag{4.59}$$

for all $p \in M$. To see how to get this, consider one of the n components of J, $J_\nu(p) : E_p \to \mathbb{R}$. In fact, $J_\nu u = \langle e_\nu, u\rangle$. Let $\beta_\nu(p) = J_\nu(p)\big|_{T_pM}$; thus β_ν is a 1-form on M.

Lemma 4.9. *Each β_ν is closed, that is, $d\beta_\nu = 0$.*

Proof. For vector fields X and Y on M, we have

$$\begin{aligned} d\beta_\nu(X,Y) &= X\cdot\beta_\nu(Y) - Y\cdot\beta_\nu(X) - \beta_\nu([X,Y]) \\ &= X\cdot\beta_\nu(Y) - Y\cdot\beta_\nu(X) - \beta_\nu(\nabla^0_X Y - \nabla^0_Y X). \end{aligned} \tag{4.60}$$

Using $\nabla_X = \nabla^0_X + II^0_X$ on sections of $E_0 = TM$, we see that this is equal to

$$\begin{aligned} X\cdot J_\nu(Y) - Y\cdot J_\nu(X) - J_\nu(\nabla_X Y - \nabla_Y X) + J_\nu(II^0_X Y - II^0_Y X) \\ = (\nabla_X J_\nu)Y - (\nabla_Y J_\nu)X + J_\nu(II^0_X Y - II^0_Y X). \end{aligned}$$

By construction, $\nabla_X J_\nu = 0$, while (4.57) says $II^0_X Y - II^0_Y X = 0$. Thus $d\beta_\nu = 0$.

Consequently, as long as M is simply connected, we can write $\beta_\nu = dx_\nu$ for some functions $x_\nu \in C^\infty(M)$. Let us therefore define the map (4.58) by $X(p) = (x_1(p), \dots, x_\nu(p))$. Thus (4.59) holds, so X is an isometric mapping. Furthermore, it is clear that $J(p)$ maps E_{1p} precisely isometrically onto the normal space $N_p \subset \mathbb{R}^n$ to $S = X(M)$ at $X(p)$, displaying II^0 as the second fundamental form of S. Thus Theorem 4.8 is established.

Let us specialize Theorem 4.8 to the case where dim $M = n-1$, so the fibers of E_1 are one-dimensional. As mentioned above, the Ricci identity (4.55) has no content in that case. We have the following special case of the fundamental theorem of surface theory.

Proposition 4.10. *Let M be an $(n-1)$-dimensional Riemann manifold; assume M is simply connected. Let there be given a symmetric tensor field $\widetilde{II}$, of type (0,2). Assume the following Gauss–Codazzi equations hold:*

$$\langle R^M(X,Y)Z,W\rangle = \widetilde{II}(Y,Z)\widetilde{II}(X,W) - \widetilde{II}(X,Z)\widetilde{II}(Y,W), \tag{4.61}$$
$$\big(\nabla^M_X \widetilde{II}\big)(Y,Z) - \big(\nabla^M_Y \widetilde{II}\big)(X,Z) = 0,$$

where ∇^M is the Levi–Civita connection of M and R^M is its Riemann curvature tensor. Then there is an isometric immersion $X : M \to \mathbb{R}^n$ such that the second fundamental form of $S = X(M) \subset \mathbb{R}^n$ is given by $\widetilde{II}$.

Exercises

1. Let $S \subset M$, with respective Levi–Civita connections ∇^0, ∇, respective Riemann tensors R_s, R, and so on, as in the text. Let $\gamma_{s,t} : [a,b] \to S$ be a two-parameter family of curves. One can also regard $\gamma_{s,t} : [a,b] \to M$. Apply the formula (3.52) for the second variation of energy in these two contexts, and compare the results, to produce another proof of Gauss' formula (4.14) for $\langle (R - R_S)(X,Y)Z,W\rangle$ when X,Y,Z,W are all tangent to S.
2. With the Ricci tensor Ric given by (3.23) and the sectional curvature $K_p(\Pi)$ by (4.32), show that, for $X \in T_pM$, of norm 1, if Ξ denotes the orthogonal complement of X in T_pM, then

$$\mathrm{Ric}(X,X) = \frac{n-1}{\mathrm{vol}\, S_p(\Xi)} \int_{S_p(\Xi)} K_p(U,X)\, dV(U),$$

where $S_p(\Xi)$ is the unit sphere in Ξ, $n =$ dim M, and $K_p(U,X) = K_p(\Pi)$, where Π is the linear span of U and X. Show that the scalar curvature at p is given by

$$S = \frac{n(n-1)}{\mathrm{vol}\, G_2} \int_{G_2} K_p(\Pi)\, dV(\Pi),$$

where G_2 is the space of 2-planes in T_pM.
3. Let γ be a curve in $\mathbb{R}^3$, parameterized by arc length. Recall the Frenet apparatus. At $p = \gamma(t)$, $T = \gamma'(t)$ spans $T_p\gamma$, and if the curvature κ of γ is nonzero, unit vectors N and B span the normal bundle $\nu_p(\gamma)$, satisfying the system of ODE

$$\begin{aligned} T' &= \kappa N, \\ N' &= -\kappa T + \tau B, \\ B' &= -\tau N, \end{aligned} \tag{4.62}$$

and furthermore, $B = T \times N,\ T = N \times B,\ N = B \times T$. Compare with Exercises 4–6 in Chap. 1, §5. Let ∇ denote the standard flat connection on $\mathbb{R}^3$, and $\nabla^0,\ \nabla^1$ the connections induced on $T(\gamma)$ and $\nu(\gamma)$, as in (4.1), (4.8). Show that

$$II(T,T) = \kappa N \tag{4.63}$$

and that

$$\begin{aligned} \nabla^1_T N &= \tau B, \\ \nabla^1_T B &= -\tau N. \end{aligned} \tag{4.64}$$

Compute the right side of the Weingarten formula

$$\nabla_T - \nabla^1_T = -(II_T)^t, \tag{4.65}$$

and show that (4.63)–(4.65) are *equivalent* to (4.62).

4. Let $S \subset \mathbb{R}^3$ be a surface, with connection ∇^S, second fundamental form II^S, and unit normal ν. Let γ be the curve of Exercise 3, and suppose γ is a curve in S. Show that

$$\begin{aligned} II^S(T,T) &= \kappa\langle N, \nu\rangle\nu \\ &= \kappa N - \nabla^S_T T. \end{aligned}$$

If A_ν denotes the Weingarten map of S, as in (4.5), show that

$$A_\nu T = \kappa T - \tau B \ \text{ and } \ N = \nu,$$

provided γ is a geodesic on S.

5. Use Theorem 4.4 to show that the Gauss curvature K of a surface $S \subset \mathbb{R}^3$ is equal to det A_ν. Use the symmetry of A_ν to show that each T_pS has an orthonormal basis T_1, T_2 such that $A_\nu T_j = \kappa_j T_j$; hence $K = \kappa_1\kappa_2$. An eigenvector of A_ν is called a *direction of principal curvature*. Show that $T \in T_pS$ is a direction of principal curvature if and only if the geodesic through p in direction T has vanishing torsion τ at p.
6. Suppose M has the property that each sectional curvature $K_p(\Pi)$ is equal to K_p, independent of Π. Show that

$$\mathcal{R} = K_p I \ \text{ in End}(\Lambda^2 T_p),$$

where $\mathcal{R}$ is as in Exercise 4 of §3. Show that K_p is constant, on each connected component of M, if dim $M \geq 3$. (*Hint*: To do the last part, use Proposition 3.3.)
7. Show that the formula (4.42) for $R - \widetilde{R}$ is equivalent to the formula (2.17). (This reiterates Exercise 5 of §2.) Also, relate (4.44) and (4.49) to (3.54).

Let M be a compact, oriented hypersurface in $\mathbb{R}^n$. Let

$$N : M \to S^{n-1}$$

be given by the outward-pointing normal. This is called the *Gauss map*.

8. If $n = 3$, show that $N^*\omega_0 = K\omega$, where ω_0 and ω are the area forms of S^2 and M, respectively, and K is the Gauss curvature of M. Note that the degree of the Gauss map is

$$\mathrm{Deg}(N) = \frac{1}{4\pi} \int_M N^* \omega_0.$$

See §19 of Chap. 1 for basic material on degrees of maps.

9. For general n, show that $N^*\omega_0 = J\omega$, with

$$J = (-1)^{n-1} \det A_N,$$

where ω and ω_0 are the volume forms and $A_N : T_pM \to T_pM$ is the Weingarten map (4.5). Consequently,

$$\mathrm{Deg}(N) = \frac{(-1)^{n-1}}{A_{n-1}} \int_M \big(\det A_N\big)\, dV, \tag{4.66}$$

where A_{n-1} is the area of S^{n-1}. (*Hint*: There is a natural identification of T_pM and $T_q(S^{n-1})$ as linear subspaces of $\mathbb{R}^n$, if $q = N(p)$. Show that the Weingarten formula gives

$$DN(p) = -A_N \in \mathrm{End}(T_pM) \approx \mathcal{L}(T_pM, T_qS^{n-1}).) \tag{4.67}$$

10. Let S be a hypersurface in $\mathbb{R}^n$, with second fundamental form $\widetilde{II}$, as in (4.15). Suppose $\widetilde{II}$ is proportional to the metric tensor, $\widetilde{II} = \lambda(x)g$. Show that λ is constant, provided S is connected. (*Hint*: Use the Codazzi equation (4.17), plus the fact that $\nabla^0 g = 0$.)
11. When S is a hypersurface in $\mathbb{R}^n$, a point p, where $\widetilde{II} = \lambda g$, is called an *umbilic point*. If every point on S is umbilic, show that S has constant sectional curvature λ^2. (*Hint*: Apply Gauss' *Theorema Egregium*, in the form (4.14).)
12. Let $S \subset \mathbb{R}^n$ be a k-dimensional submanifold ($k < n$), with induced metric g and second fundamental form II. Let ξ be a section of the normal bundle $\nu(S)$. Consider the one-parameter family of maps $S \to \mathbb{R}^n$,

$$\varphi_\tau(x) = x + \tau\xi(x), \quad x \in S,\ \tau \in (-\varepsilon, \varepsilon). \tag{4.68}$$

Let g_τ be the family of Riemannian metrics induced on S. Show that

$$\frac{d}{d\tau} g_\tau(X, Y)\bigg|_{\tau=0} = -2\langle \xi, II(X, Y)\rangle. \tag{4.69}$$

More generally, if $S \subset M$ is a submanifold, consider the one-parameter family of submanifolds given by

$$\varphi_\tau(x) = \mathrm{Exp}_x\big(\tau\xi(x)\big), \quad x \in S,\ \tau \in (-\varepsilon, \varepsilon), \tag{4.70}$$

where $\mathrm{Exp}_x : T_xM \to M$ is the exponential map, determined by the Riemannian metric on M. Show that (4.69) holds in this more general case.
13. Let $M_1 \subset M_2 \subset M_3$ be Riemannian manifolds of dimension $n_1 < n_2 < n_3$, with induced metrics. For $j < k$, denote by II^{jk} the second fundamental form of $M_j \subset M_k$ and by A^{jk} the associated Weingarten map. For $x \in M_j$, denote by N_x^{jk} the orthogonal complement of T_xM_j in T_xM_k and by ${}^{jk}\nabla^1$ the natural connection on $N^{jk}(M_j)$. Let X and Y be tangent to M_1, and let ξ be a section of $N^{12}(M_1)$. Show that

$$A_\xi^{12}X = A_\xi^{13}X.$$

Also show that

$${}^{13}\nabla_X^1 \xi = {}^{12}\nabla_X^1 \xi + II^{23}(X,\xi), \text{ orthogonal decomposition,}$$

and that

$$II^{13}(X,Y) = II^{12}(X,Y) + II^{23}(X,Y), \text{ orthogonal decomposition.}$$

Relate this to Exercises 3–5 when $n_j = j$.

14. If $S \subset M$ has codimension 1 and Weingarten map $A : T_pS \to T_pS$, show that the Gauss equation (4.14) gives

$$\big\langle (R - R_S)(X,Y)Z, W \big\rangle = \big\langle (\Lambda^2 A)(X \wedge Y), Z \wedge W \big\rangle, \quad X, Y, Z, W \in T_pS. \tag{4.71}$$

Show that (with N a unit normal to S) the scalar curvatures of M and S are related by

$$S_M - S_S = -2 \operatorname{Tr} \Lambda^2 A + 2 \operatorname{Ric}_M(N,N). \tag{4.72}$$

5. The Gauss–Bonnet theorem for surfaces

If M is a compact, oriented Riemannian manifold of dimension 2, the Gauss–Bonnet theorem says that

$$\int_M K \, dV = 2\pi \; \chi(M), \tag{5.1}$$

if K is the Gauss curvature of M and $\chi(M)$ its Euler characteristic. There is an associated formula if M has a boundary. There are a number of significant variants of this, involving, for example, the index of a vector field. We present several proofs of the Gauss–Bonnet theorem and some of its variants here.

We begin with an estimate on the effect of parallel translation about a small closed, piecewise smooth curve. This first result holds for a general vector bundle $E \to M$ with connection ∇ and curvature

$$\Omega = \frac{1}{2} R^\alpha{}_{\beta jk} \, dx_j \wedge dx_k,$$

with no restriction on $\dim M$.

Proposition 5.1. *Let γ be a closed, piecewise smooth loop on M. Assume it is parameterized by arc length for $0 \le t \le b$, $\gamma(b) = \gamma(0)$. If $u(t)$ is a section of E over γ defined by parallel transport (i.e., $\nabla_T u = 0$, $T = \dot\gamma$), then*

$$u^\alpha(b) - u^\alpha(0) = -\frac{1}{2}\sum_{j,k,\beta} R^\alpha{}_{\beta jk}\left(\int_A dx_j \wedge dx_k\right)u^\beta(0) + O(b^3), \tag{5.2}$$

where A is an oriented 2-surface in M with $\partial A = \gamma$, and the $u^\alpha(t)$ are the components of u with respect to a local frame.

Proof. If we put a coordinate system on a neighborhood of $p = \gamma(0) \in M$ and choose a frame field for E, then parallel transport is defined by

$$\frac{du^\alpha}{dt} = -\Gamma^\alpha{}_{\beta k} u^\beta \frac{dx_k}{dt}. \tag{5.3}$$

As usual, we use the summation convention. Thus

$$u^\alpha(t) = u^\alpha(0) - \int_0^t \Gamma^\alpha{}_{\beta k}(\gamma(s))u^\beta(s)\,\frac{dx_k}{ds}\,ds. \tag{5.4}$$

We hence have

$$u^\alpha(t) = u^\alpha(0) - \Gamma^\alpha{}_{\beta k}(p)u^\beta(0)(x_k - p_k) + O(t^2). \tag{5.5}$$

We can solve (5.3) up to $O(t^3)$ if we use

$$\Gamma^\alpha{}_{\beta j}(x) = \Gamma^\alpha{}_{\beta j}(p) + (x_k - p_k)\,\partial_k\Gamma^\alpha{}_{\beta j} + O\big(|x-p|^2\big). \tag{5.6}$$

Hence

$$\begin{aligned} u^\alpha(t) = u^\alpha(0) - \int_0^t &\big[\Gamma^\alpha{}_{\beta k}(p) + (x_j - p_j)\,\partial_j\Gamma^\alpha{}_{\beta k}(p)\big] \\ &\cdot\big[u^\beta(0) - \Gamma^\beta{}_{\gamma\ell}(p)u^\gamma(0)(x_\ell - p_\ell)\big]\,\frac{dx_k}{ds}\,ds + O(t^3). \end{aligned} \tag{5.7}$$

If $\gamma(b) = \gamma(0)$, we get

$$\begin{aligned} u^\alpha(b) = u^\alpha(0) &- \int_0^b x_j\,dx_k\big(\partial_j\Gamma^\alpha{}_{\beta k}\big)u^\beta(0) \\ &+ \int_0^b x_j\,dx_k\Gamma^\alpha{}_{\beta k}\Gamma^\beta{}_{\gamma j}u^\gamma(0) + O(b^3), \end{aligned} \tag{5.8}$$

the components of Γ and their first derivatives being evaluated at p. Now Stokes' theorem gives

$$\int_\gamma x_j\,dx_k = \int_A dx_j \wedge dx_k,$$

so

$$(5.9)\quad u^\alpha(b)-u^\alpha(0)=\left[-\partial_j\Gamma^\alpha{}_{\beta k}+\Gamma^\alpha{}_{\gamma k}\Gamma^\gamma{}_{\beta j}\right]\int_A dx_j\wedge dx_k\, u^\beta(0)+O(b^3).$$

Recall that the curvature is given by

$$\Omega = d\Gamma + \Gamma\wedge\Gamma,$$

that is,

$$(5.10)\qquad R^\alpha{}_{\beta jk}=\partial_j\Gamma^\alpha{}_{\beta k}-\partial_k\Gamma^\alpha{}_{\beta j}+\Gamma^\alpha{}_{\gamma j}\Gamma^\gamma{}_{\beta k}-\Gamma^\alpha{}_{\gamma k}\Gamma^\gamma{}_{\beta j}.$$

Now the right side of (5.10) is the antisymmetrization, with respect to j and k, of the quantity in brackets in (5.9). Since $\int_A dx_j\wedge dx_k$ is antisymmetric in j and k, we get the desired formula (5.2).

In particular, if $\dim M = 2$, then we can write the $\mathrm{End}(E)$-valued 2-form Ω as

$$(5.11)\qquad \Omega = \mathcal{R}\mu,$$

where μ is the volume form on M and $\mathcal{R}$ is a smooth section of $\mathrm{End}(E)$ over M. If E has an inner product and ∇ is a metric connection, then $\mathcal{R}$ is skew-adjoint. If γ is a geodesic triangle that is “fat” in the sense that none of its angles is small, (5.2) implies

$$(5.12)\qquad u(b)-u(0)=-\mathcal{R}u(0)(\text{Area } A)+O\big((\text{Area } A)^{3/2}\big).$$

If we specialize further, to oriented two-dimensional M with $E = TM$, possessing the Levi–Civita connection of a Riemannian metric, then we take $J : T_pM \to T_pM$ to be the counterclockwise rotation by 90°, which defines an almost complex structure on M. Up to a scalar this is the unique skew-adjoint operator on T_pM, and, by (3.34),

$$(5.13)\qquad \mathcal{R}u = -K\,Ju,\quad u\in T_pM,$$

where K is the Gauss curvature of M at p. Thus, in this case, (5.12) becomes

$$(5.14)\qquad u(b)-u(0)=K\,Ju(0)(\text{Area } A)+O\big((\text{Area } A)^{3/2}\big).$$

On the other hand, if a tangent vector $X_0\in T_pM$ undergoes parallel transport around a geodesic triangle, the action produced on T_pM is easily seen to be a rotation in T_pM through an angle that depends on the angle defect of the triangle. The argument can be seen by looking at Fig. 5.1. We see that the angle from X_0 to X_3 is

$$(5.15)\qquad (\pi+\alpha)-(2\pi-\beta-\gamma-\xi)-\xi=\alpha+\beta+\gamma-\pi.$$

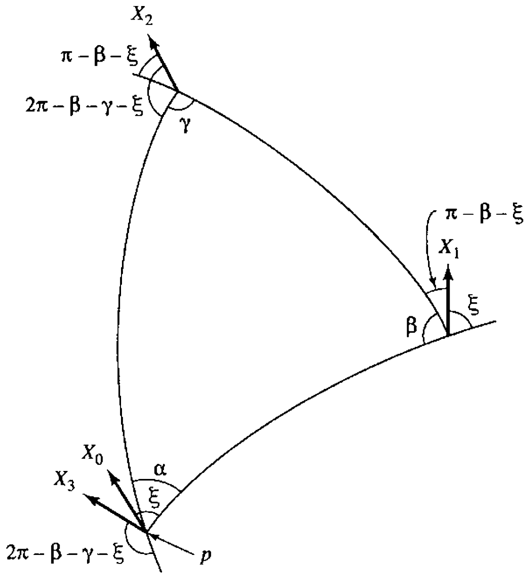

FIGURE 5.1 Parallel Transport Around a Triangle

In this case, formula (5.14) implies

$$\alpha+\beta+\gamma-\pi=\int K\,dV+O\big((\text{Area }A)^{3/2}\big). \tag{5.16}$$

We can now use a simple analytical argument to sharpen this up to the following celebrated formula of Gauss.

Theorem 5.2. *If A is a geodesic triangle in M^2, with angles α,β, and γ, then*

$$\alpha+\beta+\gamma-\pi=\int_A K\,dV. \tag{5.17}$$

Proof. Break up the geodesic triangle A into N^2 little geodesic triangles, each of diameter $O(N^{-1})$, area $O(N^{-2})$. Since the angle defects are *additive*, the estimate (5.16) implies

$$\begin{aligned}\alpha+\beta+\gamma-\pi&=\int_A K\,dV+N^2O\big((N^{-2})^{3/2}\big)\\&=\int_A K\,dV+O(N^{-1}),\end{aligned} \tag{5.18}$$

and passing to the limit as $N\to\infty$ gives (5.17).

Note that any region that is a contractible geodesic polygon can be divided into geodesic triangles. If a contractible region $\Omega \subset M$ with smooth boundary is approximated by geodesic polygons, a straightforward limit process yields the Gauss–Bonnet formula

$$\int_{\Omega} K\,dV + \int_{\partial\Omega} \kappa\,ds = 2\pi, \tag{5.19}$$

where κ is the geodesic curvature of $\partial\Omega$. We leave the details to the reader. Another proof will be given at the end of this section.

If M is a compact, oriented two-dimensional manifold without boundary, we can partition M into geodesic triangles. Suppose the triangulation of M so produced has

$$F \text{ faces (triangles)}, \quad E \text{ edges}, \quad V \text{ vertices}. \tag{5.20}$$

If the angles of the jth triangle are α_j, β_j, and γ_j, then summing all the angles clearly produces $2\pi V$. On the other hand, (5.17) applied to the jth triangle, and summed over j, yields

$$\sum_j (\alpha_j + \beta_j + \gamma_j) = \pi F + \int_M K\,dV. \tag{5.21}$$

Hence $\int_M K\,dV = (2V - F)\pi$. Since in this case all the faces are triangles, counting each triangle three times will count each edge twice, so $3F = 2E$. Thus we obtain

$$\int_M K\,dV = 2\pi(V - E + F). \tag{5.22}$$

This is equivalent to (5.1), in view of Euler's formula

$$\chi(M) = V - E + F. \tag{5.23}$$

We now derive a variant of (5.1) when M is described in another fashion. Namely, suppose M is diffeomorphic to a sphere with g handles attached. The number g is called the *genus* of the surface. The case $g = 2$ is illustrated in Fig. 5.2. We claim that

$$\int_M K\,dV = 4\pi(1 - g) \tag{5.24}$$

in this case. By virtue of (5.22), this is equivalent to the identity

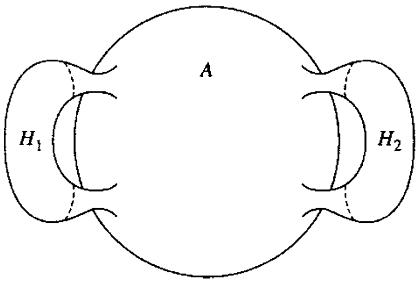

FIGURE 5.2 Surface with Handles

$$2 - 2g = V - E + F = \chi(M). \tag{5.25}$$

Direct proofs of this are possible, but we will provide a proof of (5.24), based on the fact that

$$\int_M K\, dV = C(M) \tag{5.26}$$

depends only on M, not on the metric imposed. This follows from (5.22), by forgetting the interpretation of the right side. The point we want to make is, given (5.26)—that is, the independence of the choice of metric—we can work out what $C(M)$ is, as follows.

First, choosing the standard metric on S^2, for which $K = 1$ and the area is 4π, we have

$$\int_{S^2} K\, dV = 4\pi. \tag{5.27}$$

Now suppose M is obtained by adding g handles to S^2. Since we can alter the metric on M at will, we can make sure it coincides with the metric of a sphere near a great circle, in a neighborhood of each circle where a handle is attached to the main body A, as illustrated in Fig. 5.2. If we imagine adding two hemispherical caps to each handle H_j, rather than attaching it to A, we turn each H_j into a new sphere, so by (5.27) we have

$$4\pi = \int_{H_j \cup \text{caps}} K\, dV = \int_{H_j} K\, dV + \int_{\text{caps}} K\, dV. \tag{5.28}$$

Since the caps fit together to form a sphere, we have $\int_{\text{caps}} K\, dV = 4\pi$, so for each j,

(5.29)
$$\int_{H_j} K\,dV = 0,$$

provided M has a metric such as described above. Similarly, if we add $2g$ caps to the main body A, we get a new sphere, so

(5.30)
$$4\pi = \int_{A\cup\,\text{caps}} K\,dV = \int_A K\,dV + 2g(2\pi),$$

or

(5.31)
$$\int_A K\,dV = 2\pi(2-2g).$$

Together (5.29) and (5.31) yield (5.24), and we get the identity (5.25) as a bonus.

We now give another perspective on Gauss' formula, directly dealing with the fact that TM can be treated as a complex line bundle, when M is an oriented Riemannian manifold of dimension 2. We will produce a variant of Proposition 5.1 which has no remainder term and which hence produces (5.16) with no remainder, directly, so Theorem 5.2 follows without the additional argument given above. The result is the following; again dim M is unrestricted.

Proposition 5.3. *Let $E \to M$ be a complex line bundle. Let γ be a piecewise smooth, closed loop in M, with $\gamma(0) = \gamma(b) = p$, bounding a surface A. Let ∇ be a connection on E, with curvature Ω. If $u(t)$ is a section of E over γ defined by parallel translation, then*

(5.32)
$$u(b) = \left[\exp\left(-\int_A \Omega\right)\right] u(0).$$

Proof. Pick a nonvanishing section (hence a frame field) ξ of E over S, assuming S is homeomorphic to a disc. Any section u of E over S is of the form $u = v\xi$ for a complex-valued function v on S. Then parallel transport along $\gamma(t) = (x_1(t), \dots, x_n(t))$ is defined by

(5.33)
$$\frac{dv}{dt} = -\left(\Gamma_k \frac{dx_k}{dt}\right) v.$$

The solution to this single, first-order ODE is

(5.34)
$$v(t) = \left[\exp\left(-\int_0^t \Gamma_k(\gamma(s))\frac{dx_k}{ds}\,ds\right)\right] v(0).$$

Hence

$$v(b) = \left[\exp\left(-\int_\gamma \Gamma\right)\right] v(0), \tag{5.35}$$

where $\Gamma = \sum \Gamma_k \, dx_k$. The curvature 2-form Ω is given, as a special case of (5.10), by

$$\Omega = d\Gamma, \tag{5.36}$$

and Stokes' theorem gives (5.32), from (5.35), provided A is contractible. The general case follows from cutting A into contractible pieces.

As we have mentioned, Proposition 5.3 can be used in place of Proposition 5.1, in conjunction with the argument involving Fig. 5.1, to prove Theorem 5.2.

Next, we relate $\int_M \Omega$ to the "index" of a section of a complex line bundle $E \to M$, when M is a compact, oriented manifold of dimension 2. Suppose X is a section of E over $M \setminus S$, where S consists of a finite number of points; suppose that X is nowhere vanishing on $M \setminus S$ and that, near each $p_j \in S$, X has the following form. There are a coordinate neighborhood $\mathcal{O}_j$ centered at p_j, with p_j the origin, and a nonvanishing section ξ_j of E near p_j, such that

$$X = v_j \xi_j \ \text{ on } \mathcal{O}_j, \quad v_j : \mathcal{O}_j \setminus p \to \mathbb{C} \setminus 0. \tag{5.37}$$

Taking a small counterclockwise circle γ_j about p_j, $v_j/|v_j| = \omega_j$ maps γ_j to S^1; consider the degree ℓ_j of this map, that is, the winding number of γ_j about S^1. This is the index of X at p_j, and the sum over p_j is the total index of X:

$$\text{Index}(X) = \sum_j \ell_j. \tag{5.38}$$

We will establish the following.

Proposition 5.4. *For any connection on $E \to M$, with curvature form Ω and X as above, we have*

$$\int_M \Omega = -(2\pi i) \cdot \textit{Index}(X). \tag{5.39}$$

Proof. You can replace X by a section of $E \setminus 0$ over $M \setminus \{p_j\}$, homotopic to the original, having the form (5.37) with

$$v_j = e^{i\ell_j \theta} + w_j, \tag{5.40}$$

in polar coordinates (r, θ) about p_j, with $w_j \in C^1(\mathcal{O}_j)$, $w_j(0) = 0$. Excise small disks $\mathcal{D}_j$ containing p_j; let $\mathcal{D} = \cup \mathcal{D}_j$. Then, by Stokes' theorem,

$$\int\limits_{M \setminus \mathcal{D}} \Omega = -\sum_j \int\limits_{\gamma_j} \Gamma, \tag{5.41}$$

where $\gamma_j = \partial \mathcal{D}_j$ and Γ is the connection 1-form with respect to the section X, so that, with $\nabla_k = \nabla_{D_k}$, $D_k = \partial/\partial x_k$ in local coordinates,

$$\nabla_k X = \Gamma_k X. \tag{5.42}$$

Now (5.37) gives (with no summation)

$$\Gamma_k v_j \xi_j = (\partial_k v_j + v_j \widetilde{\Gamma}_{jk}) \xi_j \tag{5.43}$$

on $\overline{\mathcal{D}}_j$, where $\widetilde{\Gamma}_{jk}\, dx_k$ is the connection 1-form with respect to the section ξ_j. Hence

$$\Gamma_k = v_j^{-1}\, \partial_k v_j + \widetilde{\Gamma}_{jk}, \tag{5.44}$$

with remainder term $\widetilde{\Gamma}_{jk} \in C^1(\mathcal{O}_j)$. By (5.40), we have

$$\int\limits_{\gamma_j} \Gamma = 2\pi i \ell_j + O(r) \tag{5.45}$$

if each $\mathcal{D}_j$ has radius $\le Cr$. Passing to the limit as the disks $\mathcal{D}_j$ shrink to p_j gives (5.39).

Since the left side of (5.39) is independent of the choice of X, it follows that the index of X depends only on E, not on the choice of such X. In Chap. 10, this formula is applied to a meromorphic section of a complex line bundle, and in conjunction with the Riemann–Roch formula it yields important information on Riemann surface theory.

In case M is a compact, oriented Riemannian 2-manifold, whose tangent bundle can be given the structure of a complex line bundle as noted above, (5.39) is equivalent to

$$\int\limits_M K\, dV = 2\pi \operatorname{Index}(X), \tag{5.46}$$

for any smooth vector field X, nonvanishing, on M minus a finite set of points. This verifies the identity

$$\operatorname{Index}(X) = \chi(M) \tag{5.47}$$

in this case.

As a further comment on the Gauss–Bonnet formula for compact surfaces, let us recall from Exercise 8 of §4 that if M is a compact, oriented surface in $\mathbb{R}^3$, with Gauss map $N : M \to S^2$, then

$$\operatorname{Deg}(N) = \frac{1}{4\pi}\int_M N^*\omega_0 = \frac{1}{4\pi}\int_M K\, dV. \tag{5.48}$$

Furthermore, in §20 of Chap. 1, Corollary 20.5 yields an independent proof that, in this case,

$$\operatorname{Deg}(N) = \frac{1}{2}\operatorname{Index}(X), \tag{5.49}$$

for any vector field X on M with a finite number of critical points. Hence (5.48)–(5.49) provide another proof of (5.1), at least for a surface in $\mathbb{R}^3$. This line of reasoning will be extended to the higher-dimensional case of hypersurfaces of $\mathbb{R}^{n+1}$, in the early part of §8, as preparation for establishing the general Chern–Gauss–Bonnet theorem.

To end this section, we provide a direct proof of the formula (5.19), using an argument parallel to the proof of Proposition 5.3. Thus, assuming that M is an oriented surface, we give TM the structure of a complex line bundle, and we pick a nonvanishing section ξ of TM over a neighborhood of $\overline{\mathcal{O}}$. Let $\gamma = \partial\mathcal{O}$ be parameterized by arc length, $T = \gamma'(s),\ 0 \le s \le b$, with $\gamma(b) = \gamma(0)$. The geodesic curvature κ of γ, appearing in (5.19), is given by

$$\nabla_T T = \kappa N, \quad N = JT. \tag{5.50}$$

If we set $T = u\xi$, where $u : \overline{\mathcal{O}} \to \mathbb{C}$, then, parallel to (5.33), we have (5.50) equivalent to

$$\frac{du}{ds} = -\sum \Gamma_k \frac{dx_k}{ds} u + i\kappa u. \tag{5.51}$$

The solution to this single, first-order ODE is (parallel to (5.34))

$$u(t) = \left[\exp\left(i\int_0^t \kappa(s)\, ds - \int_0^t \Gamma_k\big(\gamma(s)\big)\frac{dx_k}{ds}\, ds\right)\right] u(0). \tag{5.52}$$

Hence

$$u(b) = \left[\exp\left(i\int_\gamma \kappa(s)\, ds - \int_{\mathcal{O}} \Omega\right)\right] u(0). \tag{5.53}$$

By (5.13), we have

$$\Omega = -iK\,dV, \tag{5.54}$$

and since $u(b) = u(0)$, we have

$$\exp\left(i\int_\gamma \kappa(s)\,ds + i\int_{\mathcal{O}} K\,dV\right) = 1, \tag{5.55}$$

or

$$\int_{\mathcal{O}} K\,dV + \int_\gamma \kappa(s)\,ds = 2\pi\nu, \tag{5.56}$$

for some $\nu \in \mathbb{Z}$. Now if $\mathcal{O}$ were a tiny disc in M, it would be clear that $\nu = 1$. Using the contractibility of $\mathcal{O}$ and the fact that the left side of (5.56) cannot jump, we have $\nu = 1$, which proves (5.19).

Exercises

1. Given a triangulation of a compact surface M, within each triangle construct a vector field, vanishing at seven points as illustrated in Fig. 5.3, with the vertices as attractors, the center as a repeller, and the midpoints of each side as saddle points. Fit these together to produce a smooth vector field X on M. Show directly that

$$\text{Index}\,(X) = V - E + F.$$

2. Let $L \to M$ be a complex line bundle, and let u and v be sections of L with a finite number of zeros. Show directly that u and v have the same index. (*Hint*: Start with $u = fv$ on $M \setminus Z$, where Z is the union of the zero sets and $f : M \setminus Z \to \mathbb{C} \setminus 0$.)
3. Let M_1 and M_2 be n-dimensional submanifolds of $\mathbb{R}^k$. Suppose a curve γ is contained in the intersection $M_1 \cap M_2$, and assume

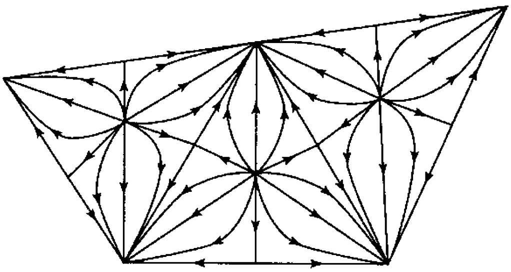

FIGURE 5.3 Vector Field on a Triangulation

$$p = \gamma(s) \Longrightarrow T_p M_1 = T_p M_2.$$

Show that parallel translations along γ in M_1 and in M_2 coincide. (*Hint*: If $T = \gamma'(s)$ and X is a vector field along γ, tangent to M_1 (hence to M_2), show that $\nabla_T^{M_1} X = \nabla_T^{M_2} X$, using Corollary 4.2.)

4. Let $\mathcal{O}$ be the region in $S^2 \subset \mathbb{R}^3$ consisting of points in S^2 of geodesic distance $< r$ from $p = (0, 0, 1)$, where $r \in (0, \pi)$ is given. Let $\gamma = \partial\mathcal{O}$. Construct a cone, with vertex at $(0, 0, \sec r)$, tangent to S^2 along γ. Using this and Exercise 3, show that parallel translation over one circuit of γ is given by

$$\text{counterclockwise rotation by } \theta = 2\pi(1 - \cos r).$$

(*Hint*: Flatten out the cone, as in Fig. 5.4. Notice that γ has length $\ell = 2\pi \ \sin r$.) Compare this calculation with the result of (5.32), which in this context implies

$$u(\ell) = \left[\exp i \int_{\mathcal{O}} K\, dV\right] u(0).$$

5. Let $\gamma : [a, b] \to \mathbb{R}^3$ be a smooth, closed curve, so $\gamma(a) = \gamma(b)$ and $\gamma'(a) = \gamma'(b)$. Assume γ is parameterized by arc length, so $\gamma'(t) = T(t)$ and $T : [a, b] \to S^2$; hence T is a smooth, closed curve in S^2. Note that the normal space to γ at $p = \gamma(t)$ is naturally identified with the tangent space to S^2 at $T(t) = q$:

$$\nu_p(\gamma) = T_q S^2.$$

a) Show that parallel translation along γ of a section of the normal bundle $\nu(\gamma)$, with respect to the connection described in Exercise 3 of §4, coincides with parallel translation along the curve T of vectors tangent to S^2. (*Hint*: Recall Exercise 3 of §1.)
b) Suppose the curvature κ of γ never vanishes, so the torsion τ is well defined, as in (4.62). Show that parallel translation once around γ acts on $\nu_p(\gamma)$ by multiplication by

$$\exp\left(-i \int_{\gamma} \tau(s)\, ds\right).$$

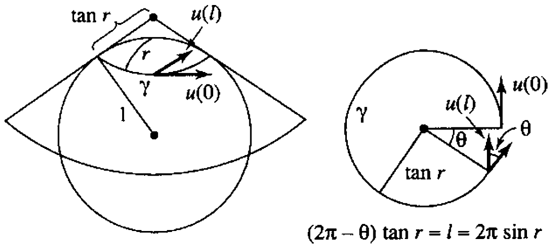

FIGURE 5.4 Cone Tangent to a Sphere

Here we use the complex structure on $\nu_p(\gamma)$ given by $JN = B,\ JB = -N$. (*Hint*: Use (4.64).) Compare the results of parts a and b.

6. The principal bundle picture

An important tool for understanding vector bundles is the notion of an underlying structure, namely that of a principal bundle. If M is a manifold and G a Lie group, then a principal G-bundle $P \xrightarrow{p} M$ is a locally trivial fibration with a G-action on P such that G acts on each fiber $P_x = p^{-1}(x)$ in a simply transitive fashion. An example is the frame bundle of an oriented Riemannian manifold M, $F(M) \to M$, where $F_x(M)$ consists of the set of ordered oriented orthonormal bases of the tangent space T_x to M at x. If $n = \dim M$, this is a principal $\mathrm{SO}(n)$-bundle.

If $P \to M$ is a principal G-bundle, then associated to each representation π of G on a vector space V is a vector bundle $E \to M$. The set E is a quotient space of the Cartesian product $P \times V$, under the equivalence relation

$$(y, v) \sim (y \cdot g, \pi(g)^{-1} v), \quad g \in G. \tag{6.1}$$

We have written the G-action on P as a right action. One writes $E = P \times_\pi V$. The space of sections of E is naturally isomorphic to a certain subspace of the space of V-valued functions on P:

$$C^\infty(M, E) \approx \{u \in C^\infty(P, V) : u(y \cdot g) = \pi(g)^{-1} u(y),\ g \in G\}. \tag{6.2}$$

We describe how this construction works for the frame bundle $F(M)$ of an oriented Riemannian manifold, which, as mentioned above, is a principal $\mathrm{SO}(n)$-bundle. Thus, a point $y \in F_x(M)$ consists of an n-tuple $(e_1, \dots, e_n)$, forming an ordered, oriented orthonormal basis of $T_x M$. If $g = (g_{jk}) \in \mathrm{SO}(n)$, the G-action is given by

$$(e_1, \dots, e_n) \cdot g = (f_1, \dots, f_n), \quad f_j = \sum_\ell g_{\ell j} e_\ell. \tag{6.3}$$

One can check that $(f_1, \dots, f_n)$ is also an oriented orthonormal basis of $T_x M$ and that $(y \cdot g) \cdot g' = y \cdot (gg')$, for $g, g' \in \mathrm{SO}(n)$. If π is the "standard" representation of $\mathrm{SO}(n)$ on $\mathbb{R}^n$, given by matrix multiplication, we claim that there is a natural identification

$$F(M) \times_\pi \mathbb{R}^n \approx TM. \tag{6.4}$$

In fact, if $y = (e_1, \dots, e_n) \in F_x(M)$ and $v = (v_1, \dots, v_n) \in \mathbb{R}^n$, the map (6.4) is defined by

$$(y, v) \mapsto \sum_j v_j e_j \in T_x M. \tag{6.5}$$

We need to show that this is constant on equivalence classes, as defined by (6.1), that is, for any $g \in \mathrm{SO}(n)$,

$$z = y \cdot g = (f_1, \dots, f_n), \quad w = \pi(g)^{-1} v \Longrightarrow \sum w_k f_k = \sum v_j e_j. \tag{6.6}$$

In fact, setting $g^{-1} = h = (h_{jk})$, we see that

$$\sum_k w_k f_k = \sum_{j,k,\ell} h_{kj} v_j g_{\ell k} e_\ell = \sum_{j,\ell} \delta_{\ell j} v_j e_\ell \tag{6.7}$$

since

$$\sum_k g_{\ell k} h_{kj} = \delta_{\ell j}, \tag{6.8}$$

and this implies (6.6).

Connections are naturally described in terms of a geometrical structure on a principal bundle. This should be expected since, as we saw in §1, a connection on a vector bundle can be described in terms of a "connection 1-form" (1.14), depending on a choice of local frame for the vector bundle.

The geometrical structure giving a connection on a principal bundle $P \to M$ is the following. For each $y \in P$, the tangent space $T_y P$ contains the subspace $V_y P$ of vectors tangent to the fiber $p^{-1}(x),\ x = p(y)$. The space $V_y P$, called the "vertical space," is naturally isomorphic to the Lie algebra $\mathfrak{g}$ of G. A connection on P is determined by a choice of complementary subspace, called a "horizontal space":

$$T_y P = V_y P \oplus H_y P, \tag{6.9}$$

with the G-invariance

$$g_*(H_y P) = H_{y \cdot g} P, \tag{6.10}$$

where $g_* : T_y P \to T_{y \cdot g} P$ is the natural derivative map.

Given this structure, a vector field X on M has a uniquely defined "lift" $\widetilde{X}$ to a vector field on P, such that $p_* \widetilde{X}_y = X_x$ $(x = p(y))$ and $\widetilde{X}_y \in H_y P$ for each $y \in P$. Furthermore, if E is a vector bundle determined by a representation of G and $u \in C^\infty(M, V)$ corresponds to a section v of E, the V-valued function $\widetilde{X} \cdot u$ corresponds to a section of E, which we denote $\nabla_X v$; ∇ is the covariant derivative on E defined by the connection on P just described. If V has an inner product and π is unitary, E gets a natural metric and ∇ is a metric connection on E.

If the π_j are representations of G on V_j, giving vector bundles $E_j \to M$ associated to a principal bundle $P \to M$ with connection, then $\pi_1 \otimes \pi_2$ is a representation of G on $V_1 \otimes V_2$, and we have a vector bundle $E \to M$, $E = E_1 \otimes E_2$. The prescription above associating a connection to E as well as to E_1 and E_2 agrees with the definition in (1.29) of a connection on a tensor product of two vector bundles. This follows simply from the derivation property of the vector field $\widetilde{X}$, acting as a first-order differential operator on functions on P.

The characterization (6.9)–(6.10) of a connection on a principal bundle $P \to M$ is equivalent to the following, in view of the natural isomorphism $V_y P \approx \mathfrak{g}$. The splitting (6.9) corresponds to a projection of $T_y P$ onto $V_y P$, hence to a linear map $T_y P \to \mathfrak{g}$, which gives the identification $V_y P \approx \mathfrak{g}$ on the linear subspace $V_y P$ of $T_y P$. This map can be regarded as a $\mathfrak{g}$-valued 1-form ξ on P, called the *connection form*, and the invariance property (6.10) is equivalent to

$$g^* \xi = Ad_{g^{-1}} \xi, \quad g \in G, \tag{6.11}$$

where $g^*\xi$ denotes the pull-back of the form ξ, induced from the G-action on P.

The Levi–Civita connection on an oriented Riemannian manifold gives rise to a connection on the frame bundle $F(M) \to M$ in the following way. Fix $y \in F(M)$, $x = p(y)$. Recall that the point y is an ordered (oriented) orthonormal basis $(e_1, \dots, e_n)$ of the tangent space $T_x M$. The parallel transport of each e_j along a curve γ through x thus gives a family of orthonormal bases for the tangent space to M at $\gamma(t)$, hence a curve $\gamma^\#$ in $F(M)$ lying over γ. The tangent to $\gamma^\#$ at y belongs to the horizontal space $H_y F(M)$, which in fact consists of all such tangent vectors as the curve γ through x is varied. This construction generalizes to other vector bundles $E \to M$ with connection ∇. One can use the bundle of orthonormal frames for E if ∇ is a metric connection, or the bundle of general frames for a general connection.

Let us restate how a connection on a principal bundle gives rise to connections on associated vector bundles. Given a principal G-bundle $P \to M$, consider a local section σ of P, over $U \subset M$. If we have a representation π of G on V, the associated vector bundle $E \to M$, and a section u of E, then we have $u \circ \sigma : U \to V$, using the identification (6.2). Given a connection on P, with connection 1-form ξ, we can characterize the covariant derivative induced on sections of E by

$$(\nabla_X u) \circ \sigma = \mathcal{L}_X (u \circ \sigma) + \Gamma(X) u \circ \sigma, \tag{6.12}$$

where $\mathcal{L}_X$ acts componentwise on $u \circ \sigma$, and

$$\Gamma(X) = (d\pi)\big(\xi_y(\widehat{X})\big), \quad y = \sigma(x), \quad \widehat{X} = D\sigma(x) X, \tag{6.13}$$

$d\pi$ denoting the derived representation of $\mathfrak{g}$ on V. That (6.12) agrees with $(\mathcal{L}_{\widetilde{X}} u) \circ \sigma$ follows from the chain rule, the fact that $\widehat{X} - \widetilde{X}$ is vertical, and the fact that if $v \in T_y P$ is vertical, then, by (6.2), $\mathcal{L}_v u = -d\pi\big(\xi(v)\big)u$. Note the similarity of (6.12) to (1.7).

Recall the curvature $R(X,Y)$ of a connection ∇ on a vector bundle $E \to M$, defined by the formula (1.10). In case $E = P \times_\pi V$, and ∇u is defined as above, we have (using the identification in (6.2))

$$R(X,Y)u = \mathcal{L}_{[\widetilde{X},\widetilde{Y}]}u - \mathcal{L}_{\widetilde{[X,Y]}}u. \tag{6.14}$$

Alternatively, using (6.12) and (6.13), we see that the curvature of ∇ is given by

$$\begin{aligned} &R(X,Y)u \circ \sigma \\ &\quad = \Big\{\mathcal{L}_X \Gamma(Y) - \mathcal{L}_Y \Gamma(X) + \big[\Gamma(X), \Gamma(Y)\big] - \Gamma\big([X,Y]\big)\Big\} u \circ \sigma. \end{aligned} \tag{6.15}$$

This is similar to (1.13). Next we want to obtain a formula similar to (but more fundamental than) (1.15).

Fix $y \in P,\ x = p(y)$. It is convenient to calculate (6.15) at x by picking the local section σ to have the property that

$$D\sigma(x) : T_x M \longrightarrow H_y P, \tag{6.16}$$

which is easily arranged. Then $\widehat{X} = \widetilde{X}$ at y, so $\Gamma(X) = 0$ at y. Hence, at x,

$$\begin{aligned} R(X,Y)u \circ \sigma &= \big\{\mathcal{L}_X \Gamma(Y) - \mathcal{L}_Y \Gamma(X)\big\} u \circ \sigma \\ &= (d\pi)\big\{\widehat{X} \cdot \xi(\widehat{Y}) - \widehat{Y} \cdot \xi(\widehat{X})\big\} u \circ \sigma \\ &= (d\pi)\big\{(d\sigma^*\xi)(X,Y) + (\sigma^*\xi)([X,Y])\big\} u \circ \sigma. \end{aligned} \tag{6.17}$$

Of course, $\sigma^*\xi = 0$ at x. Thus we see that

$$R(X,Y)u = (d\pi)\big\{(d\xi)(\widetilde{X},\widetilde{Y})\big\}u, \tag{6.18}$$

at y, and hence everywhere on P. In other words,

$$R(X,Y) = (d\pi)\big(\Omega(\widetilde{X},\widetilde{Y})\big), \tag{6.19}$$

where Ω is the $\mathfrak{g}$-valued 2-form on P defined by

$$\Omega(X^\#, Y^\#) = (d\xi)(\varkappa X^\#, \varkappa Y^\#), \tag{6.20}$$

for $X^\#, Y^\# \in T_y P$. Here, $\varkappa$ is the projection of $T_y P$ onto $H_y P$, with respect to the splitting (6.9). One calls Ω the curvature 2-form of the connection ξ on P.

If V and W are smooth vector fields on P, then

$$(d\xi)(V,W) = V \cdot \xi(W) - W \cdot \xi(V) - \xi\big([V,W]\big). \tag{6.21}$$

In particular, if $V = \widetilde{X},\ W = \widetilde{Y}$ are horizontal vector fields on P, then since $\xi(\widetilde{X}) = \xi(\widetilde{Y}) = 0$, we have

$$(d\xi)(\widetilde{X},\widetilde{Y}) = -\xi\big([\widetilde{X},\widetilde{Y}]\big). \tag{6.22}$$

Hence, given $X^\#, Y^\# \in T_y P$, we have

$$\Omega(X^\#, Y^\#) = -\xi\big([\widetilde{X},\widetilde{Y}]\big), \tag{6.23}$$

where $\widetilde{X}$ and $\widetilde{Y}$ are any horizontal vector fields on P such that $\widetilde{X} = \varkappa X^\#$ and $\widetilde{Y} = \varkappa Y^\#$ at $y \in P$. Since ξ annihilates $[\widetilde{X},\widetilde{Y}]$ if and only if it is horizontal, we see that Ω measures the failure of the bundle of horizontal spaces to be involutive.

It follows from Frobenius's theorem that, if $\Omega = 0$ on P, there is an integral manifold $S \subset P$ such that, for each $y \in S$, $T_y S = H_y P$. Each translate $S \cdot g$ is also an integral manifold. We can use this family of integral manifolds to construct local sections $v_1, \dots v_K$ of E ($K = \dim V$), linearly independent at each point, such that $\nabla v_j = 0$ for all j, given that $\Omega = 0$. Thus we recover Proposition 1.2, in this setting.

The following important result is *Cartan's formula* for the curvature 2-form.

Theorem 6.1. *We have*

$$\Omega = d\xi + \frac{1}{2}[\xi,\xi]. \tag{6.24}$$

The bracket $[\xi,\eta]$ of $\mathfrak{g}$-valued 1-forms is defined as follows. Suppose, in local coordinates,

$$\xi = \sum \xi_j \, dx_j, \quad \eta = \sum \eta_k \, dx_k, \quad \xi_j,\ \eta_k \in \mathfrak{g}. \tag{6.25}$$

Then we set

$$[\xi,\eta] = \sum_{j,k} [\xi_j,\eta_k] \, dx_j \wedge dx_k = \sum_{j<k} \big([\xi_j,\eta_k] + [\eta_j,\xi_k]\big) \, dx_j \wedge dx_k, \tag{6.26}$$

which is a $\mathfrak{g}$-valued 2-form. Equivalently, if U and V are vector fields on P,

$$[\xi,\eta](U,V) = [\xi(U),\eta(V)] + [\eta(U),\xi(V)]. \tag{6.27}$$

In particular,

$$\frac{1}{2}[\xi,\xi](U,V) = [\xi(U),\xi(V)]. \tag{6.28}$$

Note that if π is a representation of G on a vector space V and $d\pi$ the derived representation of $\mathfrak{g}$ on V, if we set $A_j = d\pi(\xi_j)$, then, for

$$d\pi(\xi) = \alpha = \sum A_j \, dx_j, \tag{6.29}$$

we have

$$\alpha \wedge \alpha = \sum_{j,k} A_j A_k \, dx_j \wedge dx_k = \frac{1}{2} \sum_{j,k} (A_j A_k - A_k A_j) \, dx_j \wedge dx_k. \tag{6.30}$$

Hence

$$\alpha \wedge \alpha = \frac{1}{2}(d\pi)[\xi, \xi]. \tag{6.31}$$

Thus we see the parallel between (6.24) and (1.15).

To prove (6.24), one evaluates each side on $(X^\#, Y^\#)$, for $X^\#, Y^\# \in T_y P$. We write $X^\# = \widetilde{X} + X_v$, with $\widetilde{X} \in H_y P$, $X_v \in V_y P$, and similarly write $Y^\# = \widetilde{Y} + Y_v$. It suffices to check the following four cases:

$$\Omega(\widetilde{X}, \widetilde{Y}), \quad \Omega(\widetilde{X}, Y_v), \quad \Omega(X_v, \widetilde{Y}), \quad \Omega(X_v, Y_v). \tag{6.32}$$

Without loss of generality, one can assume that $\widetilde{X}$ and $\widetilde{Y}$ are horizontal lifts of vector fields on M and that $\xi(X_v)$ and $\xi(Y_v)$ are constant $\mathfrak{g}$-valued functions on P. By (6.20) and (6.28), we have

$$\Omega(\widetilde{X}, \widetilde{Y}) = (d\xi)(\widetilde{X}, \widetilde{Y}), \quad \frac{1}{2}[\xi, \xi](\widetilde{X}, \widetilde{Y}) = [\xi(\widetilde{X}), \xi(\widetilde{Y})] = 0, \tag{6.33}$$

so (6.24) holds in this case. Next, clearly

$$\Omega(\widetilde{X}, Y_v) = 0, \quad [\xi(\widetilde{X}), \xi(Y_v)] = 0, \tag{6.34}$$

while

$$d\xi(\widetilde{X}, Y_v) = \widetilde{X} \cdot \xi(Y_v) - Y_v \cdot \xi(\widetilde{X}) - \xi\big([\widetilde{X}, Y_v]\big). \tag{6.35}$$

Now, having arranged that $\xi(Y_v)$ be a constant $\mathfrak{g}$-valued function on P, we have that $\widetilde{X} \cdot \xi(Y_v) = 0$. Of course, $Y_v \cdot \xi(\widetilde{X}) = 0$. Also, $[\widetilde{X}, Y_v] = -\mathcal{L}_{Y_v} \widetilde{X}$ is horizontal, by (6.10), so $\xi\big([\widetilde{X}, Y_v]\big) = 0$. This verifies (6.24) when both sides act on $(\widetilde{X}, Y_v)$, and similarly we have (6.24) when both sides act on $(X_v, \widetilde{Y})$. We consider the final case. Clearly,

$$\Omega(X_v, Y_v) = 0, \tag{6.36}$$

while

$$d\xi(X_v, Y_v) = X_v \cdot \xi(Y_v) - Y_v \cdot \xi(X_v) - \xi\big([X_v, Y_v]\big) = -\xi\big([X_v, Y_v]\big) \tag{6.37}$$

and

$$\frac{1}{2}[\xi, \xi](X_v, Y_v) = [\xi(X_v), \xi(Y_v)] = \xi\big([X_v, Y_v]\big), \tag{6.38}$$

so (6.24) is verified in this last case, and Theorem 6.1 is proved.

We next obtain a form of the Bianchi identity that will play an important role in the next section. Compare with (1.40) and (2.13).

Proposition 6.2. *We have*

$$d\Omega = [\Omega, \xi]. \tag{6.39}$$

Here, if $\Omega = \sum \Omega_{jk}\, dx_j \wedge dx_k$ in local coordinates, we set

$$\begin{aligned}[\Omega, \xi] &= \sum_{j,k,\ell} [\Omega_{jk}, \xi_\ell]\, dx_j \wedge dx_k \wedge dx_\ell \\ &= -\sum_{j,k,\ell} [\xi_\ell, \Omega_{jk}]\, dx_\ell \wedge dx_j \wedge dx_k = -[\xi, \Omega].\end{aligned} \tag{6.40}$$

To get (6.39), apply d to (6.24), obtaining (since $dd\xi = 0$)

$$d\Omega = \frac{1}{2}[d\xi, \xi] - \frac{1}{2}[\xi, d\xi] = [d\xi, \xi], \tag{6.41}$$

which differs from $[\Omega, \xi]$ by $(1/2)[[\xi, \xi], \xi]$. We have

$$[[\xi, \xi], \xi] = \sum_{j,k,\ell} [[\xi_j, \xi_k], \xi_\ell]\, dx_j \wedge dx_k \wedge dx_\ell. \tag{6.42}$$

Now cyclic permutations of (j, k, ℓ) leave $dx_j \wedge dx_k \wedge dx_\ell$ invariant, so we can replace $[[\xi_j, \xi_k], \xi_\ell]$ in (6.42) by the average over cyclic permutations of (j, k, ℓ). However, Jacobi's identity for a Lie algebra is

$$[[\xi_j, \xi_k], \xi_\ell] + [[\xi_k, \xi_\ell], \xi_j] + [[\xi_\ell, \xi_j], \xi_k] = 0,$$

so $[[\xi, \xi], \xi] = 0$, and we have (6.39).

Exercises

1. Let $P \xrightarrow{p} M$ be a principal G-bundle with connection, where M is a Riemannian manifold. Pick an inner product on $\mathfrak{g}$. For $y \in P$, define an inner product on $T_y P = V_y P \oplus H_y P$ so that if $Z \in T_y P$ has decomposition $Z = Z_v + Z_h$, then
$$\|Z\|^2 = \|\xi(Z_v)\|^2 + \|Dp(y) Z_h\|^2.$$
Show that this is a G-invariant Riemannian metric on P.
2. Conversely, if $P \xrightarrow{p} M$ is a principal G-bundle, and if P has a G-invariant Riemannian metric, show that this determines a connection on P, by declaring that, for each $y \in P$, $H_y P$ is the orthogonal complement of $V_y P$.
3. A choice of section σ of P over an open set $U \subset M$ produces an isomorphism
$$j_\sigma : C^\infty(U, E) \longrightarrow C^\infty(U, V). \tag{6.43}$$

If $\tilde{\sigma}$ is another section, there is a smooth function $g : U \to G$ such that

$$\tilde{\sigma}(x) = \sigma(x) \cdot g(x), \quad \forall\, x \in U. \tag{6.44}$$

Show that

$$j_{\tilde{\sigma}} \circ j_{\sigma}^{-1} v(x) = \pi\big(g(x)\big)^{-1} v(x). \tag{6.45}$$

4. According to (6.12), if $u \in C^\infty(U, E)$ and $v = j_\sigma u,\ \tilde{v} = j_{\tilde{\sigma}} u$, we have

$$(\nabla_X u) \circ \sigma = X \cdot v + \Gamma(X) v, \quad (\nabla_X u) \circ \tilde{\sigma} = X \cdot \tilde{v} + \widetilde{\Gamma}(X) \tilde{v}. \tag{6.46}$$

Show that

$$\widetilde{\Gamma}(X) = \pi\big(g(x)\big)^{-1} \Gamma(X) \pi\big(g(x)\big) + d\pi\big(D\lambda_{g(x)}(g(x)) \circ Dg(x)X\big), \tag{6.47}$$

where $Dg(x)X \in T_{g(x)}G,\ \lambda_g(h) = g^{-1}h,\ D\lambda_g(g) : T_g G \to T_e G \approx \mathfrak{g}$. Compare with (1.41). (*Hint*: Make use of (6.11), plus the identity $(d\pi)(Ad_{g^{-1}} A) = \pi(g)^{-1} d\pi(A) \pi(g),\ A \in \mathfrak{g}$.)

5. Show that, for X, Y vector fields on M, $\Omega(\widetilde{X}, \widetilde{Y})$ satisfies

$$\Omega(\widetilde{X}, \widetilde{Y})(y \cdot g) = \mathrm{Ad}(g)^{-1} \Omega(\widetilde{X}, \widetilde{Y}). \tag{6.48}$$

Deduce that setting

$$\Omega^b(X, Y) = \Omega(\widetilde{X}, \widetilde{Y}) \tag{6.49}$$

defines Ω^b as a section of $\Lambda^2 T^* \otimes (\mathrm{Ad}\, P)$, where $\mathrm{Ad}\, P$ is the vector bundle

$$\mathrm{Ad}\, P = P \times_{\mathrm{Ad}} \mathfrak{g}. \tag{6.50}$$

6. If ξ_0 and ξ_1 are connection 1-forms on $P \to M$, show that $t\xi_1 + (1-t)\xi_0$ is also, for any $t \in \mathbb{R}$. (*Hint*: If P_0 and P_1 are projections, show that $tP_1 + (1-t)P_0$ is also a projection, *provided* that P_0 and P_1 have the same range.)
7. Let ξ_0 and ξ_1 be two connection 1-forms for $P \to M$, and let ∇ be an arbitrary third connection on P. Consider

$$\alpha = \xi_1 - \xi_0. \tag{6.51}$$

If X is a vector field on M and $\widetilde{X}$ the horizontal lift determined by ∇, show that

$$\alpha^b(X) = \alpha(\widetilde{X}) \tag{6.52}$$

defines α^b as an element of $C^\infty(M, \Lambda^1 T^* \otimes \mathrm{Ad}\, P)$. Show that α^b is independent of the choice of ∇.
8. In the setting of Exercise 7, if Ω_j are the curvatures of the connection 1-forms ξ_j, show that

$$\Omega_1 - \Omega_0 = d\alpha + [\alpha, \xi_0] + \frac{1}{2}[\alpha, \alpha]. \tag{6.53}$$

Compare with (2.17) and (3.62). If $d^\nabla \alpha^b$ is the $(\mathrm{Ad}\, P)$-valued 2-form defined as in §2, via the connection ξ_0, relate $d^\nabla \alpha^b$ to $d\alpha + [\alpha, \xi_0]$.

7. The Chern–Weil construction

Let $P \to M$ be a principal G-bundle, endowed with a connection, as in §6. Let Ω be its curvature form, a $\mathfrak{g}$-valued 2-form on P; equivalently, there is the Ad P-valued 2-form Ω^b on M. The Chern–Weil construction gives closed differential forms on M, whose cohomology classes are independent of the choice of connection on P. These "characteristic classes" are described as follows.

A function $f : \mathfrak{g} \to \mathbb{C}$ is called "invariant" if

$$f(\mathrm{Ad}(g)X) = f(X), \quad X \in \mathfrak{g}, \ g \in G. \tag{7.1}$$

Denote by $\mathcal{I}_k$ the set of polynomials $p : \mathfrak{g} \to \mathbb{C}$ which are invariant and homogeneous of degree k. If $p \in \mathcal{I}_k$, there is associated an Ad-invariant k-linear function P on $\mathfrak{g}$, called the *polarization* of p, given by

$$P(Y_1, \dots, Y_k) = \frac{1}{k!} \frac{\partial^k}{\partial t_1 \cdots \partial t_k} p(t_1 Y_1 + \cdots + t_k Y_k), \tag{7.2}$$

such that $p(X) = P(X, \dots, X)$. Into the entries of P we can plug copies of Ω, or of Ω^b, to get $2k$-forms

$$p(\Omega) = P(\Omega, \dots, \Omega) \in \Lambda^{2k} P \tag{7.3}$$

and

$$p(\Omega^b) = P(\Omega^b, \dots, \Omega^b) \in \Lambda^{2k} M. \tag{7.4}$$

Note that if $\pi : P \to M$ is the projection, then

$$p(\Omega) = \pi^* p(\Omega^b); \tag{7.5}$$

we say $p(\Omega)$, a form on P, is "basic," namely, the pull-back of a form on M. The following two propositions summarize the major basic results about these forms.

Proposition 7.1. *For any connection ∇ on $P \to M$, $p \in \mathcal{I}_k$, the forms $p(\Omega)$ and $p(\Omega^b)$ are closed. Hence $p(\Omega^b)$ represents a deRham cohomology class*

$$[p(\Omega^b)] \in \mathcal{H}^{2k}(M, \mathbb{C}). \tag{7.6}$$

If $q \in \mathcal{I}_j$, then $pq \in \mathcal{I}_{j+k}$ and $(pq)(\Omega) = p(\Omega) \wedge q(\Omega)$. Furthermore, if $f : N \to M$ is smooth and ∇_f the connection on f^*P pulled back from ∇ on P, which has curvature $\Omega_f = f^*\Omega$, then

$$p(\Omega^b_f) = f^* p(\Omega^b). \tag{7.7}$$

Proposition 7.2. *The cohomology class (7.6) is independent of the connection on P, so it depends only on the bundle.*

The map $\mathcal{I}_* \to \mathcal{H}^{2*}(M, \mathbb{C})$ is called the *Chern–Weil homomorphism*. We first prove that $d\ p(\Omega) = 0$ on P, the rest of Proposition 7.1 being fairly straightforward. If we differentiate with respect to t at $t = 0$ the identity

$$P\big(\mathrm{Ad}(\mathrm{Exp}\, tY)X, \ldots, \mathrm{Ad}(\mathrm{Exp}\, tY)X\big) = p(X), \tag{7.8}$$

we get

$$\sum P(X, \ldots, [Y, X], \ldots, X) = 0. \tag{7.9}$$

Into this we can substitute the curvature form Ω for X and the connection form ξ for Y, to get

$$\sum P(\Omega, \ldots, [\xi, \Omega], \ldots, \Omega) = 0. \tag{7.10}$$

Now the Bianchi identity $d\Omega = -[\xi, \Omega]$ obtained in (6.8) shows that (7.10) is equivalent to $d\ p(\Omega) = 0$ on P. Since $\pi^* : \Lambda^j M \to \Lambda^j P$ is injective and (7.5) holds, we also have $d\ p(\Omega^b) = 0$ on M, and Proposition 7.1 is proved.

The proof of Proposition 7.2 is conveniently established via the following result, which also has further uses.

Lemma 7.3. *Let ξ_0 and ξ_1 be any $\mathfrak{g}$-valued 1-forms on P (or any manifold). Set $\alpha = \xi_1 - \xi_0$, $\xi_t = \xi_0 + t\alpha$, and $\Omega_t = d\xi_t + (1/2)[\xi_t, \xi_t]$. Given $p \in \mathcal{I}_k$, we have*

$$p(\Omega_1) - p(\Omega_0) = k\ d\left[\int_0^1 P(\alpha, \Omega_t, \ldots, \Omega_t)\, dt\right]. \tag{7.11}$$

Proof. Since $(d/dt)\Omega_t = d\alpha + [\xi_t, \alpha]$, we have

$$\frac{d}{dt}\, p(\Omega_t) = k\ P(d\alpha + [\xi_t, \alpha], \Omega_t, \ldots, \Omega_t). \tag{7.12}$$

It suffices to prove that the right side of (7.12) equals $k\ dP(\alpha, \Omega_t, \ldots, \Omega_t)$. This follows by the "Bianchi" identity $d\Omega_t = -[\xi_t, \Omega_t]$ and the same sort of arguments used in the proof of Proposition 7.1. Instead of (7.8), one starts with

$$P\big(\mathrm{Ad}(\mathrm{Exp}\, tY)Z, \mathrm{Ad}(\mathrm{Exp}\, tY)X, \ldots, \mathrm{Ad}(\mathrm{Exp}\, tY)X\big) = P(Z, X, \ldots, X).$$

To apply this to Proposition 7.2, let ξ_0 and ξ_1 be the connection forms associated to two connections on $P \to M$, so Ω_0 and Ω_1 are their curvature forms. Note that each ξ_t defines a connection form on P, with curvature form Ω_t. Furthermore, $\alpha = \xi_1 - \xi_0$, acting on $X^\# \in T_y P$, depends only on $\pi_* X^\# \in T_x M$ and gives rise to an Ad P-valued 1-form α^b on M. Thus the right side of (7.11) is the pull-back via π^* of the $(2k - 1)$-form

$$k\, d\left[\int_0^1 P(\alpha^b, \Omega_t^b, \dots, \Omega_t^b)\, dt\right] \tag{7.13}$$

on M, which yields Proposition 7.2.

We can also apply Lemma 7.3 to $\xi_1 = \xi$, a connection 1-form, and $\xi_0 = 0$. Then $\xi_t = t\xi$; denote $d\xi_t + (1/2)[\xi_t, \xi_t]$ by Φ_t. We have the $(2k-1)$-form on P called the *transgressed form*:

$$Tp(\Omega) = k \int_0^1 P(\xi, \Phi_t, \dots, \Phi_t)\, dt, \tag{7.14}$$

with

$$\Phi_t = t\, d\xi + \frac{1}{2}t^2[\xi, \xi]. \tag{7.15}$$

Then Lemma 7.3 gives

$$P(\Omega) = d\; Tp(\Omega); \tag{7.16}$$

that is, $p(\Omega)$ is an exact form on P, not merely a closed form. On the other hand, as opposed to $p(\Omega)$ itself, $Tp(\Omega)$ is not necessarily a basic form, that is, the pull-back of a form on M. In fact, $p(\Omega^b)$ is not necessarily an exact form on M; typically it determines a nontrivial cohomology class on M. Transgressed forms play an important role in Chern–Weil theory.

The Levi–Civita connection on an oriented Riemannian manifold of dimension 2 can be equated with a connection on the associated principal S^1-bundle. If we identify S^1 with the unit circle in $\mathbb{C}$, its Lie algebra is naturally identified with $i\mathbb{R}$, and this identification provides an element of $\mathcal{I}_1$, unique up to a constant multiple. This is of course a constant times the product of the Gauss curvature and the volume form, and the invariance of Proposition 7.2 recovers the independence (5.26) of the integrated curvature from the metric used on a Riemannian manifold of dimension 2. More generally, for any complex line bundle L over M, a manifold of any dimension, L can be associated to a principal S^1-bundle, and the Chern–Weil construction produces the class $[\Omega] \in \mathcal{H}^2(M, \mathbb{C})$. The class $c_1(L) = -(1/2\pi i)[\Omega] \in \mathcal{H}^2(M, \mathbb{C})$ is called the first Chern class of the line bundle L. In this case, the connection 1-form on P can be identified with an ordinary (complex-valued) 1-form, and it is precisely the transgressed form (7.14).

Note that if $\dim M = 2$, then (5.39) says that

$$c_1(L)[M] = \text{ Index } X,$$

for any nonvanishing section X of L over $M \setminus \{p_1, \dots, p_K\}$.

For general G, there may be no nontrivial elements of $\mathcal{I}_1$. In fact, if $p : \mathfrak{g} \to \mathbb{R}$ is a nonzero linear form, $V = \ker p$ is a linear subspace of $\mathfrak{g}$ of codimension 1, which is $\mathrm{Ad}\, G$-invariant if $p \in \mathcal{I}_1$. This means V is an ideal: $[V, \mathfrak{g}] \subset V$. Thus there are no nontrivial elements of $\mathcal{I}_1$ unless $\mathfrak{g}$ has an ideal of codimension 1. In particular, if $\mathfrak{g}$ is semisimple, $\mathcal{I}_1 = 0$.

When G is compact, there are always nontrivial elements of $\mathcal{I}_2$, namely, Ad-invariant quadratic forms on $\mathfrak{g}$. In fact, any bi-invariant metric tensor on G gives a positive-definite element of $\mathcal{I}_2$. Applying the Chern–Weil construction in this case then gives cohomology classes in $\mathcal{H}^4(M, \mathbb{C})$.

One way of obtaining elements of $\mathcal{I}_k$ is the following. Let π be a representation of G on a vector space V_π, and set

$$p_{\pi k}(X) = \operatorname{Tr} \Lambda^k d\pi(X), \quad X \in \mathfrak{g}, \tag{7.17}$$

where $d\pi(X)$ denotes the representation of $\mathfrak{g}$ on V_π. In connection with this, note that

$$\det\big(\lambda I + d\pi(X)\big) = \sum_{j=0}^{M} \lambda^{M-j} \operatorname{Tr} \Lambda^j d\pi(X), \quad M = \dim V_\pi. \tag{7.18}$$

If $P \to M$ is a principal $\mathrm{U}(n)$-bundle or $\mathrm{Gl}(n, \mathbb{C})$-bundle, and π the standard representation on $\mathbb{C}^n$, then consider

$$\det\left(\lambda - \frac{1}{2\pi i}\Omega\right) = \sum_{k=0}^{n} c_k(\Omega)\lambda^{n-k}. \tag{7.19}$$

The classes $[c_k(\Omega)] \in \mathcal{H}^{2k}(M, \mathbb{C})$ are the Chern classes of P. If $E \to M$ is the associated vector bundle, arising via the standard representation π, we also call this the kth Chern class of E:

$$c_k(E) = [c_k(\Omega)] \in \mathcal{H}^{2k}(M, \mathbb{C}). \tag{7.20}$$

The object

$$c(E) = \sum c_k(E) \in \bigoplus_{k=0}^{n} \mathcal{H}^{2k}(M, \mathbb{C}) \tag{7.21}$$

is called the total Chern class of such a vector bundle.

If $P \to M$ is a principal $\mathrm{O}(n)$-bundle, and π the standard representation on $\mathbb{R}^n$, then consider

$$\det\left(\lambda - \frac{1}{2\pi}\Omega\right) = \sum_{k=0}^{n} d_k(\Omega)\lambda^{n-k}. \tag{7.22}$$

The polynomials $d_k(\Omega)$ vanish for k odd, since $\Omega^t = -\Omega$, and one obtains Pontrjagin classes:

$$p_k(\Omega) = d_{2k}(\Omega) \in \mathcal{H}^{4k}(M, \mathbb{R}). \tag{7.23}$$

If $F \to M$ is the associated vector bundle, arising from the standard representation π, then $p_k(F)$ is defined to be (7.23).

Exercises

1. If E and F are complex vector bundles over M, we can form $E \oplus F \to M$. Show that
$$c(E \oplus F) = c(E) \wedge c(F),$$
where $c(E)$ is the total Chern class given by (7.21), that is,
$$c(E) = \det\left(I - \frac{1}{2\pi i}\Omega\right) \in \mathcal{H}^{\text{even}}(M, \mathbb{C}),$$
for a curvature 2-form arising from a connection on M.
2. Define the Chern character of a complex vector bundle $E \to M$ as the cohomology class $\text{Ch}(E) \in \mathcal{H}^{\text{even}}(M, \mathbb{C})$ of
$$\text{Ch}(\Omega) = \text{Tr}\, e^{-\Omega/2\pi i},$$
writing $\text{Tr}\, e^{-\Omega/2\pi i} \in \bigoplus_{k \geq 0} \Lambda^{2k} P$ via the power-series expansion of the exponential function. Show that
$$\text{Ch}(E \oplus F) = \text{Ch}(E) + \text{Ch}(F),$$
$$\text{Ch}(E \otimes F) = \text{Ch}(E) \wedge \text{Ch}(F)$$
in $\mathcal{H}^{\text{even}}(M, \mathbb{C})$.
3. If $F \to M$ is a real vector bundle and $E = F \otimes \mathbb{C}$ is its complexification, show that
$$p_j(F) = (-1)^j c_{2j}(E).$$
4. Using $\text{so}(4) \approx \text{so}(3) \oplus \text{so}(3)$, construct two different characteristic classes in $\mathcal{H}^4(M, \mathbb{C})$, when M is a compact, oriented, four-dimensional manifold.

8. The Chern–Gauss–Bonnet theorem

Our goal in this section is to generalize the Gauss–Bonnet formula (5.1), producing a characteristic class derived from the curvature tensor Ω of a Riemannian metric on a compact, oriented manifold M, say $e(\Omega) \in \Lambda^n(M)$, such that

$$\int_M e(\Omega) = \chi(M), \tag{8.1}$$

the right side being the Euler characteristic of M.

A clue to obtaining $e(\Omega)$ comes from the higher-dimensional generalization of the index formula (5.47), namely,

$$\text{Index}(X) = \chi(M), \tag{8.2}$$

valid for any vector field X on M with isolated critical points. The relation between these two when $\dim M = 2$ is noted at the end of §5. It arises from the relation between $\text{Index}(X)$ and the degree of the Gauss map.

Indeed, let M be a compact, n-dimensional submanifold of $\mathbb{R}^{n+k}$, X a (tangent) vector field on M with a finite number of critical points, and $\overline{\mathcal{T}}$ a small tubular neighborhood of M. By Corollary 20.5 of Chap. 1, we know that if $N : \partial\mathcal{T} \to S^{n+k-1}$ denotes the Gauss map on $\partial\mathcal{T}$, formed by the outward-pointing normals, then

$$\text{Index}(X) = \text{Deg}(N). \tag{8.3}$$

As noted at the end of §5 of this chapter, if M is a surface in $\mathbb{R}^3$, with Gauss map N_M, then $\text{Deg}(N_M) = (1/4\pi)\int_M K\, dV$, where K is the Gauss curvature of M, with its induced metric. If $\mathcal{T}$ is a small tubular neighborhood of M in this case, then $\partial\mathcal{T}$ is diffeomorphic to two oppositely oriented copies of M, with approximately the same metric tensor. The outer component of $\partial\mathcal{T}$ has Gauss map approximately equal to N_M, and the inner component has Gauss map approximately equal to $-N_M$. From this we see that (8.2) and (8.3) imply (8.1) with $e(\Omega) = (1/2\pi) K\, dV$ in this case.

We make a further comment on the relation between (8.2) and (8.3). Note that the right side of (8.3) is independent of the choice of X. Thus, any two vector fields on M with only isolated critical points, have the same index. Suppose M has a triangulation τ into n-simplices. There is a construction of a vector field X_τ, illustrated in Fig. 5.3 for $n = 2$, with the property that X_τ has a critical point at each vertex, of index $+1$, and a critical point in the middle of each j-simplex in τ, of index $(-1)^j$, so that

$$\text{Index}(X_\tau) = \sum_{j=0}^{n} (-1)^j \nu_j(M), \tag{8.4}$$

where $\nu_j(M)$ is the number of j-simplices in the triangulation τ of M. We leave the construction of X_τ in higher dimensions as an exercise.

A proof that any smooth, paracompact manifold M is triangulable is given in [Wh]. There it is shown that if you imbed M smoothly in $\mathbb{R}^N$, produce a fine triangulation of $\mathbb{R}^N$, and then perhaps jiggle the imbedding a bit, the intersection provides a triangulation of M.

Now, in view of the invariance of $\text{Index}(X)$, it follows that the right side of (8.4) is independent of the triangulation of M. Also, if M has a more general cell decomposition, we can form the sum on the right side of (8.4), where ν_j stands for the number of j-dimensional cells in M. Each cell can be divided into simplices in such a way that a triangulation is obtained, and the sum on the right side of (8.4) is unchanged under such a refinement. This alternating sum is one definition of the Euler characteristic, but as we have used another definition in §§8 and 9 of Chap. 5, namely

$$\chi(M) = \sum_{j=0}^{n} (-1)^j \dim \mathcal{H}^j(M), \tag{8.5}$$

we will temporarily denote the right side of (8.4) by $\chi_c(M)$.

Now we tackle the question of representing (8.3) as an integrated curvature, to produce (8.1). We begin with the case when M is a compact hypersurface in $\mathbb{R}^{n+1}$. In that case we have, by (4.66),

$$\operatorname{Deg}(N) = \frac{2}{A_n} \int_M \big(\det A_N\big)\, dV, \quad \text{for } n \text{ even}, \tag{8.6}$$

where A_n is the area of S^n and $A_N : T_pM \to T_pM$ is the Weingarten map. The factor 2 arises because $\partial\mathcal{T}$ consists of two copies of M. We can express det A_N directly in terms of the Riemann curvature tensor $R_{jk\ell m}$ of M, using Gauss' *Theorema Egregium.*

In fact, with respect to an oriented orthonormal basis $\{e_j\}$ of T_pM, the matrix of A_N has entries $A_{jk} = \widetilde{II}(e_j, e_k)$, and by (4.14),

$$R_{jk\ell m} = \langle R(e_\ell, e_m)e_k, e_j\rangle = \det \begin{pmatrix} A_{mk} & A_{mj} \\ A_{\ell k} & A_{\ell j} \end{pmatrix}. \tag{8.7}$$

In other words, the curvature tensor captures the action of $\Lambda^2 A_N$ on $\Lambda^2 T_pM$. If $n = 2k$ is even, we can then express det A_N as a polynomial in the components $R_{jk\ell m}$, using

$$\begin{aligned} (\det A_N)e_1 \wedge \cdots \wedge e_n &= \big(\Lambda^n A_N\big)(e_1 \wedge \cdots \wedge e_n) \\ &= (Ae_1 \wedge Ae_2) \wedge \cdots \wedge (Ae_{n-1} \wedge Ae_n). \end{aligned} \tag{8.8}$$

Now, by (8.7),

$$Ae_j \wedge Ae_k = \frac{1}{2} \sum R_{\ell m jk}\, e_\ell \wedge e_m. \tag{8.9}$$

Replacing $(1, \dots, n)$ in (8.8) with all its permutations and summing, we obtain

$$\det A_N = \frac{1}{2^{n/2} n!} \sum_{j,k} (\operatorname{sgn} j)(\operatorname{sgn} k) R_{j_1 j_2 k_1 k_2} \cdots R_{j_{n-1} j_n k_{n-1} k_n}, \tag{8.10}$$

where $j = (j_1, \dots, j_n)$ stands for a permutation of $(1, \dots, n)$. The fact that the quantity (8.10), integrated over M, is equal to $(A_n/2)\chi(M)$ when M is a hypersurface in $\mathbb{R}^{n+1}$ was first established by E. Hopf, as a consequence of his result (8.2). The content of the generalized Gauss–Bonnet formula is that for any compact Riemannian manifold M of dimension $n = 2k$, integrating the right side of (8.10) over M gives $(A_n/2)\chi(M)$.

One key point in establishing the general case is to perceive the right side of (8.10) as arising via the Chern–Weil construction from an invariant polynomial on the Lie algebra $\mathfrak{g} = \mathrm{so}(n)$, to produce a characteristic class. Now the curvature 2-form can in this case be considered a section of $\Lambda^2 T^* \otimes \Lambda^2 T^*$, reflecting the natural linear isomorphism $\mathfrak{g} \approx \Lambda^2 T^*$. Furthermore, $\Lambda^* T^* \otimes \Lambda^* T^*$ has a product, satisfying

$$(\alpha_1 \otimes \beta_1) \wedge (\alpha_2 \otimes \beta_2) = (\alpha_1 \wedge \alpha_2) \otimes (\beta_1 \wedge \beta_2). \tag{8.11}$$

If we set

$$\Omega = \frac{1}{4} \sum R_{jk\ell m} (e_j \wedge e_k) \otimes (e_\ell \wedge e_m), \tag{8.12}$$

then we form the k-fold product, $k = n/2$, obtaining

(8.13)
$$\Omega \wedge \cdots \wedge \Omega = 2^{-n} \sum_{j,k} (\mathrm{sgn}\ j)(\mathrm{sgn}\ k) R_{j_1 j_2 k_1 k_2} \cdots R_{j_{n-1} j_n k_{n-1} k_n} (\omega \otimes \omega),$$

with $\omega = e_1 \wedge \cdots \wedge e_n$. Thus, the right side of (8.10), multiplied by $\omega \otimes \omega$, is equal to $2^{n/2}/n!$ times the right side of (8.13). (Observe the distinction between the product (8.11) and the product on $\mathrm{End}(E) \otimes \Lambda^* T$, used in (7.19) and (7.22), which assigns a different meaning to $\Omega \wedge \cdots \wedge \Omega$.)

Now the Chern–Weil construction produces (8.13), with $\omega \otimes \omega$ replaced by ω, if we use the *Pfaffian*

$$\mathrm{Pf} : \mathrm{so}(n) \longrightarrow \mathbb{R}, \quad n = 2k, \tag{8.14}$$

defined as follows. Let $\xi : \mathrm{so}(n) \to \Lambda^2 \mathbb{R}^n$ be the isomorphism

$$\xi(X) = \frac{1}{2} \sum X_{jk} e_j \wedge e_k, \quad X = (X_{jk}) \in \mathrm{so}(n). \tag{8.15}$$

Then, if $n = 2k$, take a product of k factors of $\xi(X)$ to obtain a multiple of $\omega = e_1 \wedge \cdots \wedge e_n$. Then $\mathrm{Pf}\,(X)$ is uniquely defined by

$$\xi(X) \wedge \cdots \wedge \xi(X) = k!\, \mathrm{Pf}\,(X) \omega. \tag{8.16}$$

Note that if $T : \mathbb{R}^n \to \mathbb{R}^n$ is linear, then $T^* \xi(X) = \xi(T^t X T)$, so

$$\mathrm{Pf}\,(T^t X T) = (\det T) \mathrm{Pf}\,(X). \tag{8.17}$$

Now any $X \in \mathrm{so}(n)$ can be written as $X = T^t Y T$, where $T \in \mathrm{SO}(n)$, that is, T is an orthogonal matrix of determinant 1, and Y is a sum of 2×2 skew-symmetric blocks, of the form

(8.18) $$Y_\nu = \begin{pmatrix} 0 & \lambda_\nu \\ -\lambda_\nu & 0 \end{pmatrix}, \quad \lambda_\nu \in \mathbb{R}.$$

Thus $\xi(Y) = \lambda_1 e_1 \wedge e_2 + \cdots + \lambda_k e_{n-1} \wedge e_n$, so

(8.19) $$\operatorname{Pf}(Y) = \lambda_1 \cdots \lambda_k.$$

Note that $\det Y = (\lambda_1 \cdots \lambda_k)^2$. Hence, by (8.17), we have

(8.20) $$\operatorname{Pf}(X)^2 = \det X,$$

when X is a real, skew-symmetric, $n \times n$ matrix, $n = 2k$. When (8.17) is specialized to $T \in \mathrm{SO}(n)$, it implies that Pf is an invariant polynomial, homogeneous of degree k (i.e., $\operatorname{Pf} \in \mathcal{I}_k,\ k = n/2$).

Now, with Ω in (8.12) regarded as a $\mathfrak{g}$-valued 2-form, we have the left side of (8.13) equal to $(1/k!)\operatorname{Pf}(\Omega)$. Thus we are on the way toward establishing the generalized Gauss–Bonnet theorem, in the following formulation.

Theorem 8.1. *If M is a compact, oriented Riemannian manifold of dimension $n = 2k$, then*

(8.21) $$\chi(M) = (2\pi)^{-k} \int_M Pf(\Omega).$$

The factor $(2\pi)^{-k}$ arises as follows. From (8.10) and (8.13), it follows that when M is a compact hypersurface in $\mathbb{R}^{n+1}$, the right side of (8.6) is equal to $C_k \int_M \operatorname{Pf}(\Omega)$, with

(8.22) $$C_k = \frac{2^{k+1}}{A_n} \frac{k!}{n!}.$$

Now the area of the unit sphere is given by

$$A_{2k} = \frac{2\pi^{k+1/2}}{\Gamma(k+\frac{1}{2})} = \frac{2\pi^k}{\left(k-\frac{1}{2}\right)\cdots\left(\frac{1}{2}\right)},$$

as is shown in Appendix A of Chap. 3; substituting this into (8.22) gives $C_k = (2\pi)^{-k}$.

We give a proof of this which extends the proof of (5.24), in which handles are added to a surface. To effect this parallel, we consider how the two sides of (8.21) change when M is altered by a certain type of *surgery*, which we will define in the next paragraph.

First, we mention another ingredient in the proof of Theorem 8.1. Namely, the right side of (8.21) is independent of the choice of metric on M. Since different metrics produce different $\mathrm{SO}(2k)$ frame bundles, this assertion is not a simple consequence of Proposition 7.2. We will postpone the proof of this invariance until near the end of this section.

We now describe the "surgeries" alluded to above. To perform surgery on M_0, a manifold of dimension n, excise a set H_0 diffeomorphic to $S^{\ell-1} \times B^m$, with $m + \ell - 1 = n$, where $B^m = \{x \in \mathbb{R}^m : |x| < 1\}$, obtaining a manifold with boundary X, ∂X being diffeomorphic to $S^{\ell-1} \times S^{m-1}$. Then attach to X a copy of $B^\ell \times S^{m-1}$, sewing them together along their boundaries, both diffeomorphic to $S^{\ell-1} \times S^{m-1}$, to obtain M_1. Symbolically, we write

$$M_0 = X \# H_0, \quad M_1 = X \# H_1. \tag{8.23}$$

We say M_1 is obtained from M_0 by a surgery of type (ℓ, m).

We compare the way each side of (8.21) changes when M changes from M_0 to M_1. We also look at how $\chi_c(M)$, defined to be the right side of (8.4), changes. In fact, this definition easily yields

$$\chi(X\#H_1) = \chi(X\#H_0) - \chi(H_0) + \chi(H_1). \tag{8.24}$$

For notational simplicity, we have dropped the "c" subscript. It is more convenient to produce an identity involving only manifolds without boundary, so note that

$$\begin{aligned} \chi(H_0\#H_0) &= 2\chi(H_0) - \chi(\partial H_0), \\ \chi(H_1\#H_1) &= 2\chi(H_1) - \chi(\partial H_1), \end{aligned} \tag{8.25}$$

and, since $\partial H_0 = \partial H_1$, we have

$$\chi(H_1) - \chi(H_0) = \frac{1}{2}\chi(H_1\#H_1) - \frac{1}{2}\chi(H_0\#H_0), \tag{8.26}$$

hence

$$\chi(M_1) = \chi(M_0) + \frac{1}{2}\chi(H_1\#H_1) - \frac{1}{2}\chi(H_0\#H_0). \tag{8.27}$$

Note that $H_0\#H_0 = S^{\ell-1} \times S^m$, $H_1\#H_1 = S^\ell \times S^{m-1}$. To compute the Euler characteristic of these two spaces, we can use multiplicativity of χ. Note that products of cells in Y_1 and Y_2 give cells in $Y_1 \times Y_2$, and

$$\nu_j(Y_1 \times Y_2) = \sum_{i+k=j} \nu_i(Y_1)\nu_k(Y_2); \tag{8.28}$$

then from (8.4) it follows that

$$\chi(Y_1 \times Y_2) = \sum_{j\ge 0}(-1)^j \sum_{i+k=j} \nu_i(Y_1)\nu_k(Y_2) = \chi(Y_1)\chi(Y_2). \tag{8.29}$$

Using the fairly elementary result that

$$\chi(S^j) = 2 \quad \text{if } j \text{ is even}, \qquad 0 \quad \text{if } j \text{ is odd}, \tag{8.30}$$

we have $\chi(H_0 \# H_0) - \chi(H_1 \# H_1)$ equal to 4 if ℓ is odd and m even, -4 if ℓ is even and m odd, and 0 if ℓ and m have the same parity (which does not arise if $\dim M$ is even).

The change in $\chi_c(M)$ just derived in fact coincides with the change in $\chi(M)$, defined by (8.5). This follows from results on deRham cohomology obtained in Chap. 5. In fact (B.8) of Chap. 5 implies (8.24), from which (8.25)–(8.27) follow; (8.52) of Chap. 5 implies (8.29) when Y_j are smooth, compact manifolds; and (8.56) and (8.57) of Chap. 5 imply (8.30).

Thus, for $e(M) = \int_M e(\Omega)$ to change the same way as $\chi(M)$ under a surgery, we need the following properties in addition to "functoriality." We need

$$e(S^j \times S^k) = 0 \quad \text{if } j \text{ or } k \text{ is odd}, \qquad 4 \quad \text{if } j \text{ and } k \text{ are even}. \tag{8.31}$$

If $e(\Omega)$ is locally defined, we have, upon giving X, H_0, and H_1 coherent orientations,

$$\int_{M_1} e(\Omega) = \int_{M_0} e(\Omega) - \int_{H_0} e(\Omega) + \int_{H_1} e(\Omega), \tag{8.32}$$

parallel to (8.24). Place metrics on M_j that are product metrics on $(-\varepsilon, \varepsilon) \times S^{\ell-1} \times S^{m-1}$ on a small neighborhood of ∂X. If we place a metric on $H_j \# H_j$ which is symmetric with respect to the natural involution, we will have

$$\int_{H_j} e(\Omega) = \frac{1}{2} \int_{H_j \# H_j} e(\Omega), \tag{8.33}$$

provided $e(\Omega)$ has the following property. Given an oriented Riemannian manifold Y, let $Y^\#$ be the same manifold with orientation reversed, and let the associated curvature forms be denoted by Ω_Y and $\Omega_{Y^\#}$. We require

$$e(\Omega_Y) = -e(\Omega_{Y^\#}). \tag{8.34}$$

Now $e(\Omega) = \text{Pf}\,(\Omega/2\pi)$ certainly satisfies (8.34), in view of the dependence on orientation built into (8.16). To see that (8.31) holds in this case, we need only note that $S^\ell \times S^k$ can be smoothly imbedded as a hypersurface in $\mathbb{R}^{\ell+k+1}$. This can be done via imbedding $S^\ell \times I \times B^k$ into $\mathbb{R}^{\ell+k+1}$ and taking its boundary (and smoothing it out). In that case, since $\text{Pf}(\Omega/2\pi)$ is a characteristic class whose integral is independent of the choice of metric, we can use the metric induced from the imbedding. We now have (8.31)–(8.33). Furthermore, for such a hypersurface $M = H_j \# H_j$, we know that the right side of (8.21) is equal to $\chi_c(H_j \# H_j)$, by the argument preceding the statement of Theorem 8.1, and since (8.29) and (8.30) are both valid for both χ and χ_c, we also have this quantity equal to $\chi(H_j \# H_j)$.

It follows that (8.21) holds for any M obtainable from S^n by a finite number of surgeries. With one extra wrinkle we can establish (8.21) for all compact, oriented M. The idea for using this technique is one the author learned from J. Cheeger, who uses a somewhat more sophisticated variant in work on analytic torsion [Ch].

Assume M is connected. Give $M \times \mathbb{R}$ the product Riemannian metric, fix a point $p \in M$, and, with $q = (p, 0) \in M \times \mathbb{R}$, consider on $M \times \mathbb{R}$ the function $f_0(x,t) = \operatorname{dist}((x,t),q)^2$. For R sufficiently large, $f_0^{-1}(R)$ is diffeomorphic to two copies of M, under the map $(x,t) \mapsto x$. For $r > 0$ sufficiently small, $f_0^{-1}(r)$ is diffeomorphic to the sphere S^n.

Our argument will use some basic results of Morse theory. A Morse function $f : Z \to \mathbb{R}$ is a smooth function on a manifold Z all of whose critical points are nondegenerate, that is, if $\nabla f(z) = 0$, then $D^2 f(z)$ is an invertible $\nu \times \nu$ matrix, $\nu = \dim Z$. One also assumes that f takes different values at distinct critical points and that $f^{-1}(K)$ is compact for every compact $K \subset \mathbb{R}$. Now the function f_0 above may not be a Morse function on $Z = M \times \mathbb{R}$, but there will exist a smooth perturbation f of f_0 which is a Morse function. A proof is given in Proposition 4.3 of Appendix B. The new f will share with f_0 the property that $f^{-1}(r)$ is diffeomorphic to S^n and $f^{-1}(R)$ is diffeomorphic to two copies of M. Note that an orientation on M induces an orientation on $M \times \mathbb{R}$, and hence an orientation on any level set $f^{-1}(c)$ that contains no critical points. In particular, $f^{-1}(R)$ is a union of two copies of M with opposite orientations. The following is a basic tool in Morse theory.

Proposition 8.2. *If $c_1 < c_2$ are regular values of a Morse function $f : Z \to \mathbb{R}$ and there is exactly one critical point z_0, with $c_1 < f(z_0) < c_2$, then $M_2 = f^{-1}(c_2)$ is obtained from $M_1 = f^{-1}(c_1)$ by a surgery. In fact, if $D^2 f(z_0)$ has signature (ℓ, m), M_2 is obtained from M_1 by a surgery of type (m, ℓ).*

This is a consequence of the following result, known as the Morse lemma.

Theorem 8.3. *Let f have a nondegenerate critical point at $p \in Z$. Then there is a coordinate system $(x_1, \dots, x_n)$ centered at p in which*

$$f(x) = f(p) + x_1^2 + \cdots + x_\ell^2 - x_{\ell+1}^2 - \cdots - x_{\ell+m}^2 \tag{8.35}$$

near the origin, where $\ell + m = \nu = \dim Z$.

Proof. Suppose that in some coordinate system $D^2 f(p)$ is given by a nondegenerate, symmetric, $\nu \times \nu$ matrix A. It will suffice to produce a coordinate system in which

$$f(x) = f(p) + \frac{1}{2} Ax \cdot x, \tag{8.36}$$

near the origin, since going from here to (8.35) is a simple exercise in linear algebra. We will arrange (8.36) by an argument, due to Palais, similar to the proof of Darboux' theorem in Chap. 1, §14.

Begin with any coordinate system centered at p. Let

$$\omega_1 = df, \quad \omega_0 = dg, \ \text{ where } g(x) = \frac{1}{2} Ax \cdot x, \tag{8.37}$$

with $A = D^2 f(0)$ in this coordinate system. Set $\omega_t = t\omega_1 + (1-t)\omega_0$, which vanishes at p for each $t \in [0, 1]$. The nondegeneracy hypothesis on A implies that the components of each ω_t have linearly independent gradients at p; hence there exists a smooth, time-dependent vector field X_t (not unique), such that

$$\omega_t \rfloor X_t = g - f, \quad X_t(p) = 0. \tag{8.38}$$

Let $\mathcal{F}_t$ be the flow generated by X_t, with $\mathcal{F}_0 = Id$. Note that $\mathcal{F}_t$ fixes p. It is then an easy computation using (8.38), plus the identity $\mathcal{L}_X \omega = d(\omega \rfloor X) + (d\omega) \rfloor X$, that

$$\frac{d}{dt}\big(\mathcal{F}_t^* \omega_t\big) = 0. \tag{8.39}$$

Hence $\mathcal{F}_1^* \omega_1 = \omega_0$, so $f \circ \mathcal{F}_1 = g$ and the proof of Proposition 8.3 is complete.

From Theorem 8.2, it follows that, given any compact, oriented, connected M, of dimension n, a finite number of surgeries on S^n yields two copies of M, with opposite orientations, say M and $M^\#$. Hence (8.21) holds with M replaced by the disjoint union $M \cup M^\#$. But, in view of (8.34), both sides of the resulting identity are equal to twice the corresponding sides of (8.21); for χ_c this follows easily from (8.4), and for χ it follows immediately from (8.5). We hence have the Chern–Gauss–Bonnet formula and also the identity $\chi(M) = \chi_c(M)$, modulo the task of showing the invariance of the right side of (8.21) under changes of metric on M.

We turn to the task of demonstrating such invariance. Say g_0 and g_1 are two Riemannian metric tensors on M, with associated SO(n)-bundles $P_0 \to M$, $P_1 \to M$, having curvature forms Ω_0 and Ω_1. We want to show that $\text{Pf}\,(\Omega_1^b) - \text{Pf}\,(\Omega_0^b)$ is exact on M. To do this, consider the family of metrics $g_t = tg_1 + (1-t)g_0$ on M, with associated SO(n)-bundles $P_t \to M$, for $t \in [0, 1]$. These bundles fit together to produce a principal SO(n)-bundle $\widetilde{P} \to M \times [0, 1]$. We know there exists a connection on this principal bundle. Let $T = \partial/\partial t$ on $M \times [0, 1]$, and let $\widetilde{T}$ denote its horizontal lift (with respect to a connection chosen on $\widetilde{P}$). The flow generated by $\widetilde{T}$ commutes with the SO(n)-action on $\widetilde{P}$. Flowing along one unit of time then yields a diffeomorphism $\Phi : P_0 \to P_1$, commuting with the SO(n)-action, hence giving an isomorphism of SO(n)-bundles. Now applying Proposition 7.2 to the original connection on P_0 and to that pulled back from P_1 gives the desired invariance.

Before Chern's work, H. Hopf had established Theorem 8.1 when M is a compact hypersurface in $\mathbb{R}^{2k+1}$. Then C. Allendoerfer [Al] and W. Fenchel [Fen] proved (8.21) for the case when M is isometrically imbedded in $\mathbb{R}^{n+k}$, by relating the integral on the right to the integral over $\partial\mathcal{T}$ of the Gauss curvature of the boundary of a small tubular neighborhood $\mathcal{T}$ of M, and using the known result that $\chi(\partial\mathcal{T}) = 2\chi(M)$. At that time it was not known that every compact Riemannian manifold could be isometrically imbedded in Euclidean space. By other means, Allendoerfer and A.Weil [AW] proved Theorem 8.1, at least for real analytic metrics, via a triangulation and local isometric imbedding. Chern then produced an intrinsic proof of Theorem 8.1 and initiated a new understanding of characteristic classes.

In Chern's original paper [Cher], it is established that $\int_M \operatorname{Pf}(\Omega/2\pi)$ is equal to the index of a vector field X on M, by a sophisticated variant of the argument establishing Proposition 5.4, involving a differential form on the unit-sphere bundle of M related to, but more complicated than, the transgressed form (7.14). An exposition of this argument can also be found in [Poo] and in [Wil]. When dim $M = 2$, one can identify the unit-sphere bundle and the frame bundle, and in that case the form coincides with the transgressed form and the argument becomes equivalent to that used to prove Proposition 5.4. An exposition of the proof of Theorem 8.1 using tubes can be found in [Gr].

The Chern–Gauss–Bonnet theorem can also be considered as a special case of an index theorem for an elliptic differential operator. A proof in that spirit is given in Chap. 10, §7. More material on the Pfaffian is also developed there.

We mention a further generalization of the Gauss–Bonnet formula. If $E \to X$ is an SO$(2k)$-bundle over a compact manifold X (say of dimension n), with metric connection ∇ and associated curvature Ω, then $\operatorname{Pf}(\Omega/2\pi)$ is defined as a $(2k)$-form on X. This gives a class $\operatorname{Pf}(E) \in \mathcal{H}^{2k}(X)$, independent of the choice of connection on E, as long as it is a metric connection. There is an extension of Theorem 8.1, describing the cohomology class of $\operatorname{Pf}(E)$ in $\mathcal{H}^{2k}(X)$. Treatments of this can be found in [KN] and in [Stb].

Exercises

1. Verify that when Ω is the curvature 2-form arising from the standard metric on S^{2k}, then
$$\int_{S^{2k}} \operatorname{Pf}\left(\frac{\Omega}{2\pi}\right) = 2.$$
2. Generalize Theorem 8.1 to the nonorientable case. (*Hint*: If M is not orientable, look at its orientable double cover $\widetilde{M}$. Use (8.4) to show that $\chi(\widetilde{M}) = 2\chi(M)$.) Using (8.16) as a local identity, define a *measure* $\widetilde{\operatorname{Pf}}(\Omega)$ in the nonorientable case.
3. Let M be a compact, complex manifold of complex dimension n (i.e., real dimension $2n$). Denote by $\mathcal{T}$ its tangent bundle, regarded as a complex vector bundle, with fibers $\mathcal{T}_p$ of complex dimension n. Show that

$$\int_M c_n(\mathcal{T}) = \chi(M),$$

where $c_n(\mathcal{T})$ is the top Chern class, defined by (7.19) and (7.20).

4. If M_j are compact Riemannian manifolds with curvature forms Ω_j and $M_1 \times M_2$ has the product metric, with curvature form Ω, show directly that

$$\pi_1^* \mathrm{Pf}\,(\Omega_1) \wedge \pi_2^* \mathrm{Pf}\,(\Omega_2) = \mathrm{Pf}\,(\Omega),$$

where π_j projects $M_1 \times M_2$ onto M_j. If dim M_j is odd, set $\mathrm{Pf}\,(\Omega_j) = 0$. Use this to reprove (8.31) when $e(\Omega) = \mathrm{Pf}\,(\Omega)$.

5. Show directly that the right sides of (8.2) and (8.3) both vanish when M is a hypersurface of odd dimension in $\mathbb{R}^{n+1}$.

References

[AbM] R.Abraham and J.Marsden, *Foundations of Mechanics*, Benjamin/Cummings, Reading, MA, 1978

[Al] C. Allendoerfer, The Euler number of a Riemannian manifold, *Am. J. Math.* 62(1940), 243–248

[AW] C. Allendoerfer and A. Weil, The Gauss–Bonnet theorem for Riemannian polyhedra, *Trans. AMS* 53(1943), 101–129

[Ar] V. Arnold, *Mathematical Methods of Classical Mechanics*, Springer, New York, 1978

[BGM] M. Berger, P. Gauduchon, and E. Mazet, *Le Spectre d'une Variété Riemannienne*, LNM no.194, Springer, New York, 1971

[Bes] A. Besse, *Einstein Manifolds*, Springer, New York, 1987

[Car] M. doCarmo, *Riemannian Geometry*, Birkhauser, Boston, 1992

[Ch] J. Cheeger, Analytic torsion and the heat equation, *Ann. Math.* 109(1979), 259–322

[CE] J. Cheeger and D. Ebin, *Comparison Theorems in Riemannian Geometry*, North Holland, Amsterdam, 1975

[Cher] S. S. Chern, A simple intrinsic proof of the Gauss–Bonnet formula for closed Riemannian manifolds., *Ann. Math.* 45(1944), 747–752

[Chv] C. Chevalley, *Theory of Lie Groups*, Princeton University Press, Princeton, NJ, 1946

[EGH] T. Eguchi, P. Gilkey, and A. Hanson, Gravitation, Gauge Theories, and Differential Geometry, *Physics Reports*, Vol. 66, no. 6 (1980)

[Fen] W. Fenchel, On total curvature of Riemannian manifolds, *J. London Math. Soc.* 15(1940), 15–22

[FU] D. Freed and K. Uhlenbeck, *Instantons and Four-Manifolds*, Springer, New York, 1984

[Gil] P. Gilkey, *The Index Theorem and the Heat Equation*, Publish or Perish, Boston, 1974

[Gr] A. Gray, *Tubes*, Addison-Wesley, New York, 1990

[Gu] R. Gunning, *Lectures on Riemann Surfaces*, Princeton University Press, Princeton, NJ, 1967

[Gun] M. Günther, On the perturbation problem associated to isometric embeddings of Riemannian manifolds, *Ann. Global Anal. Geom.* 7(1989), 69–77

[Hi] N. Hicks, *Notes on Differential Geometry*, Van Nostrand, New York, 1965

[Jos] J. Jost, *Nonlinear Methods in Riemannian and Kahlerian Geometry*, Birkhauser, Boston, 1988

[Kaz] J. Kazdan, Prescribing the curvature of a Riemannian manifold, *CBMS Reg. Conf. Ser. in Math.*, no. 57, AMS, Providence, RI, 1985

[KN] S. Kobayashi and N. Nomizu, *Foundations of Differential Geometry*, Interscience, New York, Vol. 1, 1963; Vol. 2, 1969

[LM] H. B. Lawson and M. L. Michelson, *Spin Geometry*, Princeton University Press, Princeton, NJ, 1989

[Mil] J. Milnor, *Morse Theory*, Princeton University Press, Princeton, NJ, 1963

[Mil2] J. Milnor, *Topology from the Differentiable Viewpoint*, University Press of Virginia, Charlottesville, 1965

[MiS] J. Milnor and J. Stasheff, *Characteristic Classes*, Princeton University Press, Princeton, NJ, 1974

[Pal] R. Palais, ed., *Seminar on the Atiyah–Singer Index Theorem*, Princeton University Press, Princeton, NJ, 1963

[Poo] W. Poor, *Differential Geometric Structures*, McGraw-Hill, New York, 1981

[Stb] M. Spivak, *A Comprehensive Introduction to Differential Geometry*, Vols. 1–5, Publish or Perish Press, Berkeley, CA, 1979

[Stb] S. Sternberg, *Lectures on Differential Geometry*, Prentice Hall, Englewood Cliffs, NJ, 1964

[Str] D. Struik, *Lectures on Classical Differential Geometry*, Addison-Wesley, Reading, MA, 1964

[Wh] H. Whitney, *Geometric Integration Theory*, Princeton University Press, Princeton, NJ, 1957

[Wil] T. Wilmore, *Total Curvature in Riemannian Geometry*, Wiley, New York, 1982

Index