EIGENVALUES IN RIEMANNIAN GEOMETRY

ISAAC CHAVEL

Eigenvalues in Riemannian Geometry

This is a volume in
PURE AND APPLIED MATHEMATICS

A Series of Monographs and Textbooks

Editors: SAMUEL EILENBERG AND HYMAN BASS

A list of recent titles in this series appears at the end of this volume:

Eigenvalues in Riemannian Geometry

ISAAC CHAVEL
Department of Mathematics
The City College of the City University of New York
New York, New York

With a chapter by
BURTON RANDOL
Graduate School of the
City University of New York

With Appendix by
JOZEF DODZIUK
Queens College of the
City University of New York

1984

ACADEMIC PRESS, INC.
(Harcourt Brace Jovanovich, Publishers)
Orlando San Diego New York London
Toronto Montreal Sydney Tokyo

ACADEMIC PRESS, INC.
Orlando, Florida 32887

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DX

Library of Congress Cataloging in Publication Data

Chavel, Isaac.
Eigenvalues in Riemannian geometry.

(Pure and applied mathematics)
Includes index.
1. Geometry, Riemannian. 2. Differential equations, Partial. 3. Eigenvalues. I. Title. II. Series: Pure and applied mathematics (Academic Press)
QA3.P8 [QA649] 510s [516.92] 84-6494
ISBN 0-12-170640-0 (alk. paper)

PRINTED IN THE UNITED STATES OF AMERICA

84 85 86 87 9 8 7 6 5 4 3 2 1

For
IRVING KAMPNER
(1914–1981)
with love and affection

Contents

Preface xi

CHAPTER I
The Laplacian

1. Definitions and Preliminaries 1
2. Green's Formulas 5
3. Basic Facts for Eigenvalue Problems 7
4. The Wave and Heat Equations 11
5. Rayleigh and Max–Min Methods 13

CHAPTER II
The Basic Examples

1. Some Generalities 26
2. Tori 28
3. Weyl's Formula for Bounded Domains in $\mathbb{R}^n$ 31
4. Spheres and Real Projective Spaces 33
5. Disks in Constant Curvature Space Forms 36

CHAPTER III
λ_1 and Curvature

1. Geodesics and Curvature 55
2. Comparison Theorems for Sectional Curvature Bounded from Above 67
3. Comparison Theorems for Ricci Curvature Bounded from Below 71
4. Obata and Toponogov Theorems 82

CHAPTER IV
Isoperimetric Inequalities

1. The Co-Area Formula 85
2. The Faber–Krahn Inequality 86

3. The Cheeger, Sobolev, and Isoperimetric Constants 95
4. The Sobolev Constant and Eigenvalue, Eigenfunction, Estimates 99
5. The Constants and Estimates for the Closed Eigenvalue Problem 109

CHAPTER V
Eigenvalues and the Kinematic Measure

1. The Analytic Inequality 113
2. M. Berger's Isoembolic Inequality 117
3. Cheeger and Isoperimetric Constants, and the Kinematic Measure 124

CHAPTER VI
The Heat Kernel for Compact Manifolds

1. Duhamel's Principle and Its Consequences 136
2. The Heat Equation on $\mathbb{R}^n$ 142
3. The Minakshisundaram-Pleijel Recursion Formulas 148
4. Existence of the Heat Kernel 151

CHAPTER VII
The Dirichlet Heat Kernel for Regular Domains

1. Preliminaries 159
2. The Dirichlet Heat Kernel for Regular Domains 161
3. Duhamel's Principle 164

CHAPTER VIII
The Heat Kernel for Noncompact Manifolds

1. The Maximum Principle, and Uniqueness Theorems, for the Heat Operator 180
2. The Heat Kernel for Noncompact Manifolds 187
3. Comparison Theorems for Heat Kernels 192
4. Upper Bounds for the Heat Kernel 196

CHAPTER IX
Topological Perturbations with Negligible Effect

1. Statement and Discussion of the Results 209
2. Proof of Theorems 1–6 216

3. Using Rayleigh's Characterization of Eigenvalues 228
4. The Mathematics of Crushed Ice 233

CHAPTER X
Surfaces of Constant Negative Curvature

1. Geometry of the Hyperbolic Plane 239
2. The Heat Kernel of the Hyperbolic Plane 242
3. Löbell Surfaces and the Estimates of P. Buser 246
4. Low Eigenvalues and Short Geodesics 254
5. The Upper Half-Space Model of Hyperbolic Space 262

CHAPTER XI
The Selberg Trace Formula by Burton Randol

1. Preliminaries 266
2. The Pretrace Formula 271
3. Applications of the Pretrace Formula 278
4. The Trace Formula 288
5. Applications of the Trace Formula 293
6. Concluding Remarks 302

CHAPTER XII
Miscellanea

1. Volumes of Disks and Spheres 303
2. The Fourier Transform 304
3. The Poisson Summation Formula 306
4. The Fourier Transform and the Heat Equation 307
5. Eigenfunctions on Spheres and Hyperbolic Space 308
6. Minimal Submanifolds of Euclidean Spaces and Spheres 309
7. Normalization of Geometric Data 314
8. Geodesic Coordinates 316
9. The Levy–Gromov Isoperimetric Inequality 322
10. Heat Conduction on the Euclidean Upper Half-Space 325
11. The Maximum Principle for the Laplacian 329
12. Recent Progress on Eigenvalue and Heat Kernel Estimates 330

APPENDIX
Laplacian on Forms by Jozef Dodziuk 334

Bibliography 345
Index 359

This Page Intentionally Left Blank

Preface

The subject of the geometry of the Laplace operator has undergone a veritable explosion in the past fifteen years, characterized by a dizzying wealth and variety of subject matter and techniques. The only systematic treatment available, with a Riemannian geometric viewpoint, is the lecture-notes volume by Berger–Gauduchon–Mazet [1], which appeared in 1970. It therefore seemed to me, ten years later, that an updated introduction would be most desirable.

The basic goals of the book are: (i) to introduce the subject to those interested in discovering it, (ii) to coherently present a number of basic techniques and results, currently used in the subject, to those working in it, and (iii) to present some of the results that are attractive in their own right, and which lend themselves to a presentation not overburdened with technical machinery.

I have made no attempt to give an exhaustive definitive account of the subject with a grand overview—in fact, I have resisted any temptation to do so. More particularly, I have restricted the subject matter to the Laplacian acting on functions (except for an appendix, graciously contributed by J. Dodziuk, on the Laplacian on forms), to questions concerning eigenvalues of the Laplacian for compact manifolds and domains with compact closure, and to the study of the associated questions concerning the heat equation. (The study of the heat equation for noncompact manifolds was viewed as a natural extension and application of the study of Dirichlet heat kernels of domains with compact closure.) Even in this area, I did not present, fully, all the topics deserving a detailed presentation. In some cases I settled for remarks on the results and, in others, for references (e.g., the inequalities of Hayman and Osserman relating the inradius of a domain to its lowest Dirichlet eigenvalue). In still other instances, I simply relied on the bibliography to put the reader on the track to related topics.

I also add that there is no overarching philosophy, nor *big questions* point-of-view, to my presentation. The hallmark of the subject, to my mind, is the lively and variegated interaction of the fields of topology and

Riemannian geometry with the fields of partial differential equations, probability, and number theory; and it is precisely this quality that I have tried to capture in these pages.

The first chapter introduces the Laplacian or, more precisely, the Laplace–Beltrami operator, acting on functions on Riemannian manifolds, summarizes the basic facts of the existence of eigenvalues, and their associated Weyl formulas, and then presents the classical Rayleigh characterization of eigenvalues, followed by the max–min method.

The second chapter gives a discussion of the basic examples, emphasizing, in some detail, the Dirichlet and Neumann eigenvalues of geodesic disks in simply connected constant curvature space forms. The third chapter then develops the comparison arguments with which to estimate the lowest Dirichlet eigenvalue of a geodesic disk in a Riemannian manifold in terms of the corresponding eigenvalue in a constant curvature space form, the constant in question being an upper or lower bound of the sectional curvatures of the geodesic disk of the original manifold. These estimates are then applied to derive the theorems of Obata and Topogonov. A summary of information about the exponential map, Jacobi fields, and geodesic spherical coordinates is presented in the beginning of Chapter III.

Chapters IV and V are devoted to isoperimetric inequalities. We start with the celebrated Faber–Krahn solution to the Rayleigh conjecture and with a number of variations of the argument. Then we discuss the Cheeger and isoperimetric constants and their effect on estimating eigenvalues and eigenfunctions. Chapter V is devoted to estimating the Cheeger and isoperimetric constants in terms of Riemannian data of the underlying manifold. It starts with the contributions of M. Berger and J. L. Kazdan toward the solution of the Blaschke conjecture, which are then applied, by arguments of C. B. Croke, to the estimation of the Cheeger and isoperimetric constants.

In Chapters VI–IX we focus on the heat equation. We give the existence theorems for eigenvalues (of the closed and Dirichlet eigenvalue problems) and the convergence of temperature distributions (for large times) to steady state, following the arguments of A. N. Milgram and P. C. Rosenbloom, which, in turn, require the existence of the heat kernel, for which we use the construction of S. Minakshisundaram. (The advantage of this approach is that the most advanced spectral theorem required is the one for compact operators on Hilbert space.) Also included are the Weyl formulas for these two eigenvalue problems. We then study the heat kernel on noncompact Riemannian manifolds, starting with the uniqueness and existence theorems of J. Dodziuk, followed by the comparison theorems of J. Cheeger and S. T. Yau, and concluding with the estimates of J. Cheeger, M. Gromov, and M. Taylor.

In Chapter IX we present the work of E. A. Feldman and myself on topological perturbations of the underlying manifold with negligible spectral effects. Besides my obvious interest in these questions, I consider these results quite suitable for the book, as they draw on much of the earlier material, and as they provide an appropriate context for introducing considerations of Brownian motion to the study of the heat kernel and, thereby, the study of eigenvalues.

The next two chapters study compact surfaces of constant negative curvature. In Chapter X we study the interaction of the geometry–topology of the surface with low eigenvalues, using the arguments of P. Buser, B. Randol, and S. Wolpert–R. Schoen–S. T. Yau. These arguments require only Chapters I–IV. (In fact, one might construct a convenient introductory short-course from Chapters I–V followed by Sections 1, 3, and 4 of Chapter X.) The whole chapter contrasts sharply with the elegant and powerful theory of the Selberg trace formula, presented in Chapter XI by B. Randol. One enters, here, to a completely fresh enterprise revealing a development of the subject of eigenvalues along totally different lines from those featured earlier in the book.

The final chapter is what the title says it is: Miscellanea—those matters that did not fit smoothly into the flow of the text, but which I did not want to exclude. The Appendix is an essay, by J. Dodziuk, summarizing some of the basic features of the theory of the Laplacian on forms—most notably, the heat equation approach to index theorems.

The bibliography is far from complete, but I think that it is sufficiently serviceable to the reader wishing to continue exploring the subject. I only hope that there are no gross injustices in the inclusion–exclusion of references, and in the attribution of results.

To my knowledge, earlier surveys and collections of articles are as follows: (i) Volumes 16(1970), 23(1973), 27(1975), and 36(1980), of the Proceedings of American Mathematical Society Symposia in Pure Mathematics, with Volume 36 completely devoted to the geometry of the Laplace operator. (ii) The books Polya–Szegö [1], Bandle [1], and the survey articles of Payne [1] and Osserman [3; 4], on isoperimetric inequalities. (iii) The articles of M. Berger in Volumes 16 and 27 of the above AMS Proceedings, the survey articles of Simon–Wissner [1], Li [4], the lecture-notes Berger–Gauduchon–Mazet [1], the bibliography Bérard–Berger [1], and the broad survey article of S. T. Yau [3] on partial differential equations in differential geometry.

It is a pleasure to thank the geometers and analysts of the doctoral faculty of the City University of New York for their extended help and discussions since I came to C.U.N.Y. in 1970—J. Dodziuk, E. A. Feldman, S. Kaplan, B. Randol, and R. Sacksteder. A special thanks goes to J. Dodziuk for his appendix on the Laplacian on forms.

I am especially grateful to B. Randol for his chapter on the Selberg trace formula and am proud that his essay is included in this book.

And, finally, I am delighted to acknowledge my debt to Edgar A. Feldman, with whom I have worked on this subject for more years than it is wise to announce. There is hardly a worthwhile insight that I brought to the pages of this book in which he does not have a major share.

Riverdale, New York ISAAC CHAVEL
July 1984

CHAPTER I

The Laplacian

In this chapter we present the basic definitions and facts to be used in the subsequent chapters. Few proofs are presented here—our main interest is in the interplay between the eigenvalues of the Laplacian and the global geometric invariants of the underlying manifold, and we wish to get to this topic as quickly as possible. Thus, for example, we shall consider the heat equation later, in some detail, and content ourselves in Section 4 with the formal calculation connecting the phenomena of eigenvalues and diffusion. Only in Section 5 do we start sketching the arguments. Nearly all the background material is available in the book references. (We note that the max–min arguments are based on the classic: Courant–Hilbert [1, Vol. I, Chap. VI].)

1. DEFINITIONS AND PRELIMINARIES

We let M be an n-dimensional, $n \geq 1$, connected, C^∞, Riemannian manifold. Should M have a boundary ∂M, we shall assume that M is oriented, and that (unless otherwise noted) ∂M is also C^∞.

For each point $p \in M$, the tangent space to M at p will be denoted by M_p; and the tangent bundle, that is, the union of all the tangent spaces of M endowed with its natural differentiable structure, will be denoted by TM.

We denote the real numbers by $\mathbb{R}$. For any differentiable mapping of an interval of $\mathbb{R}$ into any manifold, differentiation with respect to the independent variable in $\mathbb{R}$ is denoted by a prime. Mappings from intervals of $\mathbb{R}$ into manifolds are usually referred to as paths.

Given p in our manifold M and a C^1 real-valued function f defined on a neighborhood of p, then to each $\xi \in M_p$ is associated the *directional derivative of f at p in the direction ξ*, denoted by ξf, and defined by

$$\xi f = (f \circ \omega)'(0), \tag{1}$$

where $\omega(t)$ is any path in M satisfying $\omega(0) = p$ and $\omega'(0) = \xi$. The map $M_p \to \mathbb{R}$ given by $\xi \mapsto \xi f$ is linear. For functions f, h one has

$$\xi(f + h) = \xi f + \xi h, \tag{2}$$

$$\xi(fh) = h(\xi f) + f(\xi h). \tag{3}$$

The Riemannian metric on M associates to each $p \in M$ an inner product on M_p, which we denote by $\langle\ ,\ \rangle$. The associated norm will be denoted by $|\ |$. The Riemannian metric is C^∞ in the sense that if X, Y are C^∞ vector fields on M, then $\langle X, Y\rangle$ is a C^∞ real-valued function on M.

DEFINITION 1. Given a real-valued C^k, $k \geq 1$, function f on M, we define the *gradient of* f, $\operatorname{grad} f$, to be the vector field on M for which

$$\langle \operatorname{grad} f, \xi\rangle = \xi f \tag{4}$$

for all $\xi \in TM$.

There is no question of the existence of $\operatorname{grad} f$, since, as noted, the map $\xi \mapsto \xi f$ is linear on each tangent space. The calculation below (cf. (22)) shows that $\operatorname{grad} f$ is a C^{k-1} vector field. One has for functions f, h,

$$\operatorname{grad}(f + h) = \operatorname{grad} f + \operatorname{grad} h, \tag{5}$$

$$\operatorname{grad}(fh) = h(\operatorname{grad} f) + f(\operatorname{grad} h). \tag{6}$$

Whereas the differentiation of functions on a manifold is naturally determined by the differentiable structure, the differentiation of vector fields, on the other hand, is not naturally determined but involves the choice of a *connection*, that is, a rule which associates to each $p \in M$, $\xi \in M_p$, and C^1 vector field X defined on a neighborhood of p, a vector $\nabla_\xi X \in M_p$ satisfying

$$\nabla_\xi(X + Y) = \nabla_\xi X + \nabla_\xi Y, \tag{7}$$

$$\nabla_\xi(fX) = (\xi f)X(p) + f(p)\nabla_\xi X, \tag{8}$$

where X, Y are C^1 vector fields, and f is a C^1 real-valued function, all defined on a neighborhood of p. It is required that $\nabla_\xi X$ be linear in $\xi \in M_p$, and that for $X, Y \in C^\infty$ we have that $\nabla_X Y$ is a C^∞ vector field. The vector $\nabla_\xi X$ is traditionally referred to as the *covariant derivative of X with respect to ξ.*

The Riemannian metric on M does determine a unique connection, called the *Levi–Civita connection,* when one adds the requirements

$$\nabla_X Y - \nabla_Y X = [X, Y], \tag{9}$$

where [,] is the Lie bracket of the indicated vector fields, and

$$\xi\langle X,Y\rangle = \langle \nabla_\xi X, Y\rangle + \langle X, \nabla_\xi Y\rangle, \tag{10}$$

for all C^1 vector fields X, Y on M, and $\xi \in TM$.

DEFINITION 2. Given a C^k, $k \geq 1$, vector field X on M, define the real-valued function the *divergence of* X, div X, by

$$(\operatorname{div} X)(p) = \operatorname{trace}(\xi \mapsto \nabla_\xi X), \tag{11}$$

where ξ ranges over M_p.

The divergence of X is a C^{k-1} function on M; and for the function f, and vector fields X, Y on M, we have

$$\operatorname{div}(X + Y) = \operatorname{div} X + \operatorname{div} Y, \tag{12}$$

$$\operatorname{div}(fX) = f(\operatorname{div} X) + \langle \operatorname{grad} f, X\rangle. \tag{13}$$

DEFINITION 3. For any C^k, $k \geq 2$, function f on M we define the function *the Laplacian of* f, Δf, by

$$\Delta f = \operatorname{div}(\operatorname{grad} f). \tag{14}$$

One has that $\Delta f \in C^{k-2}$; and for functions f, h we have

$$\Delta(f + h) = \Delta f + \Delta h, \tag{15}$$

$$\operatorname{div}(h(\operatorname{grad} f)) = h(\Delta f) + \langle \operatorname{grad} h, \operatorname{grad} f\rangle, \tag{16}$$

$$\Delta(fh) = h(\Delta f) + 2\langle \operatorname{grad} f, \operatorname{grad} h\rangle + f(\Delta h). \tag{17}$$

We now calculate the expressions of the above operators in local coordinates. Let U be an open set in M, and $x: U \to \mathbb{R}^n$ a diffeomorphism of U into $\mathbb{R}^n$, that is, a chart on M. Then associated to the chart are n coordinate vector fields, written as $\partial/\partial x^j$ or as ∂_j, $j = 1, \ldots, n$. The directional derivative determined by ∂_j satisfies

$$(\partial_j(p))f = (\partial(f \circ x^{-1})/\partial x^j)(x(p)), \tag{18}$$

where p is any point in U and f is any differentiable function defined on a neighborhood of p. For each $p \in U$, the vectors $\{\partial_1(p), \ldots, \partial_n(p)\}$ span M_p. Therefore, for

$$\xi = \sum_{j=1}^{n} \xi^j \partial_j, \tag{19}$$

and $f \in C^1$, we have

$$\xi f = \sum_j \xi^j \partial_j f. \tag{20}$$

For the given Riemannian metric, define

$$\begin{aligned} g_{jk} &= \langle \partial_j, \partial_k \rangle, & G &= (g_{jk}), \\ g &= \det G, & G^{-1} &= (g_{jk}), \end{aligned} \tag{21}$$

where $j, k = 1, \ldots, n$, and (henceforth) det denotes the determinant. Then

$$\sum_j \xi^j \partial_j f = \sum_{j,k,l} \xi^j g_{jk} g^{kl}\, \partial_l f = \left\langle \xi, \sum_{k,l} (g^{kl}\, \partial_l f) \partial_k \right\rangle$$

for all ξ. From (4) and (20) we have

$$\operatorname{grad} f = \sum_{k,l} (g^{kl}\, \partial_l f) \partial_k. \tag{22}$$

To calculate the divergence, one certainly has n^3 functions Γ_{ij}^k, $i, j, k = 1, \ldots, n$, known as *Christoffel symbols*, determined by

$$\nabla_{\partial_j} \partial_i = \sum \Gamma_{ij}^k \partial_k \tag{23}$$

on U. Thus for ξ given by (19) and X given by

$$X = \sum_j \eta^j \partial_j, \tag{24}$$

we have, using (7) and (8),

$$\nabla_\xi X = \sum_{j,k} \xi^j \left\{ \partial_j \eta^k + \sum_l \eta^l \Gamma_{lj}^k \right\} \partial_k. \tag{25}$$

Therefore

$$\operatorname{div} X = \sum_j \left\{ \partial_j \eta^j + \sum_l \eta^l \Gamma_{lj}^j \right\}; \tag{26}$$

so it remains to give an explicit calculation of Γ_{ij}^k.

Recall that if X is given by (24) and Y is given by

$$Y = \sum_j \zeta^j \partial_j, \tag{27}$$

then

$$[X, Y] = \sum_{j,k} \{ \eta^j \partial_j \zeta^k - \zeta^j \partial_j \eta^k \}\, \partial_k. \tag{28}$$

One immediately has, by setting $X = \partial_j$ and $Y = \partial_k$, that (9) is equivalent to

$$\Gamma^l_{jk} = \Gamma^l_{kj} \tag{29}$$

for all $j, k, l = 1, \ldots, n$. If one also sets $\xi = \partial_i$ then (10) becomes

$$\partial_i g_{jk} = \sum_l \{\Gamma^l_{ji} g_{lk} + g_{jl}\Gamma^l_{ki}\}, \tag{30}$$

from which one deduces, with (29), by a standard argument,

$$\Gamma^k_{ij} = \tfrac{1}{2}\sum_l g^{kl}\{\partial_i g_{lj} + \partial_j g_{il} - \partial_l g_{ij}\}. \tag{31}$$

In particular, we have, by (26) and (31),

$$\begin{aligned}\operatorname{div} X &= \sum_j \left\{\partial_j \eta^j + \eta^j \sum_{k,l} \tfrac{1}{2} g^{kl}\partial_j g_{lk}\right\}\\ &= \sum_j \{\partial_j \eta^j + \tfrac{1}{2}\eta^j \operatorname{tr}(G^{-1}\partial_j G)\}\\ &= \sum_j \{\partial_j \eta^j + \tfrac{1}{2}\eta^j \partial_j(\ln g)\}\\ &= (1/\sqrt{g})\sum_j \partial_j(\eta^j \sqrt{g}),\end{aligned}$$

that is,

$$\operatorname{dix} X = (1/\sqrt{g})\sum_j \partial_j(\eta^j\sqrt{g}). \tag{32}$$

In the above calculation we have let tr denote the trace, $\partial_j G$ the matrix obtained from G by differentiating each of the entries with respect to the jth coordinate. To go from the second line to the third, one uses the standard formula for differentiating determinants.

Finally, (22) and (32) combine to imply

$$\Delta f = (1/\sqrt{g})\sum_{j,k} \partial_j(g^{jk}\sqrt{g}\,\partial_k f). \tag{33}$$

2. GREEN'S FORMULAS

M is our given Riemannian manifold. With the Riemann metric is associated an integration theory in which (i) the function f is measurable if, for every chart $x\colon U \to \mathbb{R}^n$ on M, $f \circ x^{-1}$ is measurable on the image of U in $\mathbb{R}^n$, (ii) for every covering $\{x_\alpha\colon U_\alpha \to \mathbb{R}^n\colon \alpha \in I\}$, where I is some set, of M by charts with subordinate partition of unity $\{\phi_\alpha : \alpha \in I\}$, the *Riemannian measure on M* is given by the density

$$dV = \sum_{\alpha} \phi_\alpha \sqrt{g_\alpha}\, dx_\alpha^1 \cdots dx_\alpha^n, \tag{34}$$

where $dx_\alpha^1 \cdots dx_\alpha^n$ is the density of Lebesgue measure on $x_\alpha(U_\alpha) \subseteq \mathbb{R}^n$, and g_α is the determinant defined in (21) for the chart $x_\alpha \colon U_\alpha \to \mathbb{R}^n$. The point is that the density $\sqrt{g}\, dx^1 \cdots dx^n$ on the domain U is independent of the mapping function x. The partition of unity is then the device with which the measure is defined globally on M.

Formula (32) admits the following interpretation: Given the vector field X on M, let $\{\Phi_t\}$ denote the induced flow on M. Fix any compact set K in M, and set

$$v(t) = \int_{\Phi_t(K)} dV.$$

Then standard calculation shows that

$$v'(0) = \int_K (\operatorname{div} X)\, dV.$$

Thus $\operatorname{div} X$ measures the infinitesimal distortion of volume by the flow generated by X.

DIVERGENCE THEOREM (I). If X is a C^1 vector field on M with compact support, then

$$\int_M (\operatorname{div} X)\, dV = 0. \tag{35}$$

GREEN'S FORMULAS (I). Let $h \in C^1$, $f \in C^2$ be functions on M such that $h(\operatorname{grad} f)$ has compact support. Then

$$\int_M \{h\Delta f + \langle \operatorname{grad} h, \operatorname{grad} f\rangle\}\, dV = 0. \tag{36}$$

If we also assume that $h \in C^2$ and both f, h have compact support, then

$$\int_M \{h\Delta f - f\Delta h\}\, dV = 0. \tag{37}$$

To derive Green's formula from the divergence theorem, one simply lets $X = h(\operatorname{grad} f)$ and substitutes (16) into (35). Equation (37) follows easily from (36).

Now assume that M has boundary ∂M, with induced Riemannian metric and measure, the density of the measure being denoted by dA. Let ν denote the outward unit normal vector field on ∂M.

DIVERGENCE THEOREM (II). Let X be a vector field which is C^1 on $\overline{M}$ and with compact support on $\overline{M}$. Then

$$\int_M (\operatorname{div} X)\, dV = \int_{\partial M} \langle X, \nu\rangle\, dA. \tag{38}$$

GREEN'S FORMULAS (II). Let $h \in C^1(\overline{M})$, $f \in C^2(\overline{M})$ such that $h(\operatorname{grad} f)$ has compact support on $\overline{M}$. Then

$$\int_M \{h\Delta f + \langle \operatorname{grad} h, \operatorname{grad} f\rangle\}\, dV = \int_{\partial M} h(\nu f)\, dA. \tag{39}$$

If we also have $h \in C^2(\overline{M})$ and both f, h have compact support on $\overline{M}$, then

$$\int_M \{h\Delta f - f\Delta h\}\, dV = \int_{\partial M} \{h(\nu f) - f(\nu h)\}\, dA. \tag{40}$$

3. BASIC FACTS FOR EIGENVALUE PROBLEMS

We let $L^2(M)$ be the space of measurable functions f on M for which

$$\int_M |f|^2\, dV < +\infty.$$

On $L^2(M)$ we have the usual inner product, and induced norm, given by

$$(f, h) = \int_M fh\, dV, \qquad \|f\|^2 = (f, f) \tag{41}$$

for $f, h \in L^2(M)$. With the inner product, $L^2(M)$ is a Hilbert space.

Our fundamental interest is in the following eigenvalue problems.

Closed eigenvalue problem: Let M be compact, connected. Find all real numbers λ for which there exists a nontrivial solution $\phi \in C^2(M)$ to

$$\Delta\phi + \lambda\phi = 0. \tag{42}$$

Neumann eigenvalue problem: For $\partial M \neq \emptyset$, $\overline{M}$ compact and connected, find all real numbers λ for which there exists a nontrivial solution $\phi \in C^2(M) \cap C^1(\overline{M})$ to (42), satisfying the boundary condition

$$\nu\phi = 0 \tag{43}$$

on ∂M (recall: ν is the outward unit normal vector field on ∂M).

Dirichlet eigenvalue problem: For $\partial M \neq \emptyset$, $\overline{M}$ compact and connected, find all real numbers λ for which there exists a nontrivial solution $\phi \in C^2(M) \cap C^0(\overline{M})$ to (42), satisfying the boundary condition

$$\phi = 0 \tag{44}$$

on ∂M.

Mixed eigenvalue problem: For $\partial M \neq \emptyset$, $\overline{M}$ compact and connected, N an open submanifold of ∂M, find all real numbers λ for which there exists a nontrivial solution $\phi \in C^2(M) \cap C^1(M \cup N) \cap C^0(\overline{M})$ to (42), satisfying the boundary conditions

$$\phi = 0 \quad \text{on} \quad \partial M - N, \qquad \nu\phi = 0 \quad \text{on} \quad N. \tag{45}$$

The desired numbers λ are referred to as *eigenvalues* of Δ, and the vector space of solutions of (42) for a given eigenvalue λ [eq. (42) is linear in ϕ], its *eigenspace*. The elements of each eigenspace are called *eigenfunctions*.

THEOREM 1. For each one of the above eigenvalue problems, the set of eigenvalues consists of a sequence

$$0 \leq \lambda_1 < \lambda_2 < \cdots \uparrow +\infty,$$

and each associated eigenspace is finite dimensional. Eigenspaces belonging to distinct eigenvalues are orthogonal in $L^2(M)$, and $L^2(M)$ is the direct sum of all the eigenspaces. Furthermore, each eigenfunction is C^∞ on $\overline{M}$.

We first note that as soon as we know that the eigenfunction $\phi \in C^2(M) \cap C^1(\overline{M})$, then its eigenvalue λ must be nonnegative. Indeed one sets $f = h = \phi$ and applies the appropriate Green formula [viz., (36) or (39)] to obtain

$$\lambda = \|\phi\|^{-2} \int_M |\text{grad}\,\phi|^2\, dV \geq 0. \tag{46}$$

From (46) one has that $\lambda = 0$ implies ϕ is a constant function. Therefore in the closed and Neumann eigenvalue problems we have $\lambda_1 = 0$, and in the Dirichlet and mixed ($N \neq \partial M$) problems we have $\lambda_1 > 0$.

We also note that the orthogonality of distinct eigenspaces is a direct consequence of the Green formulas (37) and (40). Indeed, let ϕ, ψ be eigenfunctions of the respective eigenvalues λ, τ. Then

$$0 = \int_M \{\phi\Delta\psi - \psi\Delta\phi\}\, dV = (\lambda - \tau)\int_M \phi\psi\, dV$$

and the remark follows.

We refer to the dimension of each eigenspace as the *multiplicity of the eigenvalue*. It will also be convenient to (henceforth) list the eigenvalues as

$$0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \uparrow +\infty$$

with each eigenvalue repeated according to its multiplicity.

If $\phi_1, \phi_2, \ldots$ is an orthonormal sequence (in $L^2(M)$) of eigenfunctions so that ϕ_j is an eigenfunction of λ_j for each $j = 1, 2, \ldots$, then $\phi_1, \phi_2, \ldots$ is a complete orthonormal sequence of $L^2(M)$. In particular, for $f \in L^2(M)$ we have

$$f = \sum_{j=1}^{\infty} (f, \phi_j)\phi_j \tag{47}$$

in $L^2(M)$, and

$$\|f\|^2 = \sum_{j=1}^{\infty} (f, \phi_j)^2. \tag{48}$$

These last two formulas are referred to as *Parseval identities*.

Weyl's asymptotic formula (Weyl [1]): In each of the above eigenvalue problems, let $N(\lambda)$ be the number of eigenvalues, counted with multiplicity, $\leq\lambda$. Then

$$N(\lambda) \sim \omega_n(\operatorname{vol} M)\lambda^{n/2}/(2\pi)^n \tag{49}$$

as $\lambda \to +\infty$, where ω_n is the volume of the unit disk in $\mathbb{R}^n$, and $\operatorname{vol} M$ is (henceforth) the volume of M. In particular,

$$(\lambda_k)^{n/2} \sim \{(2\pi)^n/\omega_n\}\, k/\operatorname{vol} M \tag{50}$$

as $k \to +\infty$.

For now, we just consider Weyl's formula when $n = 1$. $\mathbb{R}$ will have its usual Riemannian structure, and, for the boundary value eigenvalue problems, we will have M equal to some compact interval, say, $[0, L]$. For any function f on $M = [0, L]$, $\Delta f = f''$, and Eq. (42) is

$$\phi'' + \lambda\phi = 0. \tag{51}$$

If we consider the Dirichlet eigenvalue problem on $[0, L]$ then the associated boundary condition is

$$\phi(0) = \phi(L) = 0, \tag{52}$$

from which one has, for $k = 1, 2, \ldots,$

$$\phi_k(x) = \sqrt{2/L}\,\sin(\pi k x/L) \tag{53}$$

and

$$\lambda_k = (\pi k/L)^2. \tag{54}$$

Thus we obtain equality in the Weyl formula for all k.

For the Neumann eigenvalue problem the boundary conditions are

$$\phi'(0) = \phi'(L) = 0, \tag{55}$$

from which one has

$$\lambda_k = (\pi(k - 1)/L)^2, \tag{56}$$

and the Weyl formula is again satisfied.

In the case of the mixed eigenvalue problem one has boundary conditions

$$\phi'(0) = \phi(L) = 0 \tag{57}$$

or

$$\phi(0) = \phi'(L) = 0, \tag{58}$$

and

$$\lambda_k = (\pi(k - \tfrac{1}{2})/L)^2. \tag{59}$$

Weyl's formula follows immediately.

Finally, if we set

$$\omega(x) = (L/2\pi)e^{i2\pi x/L}, \tag{60}$$

then $\omega(\mathbb{R})$ is a circle in $\mathbb{R}^2$ of length L, parametrized with respect to arc length along the circle, and the eigenfunctions of the circle consist of the L-periodic solutions of (51) (cf. Section II.2 for more details). Thus

$$\begin{aligned}\lambda_1 &= 0,\\ \lambda_{2k} &= \lambda_{2k+1} = (2\pi k/L)^2,\end{aligned} \tag{61}$$

for all $k = 1, 2, \ldots,$ and Weyl's formula follows.

4. THE WAVE AND HEAT EQUATIONS

The Weyl formula is our first example of the relation between analytic and geometric properties of the Riemannian manifold. At the same time, the eigenvalues and their eigenfunctions arise from mathematical idealizations of physical problems.

For example, if we think of M as a vibrating homogeneous membrane with fixed boundary, then for "small oscillations" the transverse vibration would be given by a function $v\colon M \times (0, \infty) \to \mathbb{R}$ satisfying the *wave equation*

$$\Delta v = (\rho/\tau)\partial^2 v/\partial t^2, \tag{62}$$

where ρ is the density and τ the tension of the membrane, with boundary condition

$$v \,|\, \partial M \times (0, \infty) = 0. \tag{63}$$

The search for solutions to (62) begins with functions of the form

$$v(x, t) = \chi(x)T(t). \tag{64}$$

Should a solution to (62) of the form (64) exist, then one is immediately led to the existence of a constant λ for which

$$T'' + (\lambda\tau/\rho)T = 0, \tag{65}$$

$$\Delta\chi'' + \lambda\chi = 0, \qquad \chi \,|\, \partial M = 0. \tag{66}$$

So λ is a Dirichlet eigenvalue of M with eigenfunction χ. From (46) we have $\lambda > 0$—of course, our physical intuition demands that the motion be oscillatory with respect to time, and the only way to achieve it in (65) is to have $\lambda > 0$.

The general solution to (65) will be

$$T(t) = A \cos\sqrt{\lambda\tau/\rho}(t - \beta), \tag{67}$$

where A, β are arbitrary constants. The number $\sqrt{\lambda\tau/\rho}/2\pi$ is the "pitch" of the vibration and is that which our ears detect. The inverse proportionality of the pitch and length of the vibrating string with fixed endpoints, deduced immediately from (54), was known to the Pythagoreans.

Since (62) is a linear equation, any finite linear combination of solutions of the form (64) is again a solution. We therefore consider the formal solution

$$v(x, t) = \sum_{k=1}^{\infty} A_k \phi_k(x) \cos\sqrt{\lambda_k \tau/\rho}(t - \beta_k),$$

where $\lambda_1 \le \lambda_2 \le \cdots$ are the Dirichlet eigenvalues of M, $\phi_1, \phi_2, \ldots$ a complete orthonormal sequence in $L^2(M)$ with ϕ_k an eigenfunction of λ_k for each k, and A_k, β_k are arbitrary constants.

Assume that at time $t = 0$ we have

$$v(x, 0) = f(x), \qquad (\partial v/\partial t)(x, 0) = 0, \tag{68}$$

where f is some given function on M, that is, we are thinking of the membrane as deformed to a given position and then released (without being pushed) at time $t = 0$. The second condition of (68) implies v can be written as

$$v(x, t) = \sum_{k=1}^{\infty} A_k \phi_k(x) \cos\sqrt{\lambda_k \tau/\rho}\, t,$$

which implies

$$f(x) = \sum_{k=1}^{\infty} A_k \phi_k(x),$$

and by (47) we have

$$A_k = (f, \phi_k),$$

that is,

$$v(x, t) = \int_M w(x, y, t) f(y)\, dV(y), \tag{69}$$

where

$$w(x, y, t) = \sum_{k=1}^{\infty} \phi_k(x) \phi_k(y) \cos\sqrt{\lambda_k \tau/\rho}\, t. \tag{70}$$

Similarly, if we think of heat diffusing through the homogeneous medium M with insulated boundary, then the temperature function $u: M \times [0, \infty) \to \mathbb{R}$ satisfies, after suitable normalization of the physical constants, the *heat equation*

$$\Delta u = \partial u/\partial t \tag{71}$$

with boundary condition

$$\nu_x u = 0 \tag{72}$$

on all of $\partial M \times (0, \infty)$. Again we seek solutions of the form

$$u(x, t) = \chi(x) T(t),$$

and are led to the equations

$$T' + \lambda T = 0, \qquad \Delta\chi + \lambda\chi = 0,$$

with boundary condition

$$\nu\chi = 0 \qquad \text{on } \partial M.$$

So λ is a Neumann eigenvalue of M with eigenfunction χ. $T(t)$ is now of the form

$$T(t) = Ae^{-\lambda t}.$$

If we wish to find the temperature function under the assumption

$$u(x, 0) = f(x),$$

where f is a given function on M, then the above arguments lead to

$$u(x, t) = \int_M p(x, y, t) f(y)\, dV(y), \tag{73}$$

where

$$p(x, y, t) = \sum_{j=1}^{\infty} e^{-\lambda_j t} \phi_j(x) \phi_j(y). \tag{74}$$

In (74) the λ_j's are Neumann eigenvalues with eigenfunctions ϕ_j.

We emphasize that the above are formal calculations with no pretense to their validity. Also the choice of boundary conditions was predicated on appeal to experience. Any of our four eigenvalue problems can be derived with the exact same manipulations from a corresponding initial-boundary (possibly empty) problem for the wave and heat equations.

5. RAYLEIGH AND MAX–MIN METHODS

For continuous vector fields X, Y on M, we define the inner product

$$(X, Y) = \int_M \langle X, Y \rangle\, dV,$$

with norm

$$\|X\|^2 = \int_M |X|^2\, dV, \tag{75}$$

and complete the resulting metric space to an L^2-space, denoted by $\mathscr{L}^2(M)$. As usual, $\mathscr{L}^2(M)$ may be identified with those measurable vector fields X (i.e., vector fields whose coefficient functions, in any chart, are measurable) for which the integral in (75) is finite. The inner product and norm extend to $\mathscr{L}^2(M)$—and $\mathscr{L}^2(M)$ is a Hilbert space.

If we are given a C^1 function f on M, and a C^1 vector field X on M with compact support, then (13) and (35) immediately imply

$$(\operatorname{grad} f, X) = -(f, \operatorname{div} X).$$

We now consider this formula for a broader range of functions.

DEFINITION 4. Given a function $f \in L^2(M)$ we say that $Y \in \mathscr{L}^2(M)$ *is a weak derivative of* f if

$$(Y, X) = -(f, \operatorname{div} X)$$

for all C^1 vector fields X with compact support on M.

It is known that there is at most one such $Y \in \mathscr{L}^2(M)$, and we may therefore write

$$Y = \operatorname{Grad} f.$$

We shall let $\mathscr{H}(M)$ denote the subspace of $L^2(M)$ (referred to as the *Sobolev space*) consisting of those functions in $L^2(M)$ possessing weak derivatives; on $\mathscr{H}(M)$ we shall define the inner product

$$(f, h)_1 = (f, h) + (\operatorname{Grad} f, \operatorname{Grad} h)$$

with associated norm

$$\|f\|_1^2 = \|f\|^2 + \|\operatorname{Grad} f\|^2. \tag{76}$$

It is also known that $\mathscr{H}(M)$ is the completion of

$$\{f \in C^\infty(M) : \|f\|_1 < +\infty\}$$

in the metric induced by (76). Moreover, since ∂M (when nonempty) is C^∞, we also have that $C^\infty(\bar{M})$ is dense in $\mathscr{H}(M)$ in the given metric.

On $\mathscr{H}(M)$ we consider the symmetric bilinear form, known as the *Dirichlet* or *energy integral*, and given by

$$D[f, h] = (\operatorname{Grad} f, \operatorname{Grad} h) \tag{77}$$

for $f, h \in \mathscr{H}(M)$.

In what follows we shall be concerned with the validity of the formula

$$(\Delta\phi, f) = -D[\phi, f], \tag{78}$$

where ϕ is usually an eigenfunction in one of our eigenvalue problems, and f is in some subspace of $\mathscr{H}(M)$.

In the *closed eigenvalue problem* we have M compact (in particular $M = \bar{M}$) and (78) is certainly valid by (36) when $\phi \in C^2(M)$ and $f \in C^\infty(M)$.

Therefore, for a fixed $\phi \in C^2(M)$, formula (78) defines a linear functional F_ϕ on $C^\infty(M)$ as a subspace of $\mathscr{H}(M)$, satisfying

$$|F_\phi(f)| \leq \|\operatorname{grad} \phi\| \, \|\operatorname{grad} f\| \leq \|\operatorname{grad} \phi\| \, \|f\|_1.$$

So F_ϕ is a bounded linear functional on $C^\infty(M) \subseteq \mathscr{H}(M)$ with norm $\leq \|\operatorname{grad} \phi\|$, and can be extended to a bounded linear functional on all of $\mathscr{H}(M)$. Thus (78) will be valid for $\phi \in C^2(M)$, $f \in \mathscr{H}(M)$.

For the *Neumann eigenvalue problem* we start with the validity of (78) for $\phi \in C^2(\overline{M})$, satisfying $\nu\phi = 0$ on ∂M, and $f \in C^\infty(\overline{M})$—the appropriate Green formula is (39). Again one extends the validity of (78) to allow $f \in \mathscr{H}(M)$.

For the *Dirichlet eigenvalue problem* we proceed as follows: (78) is valid for $\phi \in C^2(\overline{M})$, satisfying $\phi = 0$ on ∂M, and $f \in C^\infty(M)$ with compact support—again by (39). The validity of (78) will now be extended to let f be in the completion of C^∞ functions, with compact support in M, in $\mathscr{H}(M)$.

For the *mixed eigenvalue problem* we obtain the validity of (78) when $\phi \in C^2(\overline{M})$ with $\phi = 0$ on $\partial M - N$, $\nu\phi = 0$ on N, and f is in the completion of functions in $C^\infty(\overline{M})$ compactly supported in $M \cup N$.

We summarize the discussion:

DEFINITION 5. Given each of the above eigenvalue problems we define the *space of admissible functions* $\mathfrak{H}(M)$ to be $\mathscr{H}(M)$ in the case of closed, and Neumann eigenvalue problems, the completion of C^∞ functions compactly supported on M in the case of the Dirichlet problem, and the completion of the C^∞ functions compactly supported on $M \cup N$ in the mixed problem.

Before proceeding, we comment that, in what follows, not all our manifolds, with nonempty boundary, will have C^∞ boundary. In some situations we will require the discussion for when M is only piecewise C^∞, namely, (i) when M is a nodal domain on a surface (cf. below), (ii) when M is an n-dimensional rectangle in $\mathbb{R}^n$, $n \geq 2$ (cf. Section II.3), and (iii) when M is a two-dimensional geodesic triangle (cf. Sections X.3 and 4). In all these cases, the divergence theorem and resulting Green's formulas are valid (with appropriate formulation of the hypotheses) (see Whitney [1, p. 100]). All the basic facts in Section 3 about solutions to the eigenvalue problems remain unchanged except, of course, the differentiability of the solutions at the singularity of the boundary. We also note, in this case, that the functions, which are C^∞ on M and the smooth part of ∂M, are dense in $\mathscr{H}(M)$. To distinguish between the cases when ∂M is C^∞, and when ∂M is possibly only piecewise C^∞, we refer to connected M with compact closure and nonempty

C^∞ boundary as a *regular domain*, and to connected M with compact closure and nonempty piecewise C^∞ boundary as a *normal domain*.

RAYLEIGH'S THEOREM. We are given a normal domain with fixed eigenvalue problem having the function space $\mathfrak{H}(M)$, and eigenvalues

$$\lambda_1 \le \lambda_2 \le \cdots, \tag{79}$$

where each eigenvalue is repeated the number of times equal to its multiplicity. Then for any $f \in \mathfrak{H}(M)$, $f \neq 0$, we have

$$\lambda_1 \le D[f,f]/\|f\|^2 \tag{80}$$

with equality if and only if f is an eigenfunction of λ_1. If $\{\phi_1, \phi_2, \ldots\}$ is a complete orthonormal basis of $L^2(M)$ such that ϕ_j is an eigenfunction of λ_j for each $j = 1, 2, \ldots$, then for $f \in \mathfrak{H}(M)$, $f \neq 0$, satisfying

$$(f, \phi_1) = \cdots = (f, \phi_{k-1}) = 0, \tag{81}$$

we have the inequality

$$\lambda_k \le D[f, f]/\|f\|^2 \tag{82}$$

with equality if and only if f is an eigenfunction of λ_k.

PROOF: The argument is based on our earlier considerations, namely, if ϕ is an eigenfunction, and $f \in \mathfrak{H}(M)$, then (78) is valid.

For any given $f \in \mathfrak{H}(M)$ set

$$\alpha_j = (f, \phi_j).$$

For $k > 1$, (81) is equivalent to saying $\alpha_1 = \cdots = \alpha_{k-1} = 0$. So for all $k = 1, 2, \ldots$, and $r = k, k+1, \ldots$ we have

$$\begin{aligned} 0 &\le D\left[f - \sum_{j=k}^{r} \alpha_j \phi_j, f - \sum_{j=k}^{r} \alpha_j \phi_j\right] \\ &= D[f, f] - 2 \sum_{j=k}^{r} \alpha_j D[f, \phi_j] + \sum_{j,l=k}^{r} \alpha_j \alpha_l D[\phi_j, \phi_l] \\ &= D[f, f] + 2 \sum_{j=k}^{r} \alpha_j (f, \Delta\phi_j) - \sum_{j,l=k}^{r} \alpha_j \alpha_l (\phi_j, \Delta\phi_l) \\ &= D[f, f] - \sum_{j=k}^{r} \lambda_j \alpha_j^2. \end{aligned}$$

We conclude that

$$\sum_{j=k}^{\infty} \lambda_j \alpha_j^2 < +\infty$$

and

$$D[f, f] \geq \sum_{j=k}^{\infty} \lambda_j \alpha_j^2 \geq \lambda_k \sum_{j=k}^{\infty} \alpha_j^2 = \lambda_k \|f\|^2,$$

by the Parseval identities. The case of equality follows easily.

MAX–MIN THEOREM. Given $v_1, \ldots, v_{k-1} \in L^2(M)$, let

$$\mu = \inf D[f, f]/\|f\|^2$$

where f varies over the subspace (less the origin) of functions in $\mathfrak{H}(M)$ orthogonal to $v_1, \ldots, v_{k-1}$ in $L^2(M)$. Then, for eigenvalues given in (79), we have

$$\mu \leq \lambda_k.$$

Of course, if $v_1, \ldots, v_{k-1}$ are orthonormal, with each v_l an eigenfunction of λ_l, $l = 1, \ldots, k-1$, then $\mu = \lambda_k$.

PROOF: Consider the functions f of the form

$$f = \sum_{j=1}^{k} \alpha_j \phi_j,$$

where $\phi_1, \ldots, \phi_k$ are orthonormal, with each ϕ_j an eigenfunction of λ_j, $j = 1, \ldots, k$, and where f is orthogonal to $v_1, \ldots, v_{k-1}$ in $L^2(M)$, that is,

$$0 = \sum_{j=1}^{k} \alpha_j(\phi_j, v_l), \qquad l = 1, \ldots, k-1. \tag{83}$$

If we think of $\alpha_1, \ldots, \alpha_k$ as unknowns and (ϕ_j, v_l) as given coefficients, then system (83) has more unknowns than equations and a nontrivial solution of (83) must exist. But then

$$\mu\|f\|^2 \leq D[f, f] = \sum_{j=1}^{k} \lambda_j \alpha_j^2 \leq \lambda_k \|f\|^2$$

which implies the claim.

Domain monotonicity of eigenvalues (vanishing Dirichlet data): Let $\Omega_1, \ldots, \Omega_m$ be pairwise disjoint normal domains in M, whose boundaries, when intersecting ∂M, do so transversally. Given an eigenvalue problem on M, consider, for each $r = 1, \ldots, m$, the eigenvalue problem on Ω_r obtained by requiring vanishing Dirichlet data on $\partial\Omega_r \cap M$ and by leaving the original data on $\partial\Omega_r \cap \partial M$ unchanged. Arange *all* the eigenvalues of $\Omega_1, \ldots, \Omega_m$ in an increasing sequence

$$0 \leq \nu_1 \leq \nu_2 \leq \cdots$$

with each eigenvalue repeated according to its multiplicity, and let the eigenvalues of M be given as in (79). Then we have for all $k = 1, 2, \ldots,$

$$\lambda_k \leq \nu_k. \tag{84}$$

PROOF: We use the max–min method. For functions in L^2 pick $\phi_1, \ldots, \phi_{k-1}$. For $j = 1, \ldots, k$ let $\psi_j \colon \overline{M} \to \mathbb{R}$ be an eigenfunction of ν_j when restricted to the appropriate subdomain, and identically zero, otherwise. Then $\psi_j \in \mathfrak{H}(M)$, and $\psi_1, \ldots, \psi_k$ may be chosen orthonormal in $L^2(M)$. As before, there exist $\alpha_1, \ldots, \alpha_k$, not all equal to zero, satisfying

$$\sum_{j=1}^{k} \alpha_j(\psi_j, \phi_l) = 0, \qquad l = 1, \ldots, k-1.$$

Therefore the function

$$f = \sum_{j=1}^{k} \alpha_j \psi_j$$

is orthogonal to $\phi_1, \ldots, \phi_{k-1}$ in $L^2(M)$, which implies

$$\lambda_k \|f\|^2 \leq D[f, f] = \sum_{j=1}^{k} \nu_j \alpha_j^2 \leq \nu_k \|f\|^2$$

which is the claim.

COROLLARY 1. If $\Omega \subseteq M$, then for the Dirichlet eigenvalue problem on Ω, and any eigenvalue problem on M we have

$$\lambda_k(\Omega) \geq \lambda_k. \tag{85}$$

If $M - \overline{\Omega}$ is open in M, then the inequality is strict.

Domain monotonicity of eigenvalues (vanishing Neumann data): Let $\Omega_1, \ldots, \Omega_m$ be as above, and also assume

$$\overline{M} = \overline{\Omega_1} \cup \cdots \cup \overline{\Omega_m}.$$

For each $r = 1, \ldots, m$, add Neumann data to $\partial\Omega_r \cap M$ and leave original data on $\partial\Omega_r \cap \partial M$ unchanged. Again arrange *all* the eigenvalues of $\Omega_1, \ldots, \Omega_m$ in increasing order, with repetition according to multiplicity:

$$0 \leq \mu_1 \leq \mu_2 \leq \cdots.$$

Then for each $k = 1, 2, \ldots$ we have

$$\mu_k \leq \lambda_k. \tag{86}$$

Proof: Let $\Psi_l \colon \overline{M} \to \mathbb{R}$ be the eigenfunction of μ_l when Ψ_l is restricted to the appropriate subdomain, and let Ψ_l be identically zero otherwise.

Now if f is any function in $\mathfrak{H}(M)$, then $f \in \mathfrak{H}(\Omega_r)$ for every $r = 1, \ldots, m$. We can therefore argue that if f is orthogonal to $\Psi_1, \ldots, \Psi_{k-1}$ in $L^2(M)$ then

$$D[f, f] = \sum_{r=1}^{m} \int_{\Omega_r} \|\operatorname{Grad} f\|^2 \, dV \geq \sum_{r=1}^{m} \mu_k \iint_{\Omega_r} f^2 \, dV = \mu_k \|f\|^2.$$

But there exists a nontrivial

$$f = \sum_{j=1}^{k} \alpha_j \phi_j$$

orthogonal to $\Psi_1, \ldots, \Psi_{k-1}$ in $L^2(M)$. Then

$$D[f, f] \leq \lambda_k \|f\|^2,$$

which implies the claim.

Remark 1: We note the contrast in the hypotheses of the two domain monotonicity theorems for eigenvalues, the addition of vanishing Neumann data requiring a complete partition of M. To employ this result in a fixed geometric setting, with good choices of $\Omega_1, \ldots, \Omega_m$, one requires a priori some knowledge of the decomposability of M. Such an application is given in Theorem X.3.

DEFINITION 6. Let $f \colon M \to \mathbb{R} \in C^0$. Then the *nodal set* of f is the set $f^{-1}[0]$, and a *nodal domain* of f is a component on $\overline{M} \backslash f^{-1}[0]$.

COURANT'S NODAL DOMAIN THEOREM. Let (79) be our list of eigenvalues and $\{\phi_1, \phi_2, \ldots\}$ a complete orthonormal basis of $L^2(M)$ with each ϕ_j an eigenfunction of λ_j, $j = 1, 2, \ldots$. Then the number of nodal domains of ϕ_k is less than or equal to k, for every $k = 1, 2, \ldots$.

Proof: We will give proofs for two distinct cases: (i) All the nodal domains of ϕ_k are normal domains; and (ii) no assumption is made on the nodal domains, but then we only consider the closed and Dirichlet eigenvalue problems.

In case (i) we argue as follows.

Let $G_1, \ldots, G_k, G_{k+1}, \ldots$ be nodal domains of ϕ_k. For each $j = 1, \ldots, k$ define

$$\psi_j = \begin{cases} \phi_k \,|\, G_j & \text{on} \quad G_j, \\ 0 & \text{on} \quad \bar{M} - G_j. \end{cases}$$

One then obtains, as above, the existence of a nontrivial function

$$f = \sum_{j=1}^{k} \alpha_j \psi_j$$

satisfying

$$0 = (f, \phi_1) = \cdots = (f, \phi_{k-1}).$$

One verifies that $\psi_j \in \mathfrak{H}(M)$ for each $j = 1, \ldots, k$. Then Rayleigh's theorem, the max–min method, and the divergence theorem imply

$$\lambda_k \leq D[f, f]/\|f\|^2 \leq \lambda_k.$$

So f is therefore, an eigenfunction of λ_k vanishing identically on G_{k+1}. But then the maximum principle (cf. Section XII.11) implies that f vanishes identically on M—a contradiction.

Before proceeding to case (ii) we first note an immediate consequence of the nodal domain theorem.

COROLLARY 2. ϕ_1 always has constant sign; λ_1 has multiplicity equal to 1; and ϕ_2 has precisely 2 nodal domains. λ_1 is characterized as being the *only* eigenvalue with eigenfunction of constant sign.

In the special case (i), we also have, immediately,

COROLLARY 3. If a normal domain Ω in M is a nodal domain of an eigenfunction of some eigenvalue λ, then λ is the lowest eigenvalue for the eigenvalue problem of Ω with original boundary data on $\partial\Omega \cap \partial M$, and vanishing Dirichlet boundary data on $\partial\Omega \cap M$.

In the following, we will be using some elementary facts about the foliation of a noncompact manifold with compact closure M by the level surfaces of a function in $C^0(\bar{M}) \cap C^\infty(M)$ which vanishes on ∂M. We refer the reader to Section IV.1 for a summary.

DEFINITION 7. Let Ω be an arbitrary open set in a Riemannian manifold M. Define $\mathscr{H}_0(\Omega)$ to be the completion, in $\mathscr{H}(\Omega)$, of the collection of C^∞ functions on Ω which are compactly supported in Ω, with respect to the norm (76). Then the *fundamental tone of* Ω, $\lambda^*(\Omega)$, is defined by

$$\lambda^*(\Omega) = \inf D[f, f]/\|f\|^2$$

where f ranges over nonidentically vanishing functions in $\mathscr{H}_0(\Omega)$.

We note that one can show that if $u \in \mathscr{H}(\Omega) \cap C^0(\bar{\Omega})$, and $u \,|\, \partial\Omega = 0$, then $u \in \mathscr{H}_0(\Omega)$ (remark of J. L. Kazdan).

Also note that if Ω is a compact manifold (resp., a normal domain), then Rayleigh's theorem implies that $\lambda^*(\Omega)$ coincides with the lowest eigenvalue of the closed (resp., Dirichlet) eigenvalue problem.

Finally, if Ω is given by

$$\Omega = \bigcup_\alpha \Omega_\alpha,$$

where Ω_α is a domain in M, for each α, then

$$\lambda^*(\Omega) \leq \inf_\alpha \lambda^*(\Omega_\alpha).$$

We now fix M to be either (i) a compact Riemannian manifold, in which case we are considering the closed eigenvalue problem, or (ii) a regular domain, in which case we are considering the Dirichlet eigenvalue problem.

LEMMA 1. Let u be an eigenfunction with eigenvalue λ, and let Ω be a nodal domain of u. Then $u \in \mathscr{H}_0(\Omega)$, and

$$\lambda = \lambda^*(\Omega).$$

PROOF: Assume $u > 0$ on Ω, and for each $\varepsilon > 0$, set

$$\Omega_\varepsilon = \{x \in \Omega : u(x) > \varepsilon\},$$

$$u_\varepsilon = \begin{cases} u - \varepsilon & \text{on } \Omega_\varepsilon \\ 0 & \text{on } M \backslash \Omega_\varepsilon . \end{cases}$$

Then, by Sard's theorem (Narasimhan [1, p. 19 ff.]), there exists a sequence ε_j, of regular values of u, decreasing to 0 as $j \to +\infty$. Set

$$\Omega_j = \Omega_{\varepsilon_j}, \qquad u_j = u_{\varepsilon_j}.$$

Then $u_j \in \mathscr{H}_0(\Omega_j) \subseteq \mathscr{H}_0(\Omega)$, as mentioned earlier, and it is clear that $u_j \to u \,|\, \Omega$ in $\mathscr{H}(\Omega)$.

Since $\partial\Omega_j$ is C^∞ we also use the Green's formula to obtain

$$\begin{aligned}\lambda \iint_{\Omega_j} u_j u\, dV &= -\iint_{\Omega_j} u_j\, \Delta u_j\, dV \\ &= \iint_{\Omega_j} \operatorname{grad} u_j^2\, dV \\ &\geq \lambda^*(\Omega_j) \iint_{\Omega_j} u_j^2\, dV \\ &\geq \lambda^*(\Omega) \iint_{\Omega_j} u_j^2\, dV,\end{aligned}$$

which implies, by letting $j \to +\infty$,

$$\lambda \iint_{\Omega} u^2\, dV \geq \lambda^*(\Omega) \iint_{\Omega} u^2\, dV.$$

Therefore,

$$\lambda \geq \lambda^*(\Omega).$$

To show the opposite inequality, let $\varepsilon > 0$ be a regular value of u, and let $v_\varepsilon > 0$ be the eigenfunction of the Dirichlet eigenvalue $\lambda_1(\Omega_\varepsilon) = \lambda^*(\Omega_\varepsilon)$. Then

$$\begin{aligned}\lambda \iint_{\Omega_\varepsilon} v_\varepsilon u\, dV &= -\iint_{\Omega_\varepsilon} v_\varepsilon(\Delta u)\, dV \\ &= -\iint_{\Omega_\varepsilon} (\Delta v_\varepsilon) u\, dV + \int_{\partial\Omega_\varepsilon} u(\partial v_\varepsilon/\partial \nu)\, dA \\ &\leq -\iint_{\Omega_\varepsilon} (\Delta v_\varepsilon) u\, dV \\ &= \lambda^*(\Omega_\varepsilon) \iint_{\Omega_\varepsilon} v_\varepsilon u\, dV,\end{aligned}$$

which implies

$$\lambda \leq \lambda^*(\Omega_\varepsilon)$$

for all regular values $\varepsilon > 0$.

We now show

(87) $$\lim_{\varepsilon \downarrow 0} \lambda^*(\Omega_\varepsilon) = \lambda^*(\Omega),$$

which will conclude the proof of the lemma. Given any $\delta > 0$ there exists $f \in C^\infty(\Omega)$, compactly supported on Ω, such that

$$D[f, f]/\|f\|^2 \leq \lambda^*(\Omega) + \delta.$$

But there certainly exists $\varepsilon > 0$ for which

$$\operatorname{supp} f \subseteq \Omega_\varepsilon;$$

so

$$\lambda^*(\Omega_\varepsilon) \leq D[f, f]/\|f\|^2.$$

We therefore have, for given $\delta > 0$, the existence of $\varepsilon > 0$ for which

$$\lambda^*(\Omega) \leq \lambda^*(\Omega_\varepsilon) \leq \lambda^*(\Omega) + \delta.$$

Since $\lambda^*(\Omega_\varepsilon)$ is increasing with respect to ε, we obtain (87), which implies the lemma.

To prove the Courant nodal domain theorem for case (ii), one can now argue as above. However, instead of invoking the maximum principle, one must invoke the unique continuation principle (Aronsajn [1]).

Remark 2: For f an eigenfunction of the Laplacian, the regularity of the *nodal sets*, that is, $f^{-1}[0]$, has been studied in Cheng [4]. His results are that except on a closed set of lower dimension (i.e., $<n - 1$) the nodal set of f forms an $(n - 1)$-dimensional C^∞ manifold. The singular points are those at which, not only f vanishes, but the gradient of f vanishes (that is, the critical points of f). In the neighborhood of these points, the nodal set is C^1 diffeomorphic to the nodal set of a homogeneous harmonic polynomial on $\mathbb{R}^n$.

As a result of these statements, Cheng proved the general nodal domain theorem as though case (ii) never need arise. However, C. deVerdiere has noted that the regularity argument of Cheng has a gap when $n = \dim M > 2$, and that one can even give a counterexample to the argument (though not, necessarily, the result) as it stands (see Bérard–Meyer [1]). Hence the necessity of considering case (ii) in the above argument. The argument given for case (ii) is that of Bérard–Meyer [1].

When $n = 2$, that is, M is 2-dimensional, then Cheng's results are valid, and one can actually say more, namely (Cheng [4]) (i) the critical points on the nodal lines are isolated; (ii) when the nodal lines meet, they form an equiangular system—the number of lines being equal to the order of the vanishing of f; (iii) the nodal lines consist of a finite number of C^2 immersed one-dimensional closed submanifolds. Therefore, when M is compact, the nodal lines consist of a finite number of C^2 immersed circles; (iv) the geodesic curvatures of intersecting nodal lines vanish at the point of intersection. In the compact case, upper bounds on the order of the vanishing of the kth eigenfunction can be given in terms of k and the topology of the manifold, which, in turn, imply upper bounds on the multiplicity of the first nonzero eigenvalue. In particular, if M is homeomorphic to $\mathbb{S}^2$, then the nodal line of an eigenfunction, corresponding to the first nonzero eigenvalue, consists of one C^∞ simple closed curve, and the multiplicity of the eigenvalue is less than or equal to 3. The upper bound 3 is best possible as it is realized by the standard metric on $\mathbb{S}^2$ (cf. Proposition II.1). Besson [1] sharpened Cheng's original argument to give best possible upper bounds of the multiplicity of the lowest nonzero eigenvalue of the real projective plane, and two-dimensional tori. He also showed that the upper bounds do not characterize the standard metrics.

Remark 3: An old conjecture (Payne [1]) was that, given a convex domain Ω in $\mathbb{R}^n$, then the level sets, of the eigenfunction of the lowest Dirichlet eigenvalue of Ω, are convex. It was settled affirmatively in 1976 by H. J. Brascamp and E. Lieb [1], and reinvestigated, using more elementary arguments, in Cafarelli–Spruck [1]. We refer the reader to this last paper for other results and references on this and related questions.

We now consider the possibility that we have equality in Courant's nodal domain theorem for infinitely many k.

PLEIJEL'S THEOREM (Pleijel [1]). Consider M with either the closed or Dirichlet eigenvalue problem. Assume that for any domain Ω in M we have the isoperimetric inequality

$$\{\lambda^*(\Omega)\}^{n/2} \operatorname{vol} \Omega > (2\pi)^n/\omega_n. \tag{88}$$

Then, letting n_k denote the number of nodal domains of $\lambda_k(M)$, we have

$$\limsup_{k\to\infty} n_k/k < 1. \tag{89}$$

Thus equality in Courant's theorem can be achieved for only a finite number of eigenvalues.

PROOF: If Ω is a nodal domain of $\lambda_k(M)$, then

$$\lambda^*(\Omega) = \lambda_k(M),$$

which implies there exists a constant $\alpha > (2\pi)^n/\omega_n$ such that

$$\{\lambda_k(M)\}^{n/2} \operatorname{vol} M \geq n_k \alpha.$$

Thus

$$\begin{aligned}\limsup_{k\to\infty} n_k/k &\leq (1/\alpha) \lim_{k\to\infty} (1/k)\{\lambda_k(M)\}^{n/2} \operatorname{vol} M \\ &= (2\pi)^n/\alpha\omega_n \\ &< 1\end{aligned}$$

by Weyl's asymptotic formula.

Remark 4: Of course, it remains to investigate the validity of (88), so we may in fact conclude the inequality (89). For the case of Dirichlet eigenvalues, the matter has been affirmatively settled in Bérard–Meyer [1]—cf. our discussion in Remark IV.4.

One may persist by arguing that even though equality in Courant's theorem might be achieved only a finite number of times [if one has the isoperimetric inequality (88)], it is still true that given the manifold M, one can guarantee the existence of an integer k such that for all $l > k$, the lth eigenfunction has more than 2 nodal domains. This is false. A metric on $\mathbb{S}^3$ exists which provides a counterexample (see Bérard Bergery–Bourguignon [1]).

CHAPTER II

The Basic Examples

In this chapter we present the basic examples which, to a large extent, motivate the subsequent theory. In general, we provide more detail when systematic treatment is yet to be available in book form. Two comments are in order here: (i) Our presentation of hyperbolic space is limited to providing what is necessary for the work in Section 5. More detail will be provided in Chapters X and XI. (ii) Some of the techniques of estimation have greater validity than their employment in Section 5, where we are only interested in the case of constant sectional curvature. We refer there to the literature to, obviously, give credit to the authors, and to inform the reader of the broader possibilities.

1. SOME GENERALITIES

If we are given Riemannian manifolds M, N, then the product manifold has a natural Riemannian metric, determined as follows: For any $(p, q) \in M \times N$, the tangent space $(M \times N)_{(p,q)}$ is canonically isomorphic to the direct sum $M_p \oplus N_q$. For vectors $\xi, \eta \in M_p$, $\zeta, \nu \in N_q$ we define the inner product of $\xi \oplus \zeta$ and $\eta \oplus \nu$ by

$$\langle \xi \oplus \zeta, \eta \oplus \nu \rangle_{(p,q)} = \langle \xi, \eta \rangle_p + \langle \zeta, \nu \rangle_q,$$

where the subscripts indicate the tangent space in which the different inner products are to be calculated. In particular, $M_p \oplus \{0\}$ is orthogonal to $\{0\} \oplus N_q$. The associated Riemannian measure on $M \times N$ is the product measure determined by dV_M and dV_N.

For C^2 functions $F: M \times N \to R$ of the form

$$F(p, q) = f(p)h(q), \tag{1}$$

where $f: M \to R$, $h: N \to R$ are both C^2 on their respective manifolds, we have (using obvious notation)

$$(\Delta_{M \times N} F)(p, q) = (\Delta_M f)(p)h(q) + f(p)(\Delta_N h)(q).$$

Thus if f is an eigenfunction of Δ_M with eigenvalue σ and h is an eigenfunction of Δ_N with eigenvalue τ, then F is an eigenfunction of $\Delta_{M \times N}$ with eigenvalue $\sigma + \tau$. One knows (Berger–Gauduchon–Mazet [1, p. 144]) that given eigenvalue problems on M, N, the collection of eigenfunctions of the induced eigenvalue problem on $M \times N$, which come from the algebra of C^2 functions on $\overline{M \times N}$ generated by those of the form (1), is the complete set of eigenfunctions of that induced eigenvalue problem on $M \times N$.

We recall some facts about isometries. Given manifolds M, N and a C^∞ map $\Phi\colon M \to N$, associate to Φ the maps $\Phi^*\colon C^0(N) \to C^0(M)$ and $\Phi_*\colon TM \to TM$ in the usual manner:

$$\Phi^* f = f \circ \Phi, \qquad (\Phi_* \xi) f = \xi(\Phi^* f).$$

To any Riemannian metric B on N is associated, via Φ, a symmetric, positive semidefinite, (0, 2)-tensor field, $\Phi^* B$, on M, by

$$(\Phi^* B)(\xi, \eta) = B(\Phi_* \xi, \Phi_* \eta)$$

for any $\xi, \eta \in M_p$, $p \in M$. If b is a given Riemannian metric on M, we say that Φ is a *local isometry of* b *onto* B if $b = \Phi^* B$. In such a case, Φ is an immersion. We call Φ an *isometry* if Φ is a diffeomorphism and a local isometry.

Let M, N have the Riemannian metrics b, B, respectively, and let $\Phi\colon M \to N$ be a local isometry of M onto N. Then for any C^1 function f on N, we have

$$\Phi_*(\operatorname{grad}_b \Phi^* f) = \operatorname{grad}_B f.$$

For any vector field Y on M, for which $\Phi_* Y$ is a globally well-defined vector field on N, we have

$$\operatorname{div}_b Y = \Phi^*(\operatorname{div}_B \Phi_* Y).$$

And for any C^2 function f on N we have

$$\Delta_b(\Phi^* f) = \Phi^*(\Delta_B f). \tag{2}$$

If, in addition, $M = N$, $b = B$, then Φ is an isometry of M onto itself, and Φ^* acts as an orthogonal transformation of $L^2(M)$. From (2) one has that Φ^* leaves every eigenspace invariant.

Recall that a map $\mathfrak{p}\colon \tilde{M} \to M$ of Riemannian manifolds $\tilde{M}, M$ is a Riemannian covering if $\mathfrak{p}$ is a differentiable covering and a local isometry. Given any covering $\mathfrak{q}\colon \tilde{M} \to M$ of manifolds $\tilde{M}, M$, the *deck transformation group* of the covering $\mathfrak{q}$ is the group of homeomorphisms $\phi\colon \tilde{M} \to \tilde{M}$ preserving $\mathfrak{q}$, that is,

$$\mathfrak{q} \circ \phi = \mathfrak{q}.$$

It is known that the deck transformation group acts properly discontinuously on $\tilde{M}$ (i.e., to each $p \in \tilde{M}$ there exists a neighborhood U of p such that the collection of open sets $\{\phi(U)\}$ are pairwise disjoint)—in fact, the deck transformation group determines the covering. When $\mathfrak{p}$ is a Riemannian covering, the elements of the deck transformation group are isometries.

When $\mathfrak{p}: \tilde{M} \to M$ is a Riemannian covering, the functions on M may be naturally considered as functions on $\tilde{M}$ invariant under the action of the deck transformation group of $\mathfrak{p}$. Therefore, for M compact, the eigenvalues of M are precisely those eigenvalues of $\tilde{M}$ in whose eigenspace there are nontrivial eigenfunctions of $\tilde{M}$ invariant under the action of the deck transformation group of $\mathfrak{p}$. The invariant subspace of the $\tilde{M}$-eigenspace in question is precisely the M-eigenspace of the given eigenvalue. Note that if $\tilde{M}$ is not compact, then the invariant eigenfunction of $\tilde{M}$ is not in $L^2(\tilde{M})$.

2. TORI

We start with $\mathbb{R}^n$, $n \geq 1$, considered as a group under vector addition, and a *lattice* Γ, that is, a discrete subgroup of $\mathbb{R}^n$. The lattice acts on $\mathbb{R}^n$ by

$$\gamma(x) = \gamma + x$$

for $\gamma \in \Gamma$, $x \in \mathbb{R}^n$; the action is properly discontinuous, and determines the Riemannian covering $\mathfrak{p}: \mathbb{R}^n \to \mathbb{R}^n/\Gamma$. We shall assume that the rank of Γ is n, that is, there exists n linearly independent vectors $\{v_1, \ldots, v_n\}$ in Γ for which

$$\Gamma = \left\{ \sum_{j=1}^{n} \alpha^j v_j : \alpha^j \in \mathbb{Z}, j = 1, \ldots, n \right\},$$

where $\mathbb{Z}$ denotes the integers. In this case $\mathbb{R}^n/\Gamma$ is compact and is diffeomorphic to the torus $(\mathbb{S}^1)^n$, the Cartesian product of the circle $\mathbb{S}^1$ with itself n times. For convenience we denote $\mathbb{R}^n/\Gamma$ by T.

We shall find it convenient to consider the functions on T to be complex-valued; so $L^2(T)$ will, now, be a Hilbert space with Hermitian inner product

$$(f, h) = \int_T f\bar{h}\, dV.$$

The Laplacian will act on complex-valued functions by acting on their real and imaginary parts separately, namely, for real-valued functions u, v on T, we have

$$\Delta(u + iv) = \Delta u + i\Delta v.$$

One checks that the same eigenvalues are obtained, as when only admitting real-valued functions, with the same multiplicity.

To obtain a collection of eigenfunctions on T we proceed as follows: Associate to the lattice Γ, the *dual lattice*, Γ^*, given by

$$\Gamma^* = \{y \in \mathbb{R}^n : \langle x, y\rangle \in \mathbb{Z} \text{ for all } x \in \Gamma\}.$$

Then Γ^* is indeed a lattice of rank n, and to the above basis $\{v_1, \ldots, v_n\}$ of Γ is associated the dual basis $\{w_1, \ldots, w_n\}$ of Γ^*, determined by

$$\langle w_j, v_k\rangle = \delta_j^k,$$

where δ_{jk} is the Kronecker delta. Naturally,

$$(\Gamma^*)^* = \Gamma.$$

Now to each $y \in \Gamma^*$ associate the complex-valued function ϕ_y defined on $\mathbb{R}^n$, and given by

$$\phi_y(x) = e^{2\pi i\langle x, y\rangle}.$$

One easily sees that ϕ_y is invariant under the action of Γ, and

$$\Delta\phi_y(x) = -4\pi^2|y|^2\phi_y.$$

Thus ϕ_y determines an eigenfunction on T with eigenvalue

(3) $$\lambda = 4\pi^2|y|^2.$$

The functions ϕ_y determined in this manner are known to span $L^2(T)$ (Berger–Gauduchon–Mazet [1, pp. 146–148]).

We remark that if given $y_1, \ldots, y_k \in \Gamma^*$, then the functions

$$\{\phi_{y_j} : j = 1, \ldots, k\}$$

are linearly independent. Indeed, if $k = 1$ then all is well. Assume now that for given $l > 1$ our remark is true for any choice of $l - 1$ distinct elements of Γ^*. Should the l distinct elements $y_1, \ldots, y_l \in \Gamma^*$ satisfy

$$\sum_{j=1}^{l} \beta_j\phi_{y_j} = 0$$

for a given choice of complex numbers $\beta_1, \ldots, \beta_l$, then, since

$$\phi_{y_r}\phi_{y_s} = \phi_{y_r + y_s}$$

for all $y_r, y_s \in \Gamma^*$, we would have

$$0 = \sum_{j=1}^{l} \beta_j\phi_{y_j - y_l} = \beta_l + \sum_{j=1}^{l-1} \beta_j\phi_{y_j - y_l}.$$

Now calculate the Laplacian of both sides; then

$$0 = \sum_{j=1}^{l-1} \beta_j |y_j - y_l|^2 \phi_{y_j - y_l},$$

from which one easily concludes: $\beta_1 = \cdots = \beta_l = 0$.

Thus for a given $\lambda > 0$, the eigenspace of λ has dimension equal to the number of solutions $y \in \Gamma^*$ of (3), and the summatory function $N(\lambda)$ is equal to the number of elements of Γ^* inside the closed disk in $\mathbb{R}^n$, about the origin, of radius $\sqrt{\lambda}/2\pi$.

THEOREM (Weyl's asymptotic formula). For the torus T we have

$$N(\lambda) \sim \omega_n \lambda^{n/2} (\text{vol}\, T)/(2\pi)^n \tag{4}$$

as $\lambda \to +\infty$.

PROOF: Let $\mathbb{B}^n(r)$ denote the open disk in $\mathbb{R}^n$, centered at the origin and having radius r, and let $\mathcal{N}^*(r)$ be the number of points of Γ^* in $\overline{\mathbb{B}^n(r)}$. Then (3) implies that (4) is equivalent to

$$\mathcal{N}^*(r) \sim \omega_n r^n (\text{vol}\, T) \tag{5}$$

as $r \to +\infty$.

Let $\{v_1, \ldots, v_n\}$ be the basis of Γ referred to at the beginning of our discussion. To each

$$v = \sum_{j=1}^{n} \alpha^j v_j$$

in Γ associate the parallelopiped $P(v)$ defined by

$$P(v) = \left\{ x = \sum_{j=1}^{n} \xi^j v_j : \alpha^j < \xi^j < \alpha^j + 1, j = 1, \ldots, n \right\}.$$

Then $P(v)$ is a *fundamental domain* of T, that is, no two distinct points of $P(v)$ represent the same point of T, and the image of the closed parallelopiped $\overline{P(v)}$, under the covering, is all of T. Note that

$$\text{vol}\, T = \text{vol}\, P(v)$$

for every $v \in \Gamma$. We simply refer to $P(v)$ as a *copy* of T.

Also note that if $A: \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation for which $\Gamma = A(\mathbb{Z}^n)$, then $\Gamma^* = (A^*)^{-1}(\mathbb{Z}^n)$, where A^* is the adjoint of A. Therefore, if T^* is the torus determined by Γ^* then we have

$$\text{vol}\, T = |\det A| = (\text{vol}\, T^*)^{-1};$$

so (5) is equivalent to

$$\mathcal{N}^*(r) \sim \omega_n r^n/(\text{vol}\, T^*) \tag{6}$$

as $r \to +\infty$.

For each $w \in \Gamma^*$, let $P^*(w)$ be the copy of T^* determined by w, and d its diameter. Note that d is independent of $w \in \Gamma^*$. If we let $\mathscr{P}^*(r)$ denote the number of copies of T^* contained in $\overline{\mathbb{B}^n(r)}$, then one easily sees that

$$\mathscr{P}^*(r) \le \mathcal{N}^*(r) \le \mathscr{P}^*(r + d).$$

Now the open polyhedron $C^*(r)$, determined by all copies of T^* contained in $\overline{\mathbb{B}^n(r)}$, has volume equal to $(\text{vol}\, T^*)\mathscr{P}^*(r)$. Thus

$$(\text{vol}\, T^*)\mathscr{P}^*(r) \le \omega_n r^n.$$

On the other hand, if

$$\beta(r) = \min\{|y| : y \in \Gamma^* \cap \partial C^*(r)\},$$

then

$$\beta(r) > r - d,$$

which implies $C^*(r) \supseteq \overline{\mathbb{B}^n(r - d)}$. Thus

$$(\text{vol}\, T^*)\mathscr{P}^*(r) \ge \omega_n(r - d)^n.$$

By putting the inequalities together, we have

$$\frac{\omega_n(r - d)^n}{\text{vol}\, T^*} \le \mathcal{N}^*(r) \le \frac{\omega_n(r + d)^n}{\text{vol}\, T^*},$$

and (6) follows.

3. WEYL'S FORMULA FOR BOUNDED DOMAINS IN $\mathbb{R}^n$

Recall from Section I.3 that if $M = (0, \alpha) \subseteq \mathbb{R}$ then the Dirichlet eigenvalues of M are given by [cf. (I.54)]

$$\lambda_k = k^2\pi^2/\alpha^2, \qquad k = 1, 2, \ldots.$$

So if Ω is the open n-dimensional rectangle $\Omega = (0, \alpha_1) \times \cdots \times (0, \alpha_n) \subseteq \mathbb{R}^n$, then, for the Dirichlet eigenvalue problem on Ω, the eigenvalues are given by the collection

$$\{\pi^2(k_1^2/\alpha_1^2 + \cdots + k_n^2/\alpha_n^2)\}$$

where each k_j, $j = 1, \ldots, n$, ranges over the positive integers.

To calculate $N(\lambda)$ for Ω, let $\{e_1, \ldots, e_n\}$ be the standard basis of $\mathbb{R}^n$, Γ the lattice generated by $\{\alpha_1 e_1, \ldots, \alpha_n e_n\}$. Thus Γ^* is generated by $\{e_1/\alpha_1, \ldots, e_n/\alpha_n\}$, and

$$N(\lambda) \sim \mathcal{N}^*(\sqrt{\lambda}/\pi)/2^n \sim \omega_n(\sqrt{\lambda}/\pi)^n(\operatorname{vol}\Omega)/2^n = \omega_n\lambda^{n/2}(\operatorname{vol}\Omega)/(2\pi)^n.$$

Note that, in the first line, when we divided by 2^n we ignored the intersection of Γ^* with each of the coordinate hyperplanes of $\mathbb{R}^n$, since each such intersection forms an $(n-1)$-dimensional lattice in $\mathbb{R}^n$ and therefore does not affect the limit of $N(\lambda)/\lambda^{n/2}$ as $\lambda \to +\infty$. So the Weyl formula is valid for n-dimensional rectangles with Dirichlet eigenvalue problem.

A similar argument yields Weyl's formula for n-dimensional rectangles with the Neumann eigenvalue problem.

An immediate consequence is that for any bounded domain Ω in $\mathbb{R}^n$, with any of our boundary value problems, the associated summatory function $N(\lambda)$ satisfies

$$\liminf N(\lambda)/\lambda^{n/2} \geq \omega_n(\operatorname{vol}\Omega)/(2\pi)^n \tag{7}$$

as $\lambda \to +\infty$. Indeed, let $G_1, \ldots, G_l$ be pairwise disjoint open n-rectangles, relatively compact in Ω, and let $N_j(\lambda)$ be the number of Dirichlet eigenvalues of G_j which are less than or equal to λ. Then the max–min arguments of Section I.5 (namely, the domain monotonicity of eigenvalues for Dirichlet data) imply

$$N(\lambda) \geq \sum_{j=1}^{l} N_j(\lambda),$$

which implies

$$\liminf N(\lambda)/\lambda^{n/2} \geq \sum_{j=1}^{l} \liminf N_j(\lambda)/\lambda^{n/2} = \{\omega_n/(2\pi)^n\} \sum_{j=1}^{l} \operatorname{vol} G_j$$

for all such choices of $G_1, \ldots, G_l$. Therefore (7) follows.

We only sketch the argument for

$$\limsup N(\lambda)/\lambda^{n/2} \leq \omega_n(\operatorname{vol}\Omega)/(2\pi)^n \tag{8}$$

for the Dirichlet eigenvalue problem. Let $G_1, \ldots, G_l$ be pairwise disjoint open n-rectangles with

$$\Omega \subseteq \operatorname{int}(\overline{G_1} \cup \cdots \cup \overline{G_l}),$$

where int( ) denotes the interior. Let $M(\lambda)$ be the summatory function of the Neumann eigenvalues of $\operatorname{int}(G_1 \cup \cdots \cup G_l)$; and for each $j = 1, \ldots, l$, let

$M_j(\lambda)$ be the summatory of the Neumann eigenvalues of G_j. Then the max–min arguments imply

$$N(\lambda) \le M(\lambda) \le \sum_{j=1}^{l} M_j(\lambda),$$

which implies

$$\limsup N(\lambda)/\lambda^{n/2} \le \{\omega_n/(2\pi)^n\} \sum_{j=1}^{l} \operatorname{vol} G_j$$

for all such choices of $G_1, \ldots, G_l$. Inequality (8) then follows, and, with it, the Weyl formula for Dirichlet eigenvalue problems on bounded domains in $\mathbb{R}^n$.

For the proof of (8) for the Neumann eigenvalue problem, compare Courant–Hilbert [1, Vol. I, pp. 432–434].

Remark 1: The eigenvalues of the equilateral triangle have been studied by Lee–Crandall [1], and Pinsky [3, 4]; for Euclidean and spherical domains associated to crystallographic groups (cf. Bérard [1] and Bérard–Besson [1]).

More generally, in another direction, it has been conjectured by G. Polya [1] that for Dirichlet eigenvalues of regular domains in $\mathbb{R}^n$, one not only has the Weyl asymptotic formula, one also has one-sided inequalities for all the Dirichlet eigenvalues, namely,

$$(\lambda_k)^{n/2} \ge \{(2\pi)^n/\omega_n\}\, k/\operatorname{vol} M$$

for all $k \ge 1$. The conjectured has yet to be proved, but progress has been made on the question. Compare Polya [1], Lieb [1], and Li–Yau [3].

4. SPHERES AND REAL PROJECTIVE SPACES

We write

$$\mathbb{S}^n(r) = \{x \in \mathbb{R}^{n+1} : |x| = r\}, \qquad \mathbb{S}^n = \mathbb{S}^n(1),$$

and introduce *spherical coordinates* P: $[0, \infty) \times \mathbb{S}^n \to \mathbb{R}^{n+1}$, *in* $\mathbb{R}^{n+1}$, by

$$x = P(r, \xi) = r\xi.$$

On $\mathbb{R}^{n+1} - \{0\}$ we have the inverse map

$$Q(x) = (|x|, x/|x|),$$

and if $u\colon U \to \mathbb{R}^n$ is any chart on $\mathbb{S}^n$ then a chart v is determined on the open cone in $\mathbb{R}^{n+1}$, $P((0, \infty) \times U)$, by the formula

$$v(x) = (|x|, u(x/|x|)) = (r, u(\xi)).$$

To calculate the Laplacian of $\mathbb{R}^{n+1}$ in spherical coordinates, we note that

$$\partial x/\partial r = \xi, \qquad \partial x/\partial u^j = r\, \partial\xi/\partial u^j,$$

from which one has

$$|\partial x/\partial r| = 1, \qquad \langle \partial x/\partial r, \partial x/\partial u^j \rangle = 0, \tag{9}$$

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

$$\langle \partial x/\partial u^j, \partial x/\partial u^k \rangle = r^2 \langle \partial\xi/\partial u^j, \partial\xi/\partial u^k \rangle, \tag{10}$$

for all $j, k = 1, \ldots, n$. Thus if G is the matrix of the Riemannian metric on $\mathbb{R}^{+1}$ associated to the chart v, and H the matrix associated to the chart u on $\mathbb{S}^n$, then

$$g_{rr} = 1, \qquad g_{rj} = 0, \tag{11}$$

$$g_{jk}(r\xi) = r^2 h_{jk}(\xi), \tag{12}$$

and

$$\sqrt{g}(r\xi) = r^n \sqrt{h}(\xi). \tag{13}$$

Note that (10), (12), and (13) also give the calculation of the Riemannian metric of $\mathbb{S}^n(r)$ in terms of the one on $\mathbb{S}^n$.

One immediately has for a function F,

$$\begin{aligned} \Delta_{\mathbb{R}^{n+1}} F &= r^{-n}\partial_r(r^n \partial_r F) + \Delta_{\mathbb{S}^n(r)}(F \,|\, \mathbb{S}^n(r)) \\ &= r^{-n}\partial_r(r^n \partial_r F) + r^{-2}\Delta_{\mathbb{S}^n}(F \,|\, \mathbb{S}^n(r)) \end{aligned}$$

where $\Delta_{\mathbb{R}^{n+1}}, \Delta_{\mathbb{S}^n(r)}, \Delta_{\mathbb{S}^n}$ are the Laplacians on the indicated Riemannian manifolds—by $\Delta_{\mathbb{S}^n}(F \,|\, \mathbb{S}^n(r))$ we mean that $F \,|\, \mathbb{S}^n(r)$ is to be considered as a function on $\mathbb{S}^n$, and the Laplacian is to be calculated with respect to the Riemannian metric of $\mathbb{S}^n$.

If F has the form

$$F(x) = R(r)\, G(\xi),$$

then

$$\Delta_{\mathbb{R}^{n+1}} F = r^{-n}(r^n R')' G + r^{-2} R\, \Delta_{\mathbb{S}^n} G.$$

In particular, if for some nonnegative integer k

$$F(x) = r^k G(\xi),$$

we then have

$$\Delta_{\mathbb{R}^{n+1}} F = r^{k-2}\{\Delta_{\mathbb{S}^n} G + k(k+n-1)G\}.$$

So F is also harmonic on $\mathbb{R}^{n+1}$ (i.e., its Laplacian vanishes identically) if and only if G is an eigenfunction on $\mathbb{S}^n$ with eigenvalue $k(k+n-1)$. When G is considered as a function on $\mathbb{S}^n(r)$, then it is an eigenfunction of $\Delta_{\mathbb{S}^n(r)}$ with eigenvalue $k(k+n-1)/r^2$.

It is known that all eigenfunctions of the sphere are obtained in this manner. More precisely, the space of homogeneous harmonic polynomials on $\mathbb{R}^{n+1}$ of degree k, when restricted to $\mathbb{S}^n$, constitute the eigenspace of the kth distinct eigenvalue

$$\bar{\lambda}_k = k(k+n-1), \tag{14}$$

where, *now*, $k = 0, 1, 2, \ldots$, that is, we have labeled the eigenvalues as starting from $\bar{\lambda}_0 = 0$. The multiplicity of $\bar{\lambda}_k$ is

$$\binom{n+k}{k} - \binom{n+k-1}{k-1}.$$

Compare Berger–Gauduchon–Mazet [1, p. 159 ff.] and Stein–Weiss [1, p. 137 ff.] for details.

We note for future reference.

PROPOSITION 1. An $L^2(\mathbb{S}^n)$-orthogonal basis of the eigenspace of

$$\bar{\lambda}_1\,(\mathbb{S}^n) = n \tag{15}$$

is given by the $\mathbb{R}^{n+1}$-coordinate functions

$$\{x^A \,|\, \mathbb{S}^n : A = 1, \ldots, n+1\}. \tag{16}$$

Remark 2: If M is a Riemannian manifold with isometry $\Phi\colon M \to M$, and f is a solution to

$$\Delta u + \lambda u = 0,$$

then, by (2), $\Phi^* f$ is also a solution, that is, Φ^* preserves the eigenspaces of Δ. The existence of a large isometry group of M will therefore tend to imply high multiplicities for the eigenvalues. It was thought that the standard spheres, with their highest degree of symmetry, would therefore exhibit the highest multiplicity for $\bar{\lambda}_1$. As mentioned in Remark I.2, this is indeed correct in the 2-dimensional case. The result, however, is false in 3-dimensions, namely, Urakawa [1] has exhibited Riemannian metrics on $\mathbb{S}^3$ for which the multiplicity of $\bar{\lambda}_1$ is 7—not 4, as one might expect. Compare also Bérard Bergery–Bourguignon [1].

We now consider the real projective spaces. The n-dimensional real projective space $\mathbb{P}^n$ is obtained by identifying antipodal points of $\mathbb{S}^n$, that is, $\mathbb{S}^n$ covers $\mathbb{P}^n$ with deck transformation group $\mathbb{Z}_2 = \{I_{\mathbb{R}^{n+1}} \mid \mathbb{S}^n, -I_{\mathbb{R}^{n+1}} \mid \mathbb{S}^n\}$, where $I_{\mathbb{R}^{n+1}}$ is the identity map of $\mathbb{R}^{n+1}$. The space of functions on $\mathbb{P}^n$ is therefore identified with those functions f on $\mathbb{S}^n$ for which

$$f(x) = f(-x)$$

for all $x \in \mathbb{S}^n$. Thus

$$\bar{\lambda}_k(\mathbb{P}^n) = \bar{\lambda}_{2k}(\mathbb{S}^n) = 2k(2k + n - 1),$$

with eigenspace consisting of homogeneous harmonic polynomials on $\mathbb{R}^{n+1}$, of degree $2k$, restricted to $\mathbb{S}^n$.

5. DISKS IN CONSTANT CURVATURE SPACE FORMS

We shall be more informal in the calculations that follow, namely, for a given Riemannian manifold, and chart $x: U \to \mathbb{R}^n$, it is traditional to write the Riemannian metric in the chart as

$$ds^2 = \sum_{j,k} g_{jk}(x)\, dx^j\, dx^k.$$

When changing coordinates one substitutes formally into the differential expressions—and all is well.

Thus for the usual metric in $\mathbb{R}^n$ we write

$$ds^2 = \sum_{j=1}^{n} (dx^j)^2 =: |dx|^2,$$

where x is the standard chart on $\mathbb{R}^n$. Upon introducing spherical coordinates about any $p \in \mathbb{R}^n$,

$$x = p + t\xi,$$

where $t \in [0, \infty)$, $\xi \in \mathbb{S}^{n-1}$, we write

$$\begin{aligned} dx &= (dt)\xi + t\, d\xi, \\ |dx|^2 &= (dt)^2 |\xi|^2 + 2t(dt)\langle \xi, d\xi \rangle + t^2 |d\xi|^2 \\ &= (dt)^2 + t^2 |d\xi|^2, \end{aligned}$$

wherein we let $|d\xi|^2$ denote the Riemannian metric on $\mathbb{S}^{n-1}$. So we write in spherical coordinates on $\mathbb{R}^n$,

$$ds^2 = (dt)^2 + t^2|d\xi|^2.$$

To calculate the Riemannian metric of $\mathbb{S}^n$ relative to *spherical coordinates in* $\mathbb{S}^n$, we proceed as follows: Fix any $p \in \mathbb{S}^n$. Identify in the obvious manner $(\mathbb{S}^n)_p$, the tangent space to $\mathbb{S}^n$ at p, with the orthogonal complement of $\mathbb{R}p$ in $\mathbb{R}^{n+1}$, whose intersection with $\mathbb{S}^n$ would be the "equator of p" that is, $\mathbb{S}^{n-1}$. For $x \in \mathbb{S}^n$ set

$$x = (\cos t)p + (\sin t)\xi,$$

where $t \in [0, \pi]$, $\xi \in \mathbb{S}^{n-1}$. Then

$$dx = \{-(\sin t)p + (\cos t)\xi\}\, dt + (\sin t)\, d\xi,$$

and

$$|dx|^2 = (dt)^2 + (\sin^2 t)|d\xi|^2.$$

More generally, if we were calculating the Riemannian metric of $\mathbb{S}^n(\rho)$ then the appropriate scaling would produce

$$ds^2 = (dt)^2 + \rho^2 \sin^2(t/\rho)|d\xi|^2,$$

where $t \in [0, \pi\rho]$, $\xi \in \mathbb{S}^{n-1}$.

Our third example is known as *hyperbolic space*, namely, on $\mathbb{B}^n(\rho) \subseteq \mathbb{R}^n$ we define the Riemannian metric

$$ds^2 = 4|dx|^2/\{1 - |x/\rho|^2\}^2. \tag{17}$$

If we now define spherical coordinates about $x = 0$ by

$$x = r\xi, \qquad r = \rho \tanh(t/2\rho),$$

where $r \in [0, \rho]$, $t \in [0, \infty)$, $\xi \in \mathbb{S}^{n-1}$, then the same type of calculation as above produces

$$ds^2 = (dt)^2 + \rho^2 \sinh^2(t/\rho)|d\xi|^2. \tag{18}$$

As is well known, the three examples describe the simply connected spaces with constant sectional curvature κ. For $\kappa = 0$ we have $M = \mathbb{R}^n$, for $\kappa > 0$ we have $M = \mathbb{S}^n(\rho)$ with $\kappa = 1/\rho^2$, and for $\kappa < 0$ we have $\mathbb{B}^n(\rho)$ endowed with the metric (17), and satisfying $\kappa = -1/\rho^2$. Our emphasis on the geometry of these spaces will come later—here our emphasis is on calculation.

Note that our spherical coordinates are valid about any point in M in the first two cases, but it is not obvious that, in the third case, such a coordinate system may be introduced about *any* point of the space. To prove that such coordinates may be defined about any point in the hyperbolic space, we consider a different model.

Write a point in $\mathbb{R}^{n+1}$ as (x, τ), where $x \in \mathbb{R}^n$, $\tau \in \mathbb{R}$, and endow $\mathbb{R}^{n+1}$ with the nondegenerate quadratic form

$$|(x, \tau)|_*^2 = |x|^2 - \tau^2, \tag{19}$$

where $|x|$ is the Euclidean norm in $\mathbb{R}^n$. If we restrict the induced pseudo-Riemannian metric on $\mathbb{R}^{n+1}$ to the n-dimensional hypersurface

$$|x|^2 - \tau^2 = -\rho^2, \qquad \tau > 0, \tag{20}$$

for some fixed $\rho > 0$, then

$$ds^2 = |dx|^2 - \langle x, dx\rangle^2/\{\rho^2 + |x|^2\}, \tag{21}$$

which is positive definite by the Cauchy–Schwarz inequality. Now introduce spherical coordinates

$$x = r\xi,$$

where $r \in [0, \infty)$, $\xi \in \mathbb{S}^{n-1}$, that is, the coordinates will be centered about $(x = 0, \tau = \rho)$ on the hypersurface; then we obtain from (21)

$$ds^2 = \frac{\rho^2(dr)^2}{\rho^2 + r^2} + r^2|d\xi|^2.$$

Upon setting

$$r = \rho \sinh t/\rho$$

and substituting, we obtain (18). The above calculations determine an isometry between the two Riemannian manifolds.

At first glance, little has been accomplished since, again, the coordinates are centered about a specific point. But here it is easy to show that the orthogonal group $O(n, 1)$ preserving the quadratic form (19), not only leaves the hypersurface invariant—and is therefore an isometry of the hypersurface onto itself, but also acts transitively on the hypersurface, that is, given any two points p_1, p_2 in the hypersurface, there exists an element in $O(n, 1)$ which determines an isometry of the hypersurface and maps p_1 onto p_2. Using the transitivity of the action of $O(n, 1)$, one obtains the existence of a coordinate system for which (18) is valid about any point in the hyperbolic space.

It will be convenient to summarize the above discussion as

THEOREM 1. For each fixed $\kappa \in \mathbb{R}$, let $\mathbb{M}_\kappa$ be the simply connected Riemannian manifold of constant sectional curvature κ as described above. About each point $p \in \mathbb{M}_\kappa$ there exists a coordinate system

$(t, \xi) \in [0, \pi/\sqrt{\kappa}) \times \mathbb{S}^{n-1}$ (where, when $\kappa \leq 0$, we are letting $\pi/\sqrt{\kappa}$ denote $+\infty$), relative to which the Riemannian metric reads as

$$ds^2 = (dt)^2 + \mathrm{S}_\kappa^2(t)|d\xi|^2, \tag{22}$$

where $\mathrm{S}_\kappa(t)$ is the solution to the differential equation

$$\psi'' + \kappa\psi = 0, \tag{23}$$

satisfying the initial conditions

$$\mathrm{S}_\kappa(0) = 0, \qquad \mathrm{S}_\kappa'(0) = 1. \tag{24}$$

Of course,

$$\mathrm{S}_\kappa(t) = \begin{cases} (1/\sqrt{\kappa}) \sin\sqrt{\kappa}t, & \kappa > 0, \\ t, & \kappa = 0, \\ (1/\sqrt{-\kappa}) \sinh\sqrt{-\kappa}t, & \kappa < 0. \end{cases} \tag{25}$$

In what follows later on, it will be convenient to let $\mathrm{C}_\kappa(t)$ denote the solution of (23) satisfying the initial conditions

$$\mathrm{C}_\kappa(0) = 1, \qquad \mathrm{C}_\kappa'(0) = 0, \tag{26}$$

that is,

$$\mathrm{C}_\kappa(t) = \begin{cases} \cos t, & \kappa > 0, \\ 1, & \kappa = 0, \\ \cosh\sqrt{-\kappa}t, & \kappa < 0. \end{cases} \tag{27}$$

Then

$$\mathrm{S}_\kappa' = \mathrm{C}_\kappa, \qquad \mathrm{C}_\kappa' = -\kappa\mathrm{S}_\kappa, \qquad \mathrm{C}_\kappa^2 + \kappa\mathrm{S}_\kappa^2 = 1,$$
$$(\mathrm{C}_\kappa/\mathrm{S}_\kappa)' = (\mathrm{S}_\kappa'/\mathrm{S}_\kappa)' = -\mathrm{S}_\kappa^{-2}.$$

What we have done thus far is construct geodesic spherical coordinates about any point $p \in \mathbb{M}_\kappa$ without using the general theory—we shall turn to it later. For the moment we comment on some metric properties of the geodesic spherical coordinates.

Recall that for any Riemannian manifold M, and path $\omega: [\alpha, \beta] \to M$, the *length of* ω is defined as

$$L(\omega) = \int_\alpha^\beta |\omega'|,$$

and that for any two points $p, q \in M$, the *distance* $d(p, q)$ between them is defined as

$$d(p, q) = \inf_\omega L(\omega),$$

where ω ranges over all continuous, piecewise C^1 paths $\omega\colon [\alpha, \beta] \to M$ for which $\omega(\alpha) = p$, $\omega(\beta) = q$. With this distance function, M is known to be a metric space whose topology coincides with that of the underlying manifold structure. One easily sees from (22) that if the coordinate system is centered at p, and $q = q(t, \xi)$ denotes any point in M, then

$$d(p, q) = t.$$

Thus

$$\mathsf{B}(p; \delta) \equiv \{q = q(t, \xi) : 0 \le t < \delta\}$$

is the open metric disk in $\mathbb{M}_\kappa$ of radius δ, centered at p, and its boundary, $\mathsf{S}(p; \delta)$, the corresponding sphere. When, as in our discussion below, p is fixed, we simply write $\mathsf{B}(\delta)$, $\mathsf{S}(\delta)$ for $\mathsf{B}(p; \delta)$, $\mathsf{S}(p; \delta)$, respectively.

It is standard that relative to the given metrics, $\mathbb{M}_\kappa$ is a complete metric space.

We now have $n \ge 2$, and Δ the Laplace operator on $\mathbb{M}_\kappa$. Let $\Box$ be the Laplace operator on $\mathbb{S}^{n-1}$. For any $F\colon \mathbb{M}_\kappa \to \mathbb{R} \in C^2$, with

$$F(q(t, \xi)) = f(t, \xi), \tag{28}$$

we have by direct calculation

$$(\Delta F)(q(t, \xi)) = \mathsf{S}_\kappa^{1-n}\partial_t(\mathsf{S}_\kappa^{n-1}\partial_t f) + \mathsf{S}_\kappa^{-2}\Box_\xi\, \mathrm{f}, \tag{29}$$

where, when writing $\Box_\xi\, \mathrm{f}$, we mean that $f \,|\, \mathsf{S}(t)$ is to be considered as a function on $\mathbb{S}^{n-1}$ with associated Laplacian $\Box$. If f has the form

$$f(t, \xi) = T(t)G(\xi), \tag{30}$$

then

$$\Delta F = \mathsf{S}_\kappa^{1-n}\{\mathsf{S}_\kappa^{n-1}T'\}'G + \mathsf{S}_\kappa^{-2}T\Box G, \tag{31}$$

where the prime $'$ is differentiation with respect to t. If, in addition, F satisfies

$$\Delta F + \lambda F = 0,$$

then explicit calculation shows that there exists a constant ν such that

$$\Box G + \nu G = 0,$$

$$(\mathsf{S}_\kappa^{n-1}T')' + \{\lambda - \nu\mathsf{S}_\kappa^{-2}\}\mathsf{S}_\kappa^{n-1}T = 0; \tag{32}$$

an equivalent form of (32) is

$$T'' + (n - 1)(\mathsf{C}_\kappa/\mathsf{S}_\kappa)T' + \{\lambda - (\nu/\mathsf{S}_\kappa^2)\}T = 0. \tag{33}$$

In particular, ν is an eigenvalue of $\mathbb{S}^{n-1}$ with eigenfunction G. From (14) we have (the dimension now is $n-1$) that the distinct eigenvalues of $\mathbb{S}^{n-1}$ are given by

$$\nu_l = l(l+n-2), \qquad l = 0, 1, 2, \ldots. \tag{34}$$

A standard argument implies that if F is to be differentiable at p, then

$$T'(0) = 0 \tag{35}$$

when $l = 0$, and

$$T(t) \sim (\mathrm{const})t^l \tag{36}$$

as $t \downarrow 0$, when $l = 1, 2, \ldots$.

We now want to consider the Dirichlet and Neumann eigenvalue problems on $\mathsf{B}(\delta)$. The vanishing of the respective boundary data on $\mathsf{S}(\delta)$ is now given by

$$T(\delta) = 0, \tag{37}$$

$$T'(\delta) = 0, \tag{38}$$

respectively. Fix one of the above eigenvalue problems. Then for each $l = 0, 1, \ldots$, the classical 1-dimensional arguments (compare Courant–Hilbert [1, Vol. I, Chap. 5] and Coddington–Levinson [1, Chaps. 7, 8]) yield that the collection of real λ, for which there is a nontrivial solution of (32), with $\nu = \nu_l$, satisfying the given boundary data, consists of a sequence

$$0 \le \lambda_{l,1} < \lambda_{l,2} < \cdots \uparrow +\infty,$$

and for each $j = 1, 2, \ldots, \lambda_{l,j}$ determines only a 1-dimensional space of solutions. One easily checks, using integration-by-parts, that if $j \neq k$ and $T_{l,j}$, $T_{l,k}$ are solutions of (32), with $\nu = \nu_l$, for $\lambda_{l,j}$, $\lambda_{l,k}$, respectively, then $T_{l,j}$ and $T_{l,k}$ are orthogonal in the L^2 space on $(0, \delta)$ whose measure has the density $\mathrm{S}_\kappa^{n-1}(t)\,dt$. In what follows we normalize $T_{l,j}$ to satisfy

$$\int_0^\delta T_{l,j}^2(t)\mathrm{S}_\kappa^{n-1}(t)\,dt = 1.$$

We now wish to show that the function-space, L, consisting of the span, in $L^2(\mathsf{B}(\delta))$, of all eigenfunctions of $\mathsf{B}(\delta)$ obtained by the above procedure, is dense in $L^2(\mathsf{B}(\delta))$. A moment's thought would convince the reader that it suffices to show that any eigenfunction of our eigenvalue problem is in the subspace L.

Let F be such an eigenfunction with eigenvalue λ. Represent F in the geodesic spherical coordinates by (28). Then for each fixed $t \in [0, \delta]$ we have the decomposition in $L^2(\mathbb{S}^{n-1})$,

$$f(t, \xi) = \sum_{l=0}^{\infty} a_l(t) G_l(\xi),$$

where G_l is an eigenfunction of ν_l on $\mathbb{S}^{n-1}$, with $L^2(\mathbb{S}^{n-1})$-norm equal to 1. Thus

$$\int_{\mathbb{S}^{n-1}} f^2(t, \xi)\, dA(\xi) = \sum_{l=0}^{\infty} a_l^2(t), \tag{39}$$

where we are letting dA denote Riemannian measure on $\mathbb{S}^{n-1}$, and

$$a_l(t) = \int_{\mathbb{S}^{n-1}} f(t, \xi) G_l(\xi)\, dA(\xi).$$

Therefore we have

$$\begin{aligned} -\nu_l a_l(t) &= \int_{\mathbb{S}^{n-1}} f(t, \xi) \square G_l(\xi)\, dA(\xi) \\ &= \int_{\mathbb{S}^{n-1}} (\square_\xi f)(t, \xi) G_l(\xi)\, dA(\xi) \\ &= -\lambda \mathsf{S}_\kappa^2(t) a_l(t) - \mathsf{S}_\kappa^{n-3}(t) \{\mathsf{S}_\kappa^{n-1}(t) a_l'(t)\}' \end{aligned}$$

for each $l = 0, 1, \ldots$ by (29), that is, $a_l(t)$, $\nu = \nu_l$, and λ provide a solution of the 1-dimensional eigenvalue problem given by (32).

Thus there exist $\alpha_l \in \mathbb{R}$, $j \in \{1, 2, \ldots\}$, such that

$$\lambda = \lambda_{l,j}, \qquad a_l(t) = \alpha_l T_{l,j}.$$

The $L^2(\mathsf{B}(\delta))$-sum

$$\sum_{l=0}^{\infty} \alpha_l T_{l,j}(t) G_l(\xi)$$

is the projection of F onto the subspace L; but by (39) we have

$$\|F\|^2 = \int_0^\delta \mathsf{S}_\kappa^{n-1}(t)\, dt \int_{\mathbb{S}^{n-1}} f^2(t, \xi)\, dA(\xi) = \sum_{l=0}^{\infty} \alpha_l^2,$$

that is, $F \in L$, which was our claim.

We now let $\lambda(\delta)$ denote the lowest Dirichlet eigenvalue of $\mathsf{B}(\delta)$, and $\mu(\delta)$ the lowest nonzero Neumann eigenvalue of $\mathsf{B}(\delta)$. Our interest is in the eigenfunctions of $\lambda(\delta)$, $\mu(\delta)$, the relative size of $\lambda(\delta)$, $\mu(\delta)$, and their behavior as $\delta \to \pi/\sqrt{\kappa}$.

Since the eigenfunction of $\lambda(\delta)$ cannot vanish on $\mathsf{B}(\delta)$ it must correspond to the lowest eigenvalue λ of (32) for $\nu = 0$—in particular, the eigenfunction of $\lambda(\delta)$ is radial.

To consider the eigenspace of $\mu(\delta)$, a little more is required.

PROPOSITION 2. Given

$$(\mathrm{S}_\kappa^{n-1}\phi')' + (\alpha - \sigma\mathrm{S}_\kappa^{-2})\mathrm{S}_\kappa^{n-1}\phi = 0,$$

$$(\mathrm{S}_\kappa^{n-1}\psi')' + (\beta - \tau\mathrm{S}_\kappa^{-2})\mathrm{S}_\kappa^{n-1}\psi = 0,$$

with ϕ, ψ bounded near $t = 0$. Then

$$\{\phi\psi' - \psi\phi'\}(t)\mathrm{S}_\kappa^{n-1}(t) = \int_0^t \{\alpha - \beta + (\tau - \sigma)\mathrm{S}_\kappa^{-2}\}\mathrm{S}_\kappa^{n-1}\phi\psi.$$

The proof is straightforward.

Now consider the Neumann eigenvalue problem on $\mathsf{B}(\delta)$. For any given $l = 1, 2, \ldots$ it is standard (Coddington–Levinson [1, Chap. 8]) that $T_{l,j}'$ has precisely $j - 1$ zeros in $(0, \delta)$; in particular, $T_{l,1}'$ never vanishes on $(0, \delta)$. One now uses this fact, in conjunction with Proposition 2, to conclude that $\lambda_{l,1} < \lambda_{k,1}$ whenever $l < k$. So determining $\mu(\delta)$ is reduced to choosing between $\lambda_{0,2}$ and $\lambda_{1,1}$ (remember: $\lambda_{0,1} = 0$).

PROPOSITION 3. Let $T(t)$ be any solution of

$$(\mathrm{S}_\kappa^{n-1}T')' + \lambda\mathrm{S}_\kappa^{n-1}T = 0. \tag{40}$$

Then for

$$\mathscr{T} = T'$$

we have

$$(\mathrm{S}_\kappa^{n-1}\mathscr{T}')' + \{\lambda - (n-1)\mathrm{S}_\kappa^{-2}\}\mathrm{S}_\kappa^{n-1}\mathscr{T} = 0. \tag{41}$$

We also have that $\mathscr{T}\,|(0, \beta] < 0$ whenever we are given that $T\,|(0, \beta) > 0$, $\lambda > 0$.

The derivation of (41) from (40) is straightforward, and the second claim of the proposition is a direct consequence of

$$(\mathrm{S}_\kappa^{n-1}T')(t) = -\lambda\int_0^t \mathrm{S}_\kappa^{n-1}T.$$

To determine whether $\mu(\delta)$ is $\lambda_{0,2}$ or $\lambda_{1,1}$, set $T_0 = T_{0,2}$, $\mathscr{T}_0 = T_0'$, and $\mathscr{T}_1 = T_{1,1}$ then T_0' and $\mathscr{T}_1'$ never vanish on $(0, \delta)$ (see Coddington–Levinson [1, Chap. 8]). But

$$\mathscr{T}_0(0) = \mathscr{T}_0(\delta) = 0$$

implies there exists $x_0 \in (0, \delta)$ for which $\mathscr{T}_0'(x_0) = 0$. From Proposition 2, one now easily shows that $\lambda_{1,1} < \lambda_{0,2}$. Thus $\mu(\delta) = \lambda_{1,1}$.

We summarize the discussion as

THEOREM 2. If $\mathsf{B}(\delta)$ is the n-disk of radius $\delta > 0$ in $\mathbb{M}_\kappa$, then the lowest Dirichlet eigenvalue, $\lambda(\delta) > 0$, of $\mathsf{B}(\delta)$ has eigenfunction F of the form

$$F(q(t, \xi)) = T(t),$$

where T is a solution of (40), with $\lambda = \lambda(\delta)$, satisfying

$$T'(0) = T(\delta) = 0, \qquad T \mid [0, \delta) \neq 0.$$

The lowest nonzero Neumann eigenvalue, $\mu(\delta) > 0$, of $\mathsf{B}(\delta)$ has multiplicity equal to n. Any eigenfunction $\mathscr{F}$ of $\mu(\delta)$ has the form

$$\mathscr{F}(q(t, \xi)) = \mathscr{T}(t) G(\xi),$$

where G is an eigenfunction of $\overline{\lambda_1}\ (\mathbb{S}^{n-1}) = n - 1$, and $\mathscr{T}$ is a solution of (41), with $\lambda = \mu(\delta)$, satisfying

$$\mathscr{T}(0) = \mathscr{T}'(\delta) = 0, \qquad \mathscr{T}' \mid [0, \delta) \neq 0.$$

Thus, by Proposition 1, an $L^2(\mathsf{B}(\delta))$-orthogonal basis of the eigenspace of $\mu(\delta)$ is given by setting

$$y = t\xi, \qquad \|y\| = t, \qquad \mathscr{F}_j(y) = y^j \mathscr{T}(\|y\|)/\|y\|,$$

where $j = 1, \ldots, n$.

THEOREM 3. If $\kappa \leq 0$ then

$$\mu(\delta) < \lambda(\delta) \tag{42}$$

for all $\delta > 0$. If $\kappa > 0$ then equality (42) is valid for all $\delta < \pi/2\sqrt{\kappa}$. For $\delta = \pi/2\sqrt{\kappa}$ we have

$$\mu(\pi/2\sqrt{\kappa}) = \lambda(\pi/2\sqrt{\kappa}) = n\kappa, \tag{43}$$

and for $\delta \in (\pi/2\sqrt{\kappa}, \pi/\sqrt{\kappa}]$ inequality (42) is to be reversed.

PROOF: We assume, for our convenience, that for $T, \mathscr{T}$ of Theorem 2 we have $T > 0$, $\mathscr{T}' > 0$ on all of $(0, \delta)$.

If we let $V = T'$ then V is a solution of (41), with $\lambda = \lambda(\delta)$, satisfying

$$V(0) = 0, \qquad V \mid (0, \delta] < 0,$$

by Proposition 3. Since (40) now reads as

$$V' + (n-1)(\mathbf{C}_\kappa/\mathbf{S}_\kappa)V + \lambda(\delta)T = 0, \tag{44}$$

we have, on one hand, $V' < 0$ on some neighborhood of $t = 0$. On the other hand, if $\mathbf{C}_\kappa(\delta) > 0$ then $V'(\delta) > 0$, from which we conclude that there exists $t_0 \in (0, \delta)$ for which $V'(t_0) = 0$. Inequality (42) would then follow from Proposition 2 (with $\tau = \sigma = (n-1)$, $\alpha = \lambda(\delta)$, $\phi = V$, $\beta = \mu(\delta)$, $\psi = \mathscr{T}$, and $t = t_0$).

So inequality (42) is a consequence of $\mathbf{C}_\kappa(\delta) > 0$. This is always true when $\kappa \le 0$, and when $\kappa > 0$, $\delta < \pi/2\sqrt{\kappa}$—thus (42) is valid in these cases.

If $\kappa > 0$, $\delta = \pi/2\sqrt{\kappa}$, then $T(t) = \cos\sqrt{\kappa}t$, $\mathscr{T}(t) = \sin\sqrt{\kappa}t$ are the desired respective eigenfunctions, and (43) is verified directly.

If $\kappa > 0$, $\delta > \pi/2\sqrt{\kappa}$, and $\mu(\delta) \le \lambda(\delta)$, then $V' = T''$ would have its first zero, $t = t_1$, in $(0, \delta]$ (use Proposition 2). From (44) we would conclude that $t_1 \in (0, \pi/2\sqrt{\kappa})$. But

$$\lambda(\delta) \le \lambda(\pi/2\sqrt{\kappa}) = n\kappa$$

implies that

$$\{(V\mathbf{C}_\kappa - V'\mathbf{S}_\kappa)\mathbf{S}^{n-1}\}(t) = \{\lambda(\delta) - n\kappa\}\int_0^t V\mathbf{S}_\kappa^n$$

is nonnegative on $(0, \delta)$. Therefore at $t = t_1$ we have $V(t_1) \ge 0$—a contradiction. Thus for $\kappa > 0$, $\delta > \pi/2\sqrt{\kappa}$, we have $\mu(\delta) > \lambda(\delta)$.

THEOREM 4. If $\kappa = 0$, that is, if $\mathbb{M}_\kappa = \mathbb{R}^n$, then there exist positive constants $\mathrm{c}_D, \mathrm{c}_N$ such that

$$\lambda(\delta) = \mathrm{c}_D^2/\delta^2, \qquad \mu(\delta) = \mathrm{c}_N^2/\delta^2$$

for all $\delta > 0$.

PROOF: For $\kappa = 0$ we have

$$\mathbf{S}_\kappa(t) = t, \qquad \mathbf{C}_\kappa(t) = 1,$$

so the differential equation under study is

$$y'' + \frac{(n-1)}{t}y' + \left\{\lambda - \frac{l(l+n-2)}{t^2}\right\}y = 0.$$

Let

$$\tau = \sqrt{\lambda}t, \qquad y(t) = z(\tau);$$

then our equation becomes

$$z'' + \frac{(n-1)}{\tau}z' + \left\{1 - \frac{l(l+n-2)}{\tau^2}\right\}z = 0. \tag{45}$$

For $l = 0$, let c_D be the first zero of $z(\tau)$ satisfying (45) with initial conditions: $z'(0) = 0$, $z(0) = 1$. Then one easily has

$$\lambda(\delta) = c_D^2/\delta^2.$$

Similarly, for $l = 1$, let c_N be the first zero of $z'(\tau)$, where $z(\tau)$ is the solution of (45) with initial conditions: $z(0) = 0$, $z'(0) = 1$. Then one easily sees that

$$\mu(\delta) = c_N^2/\delta^2,$$

which was the claim.

To obtain more precise information about $z(\tau)$ let

$$J(\tau) = \tau^{n/2-1}z(\tau).$$

Then (45) becomes

$$J'' + \frac{1}{\tau}J' + \left\{1 - \frac{(n+2l-2)^2}{4\tau^2}\right\}J = 0;$$

so J is a bounded Bessel function of order equal to $n/2 + l - 1$. Thus except for a multiplicative constant,

$$J = J_{n/2+l-1}.$$

In particular, c_D is the first zero of $J_{n/2-1}$, and c_N is the first zero of $J'_{n/2}$.

THEOREM 5. Let $\kappa < 0$, that is, $\mathbb{M}_\kappa$ is the hyperbolic space of constant negative sectional curvature κ. Then

$$\lambda(\delta) \geq -(n-1)^2\kappa/4 \tag{46}$$

for all $\delta > 0$, and

$$\lim \lambda(\delta) = -(n-1)^2\kappa/4 \tag{47}$$

as $\delta \to +\infty$ (McKean [1]). We also have

$$\lim \mu(\delta) = 0 \tag{48}$$

as $\delta \to +\infty$ (B. Randol, private communication).

PROOF: Here we have

$$\mathrm{S}_\kappa(t) = (\sinh\sqrt{-\kappa}t)/\sqrt{-\kappa}, \qquad \mathrm{C}_\kappa(t) = \cosh\sqrt{-\kappa}t.$$

On the one hand, the Cauchy–Schwarz inequality implies that for any C^1 function $\phi(t)$ on $[0, \delta]$ we have

$$-\int_0^\delta \phi\phi'\mathrm{S}_\kappa^{n-1} \leq \left\{\int_0^\delta \phi^2\mathrm{S}_\kappa^{n-1}\right\}^{1/2}\left\{\int_0^\delta \phi'^2\mathrm{S}_\kappa^{n-1}\right\}^{1/2}.$$

On the other hand, $\phi(\delta) = 0$ implies that

$$\begin{aligned}-\int_0^\delta \phi\phi'\mathrm{S}_\kappa^{n-1} &= \int_0^2 (\phi^2/2)(\mathrm{S}_\kappa^{n-1})', \\ &= (n-1)\int_0^\delta (\phi^2/2)\mathrm{C}_\kappa\mathrm{S}_\kappa^{n-2} \\ &\geq \sqrt{-\kappa}(n-1)/2\int_0^\delta \phi^2\mathrm{S}_\kappa^{n-1}.\end{aligned}$$

Therefore

$$-(n-1)^2(\kappa/4)\int_0^\delta \phi^2\mathrm{S}_\kappa^{n-1} \leq \int_0^\delta \phi'^2\mathrm{S}_\kappa^{n-1}. \tag{49}$$

It remains to note that (49) will imply (46). Indeed, given $f \in C_c^\infty(\mathsf{B}(\delta))$, then

$$\begin{aligned}-(n-1)^2(\kappa/4)\int_{\mathsf{B}(\delta)} f^2 &= -(n-1)^2(\kappa/4)\int_{\mathbb{S}^{n-1}} dA(\xi)\int_0^\delta f^2(q(t,\xi))\mathrm{S}_\kappa^{n-1}(t)\,dt \\ &\leq \int_{\mathbb{S}^{n-1}} dA(\xi)\int_0^\delta (\partial_t(f(q(t,\xi))))^2\mathrm{S}_\kappa^{n-1}(t)\,dt \\ &\leq \int_{\mathbb{S}^{n-1}} dA(\xi)\int_0^\delta |\operatorname{grad} f|^2(q(t,\xi))\mathrm{S}_\kappa^{n-1}(t)\,dt \\ &= \int_{\mathsf{B}(\delta)} |\operatorname{grad} f|^2,\end{aligned}$$

and (46) follows by Rayleigh's theorem (see Pinsky [1]).

To establish (47) we set, for our own convenience, $\kappa = -1$. Then Eq. (40) becomes

$$T'' + (n-1)(\coth t)T' + \lambda T = 0. \tag{50}$$

The idea of the proof is that for large values of t, $\coth t \sim 1$; so the behavior of T should be approximated, for large values of t, by $s(t)$ satisfying

$$s'' + (n-1)s' + \lambda s = 0,$$

that is,

$$s(t) = e^{\alpha t}$$

with

$$\alpha = -(n-1)/2 \pm i\{\lambda - (n-1)^2/4\}^{1/2}.$$

We therefore let f, defined on $\overline{\mathsf{B}(\delta)}$, be given by

$$f(q(t,\xi)) = \phi(t)$$

with

$$\phi(t) = \begin{cases} e^{-(n-1)t/2}\sin(2\pi(t-\delta/2)/\delta) & \text{if } \quad t \in [\delta/2, \delta], \\ 0 & \text{otherwise.} \end{cases}$$

Then f is an admissible function for the Dirichlet eigenvalue problem on $\mathsf{B}(\delta)$, and, on $(\delta/2, \delta)$, ϕ satisfies

$$\phi'' + (n-1)\phi' + [(n-1)^2/4 + 4\pi^2/\delta^2]\phi = 0.$$

Now

$$\begin{aligned} &\int_{\delta/2}^{\delta} \phi'^2 \sinh^{n-1} \\ &\quad = \int_{\delta/2}^{\delta} \{-(n-1)\phi\phi'(\coth - 1) + [(n-1)^2/4 + 4\pi^2/\delta^2]\phi^2\} \sinh^{n-1}, \end{aligned}$$

which implies

$$\begin{aligned} \int_{\delta/2}^{\delta} \{\phi'^2 - [(n-1)^2/4 + 4\pi^2/\delta^2]\phi^2\} \sinh^{n-1} &= -\int_{\delta/2}^{\delta} (n-1)\phi\phi'(\coth - 1)\sinh^{n-1} \\ &\le \int_{\delta/2}^{\delta} (n-1)|\phi||\phi'||\coth - 1| \sinh^{n-1}. \end{aligned}$$

By the previous argument, Cauchy's inequality, and

$$|\phi'| = |\operatorname{grad} f|,$$

one obtains

$$\begin{aligned} &\|\operatorname{grad} f\|^2 - [(n-1)^2/4 + 4\pi^2/\delta^2]\,\|f\|^2 \\ &\quad \le (n-1)(\coth \delta/2 - 1)\,\|f\|\,\|\operatorname{grad} f\|. \end{aligned}$$

If we set $x = \|\text{grad}\, f\|/\|f\|$, $A = [(n-1)^2/4 + 4\pi^2/\delta^2]$, $B = (n-1)(\coth \delta/2 - 1)$, then the above inequality reads as

$$x^2 - Bx \leq A,$$

from which one has

$$x \leq B/2 + \{A + (B^2/4)\}^{1/2}$$

Thus, by Rayleigh's theorem,

$$\begin{aligned}\sqrt{\lambda(\delta)} \leq{} & (n-1)(\coth \delta/2 - 1)/2 \\ & + \{(n-1)^2/4 + 4\pi^2/\delta^2 + (n-1)^2(\coth \delta/2 - 1)^2/4\}^{1/2},\end{aligned}$$

Now let $\delta \to +\infty$ and (47) follows (see Pinsky [2]). In the next chapter we give a sharper technique due to M. E. Gage [1].

We now prove (48). We still have $\kappa = -1$. We shall use our original model of hyperbolic space—thus the space is given as the interior of the unit disk $\mathbb{B}^n$ in $\mathbb{R}^n$ centered at the origin and having the Riemannian metric given by

$$ds^2 = 4|dx|^2/\{1 - |x|^2\}^2.$$

The disk $\mathsf{B}(\delta)$ is represented now by the disk $\mathbb{B}^n(\tau)$ in $\mathbb{R}^n$ with

$$\tau = \tanh \delta/2.$$

Our method is to estimate $\mu(\delta)$ with a test function.

Pick

$$\phi(x) = x^1.$$

Then

$$\int_{B(\delta)} \phi\, dV = \int_{\mathbb{B}^n(\tau)} \frac{2^n x^1\, dx^1 \cdots dx^n}{\{1 - |x|^2\}^n} = 0,$$

and

$$\begin{aligned}\int_{\mathsf{B}(\delta)} |\text{grad}\, \phi|^2\, dV &= \int_{\mathbb{B}^n(\tau)} \frac{2^{n-2}\, dx^1 \cdots dx^n}{\{1 - |x|^2\}^{n-2}} \\ &= \text{const} \int_0^\tau \frac{r^{n-1}}{\{1 - r^2\}^{n-2}}\, dr \\ &\leq \text{const} \begin{cases} 1, & n = 2, \\ |\ln(1-\tau)|, & n = 3, \\ (1-\tau)^{3-n}, & n > 3. \end{cases}\end{aligned}$$

On the other hand, for $\tau > \frac{1}{2}$, we have

$$\begin{aligned}\int_{\mathbb{B}^n(\tau)} \phi^2\, dV &= 2^n \int_{\mathbb{B}^n(\tau)} \frac{(x^1)^2}{\{1 - |x|^2\}^n}\, dx^1 \cdots dx^n \\ &\geq 2^n \int_{\{\mathbb{B}^n(\tau) - \mathbb{B}^n(1/2)\} \cap \{x^1 \geq |x|/\sqrt{2}\}} \frac{(x^1)^2}{\{1 - |x|^2\}^n}\, dx^1 \cdots dx^n \\ &\geq \text{const} \int_{1/2}^{r} \frac{r^{n+1}}{\{1 - r^2\}^n}\, dr \\ &\geq \text{const}(1 - \tau)^{1-n}.\end{aligned}$$

Thus, by Rayleigh's theorem, we have

$$\mu(\delta) \leq \text{const} \begin{cases} 1 - \tanh \delta/2, & n > 2, \\ (1 - \tanh \delta/2)^2\, |\ln(1 - \tanh \delta/2)|, & n = 3, \\ (1 - \tanh \delta/2)^2, & n > 3, \end{cases}$$

for large δ, and the theorem follows.

THEOREM 6. Let $\kappa > 0$, that is, $\mathbb{M}_\kappa = \mathbb{S}^n(1/\sqrt{\kappa})$. Then

(51) $$\lim_{\delta \to \pi/\sqrt{\kappa}} \lambda(\delta) = 0$$

and

(52) $$\mu(\delta) \geq (n - 1)\kappa$$

for all $\delta \in (0, \pi/\sqrt{\kappa})$.

PROOF: Here

$$\mathbf{S}_\kappa(t) = (\sin\sqrt{\kappa t})/\sqrt{\kappa}, \qquad \mathbf{C}_\kappa(t) = \cos\sqrt{\kappa t},$$

and, for our convenience, we shall set $\kappa = 1$. Then the Riemannian metric in our coordinate system is given by

$$ds^2 = dt^2 + \sin^2 t\, |d\xi|^2.$$

We prove (52) first. To this end, let the indices j, k vary from 1 to n, and the indices α, β from 1 to $n - 1$. By Theorem 2, a function $\mathscr{T}(t)$ is determined such that the eigenfunctions of $\mu(\delta)$ are given by

$$\mathscr{F}_j(q(t, \xi)) = \mathscr{T}(t)\xi^j,$$

$j = 1, \ldots, n$. Let u denote a chart on $\mathbb{S}^{n-1}$, with the Riemannian metric of $\mathbb{S}^{n-1}$ given, in the chart by

$$|d\xi|^2 = \sum_{\alpha, \beta} h_{\alpha\beta}(\xi)\, du^\alpha\, du^\beta.$$

Then for each $j = 1, \ldots, n$ we have, by direct calculation,

$$|\operatorname{grad} \mathscr{F}_j|^2 = \mathscr{T}'^2(\xi^j)^2 + (\sin t)^{-2}\mathscr{T}^2 \sum_{\alpha,\beta} \frac{\partial \xi^j}{\partial u^\alpha} h^{\alpha\beta} \frac{\partial \xi^j}{\partial u^\beta}.$$

Since $\mathscr{F}_j$ is an eigenfunction of $\mu(\delta)$ we have

$$\int_{B(\delta)} |\operatorname{grad} \mathscr{F}_j|^2 \, dV = \mu(\delta) \int_{B(\delta)} (\mathscr{F}_j)^2$$

for each $j = 1, \ldots, n$. Adding the equations we obtain

$$\begin{aligned}
\mu(\delta) \int_{B(\delta)} \sum_j (\mathscr{F}_j)^2 \, dV &= \int_{B(\delta)} \sum_j |\operatorname{grad} \mathscr{F}_j|^2 \, dV \\
&= \int_{B(\delta)} \{\mathscr{T}'^2 + (n-1)(\sin t)^{-2}\mathscr{T}^2\} \, dV \\
&\geq \int_{B(\delta)} (n-1)\mathscr{T}^2 \, dV \\
&= (n-1) \int_{B(\delta)} \sum_j (\mathscr{F}_j)^2 \, dV,
\end{aligned}$$

which implies (52).

To prove (51) set

$$\phi(t) = \begin{cases} 1 - \ln(\pi - t)/\ln(\pi - \delta), & n = 2, \\ 1 - \{(\pi - \delta)/(\pi - t)\}^{n-2}, & n > 2, \end{cases}$$

where $0 \leq t \leq \delta$.

For $n = 2$ we have

$$\phi'(t) = \{(\pi - t)\ln(\pi - \delta)\}^{-1}$$

and

$$\begin{aligned}
&\int_0^\delta \phi'^2 \sin t \, dt \\
&\quad = \ln^{-2}(\pi - \delta) \int_0^\delta (\pi - t)^{-2} \sin t \, dt \\
&\quad \leq \ln^{-2}(\pi - \delta) \int_0^\sigma (\pi - t)^{-2} \sin t \, dt + \ln^{-2}(\pi - \delta) \int_\sigma^\delta (\pi - t)^{-1} \, dt \\
&\quad = \ln^{-2}(\pi - \delta) \int_0^\sigma (\pi - t)^{-2} \sin t \, dt - \ln^{-2}(\pi - \delta) \ln\{(\pi - \delta)/(\pi - \sigma)\}
\end{aligned}$$

for every $\sigma \in (0, \delta)$. On the other hand,

$$\int_0^\delta \phi^2 \sin t\,dt \geq \ln^{-2}(\pi - \delta)\ln^2\{(\pi - \delta)/(\pi - \sigma)\} \int_0^\sigma \sin t\,dt.$$

Thus Rayleigh's principle implies that, for every $\sigma \in (0, \pi)$, we have

$$\lambda(\delta) \leq [\ln^{-2}\{(\pi - \delta)/(\pi - \sigma)\}]\left[\int_0^\sigma \sin t\,dt\right]^{-1} \int_0^\sigma (\pi - t)^{-2} \sin t\,dt$$
$$- [\ln^{-1}\{(\pi - \delta)/(\pi - \sigma)\}]\left[\int_0^\sigma \sin t\,dt\right]^{-1}$$

for every $\delta \in (\sigma, \pi)$. One now easily has

$$\limsup_{\delta \to \pi}\{\ln 1/(\pi - \delta)\}\,\lambda(\delta) \leq \left\{\int_0^\pi \sin t\,dt\right\}^{-1} \tag{53}$$

and (51) follows for $n = 2$.

For $n \geq 3$ we have

$$\phi'(t) = (2 - n)(\pi - \delta)^{n-2}(\pi - t)^{1-n},$$

and

$$\int_0^\delta \phi'^2 \sin^{n-1} = (n - 2)^2(\pi - \delta)^{2(n-2)} \int_0^\delta (\pi - t)^{-2(n-1)} \sin^{n-1} t\,dt$$
$$\leq (n - 2)^2(\pi - \delta)^{2(n-2)} \int_0^\delta (\pi - t)^{-(n-1)}\,dt$$
$$\leq (n - 2)(\pi - \delta)^{n-2}.$$

Also, for all $\delta \in (\sigma, \pi)$, $\sigma \in (0, \pi)$,

$$\int_0^\delta \phi^2 \sin^{n-1} \geq [1 - \{(\pi - \delta)/(\pi - \sigma)\}^{n-2}]^2 \int_0^\sigma \sin^{n-1}$$

from which we have, by Rayleigh's theorem,

$$\limsup_{\delta \to \pi}(\pi - \delta)^{2-n}\lambda(\delta) \leq (n - 2)\left\{\int_0^\pi \sin^{n-1} t\,dt\right\}^{-1}. \tag{54}$$

Thus (51) is valid for $n > 2$, as well, and the theorem is proven.

Result (51) is part of a more ambitious result, namely, if the eigenvalues of $\mathbb{S}^n$ are denoted by $\{0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots\}$ with repetitions according to multiplicity, and the Dirichlet eigenvalues of $\mathsf{B}(\delta)$ are denoted by

$\{0 < \lambda_1(\delta) \le \lambda_2(\delta) \le \lambda_3(\delta) \le \cdots\}$ with repetitions according to multiplicity, then

$$\lim_{\delta \to \pi} \lambda_j(\delta) = \lambda_{j-1}. \tag{55}$$

for all $j = 1, 2, \ldots$. Result (55) is due to H. M. MacDonald [1] (also cf. Hobson [1, pp. 403–408]), and was generalized for domains in $\mathbb{R}^n$ in Rauch–Taylor [2], and for domains in arbitrary compact Riemannian manifolds in Chavel–Feldman [4] (we treat this result and other related questions in detail in Chapter IX).

The proof of (51) that we have given is due to Del Grosso–Marchetti [1]. It has the advantage of providing a sharp estimate of $\lambda(\delta)$, namely, it is proved in Del Grosso–Marchetti [1] that

$$\left\{\int_0^\pi \sin^{n-1} t\, dt\right\} \lambda(\delta) \sim \begin{cases} \{\ln 1/(\pi - \delta)\}^{-1}, & n = 2, \\ (n-2)(\pi - \delta)^{n-2}, & n > 2, \end{cases} \tag{56}$$

as $\delta \to \pi$. To prove (56) we require the following:

LEMMA 1. Let Ω be a normal domain in a Riemannian manifold with lowest Dirichlet eigenvalue λ. Let $v \in C^2(\Omega) \cap C^0(\bar{\Omega})$ satisfy

$$\Delta v = -1 \tag{57}$$

with $v \,|\, \partial\Omega = 0$. Then

$$\lambda \ge (\max_\Omega v)^{-1}. \tag{58}$$

PROOF: A proof is given in Del Grosso–Marchetti [1] via probability, but a simpler one is available. Let u be an eigenfunction of λ. Then u does not vanish in Ω—we shall assume it is positive on all of Ω. Then

$$\int_\Omega u = -\int_\Omega (\Delta v) u = -\int_\Omega v(\Delta u) = \lambda \int v u \le \lambda \{\max_\Omega v\} \int_\Omega u,$$

which implies (58).

We now prove (56). To solve (57) on $\mathsf{B}(\delta)$ with $v \,|\, \mathsf{S}(\delta) = 0$, one easily derives

$$v_\delta(t) = \int_t^\delta \sin^{1-n} \tau \, d\tau \int_0^\tau \sin^{n-1} r\, dr.$$

By (54), and Lemma 1, it suffices to prove, in the case $n > 2$, that

$$\limsup_{\delta \to \pi} (\pi - \delta)^{n-2}(\max v_\delta) \leq \frac{1}{n-2}\int_0^\pi \sin^{n-1} t \, dt \tag{59}$$

(we leave the case $n = 2$ to the reader). Note that

$$\max v_\delta = v_\delta(0).$$

Given any $\varepsilon > 0$, there exists $\sigma_0 \in (0, \pi)$ such that

$$(1 - \varepsilon)(\pi - \tau) \leq \sin \tau$$

for all $\tau \in (\sigma_0, \pi)$. Thus for any fixed $\sigma \in (\sigma_0, \pi)$ we have

$$\begin{aligned} \max v_\delta &= \int_0^\delta \sin^{1-n}\tau \, d\tau \int_0^\tau \sin^{n-1} r \, dr \\ &= \int_0^\sigma \sin^{1-n}\tau \, d\tau \int_0^\tau \sin^{n-1} r \, dr + \int_\sigma^\delta \sin^{1-n}\tau \, d\tau \int_0^\tau \sin^{n-1} r \, dr. \end{aligned}$$

Since the first integral is bounded, independently of $\pi - \delta$, we need only consider

$$\begin{aligned} &(\pi - \delta)^{n-2} \int_\sigma^\delta \sin^{1-n}\tau \, d\tau \int_0^\tau \sin^{n-1} r \, dr \\ &\leq \left\{\int_0^\pi \sin^{n-1}\tau \, d\tau\right\}(\pi - \delta)^{n-2} \int_\sigma^\delta \sin^{1-n}\tau \, d\tau \\ &\leq \left\{\int_0^\pi \sin^{n-1}\tau \, d\tau\right\}(\pi - \delta)^{n-2}(1 - \varepsilon)^{1-n} \int_\sigma^\delta (\pi - \tau)^{1-n} \, d\tau \\ &\leq \frac{(1 - \varepsilon)^{1-n}}{(n-2)}\left\{\int_0^\pi \sin^{n-1}\tau \, d\tau\right\}. \end{aligned}$$

This, then, implies (59), which in turn implies (56).

Remark 3: Other discussions of estimating the lowest Dirichlet eigenvalue of a geodesic disk in $\mathbb{S}^n$ can be found in Barbosa–doCarmo [2], Friedland–Hayman [1], Matsuzawa–Tanno [1], Pinsky [5], and Sato [1].

CHAPTER III

λ_1 and Curvature

In this chapter we extend the study of eigenvalues to Riemannian manifolds whose curvature may not be constant, but is, nevertheless, bounded. The idea is to compare the geometric and physical quantities of a given Riemannian manifold with the respective ones in a space form of constant curvature, where the constant is a lower or upper bound of the curvature of the given Riemannian manifold.

Geometric comparison theorems go back to O. Bonnet and S. B. Myers, for estimates of the diameter, to H. E. Rauch for estimates of the growth of Jacobi fields, and to R. Bishop for estimates of the growth of volume elements of geodesic spheres. Here we present the work of M. Obata [1] and S. Y. Cheng [1, 2] on estimating λ_1 in terms of bounds on the curvature. Our treatment contains a summary of basic facts about geodesics, the exponential map, curvature, Jacobi fields, and geodesic spherical coordinates, prior to our dealing with the eigenvalue comparison theorems. Details and other developments of the background material may be found in Berger–Gauduchon–Mazet [1], Chavel [1], Cheeger–Ebin [1], and Gromoll–Klingenberg–Meyer [1].

Just a word on notation: *For the rest of the book*, given a normal domain Ω we let $\lambda(\Omega)$ denote the lowest Dirichlet eigenvalue of Ω, and $\mu(\Omega)$ the lowest nonzero Neumann eigenvalue of Ω. For a compact Riemannian manifold M, we let $\lambda(M)$ denote the lowest nonzero eigenvalue of M.

1. GEODESICS AND CURVATURE

M is our fixed Riemannian manifold, TM its tangent bundle with projection map $\pi\colon TM \to M$, that is, if $\xi \in M_p$ then $\pi(\xi) = p$. Let ∇ denote the Levi–Civita connection of the Riemannian metric.

Let $\omega\colon (\alpha, \beta) \to M$ be a C^1 path in M. A *vector field X along ω* is a map $X\colon (\alpha, \beta) \to TM$ for which $\pi \circ X = \omega$, that is, $X(t) \in M_{\omega(t)}$ for all t. To define the *derivative of a C^1 vector field along ω*, $\nabla_t X$, let $x\colon U \to \mathbb{R}^n$ be a chart on

M, containing $\omega((\alpha, \beta))$, and define the Christoffel symbols Γ_{ij}^k as in (I.23), namely,

$$\nabla_{\partial_j}\partial_i = \sum_k \Gamma_{ij}^k \partial_k, \tag{1}$$

where $\partial_1, \ldots, \partial_n$ are the coordinate vector fields associated with the chart. Set

$$\omega^j = x^j \circ \omega, \tag{2}$$

write X as

$$X = \sum_j \eta^j(\partial_j \circ \omega),$$

and define

$$\nabla_t X = \sum_l \left\{\eta^{l\prime} + \sum_{j,k} (\Gamma_{jk}^l \circ \omega)\eta^j \omega^{k\prime}\right\}(\partial_l \circ \omega). \tag{3}$$

One checks that definition (3) is independent of the choice of chart on U, and thereby determines a well-defined vector field $\nabla_t X$ along ω (even if ω is not contained in the domain of one chart on M). Also, one has for C^1 vector fields X, X_1, X_2 along ω, and $f\colon (\alpha, \beta) \to \mathbb{R} \in C^1$,

$$\nabla_t(X_1 + X_2) = \nabla_t X_1 + \nabla_t X_2, \tag{4}$$

$$\nabla_t(fX) = f'X + f\nabla_t X, \tag{5}$$

$$\langle X_1, X_2\rangle' = \langle \nabla_t X_1, X_2\rangle + \langle X_1, \nabla_t X_2\rangle. \tag{6}$$

DEFINITION 1. Let $\omega\colon (\alpha, \beta) \to M$ be a C^1 path in M. We say that the vector field X along ω is *parallel* if

$$\nabla_t X = 0$$

on all of (α, β).

By (4), (5) we have, that given ω, the set of parallel vector fields along ω is a vector space over $\mathbb{R}$. From (3) one has, via the theory of linear ordinary differential equations, that to each $t_0 \in (\alpha, \beta)$, $\xi \in M_{\omega(t_0)}$, there exists a unique parallel vector field X_ξ along ω satisfying $X_\xi(t_0) = \xi$. In particular, the space of parallel vector fields along ω is finite dimensional, and has dimension equal to that of M.

Thus we can construct isomorphisms between the tangent spaces to M at different points of ω, namely, let $t, s \in (\alpha, \beta)$, and for $\xi \in M_{\omega(t)}$ let X_ξ be the parallel vector field along ω satisfying $X_\xi(t) = \xi$. Now set

$$\tau_{t,s}(\xi) = X_\xi(s).$$

Then $\tau_{t,s}$ is an orthogonal map to $M_{\omega(t)}$ onto $M_{\omega(s)}$, and is called *parallel translation along* ω *from* $M_{\omega(t)}$ *to* $M_{\omega(s)}$.

DEFINITION 2. A path $\omega: (\alpha, \beta) \to M \in C^2$ is called a *geodesic* if we have

$$\nabla_t \omega' = 0 \tag{7}$$

on all of (α, β).

To write the equation for a geodesic in a chart, let $x: U \to \mathbb{R}^n$ be a chart on M, Γ_{ij}^k as in (1), and ω^j as in (2). Then (7) reads as

$$\omega^{l''} + \sum_{j,k} (\Gamma_{jk}^l \circ \omega)\omega^{j'}\omega^{k'} = 0. \tag{8}$$

Note that if ω is C^1, piecewise C^2, and ω satisfies (8) when it is C^2, then ω is C^∞.

To every $\xi \in TM$ there exists a maximal open interval I_ξ in $\mathbb{R}$ about the origin, and a unique geodesic $\gamma_\xi: I_\xi \to M$ satisfying

$$\gamma_\xi(0) = \pi(\xi), \qquad \gamma_\xi'(0) = \xi.$$

ASSUMPTION. For the rest of this chapter we assume that M is *geodesically complete*, that is, $I_\xi = \mathbb{R}$ for all $\xi \in TM$.

One easily checks that

$$\gamma_\xi(\alpha t) = \gamma_{\alpha\xi}(t)$$

for all α, t in $\mathbb{R}$, and that

$$|\gamma_\xi'| = |\xi|,$$

that is, γ_ξ has constant speed.

It is helpful to define the *exponential map* exp: $TM \to M$ by

$$\exp \xi = \gamma_\xi(1)$$

—thus

$$\gamma_\xi(t) = \exp t\xi$$

for all $t \in \mathbb{R}$, $\xi \in TM$. It is known that the exponential map is C^∞ and has maximal rank on the image of the zero section of M in TM (by the *zero section* we mean the identically zero vector field on M, viewed as a map $M \to TM$). More precisely, if for $p \in M$ we define

$$\exp_p = \exp | M_p,$$

and we identify M_p with *its* tangent space at each of its points in the usual manner, then

$$(9) \qquad (\exp_p)_{*|0}\,\xi = \xi$$

where (i) $(\exp_p)_{*|0}$ is the restriction of $(\exp_p)_*$ to the tangent space of M_p at the origin $(M_p)_0$, (ii) ξ on the left-hand side of (9) is viewed as an element in $(M_p)_0$, and (iii) ξ on the right-hand side of (9) is considered as an element of M_p. Finally, the map

$$\pi \times \exp\colon TM \to M \times M,$$

given by

$$(\pi \times \exp)(\xi) = (\pi(\xi), \exp \xi),$$

is C^∞, and there exists a neighborhood W of the image of the zero section of M in TM such that $(\pi \times \exp)\,|\,W$ is a diffeomorphism of W onto its image—an open neighborhood of the diagonal in $M \times M$.

As mentioned in Section II.5, for any continuous, piecewise C^1, path $\omega\colon [\alpha, \beta] \to M$, the *length of* ω, $L(\omega)$, is defined by

$$L(\omega) = \int_\alpha^\beta |\omega'|;$$

and for any two points p, q in M, the *distance from p to q*, $d(p, q)$, is defined by

$$d(p, q) = \inf L(\omega),$$

where ω ranges over continuous, piecewise C^1, paths $\omega\colon [\alpha, \beta] \to M$ for which $\omega(\alpha) = p$, $\omega(\beta) = q$. The distance function turns M into a metric space.

One has the following important facts:

THEOREM 1. (i) If ω is a continuous, piecewise C^1 path in M joining p to q, and $L(\omega) = d(p, q)$, then the image of ω is that of a geodesic. If, in addition, ω has constant speed, then ω itself is a C^∞ geodesic.

(ii) (de Rham) When M is geodesically complete, then given any $p, q \in M$, there exists at least one geodesic γ in M connecting p to q with $L(\gamma) = d(p, q)$.

(iii) (Hopf, Rinow) The geodesic completeness of M is equivalent to the completeness of M as a metric space, which, in turn, is equivalent to the statement that a subset of M is compact if and only if it is closed and bounded.

(iv) M is complete whenever it is compact.

We note that the main tools for deriving Theorem 1 are (9), and

GAUSS'S LEMMA. For $p \in M$, let $\mathfrak{S}_p$ denote the unit sphere in M_p about the origin, and let $\mathfrak{S}(p; \delta)$ be the sphere in M_p, of radius δ, about the origin. Then for any $\xi \in \mathfrak{S}_p$, $\zeta \in (\mathfrak{S}(p; t))_{t\xi}$, we have

$$\langle \gamma_\xi'(t), (\exp_p)_{*|t\xi}\zeta \rangle = 0.$$

An equivalent way of stating the result is: if ξ, η are orthonormal vectors in M_p,

$$v(t, \theta) = \exp t\{(\cos \theta)\xi + (\sin \theta)\eta\},$$

with

$$\partial_t v = v_*(\partial_t), \qquad \partial_\theta v = v_*(\partial_\theta),$$

then

$$\langle \partial_t v, \partial_\theta v \rangle = 0.$$

We now turn to the Riemann curvature tensor. For vector fields X, Y, Z on M, define

$$R(X, Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z - \nabla_{[Y,X]} Z. \tag{10}$$

It is known that $(R(X, Y)Z)(p)$ is completely determined by the values of X, Y, Z at p, and that (10) therefore determines a multilinear map $R\colon M_p \times M_p \times M_p \to M_p$ for every $p \in M$. R is called the *Riemann curvature tensor* of ∇. One has the identities

$$R(\xi, \eta)\zeta + R(\eta, \xi)\zeta = 0, \tag{11}$$

$$R(\xi, \eta)\zeta + R(\zeta, \xi)\eta + R(\eta, \zeta)\xi = 0, \tag{12}$$

$$\langle R(\xi, \eta)\zeta, \nu \rangle - \langle R(\zeta, \nu)\xi, \eta \rangle = 0, \tag{13}$$

$$\langle R(\xi, \eta)\zeta, \nu \rangle + \langle R(\xi, \eta)\nu, \zeta \rangle = 0, \tag{14}$$

for all $\xi, \eta, \zeta, \nu \in M_p$, $p \in M$. Note that R vanishes identically when $\dim M = 1$, so all discussions involving the Riemann curvature tensor will, automatically, assume $\dim M \geq 2$.

If ξ, η are linearly independent vectors of M_p, then

$$K(\xi, \eta) = \frac{\langle R(\xi, \eta)\xi, \eta \rangle}{|\xi|^2 |\eta|^2 - \langle \xi, \eta \rangle^2}$$

only depends on the two-dimensional subspace determined by ξ, η. $K(\xi, \eta)$ is referred to as the *Riemann sectional curvature of the 2-plane determined by*

ξ, η. If $G_2(M)$ denotes the collection of all two-dimensional spaces tangent to M, then $G_2(M)$ can be provided with a differentiable structure, in a natural manner, and K: $G_2(M) \to \mathbb{R}$ will then be C^∞.

If $\dim M = 2$, then $G_2(M) = M$, and K is called the *Gauss curvature of* M. For $p \in M$, $\xi, \eta, \zeta \in M_p$, we have, in the 2-dimensional case,

$$R(\xi, \eta)\zeta = K(p)\{\langle \xi, \zeta \rangle \eta - \langle \eta, \zeta \rangle \xi\}. \tag{15}$$

Similarly, for arbitrary dimension ≥ 2, the fact that K is a constant function on $G_2(M)$, equal to κ, is equivalent to saying that

$$R(\xi, \eta)\zeta = \kappa\{\langle \xi, \zeta \rangle \eta - \langle \eta, \zeta \rangle \xi\} \tag{16}$$

for all $\xi, \eta, \zeta \in M_p$, $p \in M$.

For $p \in M$, the *Ricci tensor* Ric: $M_p \times M_p \to \mathbb{R}$ is defined by

$$\mathrm{Ric}(\xi, \eta) = \mathrm{tr}(\zeta \mapsto R(\xi, \zeta)\eta),$$

and the *scalar curvature* S is defined to be the trace of Ric with respect to the Riemannian metric. Thus for any orthonormal basis of M_p, $\{e_1, \ldots, e_n\}$, we have

$$\mathrm{Ric}(\xi, \eta) = \sum_{j=1}^{n} \langle R(\xi, e_j)\eta, e_j \rangle$$

—in particular, Ric is a symmetric bilinear form on M_p. If $\xi = |\xi| e_n$, then

$$\mathrm{Ric}(\xi, \xi) = \left\{\sum_{j=1}^{n-1} K(e_j, \xi)\right\} |\xi|^2,$$

and for general $\{e_1, \ldots, e_n\}$ we have

$$S = \sum_{j \neq k} K(e_j, e_k).$$

We note that although our discussion of curvature may be the appropriate generalization, arising from the *theorema egregium*, of the intrinsic character of the Gauss curvature of a surface in 3-space, its *geometric* content is not readily apparent from the above formulas. The widely known interpretation, in the 2-dimensional case, is given by the

GAUSS–BONNET FORMULA AND THEOREM. Let M be an oriented 2-dimensional Riemannian manifold, and Ω a normal domain in M, homeomorphic to a 2-disk, with boundary Γ having corners with exterior angles $\beta_1, \ldots, \beta_l \in (-\pi, \pi)$. Then

$$\iint_\Omega K\, dA + \int_\Gamma \kappa_g\, ds = 2\pi - \sum_{j=1}^{l} \beta_j,$$

where κ_g is the geodesic curvature of Γ, ds the 1-dimensional volume element (i.e., arc length) of Γ, and dA the 2-dimensional volume element (i.e., area) of Ω. In particular, if Γ consists of 3 geodesic segments, and $\alpha_j = \pi - \beta_j$ are the interior angles, then

$$\pi + \iint_\Omega K\,dA = \sum_{j=1}^{3} \alpha_j.$$

If M is a compact 2-dimensional Riemannian manifold with Euler characteristic $\chi(M)$, then

$$\iint_M K\,dA = 2\pi\chi(M).$$

The interpretation of curvature that we shall emphasize here is its relation to geodesics.

First, let M be our given complete Riemannian manifold, with Levi–Civita connection ∇, and let ϕ: $N \to M$ be a C^∞ map of the manifold N into M. Then a *vector field X along ϕ* is defined to be a map $X: N \to TM$ satisfying $\pi \circ X = \phi$, that is, $X(p) \in M_{\phi(p)}$ for all $p \in N$. To *differentiate X along ϕ with respect to $\xi \in TN$*, let $\omega: (-\varepsilon, \varepsilon) \to N$ satisfy $\omega(0) = \pi(\xi)$, $\omega'(0) = \xi$, and define

$$\nabla_\xi X = (\nabla_t(X \circ \omega))(0).$$

Then $\nabla_\xi X$ is well defined and satisfies

$$\begin{aligned} \nabla_\xi(X_1 + X_2) &= \nabla_\xi X_1 + \nabla_\xi X_2, \\ \nabla_\xi(fX) &= (\xi f)X + f\nabla_\xi X, \\ \xi\langle X_1, X_2\rangle &= \langle \nabla_\xi X_1, X_2\rangle + \langle X_1, \nabla_\xi X_2\rangle \end{aligned}$$

for all $f: N \to \mathbb{R}$, vector fields X, X_1, X_2 along ϕ, and $\xi \in TN$.

Now let $\varepsilon_0 > 0$, $v: (\alpha, \beta) \times (-\varepsilon_0, \varepsilon_0) \to M \in C^\infty$, and write

$$v = v(t, \varepsilon), \qquad \partial_t v = v_*(\partial_t), \qquad \partial_\varepsilon v = v_*(\partial_\varepsilon).$$

Then $\partial_t v$, $\partial_\varepsilon v$ are vector fields along v satisfying

$$\nabla_\varepsilon \partial_t v - \nabla_t \partial_\varepsilon v = 0, \tag{17}$$

$$\nabla_\varepsilon \nabla_t - \nabla_t \nabla_\varepsilon = R(\partial_t v, \partial_\varepsilon v), \tag{18}$$

where ∇_t, ∇_ε is differentiation with respect to ∂_t, ∂_ε, respectively. For every ε, we set

$$\omega_\varepsilon(t) = v(t, \varepsilon).$$

THEOREM 2 (Jacobi). If ω_ε is a geodesic, for every ε, then

$$\nabla_t^2\partial_\varepsilon v + R(\partial_t v, \partial_\varepsilon v)\partial_t v = 0.$$

PROOF: We have on all of $(\alpha, \beta) \times (-\varepsilon_0, \varepsilon_0)$,

$$\begin{aligned} 0 = \nabla_t\partial_t v &= \nabla_\varepsilon\nabla_t\partial_t v \\ &= \nabla_t\nabla_\varepsilon\partial_t v + R(\partial_t v, \partial_\varepsilon v)\partial_t v \\ &= \nabla_t^2\partial_\varepsilon v + R(\partial_t v, \partial_\varepsilon v)\partial_t v. \end{aligned}$$

Thus the curvature describes the behavior of infinitesimally neighboring geodesics of a given geodesic.

DEFINITION 3. Given a geodesic $\gamma = \gamma(t)$, we define a *Jacobi field along* γ, Y, to be a C^∞ vector field along γ satisfying *Jacobi's equation*:

$$\text{(J)} \qquad \nabla_t^2 Y + R(\gamma', Y)\gamma' = 0.$$

The set J of Jacobi fields along γ is a vector space over $\mathbb{R}$ of dimension equal to $2(\dim M)$. More particularly, one has: given any $t_0 \in \mathbb{R}$, $\xi, \eta \in M_{\gamma(t_0)}$, there exists a unique $Y \in J$ satisfying $Y(t_0) = \xi$, $(\nabla_t Y)(t_0) = \eta$. If $Y \in J$ is not identically zero, then $Y(t)$, $(\nabla_t Y)(t)$ cannot vanish simultaneously for any given t.

If $X, Y \in J$, then the Wronskian of X and Y is constant, that is,

$$\langle\nabla_t X, Y\rangle - \langle X, \nabla_t Y\rangle = \text{const}.$$

In particular, for any $Y \in J$ there exist constants α, β such that

$$\text{(19)} \qquad \langle Y, \gamma'\rangle = \alpha t + \beta.$$

One immediately concludes that

$$J^\perp =: \{Y \in J : \langle Y, \gamma'\rangle = 0 \text{ on all of } \gamma\}$$

is a subspace of J with codimension equal to 2.

Given the geodesic γ, and $\alpha < \beta$, we say that $\gamma(\beta)$ *is conjugate* to $\gamma(\alpha)$ *along* γ if there exists $Y \in J$, not identically zero, such that $Y(\alpha) = Y(\beta) = 0$ (of course, such a Y would be in $J^\perp$). When $\gamma(\beta)$ is conjugate to $\gamma(\alpha)$ along γ, then a theorem of Jacobi states that

$$d(\gamma(\alpha), \gamma(t)) < (t - \alpha)|\gamma'|$$

for all $t > \beta$, that is, γ *cannot minimize distance past a conjugate point.*

For any geodesic γ in M, and $t \in \mathbb{R}$, we let $M_t^\perp$ denote the orthogonal complement of $\gamma'(t)$ in $M_{\gamma(t)}$, and define the *curvature operator*, $\mathsf{R}(t)\colon M_t^\perp \to M_t^\perp$, by

$$\mathsf{R}(t)\xi = R(\gamma'(t), \xi)\gamma'(t).$$

Note that $\mathsf{R}(t)$ is self-adjoint, by (13). Furthermore, M has constant sectional curvature κ if and only if

$$\mathsf{R}(t) = \kappa|\gamma'|^2 I, \tag{20}$$

where I is the identity map of $M_t^\perp$, for all geodesics γ, and $t \in \mathbb{R}$. If $|\gamma'| = 1$, and $Y \in J^\perp$, then (20) implies that Y is given by

$$Y(t) = a(t)\mathbf{C}_\kappa(t) + b(t)\mathbf{S}_\kappa(t), \tag{21}$$

where $a(t), b(t)$ are parallel vector fields along γ which are pointwise orthogonal to γ, and $\mathbf{S}_\kappa, \mathbf{C}_\kappa$ are given by (II.25), (II.27), respectively. Thus, when M has constant sectional curvature κ, two points are conjugate along a given geodesic if and only if: $\kappa > 0$, and the distance between the points along the geodesic is an integral multiple of $\pi/\sqrt{\kappa}$. One can easily picture the phenomenon on the sphere of radius $1/\sqrt{\kappa}$ (its geodesics are the great circles, and the curvature is κ).

We now describe the relation of Jacobi fields to the exponential map.

PROPOSITION 1. Let $p \in M$, $\xi \in M_p$, and $\eta \in M_p$. To calculate $(\exp_p)_{*|t\xi}\eta$ (where we are thinking of η as an element of the tangent space $(M_p)_{t\xi}$), set $\gamma(t) = \exp_p t\xi$, and let $Y(t)$ be the Jacobi field along γ determined by the initial conditions

$$Y(0) = 0, \qquad (\nabla_t Y)(0) = \eta.$$

Then

$$(\exp_p)_{*|t\xi}\eta = (1/t)Y(t).$$

PROOF: Pick a path $\zeta(\varepsilon)$ in M_p, with

$$\zeta(0) = \xi, \qquad \zeta'(0) = \eta,$$

and consider the 1-parameter family of geodesics

$$v(t, \varepsilon) = \exp t\zeta(\varepsilon).$$

Then, by Theorem 2, the vector field $Z(t)$ along γ given by

$$Z(t) = (\partial_\varepsilon v)(t, 0)$$

is a Jacobi field, and

$$Z(t) = (\exp_p)_{*|t\xi} t\zeta'(0) = t(\exp_p)_{*|t\xi}\eta.$$

So we wish to verify that Z has the initial conditions that determined Y.

Certainly, $Z(0) = 0$; and (17) implies

$$(\nabla_t \partial_\varepsilon v)(0, \varepsilon) = \zeta'(\varepsilon).$$

Thus

$$(\nabla_t Z)(0) = (\nabla_t \partial_\varepsilon v)(0, 0) = \eta,$$

and the claim is proven.

COROLLARY 2. The kernel of $(\exp_p)_{*|\xi}$ is isomorphic to the subspace of Jacobi fields along $\gamma(t) = \exp t\xi$, vanishing at p and $\exp \xi$.

We now turn to notions involving cut points. For each $\xi \in \mathfrak{S}_p$ we define

$$\mathrm{c}(\xi) = \sup\{t > 0 : d(p, \gamma_\xi(t)) = t\}.$$

One knows, by (9) and Gauss's lemma, that $\mathrm{c}(\xi) > 0$ for all ξ. Note that if $t \in (0, \mathrm{c}(\xi))$, then $d(p, \gamma_\xi(t)) = t$—for if there exists $t \in (0, \mathrm{c}(\xi))$ for which $d(p, \gamma_\xi(t)) < t$, then for any ε in $(0, \mathrm{c}(\xi) - t)$ we would have

$$d(p, \gamma_\xi(t + \varepsilon)) \leq d(p, \gamma_\xi(t)) + d(\gamma_\xi(t), \gamma_\xi(t + \varepsilon)) < t + \varepsilon;$$

if $\mathrm{c}(\xi)$ is finite, then the argument also applies to $\varepsilon = \mathrm{c}(\xi) - t$. In either case, we would be contradicting the definition of $\mathrm{c}(\xi)$.

Also, for $t \in (0, \mathrm{c}(\xi))$, γ_ξ must be the only minimizing geodesic connecting p to $\gamma_\xi(t)$. Otherwise, given any ε in $(0, \mathrm{c}(\xi) - t)$, there would exist a broken geodesic, of length $t + \varepsilon = d(p, \gamma_\xi(t + \varepsilon))$, joining p and $\gamma_\xi(t + \varepsilon)$. But a broken geodesic cannot minimize arc length between two points.

Thus, if we set

$$\mathfrak{D}_p = \{t\xi \in M_p : 0 \leq t < \mathrm{c}(\xi), \xi \in \mathfrak{S}_p\}$$

and

$$\mathsf{D}_p = \exp \mathfrak{D}_p,$$

then $\exp_p$ maps $\mathfrak{D}_p$ diffeomorphically onto D_p, and $\overline{\mathfrak{D}_p}$ is mapped onto all of M. Furthermore, one knows that $\mathrm{c}(\xi)$ is continuous on $\mathfrak{S}_p$, and that $\partial\mathfrak{D}_p$ has zero n-dimensional Lebesgue measure in M_p. Its image in M is $\partial\mathsf{D}_p$, and $\partial\mathsf{D}_p$ also has zero n-dimensional Riemannian measure in M.

It is customary to refer to $\gamma_\xi(\mathrm{c}(\xi))$, when $\mathrm{c}(\xi) < +\infty$, as the *cut point of p along* γ_ξ, and to refer to $\partial\mathsf{D}_p$ as the *cut locus of p in M*.

DEFINITION 4. For each $p \in M$, $\delta > 0$, set

$$\begin{aligned} \mathfrak{B}(p;\delta) &= \{\xi \in M_p : |\xi| < \delta\}, \\ \mathfrak{S}(p;\delta) &= \{\xi \in M_p : |\xi| = \delta\}, \\ \mathfrak{S}_p &= \mathfrak{S}(p;1), \\ \mathsf{B}(p;\delta) &= \{q \in M : d(p,q) < \delta\}, \\ \mathsf{S}(p;\delta) &= \{q \in M : d(p,q) = \delta\}. \end{aligned}$$

Note that we always have

$$\mathsf{B}(p;\delta) = \exp \mathfrak{B}(p;\delta),$$

and

$$\mathsf{S}(p,\delta) \cap \mathsf{D}_p = \exp \mathfrak{S}(p;\delta) \cap \mathfrak{D}_p.$$

We now work with *geodesic spherical coordinates on* D_p, induced by $\exp_p | \mathfrak{D}_p$. Assume we are given a coordinate system

$$\xi = \xi(u)$$

on $\mathfrak{S}_p$, where u varies over a domain in $\mathbb{R}^{n-1}$. A coordinate system is then determined on D_p by

$$v(t,u) = \exp t\xi(u).$$

We then have (for $\partial_t v = v_*(\partial_t)$)

$$\partial_t v(\exp t\xi) = \gamma_\xi'(t).$$

We write $\partial_\alpha v$ for $v_*(\partial/\partial u^\alpha)$, and $\partial_\alpha \xi$ for $\xi_*(\partial/\partial u^\alpha)$, where $\alpha = 1, \ldots, n-1$. From Proposition 1, one has that for every $\xi \in \mathfrak{S}_p$, the vector field along γ_ξ, Y_α, given by

$$Y_\alpha(t;\xi) = \partial_\alpha v(\exp t\xi),$$

is the Jacobi field along γ_ξ determined by the initial conditions

$$Y_\alpha(0;\xi) = 0, \qquad (\nabla_t Y_\alpha)(0;\xi) = \partial_\alpha \xi,$$

with $\partial_\alpha \xi$ orthogonal to ξ. Thus

$$|\partial_t v| = 1, \qquad \langle \partial_t v, \partial_\alpha v \rangle = 0 \tag{22}$$

by (19), which is the content of Gauss's lemma.

Our first comment is that if M has constant sectional curvature κ, then (20) is valid, from which one concludes, using (21), that

$$Y_\alpha(t;\xi) = \mathbf{S}_\kappa(t)\tau_t(\partial_\alpha \xi),$$

where τ_t denotes parallel translation along γ_ξ. Thus

$$\langle \partial_\alpha v, \partial_\beta v\rangle(\exp t\xi) = \mathbf{S}_\kappa^2(t)\langle \partial_\alpha \xi, \partial_\beta \xi\rangle,$$

which implies

$$ds^2 = (dt)^2 + \mathbf{S}_\kappa^2(t)|d\xi|^2, \tag{23}$$

that is, M is locally isometric to $\mathbb{M}_\kappa$, the complete simply connected model spaces of Section II.5 with constant sectional curvature κ. It is a standard argument that M is actually covered by $\mathbb{M}_\kappa$, and that if M is simply connected, then M is globally isometric to $\mathbb{M}_\kappa$. Thus for any given constant κ, there exists, up to isometry, only one n-dimensional, complete, simply connected Riemannian manifold of constant sectional curvature κ.

For the case of varying sectional curvature we need a more general method of dealing with $\langle \partial_\alpha v, \partial_\beta v\rangle$.

DEFINITION 5. For each $\xi \in \mathfrak{S}_p$, let $\xi^\perp$ be the orthogonal complement of $\{\mathbb{R}\xi\}$ in M_p, and let τ_t: $M_p \to M_{\exp t\xi}$ denote parallel translation along γ_ξ. We define the path of linear transformations

$$\mathscr{A}(t; \xi): \xi^\perp \to \xi^\perp$$

by

$$\mathscr{A}(t; \xi)\eta = (\tau_t)^{-1}Y(t),$$

where $Y(t)$ is the Jacobi field along γ_ξ determined by the initial conditions

$$Y(0) = 0, \qquad (\nabla_t Y)(0) = \eta.$$

Explicit calculation then verifies

PROPOSITION 2. For $\eta \in \xi^\perp$ set

$$\mathscr{R}(t)\eta = (\tau_t)^{-1}\mathsf{R}(t)(\tau_t\eta);$$

then $\mathscr{R}(t)$ is a self-adjoint map of $\xi^\perp$ and $\mathscr{A}(t; \xi)$ is the path of linear transformations satisfying

$$\mathscr{A}'' + \mathscr{R}\mathscr{A} = 0 \tag{$\mathscr{J}$}$$

with initial conditions

$$\mathscr{A}(0; \xi) = 0, \qquad \mathscr{A}'(0; \xi) = I. \tag{24}$$

DEFINITION 6. We set

$$\sqrt{\mathfrak{g}}(t; \xi) = \det \mathscr{A}(t; \xi).$$

Now return to our geodesic spherical coordinates, with the given notation. Then

$$\partial_\alpha v(\exp t\xi) = \tau_t \mathscr{A}(t;\xi)\partial_\alpha \xi \tag{25}$$

from which we have

$$\langle \partial_\alpha v, \partial_\beta v\rangle(\exp t\xi) = \langle \mathscr{A}(t;\xi)\partial_\alpha \xi, \mathscr{A}(t;\xi)\partial_\beta \xi\rangle \tag{26}$$

for $\alpha, \beta = 1, \ldots, n-1$; we therefore write

$$ds^2 = (dt)^2 + |\mathscr{A}(t;\xi)\,d\xi|^2. \tag{27}$$

Note that, for the case of constant sectional curvature κ, we have

$$\mathscr{A}(t;\xi) = \mathbf{S}_\kappa(t)I. \tag{28}$$

For the volume elements we have

$$\{\det(\langle \partial_\alpha v, \partial_\beta v\rangle)\}^{1/2}(\exp t\xi) = \sqrt{\mathfrak{g}}(t;\xi)\{\det(\langle \partial_\alpha \xi, \partial_\beta \xi\rangle)\}^{1/2}. \tag{29}$$

Thus, if $d\mu_p$ denotes the $(n-1)$-dimensional volume element on $\mathfrak{S}_p$, and $dA(\exp t\xi)$ the $(n-1)$-dimensional volume element of $\mathsf{S}(p;t) \cap \mathsf{D}_p$, then

$$dA(\exp t\xi) = \sqrt{\mathfrak{g}}(t;\xi)\,d\mu_p(\xi).$$

The n-dimensional volume element of M, dV, is given on D_p by

$$dV(\exp t\xi) = dt\,dA(\exp t\xi) = \sqrt{\mathfrak{g}}(t;\xi)\,dt\,d\mu_p(\xi).$$

2. COMPARISON THEOREMS FOR SECTIONAL CURVATURE BOUNDED FROM ABOVE

H. E. RAUCH'S COMPARISON THEOREM. Suppose we are given a geodesic γ in M, with $|\gamma'| = 1$, and

$$\mathsf{R} \leq \kappa I \tag{30}$$

on all of γ, for some fixed constant κ, that is, for every $t \in \mathbb{R}$, the eigenvalues of $\mathsf{R}(t)$ are less than or equal to κ. For any $Y \in J^\perp$, set

$$\psi_\kappa = |Y|'(0)\mathbf{S}_\kappa + |Y|(0)\mathbf{C}_\kappa.$$

Then, on any interval $(0, \beta)$ for which we have $\psi_\kappa > 0$ on the complete interval, we also have

$$|Y|'/|Y| \geq \psi_\kappa'/\psi_\kappa, \qquad \{|Y|/\psi_\kappa\}' \geq 0, \tag{31}$$

and

$$|Y| \geq \psi_\kappa. \tag{32}$$

Equality occurs in (31) (resp., (32)) at $t_0 \in (0, \beta)$ (resp., $(0, \beta]$) if and only if there exists a unit vector field E, parallel along γ and pointwise orthogonal to γ, such that

$$Y = \psi_\kappa E$$

on $[0, t_0]$.

PROOF: We have

$$|Y|' = \langle Y, \nabla_t Y\rangle |Y|^{-1},$$

which implies

$$\begin{aligned} |Y|'' &= \langle Y, \nabla_t Y\rangle' |Y|^{-1} + \langle Y, \nabla_t Y\rangle \{|Y|^{-1}\}' \\ &= |Y|^{-3}\{|\nabla_t Y|^2 |Y|^2 - \langle Y, \nabla_t Y\rangle^2 - \langle Y, \mathrm{R}Y\rangle |Y|^2\} \\ &\geq -\kappa |Y| \end{aligned}$$

by (30), and the Cauchy–Schwarz inequality. Thus

$$|Y|'' + \kappa |Y| \geq 0,$$

from which one concludes

$$\{\psi_\kappa |Y|' - \psi_\kappa' |Y|\}' \geq 0.$$

Inequalities (31), (32) now follow easily. The case of equality is also handled easily.

The first consequence of the Rauch theorem is the Hadamard–Cartan theorem, namely, if κ in (30) is nonpositive, then no two points along γ are conjugate. Thus, for a complete Riemannian manifold with nonpositive curvature, no two points are conjugate along any geodesic. By Proposition 1, the exponential map $\exp_p$ is of maximal rank on all of M_p, for every $p \in M$. One then argues that $\exp_p$ is a covering for every $p \in M$. The simply connected covering of M is, therefore, diffeomorphic to $\mathbb{R}^n$. If M, itself, is simply connected, then $\exp_p$ is a diffeomorphism, for every $p \in M$, in which case we would have: any two distinct points of M are joined by a unique minimizing geodesic.

If, on the other hand, the constant κ in (30) is positive, then one has the Morse–Schonberg theorem, namely, if two points are conjugate along κ, then their distance is greater than or equal to $\pi/\sqrt{\kappa}$.

BISHOP'S COMPARISON THEOREM (I). Assume γ is as in the Rauch theorem, with (30) valid on all of γ. Let $\mathscr{A}(t)$ be the solution of $(\mathscr{J})$

along γ satisfying the initial conditions (24), and set $\sqrt{\mathfrak{g}}(t) = \det \mathscr{A}(t)$. Then if $\mathbf{S}_\kappa$ does not vanish on $(0, \beta)$, then

$$\{\sqrt{\mathfrak{g}}/\mathbf{S}_\kappa^{n-1}\}' \geq 0 \tag{33}$$

on $(0, \beta)$, and

$$\sqrt{\mathfrak{g}} \geq \mathbf{S}_\kappa^{n-1} \tag{34}$$

on $(0, \beta]$.

Equality occurs in (33) (resp., (34)) at $t_0 \in (0, \beta)$ (resp., $(0, \beta]$) if and only if

$$\mathscr{R} = \kappa I, \qquad \mathscr{A} = \mathbf{S}_\kappa I$$

on all of $[0, t_0]$.

PROOF: We set

$$\mathscr{B} = \mathscr{A}^*\mathscr{A},$$

and note that $\mathscr{B}(t)$ is self-adjoint, positive definite on $(0, \beta)$, with

$$\det \mathscr{B} = (\det \mathscr{A})^2.$$

For any given $\alpha \in (0, \beta)$, fix an orthonormal basis $\{e_1, \ldots, e_{n-1}\}$ of $\{\gamma'(0)\}^\perp$ consisting of eigenvectors of $\mathscr{B}(\alpha)$, and define

$$\eta_j(t) = \mathscr{A}(t)e_j,$$

$j = 1, \ldots, n - 1$. Then one has

$$\begin{aligned}(\ln \det \mathscr{A})'(\alpha) &= \tfrac{1}{2}(\ln \det \mathscr{B})'(\alpha)\\ &= \tfrac{1}{2}(\operatorname{tr} \mathscr{B}'\mathscr{B}^{-1})(\alpha)\\ &= \sum_j (\langle \eta_j, \eta_j'\rangle/|\eta_j|^2)(\alpha)\\ &\geq (n-1)(\mathbf{C}_\kappa/\mathbf{S}_\kappa)(\alpha),\end{aligned}$$

by the Rauch comparison theorem.

The theorem now follows easily.

Geometric consequences of the above theorem are the following:

THEOREM 3 (Bishop). Let M be a complete, n-dimensional Riemannian manifold, all of whose sectional curvatures are less than or equal to a given constant κ. Then for any $p \in M$, and $\delta > 0$ for which

$$\mathsf{B}(p; \delta) \subseteq \mathsf{D}_p, \tag{35}$$

the volume of $\mathsf{B}(p;\delta)$ is greater than or equal to volume of disk of radius $\delta > 0$ in the n-dimensional simply connected space form $\mathbb{M}_\kappa$. Equality is achieved if and only if the two disks are isometric.

THEOREM 4 (McKean [1]). Assume that κ in Theorem 3 is negative, and that $\mathsf{B}(p;\delta)$ satisfies (35). Then

$$\lambda(\mathsf{B}(p;\delta)) \geq -(n-1)^2\kappa/4. \tag{36}$$

In particular, if M is simply connected, then

$$\lambda(\Omega) \geq -(n-1)^2\kappa/4 \tag{37}$$

for any normal domain Ω in M.

We leave it to the reader to supply a proof of (36)—it is an immediate generalization of the proof of the first claim of Theorem II.5, that is, (II.46). The second claim follows from the Hadamard–Cartan theorem.

THEOREM 5 (Cheng [1]). Let M be as in Theorem 3, $p \in M$, and $\delta > 0$ for which (35) is valid. Let $\lambda_\kappa(\delta)$ denote the lowest Dirichlet eigenvalue of the disk of radius δ in $\mathbb{M}_\kappa$. Then

$$\lambda(\mathsf{B}(p;\delta)) \geq \lambda_\kappa(\delta). \tag{38}$$

LEMMA 1 (Barta [1]). Let Ω be a normal domain in a Riemannian manifold, and $f \in C^2(\Omega) \cap C^0(\bar{\Omega})$, with $f \mid \Omega > 0$ and $f \mid \partial\Omega = 0$. Then

$$\inf_\Omega(\Delta f/f) \leq -\lambda(\Omega) \leq \sup_\Omega(\Delta f/f).$$

PROOF: Let ϕ be an eigenfunction of $\lambda(\Omega)$ with $\phi \mid \Omega > 0$ and $\phi \mid \partial\Omega = 0$, and set

$$h = \phi - f.$$

Then

$$-\lambda(\Omega) = (\Delta\phi)/\phi = (\Delta f)/f + \{f\Delta h - h\Delta f\}/f(f+h).$$

Since $f(f+h) \mid \Omega > 0$, and

$$\int_\Omega \{f\Delta h - h\Delta f\}\, dV = 0,$$

the claim follows.

PROOF OF THEOREM 5: Let $T\colon [0,\delta] \to [0,\infty)$ be the radial eigenfunction of $\lambda_\kappa(\delta)$ as discussed in Section II.5. Then

$$T'' + (n-1)(\mathbf{C}_\kappa/\mathbf{S}_\kappa)T' + \lambda_\kappa(\delta)T = 0$$

with $T'(0) = T(\delta) = 0$, $T\,|\,[0,\delta) > 0$. Then Proposition II.3 implies $T'\,|\,(0,\delta] < 0$.

Introduce geodesic spherical coordinates on $\mathsf{B}(p;\delta) \subseteq \mathsf{D}_p \subseteq M$, as discussed in Section 1, and let $F\colon \overline{\mathsf{B}(p;\delta)} \to [0,\infty)$ be given by

$$F(\exp t\xi) = T(t) \tag{39}$$

for $(t,\xi) \in [0,\delta] \times \mathfrak{S}_p$. Then

$$\begin{aligned}\frac{\Delta F}{F}(\exp t\xi) &= \frac{\{\sqrt{\mathfrak{g}}(t;\xi)T'\}'}{\sqrt{\mathfrak{g}}(t;\xi)T}\\ &= \frac{1}{T}\left\{T'' + \frac{\sqrt{\mathfrak{g}}'(t;\xi)}{\sqrt{\mathfrak{g}}(t;\xi)}T'\right\}\\ &\le \frac{1}{T}\left\{T'' + (n-1)\frac{\mathbf{C}_\kappa}{\mathbf{S}_\kappa}T'\right\} = -\lambda_\kappa(\delta);\end{aligned}$$

note that $\sqrt{\mathfrak{g}}(t;\xi)$ is given by Definition 6, and that the first line is the result of direct calculation, using (22), (29); in the second line, $\sqrt{\mathfrak{g}}'(t;\xi)$ denotes the derivative of $\sqrt{\mathfrak{g}}(t;\xi)$ with respect to t; and one goes from the second line to the third by the Bishop comparison theorem (I).

Thus

$$-\lambda(\mathsf{B}(p;\delta)) \le \sup(\Delta F)/F \le -\lambda_\kappa(\delta),$$

which implies the claim.

3. COMPARISON THEOREMS FOR RICCI CURVATURE BOUNDED FROM BELOW

BISHOP'S COMPARISON THEOREM (II). Let M be a complete Riemannian manifold, and κ a constant for which

$$\mathrm{Ric}(\xi,\xi) \ge \kappa(n-1)|\xi|^2 \tag{40}$$

for all $\xi \in TM$. For each unit tangent vector ξ let $\mathscr{A}(t;\xi)$, $\sqrt{\mathfrak{g}}(t;\xi)$ be as given in Definition 5, 6, respectively. If, for a given unit tangent vector ξ, we have a constant $\beta > 0$ such that $\sqrt{\mathfrak{g}}(t;\xi) > 0$ whenever $t \in (0,\beta)$, then

$$\{\sqrt{\mathfrak{g}}(t;\xi)/\mathbf{S}_\kappa^{n-1}\}' \le 0 \tag{41}$$

on $(0, \beta)$, and

$$\sqrt{\mathfrak{g}}(t; \xi) \leq \mathrm{S}_\kappa^{n-1} \tag{42}$$

on $(0, \beta]$.

Equality is achieved in (41) (resp., (42)) at $t_0 \in (0, \beta)$ (resp., $(0, \beta]$) if and only if

$$\mathscr{R} = \kappa I, \qquad \mathscr{A}(t; \xi) = \mathrm{S}_\kappa(t) I$$

on $[0, t_0]$.

PROOF: Let $V = \xi^\perp$. If T: V → V is a self-adjoint linear transformation of V, then the Cauchy–Schwarz inequality implies that

$$\operatorname{tr} T^2 \geq (\operatorname{tr} T)^2/(n-1), \tag{43}$$

with equality if and only if T is scalar.

For two paths of linear transformations of V, $A(t)$, $B(t)$, we define their *Wronskian* $W(A, B)$ by

$$W(A, B) = (A')^* B - A^* B'.$$

Note that if $\boldsymbol{A}$, $\boldsymbol{B}$ are solutions of $(\mathscr{J})$, then $W(\boldsymbol{A}, \boldsymbol{B})$ is a constant. One then has

$$W(\mathscr{A}, \mathscr{A}) = 0.$$

We are now given $\beta > 0$ as in the statement of the theorem. Set

$$U = \mathscr{A}'\mathscr{A}^{-1}$$

on $(0, \beta)$. Then

$$U^* - U = (\mathscr{A}^{-1})^* W(\mathscr{A}, \mathscr{A})\mathscr{A}^{-1} = 0;$$

so U is self-adjoint. Explicit calculation shows that U satisfies the *matrix Riccati equation*

$$U' + U^2 + \mathscr{R} = 0. \tag{$\mathscr{R}$}$$

Thus (40) and (43) imply

$$0 = \operatorname{tr} U' + \operatorname{tr} U^2 + \operatorname{tr} \mathscr{R} \geq (\operatorname{tr} U)' + (\operatorname{tr} U)^2/(n-1) + \kappa(n-1),$$

that is,

$$\phi =: \operatorname{tr} U = \sqrt{\mathfrak{g}}'/\sqrt{\mathfrak{g}}$$

satisfies the differential inequality

$$\phi' + \phi^2/(n-1) + \kappa(n-1) \leq 0. \tag{44}$$

Note that if

$$\psi = (n-1)\mathbf{C}_\kappa/\mathbf{S}_\kappa,$$

then

$$\psi' + \psi^2/(n-1) + \kappa(n-1) = 0.$$

Therefore consider the continuous function $\theta(t)$ defined on $[0, \beta) \cap [0, \pi/\sqrt{\kappa})$ (where $\pi/\sqrt{\kappa} = +\infty$ when $\kappa \le 0$) by

$$\theta(0) = 0, \qquad \phi(t) = \psi(\theta(t))$$

when $t \neq 0$.

One easily concludes that

$$\theta' \ge 1$$

on $[0, \beta) \cap [0, \pi/\sqrt{\kappa})$, from which one has

$$\theta(t) \ge t, \qquad \phi(t) \le \psi(t) = (n-1)(\mathbf{C}_\kappa/\mathbf{S}_\kappa)(t),$$

that is, (41) and (42) are valid on $(0, \beta) \cap (0, \pi/\sqrt{\kappa})$. But the validity of (42) on $(0, \beta) \cap (0, \pi/\sqrt{\kappa})$ implies $\beta \le \pi/\sqrt{\kappa}$; so (41) (resp., (42)) is valid on all of $(0, \beta)$ (resp., $(0, \beta]$).

We postpone, for a moment, considering the case of equality and draw some immediate conclusions from $\beta \le \pi/\sqrt{\kappa}$.

BONNET–MYERS THEOREM. Let M be a complete Riemannian manifold with Ricci curvature satisfying (40) on all of TM, with $\kappa > 0$. Then M is compact, has finite fundamental group, and has diameter $d(M)$ satisfying

$$d(M) \le \pi/\sqrt{\kappa}.$$

Indeed, M is bounded since, for any unit tangent vector ξ, $\sqrt{\mathfrak{g}}(t; \xi)$ has a zero on $(0, \pi/\sqrt{\kappa}]$, which implies there exists a nontrivial Jacobi field vanishing at $t = 0$ and at some $\beta_0 \in (0, \pi/\sqrt{\kappa}]$. But, as we mentioned, geodesics do not minimize distance past conjugate points. Thus M is bounded, with diameter less than or equal to $\pi/\sqrt{\kappa}$. Once M is bounded and complete, the Hopf–Rinow theorem states that M is compact. Since the universal cover of a complete Riemannian manifold is always complete, the universal cover of M is complete and its Ricci curvature satisfies (40) on all of TM. Thus the universal cover of M is compact, and the fundamental group of M is finite.

We now consider the case of equality in the Bishop theorem (II) (the case of equality in the Bonnet–Myers theorem is the Toponogov theorem below). Note that equality in (42) at some $t_0 \in (0, \beta]$ implies equality in (41) on all of $(0, t_0)$). So it suffices to consider the case of equality in (41) at some $t_0 \in (0, \beta)$.

If we are given equality in (41) at some $t_0 \in (0, \beta)$, then $\theta(t_0) = t_0$, which implies $\theta' = 1$ on all of $[0, t_0]$, that is, $\theta(t) = t$ on all of of $[0, t_0]$. Thus

$$\phi(t) = (n - 1)(\mathbf{C}_\kappa/\mathbf{S}_\kappa)(t)$$

on all of $[0, t_0]$, and we have equality in (44) on all of $[0, t_0]$. This, in turn, implies that $U(t)$ is scalar on all of $[0, t_0]$, and

$$\mathscr{A}' = (\mathbf{C}_\kappa/\mathbf{S}_\kappa)\mathscr{A}$$

on all of $[0, t_0]$. But, now, explicit calculation implies

$$\mathscr{A}'' + \kappa\mathscr{A} = 0$$

on $[0, t_0]$, and the theorem follows.

THEOREM 6 (Bishop). Let M be complete, with Ricci curvature satisfying (40) on all of TM. Then for any $\delta > 0$, $p \in M$, the volume of $\mathsf{B}(p; \delta)$ is less than or equal to the volume of the disk of the same dimension, having constant sectional curvature κ and radius δ. Equality is achieved if and only if the two disks are isometric.

If $\kappa > 0$, and $\dim M = n$, then

$$\operatorname{vol} M \le (\operatorname{vol} \mathbb{S}^n)/\kappa^{n/2},$$

with equality if and only if M is isometric to the n-sphere of constant sectional curvature κ.

Remark 1: We note the contrast in the two Bishop comparison theorems, the first one requiring the Rauch estimate on the growth of a Jacobi field vanishing at the initial point of the geodesic, and the second one working directly with the matrix Ricatti equation. For sectional curvature bounded from below, one also has a Rauch estimate on the growth of Jacobi fields which vanish at the initial point of the geodesic. Compare H. E. Rauch's [1] original paper, and M. Berger's survey [7] of the field of pinching theorems inspired by it.

THEOREM 7 (Cheng [2]). Let M be complete with Ricci curvature satisfying (40) on all of TM. Then for any $\delta > 0$, $p \in M$, we have

$$\lambda^*(\mathsf{B}(p; \delta)) \le \lambda_\kappa(\delta) \tag{45}$$

with equality if and only if $\mathsf{B}(p; \delta)$ is isometric to the disk in $\mathbb{M}_\kappa$ of radius δ.

PROOF: Note that we have used $\lambda^*(\mathsf{B}(p;\delta))$ instead of $\lambda(\mathsf{B}(p;\delta))$ since we dispensed with the hypothesis that $\overline{\mathsf{B}(p;\delta)} \subseteq \mathsf{D}_p$, and can, therefore, no longer assume that $\mathsf{B}(p;\delta)$ is a regular domain in M. Thus the theorem is understood to be claiming that, given any $\varepsilon > 0$, there exists a function F, approximated, relative to (I.76), by functions in $C_c^\infty(\mathsf{B}(p;\delta))$, such that $F \not\equiv 0$, and

$$D[F,F] \le \{\lambda_\kappa(\delta) + \varepsilon\}\|F\|^2.$$

The function F is constructed as follows: Let $T(t)$ be the eigenfunction of $\lambda_\kappa(\delta)$, and for $(t,\xi) \in [0,\infty) \times \mathfrak{S}_p$, satisfying $t\xi \in \overline{\mathfrak{B}(p;\delta)} \cap \overline{\mathfrak{D}}_p$, define

$$F(\exp t\xi) = T(t).$$

We note that F is well defined, since if two minimizing geodesics, emanating from p, intersect at q then both geodesics have the same length. F is continuous, since the function $\mathrm{c}(\xi)$, $\xi \in \mathfrak{S}_p$, is continuous. Also, for $t\xi \in \mathfrak{D}_p$, we have

$$|(\operatorname{grad} F)(\exp t\xi)| = |T'(t)|;$$

so $\operatorname{grad} F$ has bounded length on $\mathsf{B}(p;\delta)$. Since $\operatorname{grad} F$ is continuous everywhere except, possibly, on $\partial\mathsf{D}_p \cap \overline{\mathsf{B}(p;\delta)}$—a set of Riemannian n-measure 0—we conclude that $F \in \mathscr{H}(\mathsf{B}(p;\delta))$. We now show that F can be approximated, in the metric (I.76), by functions in $\mathscr{H}(\mathsf{B}(p;\delta))$ which are compactly supported on $\mathsf{B}(p;\delta)$. Clearly, one will then be able to approximate F with functions in $C_c^\infty(\mathsf{B}(p;\delta))$.

Let $L\colon [0,\infty) \to \mathbb{R} \in C$ with $L'(0) = 0$, and $\operatorname{supp} L \subseteq [0,\delta_1]$ for some $\delta_1 < \delta$; and let $G\colon \overline{\mathsf{B}(p;\delta)} \to \mathbb{R}$ be defined by

$$G(\exp t\xi) = L(t)$$

for all $t\xi \in \bar{\mathsf{D}}_p \cap \overline{\mathsf{B}(p;\delta)}$. Then $G \in \mathscr{H}(\mathsf{B}(p;\delta))$, as F is, and has compact support. Also, we have

$$\begin{aligned}\|F - G\|^2 &= \int_{\mathfrak{S}_p} d\mu_p(\xi)\int_0^{\min\{\mathrm{c}(\xi),\delta\}} (T-L)^2(t)\sqrt{\mathrm{g}}(t;\xi)\,dt\\ &\le \int_{\mathfrak{S}_p} d\mu_p(\xi)\int_0^{\min\{\mathrm{c}(\xi),\delta\}} (T-L)^2(t)\mathsf{S}_\kappa^{n-1}(t)\,dt\\ &\le \int_{\mathfrak{S}_p} d\mu_p(\xi)\int_0^{\delta} (T-L)^2(t)\mathsf{S}_\kappa^{n-1}(t)\,dt\end{aligned}$$

which is the $L^2(\mathsf{B}_\kappa(\delta))$ distance of the functions on $\mathsf{B}_\kappa(\delta)$ determined by T and L. A similar estimate holds for $\|\operatorname{grad}(F-G)\|^2$. So any degree of approximation, achieved in $\mathscr{H}(\mathsf{B}_\kappa(\delta))$, is achieved, automatically, in $\mathscr{H}(\mathsf{B}(p;\delta))$. Thus F is an admissible function for $\lambda(\mathsf{B}(p;\delta))$.

Let

$$b(\xi) = \min(\mathrm{c}(\xi), \delta).$$

Then, as mentioned,

$$\|\operatorname{grad} F\|^2 = \int_{\mathfrak{S}_p} d\mu_p(\xi) \int_0^{b(\xi)} T'^2 \sqrt{\mathfrak{g}}(t;\xi)\, dt,$$

and

$$\|F\|^2 = \int_{\mathfrak{S}_p} d\mu_p(\xi) \int_0^{b(\xi)} T^2 \sqrt{\mathfrak{g}}(t;\xi)\, dt,$$

and it suffices to establish

$$\int_0^{b(\xi)} T'^2 \sqrt{\mathfrak{g}}(t;\xi)\, dt \le \lambda_\kappa(\delta) \int_0^{b(\xi)} T^2 \sqrt{\mathfrak{g}}(t;\xi)\, dt \tag{46}$$

for every $\xi \in \mathfrak{S}_p$. Well,

$$\begin{aligned}
\int_0^{b(\xi)} T'^2 \sqrt{\mathfrak{g}}(t;\xi)\, dt &= TT'\sqrt{\mathfrak{g}}(t;\xi)\Big|_0^{b(\xi)} - \int_0^{b(\xi)} T\{T'\sqrt{\mathfrak{g}}(t;\xi)\}'\, dt \\
&= (TT')(b(\xi))\sqrt{\mathfrak{g}}(b(\xi);\xi) - \int_0^{b(\xi)} T\{T'\sqrt{\mathfrak{g}}(t;\xi)\}'\, dt \\
&\le -\int_0^{b(\xi)} T\{T'' + T'\sqrt{\mathfrak{g}}'(t;\xi)/\sqrt{\mathfrak{g}}(t;\xi)\}\sqrt{\mathfrak{g}}(t;\xi)\, dt \\
&\le -\int_0^{b(\xi)} T\{T'' + (n-1)(\mathrm{C}_\kappa/\mathrm{S}_\kappa)T'\}\sqrt{\mathfrak{g}}(t;\xi)\, dt \\
&= \lambda_\kappa(\delta)\int_0^{b(\xi)} T^2\sqrt{\mathfrak{g}}(t;\xi)\, dt,
\end{aligned}$$

which is the claim (46). Note that we have used the facts $T\,|\,[0,\delta) > 0$, $T'\,|\,(0,\delta] < 0$, and the Bishop comparison theorem (II).

The case of equality is easily handled.

Remark 2: Assume that in Cheng's theorem we have $\kappa > 0$. Then the theorem only has content when $\delta < \pi/\sqrt{\kappa}$. For by the Bonnet–Myers theorem (Section III.3), $d(M) \le \pi/\sqrt{\kappa}$. Therefore, if $\delta > \pi/\sqrt{\kappa}$ we have $\mathsf{B}(p;\delta) = M$ and $\lambda(\mathsf{B}(p;\delta) = 0$. If $\delta = \pi/\sqrt{\kappa}$ then one also has $\lambda(\mathsf{B}(p;\delta) = \lambda_\kappa(\delta) = 0$ by Theorem II.6. Of course, when $d(M) = \pi/\sqrt{\kappa}$, the subsequent argument of the Toponogov theorem shows that M is isometric to $\mathbb{M}_\kappa = \mathbb{S}^n(1/\sqrt{\kappa})$.

Remark 3: We note that Cheng's theorem is a counterexample to the metaphysical principle: *Large spaces–small eigenvalues, and small spaces–large eigenvalues.* We refer to the rubric as a metaphysical principle because (i) *large spaces* and *small spaces* are not well defined, (ii) the rubric is valid in sufficiently many situations, where there might be agreement as how to declare a space larger or smaller, as to serve as a guide to intuition in unknown situations, (iii) it therefore serves as a barometer of the striking character of a result serving as a counterexample to the principle, and therefore, (iv) a commitment to the validity of such a principle leads to either (a) a more penetrating study into the determination of geometric size or (b) declaring that large eigenvalues might serve—in appropriate situations—as a definition for a small space.

Many of the results described in this book reflect the validity of this principle. Compare the normalization of geometric data (Section XII.7), the Weyl formula (Section I.3), and especially the domain monotonicity of eigenvalues with vanishing Dirichlet boundary data (Section I.5), for early examples of this principle. The principle is also a very helpful guide to appreciating the results of Chapter IX. But here, the Rauch and Bishop theorems say that increasing the curvature decreases the size of a geodesic disk of fixed radius, and the Cheng theorems state that increasing the curvature *also* decreases the lowest Dirichlet eigenvalue (contrary to our expectation that it be increased).

We note that the Obata theorem in the next section states that if M is a compact Riemannian manifold of dimension $n \geq 2$, with Ricci curvatures bounded below by the constant $(n-1)\kappa$, $\kappa > 0$, then $\lambda_1(M) \geq n\kappa = \lambda_1(\mathbb{M}_\kappa)$—the result consistent with the metaphysical principle. Interestingly, the strong form of the theorem states that $\lambda_1(M) = n\kappa$ if and only if M is isometric to $\mathbb{M}_\kappa$—our proof uses Cheng's theorem, in violation of the principle.

COROLLARY 3. Let M be as in Theorem 7, and compact. Denote the list of eigenvalues of M by

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots.$$

Then

$$\lambda_j(M) \leq \lambda_\kappa(d(M)/2j) \tag{47}$$

for all $j = 1, 2, \ldots$.

PROOF: Pick points p, q in M such that

$$d(p, q) = d(M).$$

Join p and q by a geodesic $\gamma\colon [0, d(M)] \to M$, $|\gamma'| = 1$, with $\gamma(0) = p$, $\gamma(d(M)) = q$. Define

$$p_l = \gamma((l-1)d(M)/j)$$

for $l = 1, \ldots, j+1$. Then the disks $\{B(p_l; d(M)/2j) : l = 1, \ldots, l+1\}$ are pairwise disjoint. The max–min argument, for Dirichlet eigenvalues, and Cheng's theorem, implies (47).

Remark 4: We gave the particular estimate of Corollary 3 for two reasons: (i) For $\lambda_1(M)$ it helps characterize the case of equality in the Obata theorem (Section 4), and (ii) it theoretically gives an upper bound of all eigenvalues in terms of lower bounds on the Ricci curvature and diameter of M.

However, for large values of j, we have

$$\lambda_\kappa(d(M)/2j) \sim 4c_D j^2/d^2(M),$$

by (XII.34), where c_D is the lowest Dirichlet eigenvalue of the unit disk in $\mathbb{R}^n$. So the Cheng estimate (47) gives

$$\lambda_j(M) \le \text{const} \cdot j^2$$

for large j, with the constant depending on n and M.

To get a better estimate one uses the following result of Gromov [1] (our thanks to J. Dodziuk for this alternative approach): For any given $\varepsilon > 0$, let $N(\varepsilon)$ be the maximal number of pairwise disjoint geodesic disks in M all having radius $\varepsilon > 0$. Then Bishop's theorem implies

$$N(\varepsilon) \ge \mathrm{vol}(M)/\mathrm{vol}(B_\kappa(2\varepsilon)).$$

The max–min argument for Dirichlet eigenvalues, and Cheng's theorem, imply that (if $[x]$ denotes the largest integer $\le x$, then)

$$\lambda_{[\mathrm{vol}(M)/\mathrm{vol}(B_\kappa(2\varepsilon))]}(M) \le \lambda_\kappa(\varepsilon).$$

From (XII.34) one can now obtain

$$\lambda_k(M) \le \text{const} \cdot \{k/\mathrm{vol}(M)\}^{2/n},$$

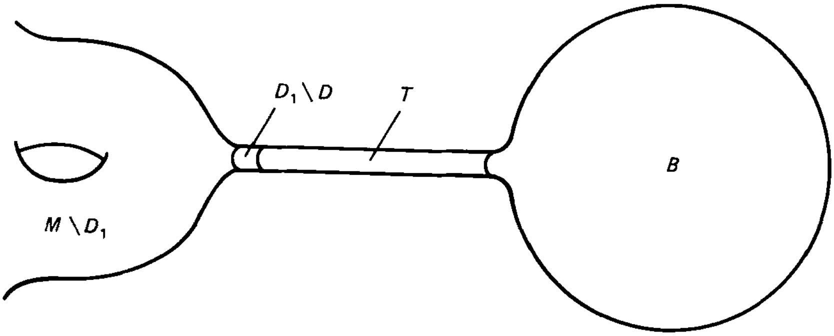


Fig. 1

where the constant depends only on n and κ, for sufficiently large k—an estimate much closer to that given by the Weyl formula (I.50). A sharper version of this argument can be found in Section 4 of Li–Yau [1].

Remark 5: We also note that Li–Yau [1] also give a *lower* bound for $\lambda(M)$ in terms of the diameter and lower bound of the Ricci curvature of M. Compare Theorem V.4 and Section XII.9.

Of course, one cannot expect to bound $\lambda(M)d^2(M)$ from below by a positive constant depending on, at most, the underlying topology of M. Indeed, given any fixed Riemannian metric on M, let D be a disk, and change the metric in a neighborhood D_1 of D so as to form the "dumbbell" $T \cup B$, where T contains a flat cylinder, isometric to $\mathbb{S}^{n-1}(\varepsilon) \times [-l, l]$, and B has the same volume as $M \setminus D_1$ but is otherwise arbitrary (cf. Fig. 1). It is easy to show that for the new metric

$$\lambda(M) \leq \text{const} \cdot \varepsilon^{n-1}/l\ (\text{vol}(M))$$

where the constant is independent of ε and l. One can show that by suitable choice of ε and l, one can arrange that both

$$\lambda(M)d^2(M), \qquad \lambda(M)\,\text{vol}^{2/n}(M)$$

are as small as one desires.

For a survey on bounds for $\lambda(M)$ (see Li [4]).

COROLLARY 4 (Cheng [5]). If M is a compact two-dimensional Riemannian manifold, of genus g, whose Gauss curvature is greater than or equal to κ on all of M, with $\kappa < 0$, then λ_1 satisfies

$$\lambda_1 \leq \lambda_\kappa(\{\text{arccosh}(2g - 1)\}/2\sqrt{-\kappa}). \tag{48}$$

PROOF: The Gauss–Bonnet and Bishop theorems imply, for A, the area of M, since $\chi(M) = 2(1 - g)$,

$$\begin{aligned} g - 1 \le -\kappa A/4\pi &\le -\kappa 2\pi\left\{\int_0^{d(M)} \mathrm{S}_\kappa\right\}\Big/4\pi \\ &= \tfrac{1}{2}\{\cosh(\sqrt{-\kappa}\,d(M)) - 1\}. \end{aligned}$$

Thus

$$d(M) \ge (1/\sqrt{-\kappa})\operatorname{arccosh}(2g - 1)$$

and (48) follows from (47) for $j = 1$.

We now present Gage's estimate for $\lambda_\kappa(\delta)$ for $\kappa \le 0$. For our convenience, we only consider the case where $n = 2$. We present the estimate here, instead of after Theorem II.5, because the integration-by-parts, used in this argument, is the same as that used in Cheng's argument.

THEOREM 8 (Gage [1]). For $n = 2$, and $\kappa \le 0$ we have

$$\lambda_\kappa(\delta) \le \frac{-\kappa}{4} + \frac{\pi^2}{\delta^2} - \frac{1}{4\mathrm{S}_\kappa^2(\delta)}. \tag{49}$$

PROOF: First note that for any given constants α, ε, and

$$f(t) = \mathrm{S}_\kappa^{-1/2}(t)\cos(\alpha t - \varepsilon),$$

we have

$$f'(t) = -\mathrm{S}_\kappa^{-1/2}\{(\mathrm{C}_\kappa/2\mathrm{S}_\kappa)\cos(\alpha t - \varepsilon) + \alpha\sin(\alpha t - \varepsilon)\},$$

and

$$f'' + (\mathrm{C}_\kappa/\mathrm{S}_\kappa)f' + \{(-\kappa/4) + \alpha^2 - (1/4\mathrm{S}_\kappa^2)\}f = 0.$$

To estimate $\lambda_\kappa(\delta)$ from above, consider the function

$$\phi(t) = \begin{cases} \mathrm{S}_\kappa^{-1/2}(\sigma)\cos(\alpha\sigma - \varepsilon), & 0 \le t \le \sigma, \\ \mathrm{S}_\kappa^{-1/2}(t)\cos(\alpha t - \varepsilon), & \sigma \le t \le \delta, \end{cases}$$

where σ, α, ε are constants to be chosen. We shall pick them to ensure

$$\phi\,|\,[\sigma, \delta) > 0, \qquad \phi(\delta) = 0, \tag{50}$$

$$\phi'(\sigma) = 0, \qquad \phi'\,|\,(\sigma, \delta] < 0. \tag{51}$$

Fix an arbitrary τ in $(0, 1)$, and set

$$\sigma = \tau\delta.$$

For any ε in $(0, \pi(1+\tau)/2(1-\tau))$, set

$$\alpha = (1/\delta)(\pi/2 + \varepsilon). \tag{52}$$

Then

$$\alpha < \pi/\delta(1-\tau), \qquad \alpha\sigma - \varepsilon > -\pi/2, \qquad \alpha\delta - \varepsilon = \pi/2.$$

So (50) is satisfied for any ε in $(0, \pi(1+\tau)/2(1-\tau))$.

We now wish to determine ε to satisfy (51). Set

$$\psi(t; \varepsilon) = (\mathbf{C}_\kappa/2\mathbf{S}_\kappa)\cos(\alpha t - \varepsilon) + \alpha \sin(\alpha t - \varepsilon).$$

If α is related to ε by (52), then

$$\psi(\tau\delta; 0) > 0, \qquad \psi(\tau\delta; \pi(1+\tau)/2(1-\tau)) < 0.$$

One easily has the existence of ε, in the stated interval, for which (51) is valid.

Now let Λ be defined by

$$\Lambda = \frac{-\kappa}{4} + \frac{\pi^2}{\delta^2(1-\tau)^2} - \frac{1}{4\mathbf{S}_\kappa^2(\delta)}.$$

Then

$$\phi'' + (\mathbf{C}_\kappa/\mathbf{S}_\kappa)\phi' + \Lambda\phi \geq 0$$

on all of (σ, δ). Thus

$$\begin{aligned}\int_0^\delta \phi'^2 \mathbf{S}_\kappa &= \int_\sigma^\delta \phi'^2 \mathbf{S}_\kappa \\ &= -\int_\sigma^\delta \phi\{\phi'' + (\mathbf{C}_\kappa/\mathbf{S}_\kappa)\phi'\}\mathbf{S}_\kappa \\ &\leq \Lambda \int_\sigma^\delta \phi^2 \mathbf{S}_\kappa \\ &\leq \Lambda \int_0^\delta \phi^2 \mathbf{S}_\kappa.\end{aligned}$$

Thus, Rayleigh's principle implies

$$\lambda_\kappa(\delta) \leq \frac{-\kappa}{4} + \frac{\pi^2}{\delta^2(1-\tau)^2} - \frac{1}{4\mathbf{S}_\kappa^2(\delta)}$$

for all $\tau \in (0, 1)$, which implies (49).

Remark 6: Using estimates on the zeros of Legendre functions, Cheng [2] showed that there exist constants, depending only on n, such that

$$\lambda_\kappa(\delta) \leq \text{const} \cdot \max\{0, -\kappa\} + \text{const}/\delta^2.$$

Gage's approach, through test functions, is also valid in all dimensions and yields a sharpening of Cheng's original results.

4. OBATA AND TOPONOGOV THEOREMS

If V is an n-dimensional, real inner product space, and $T\colon V \to V$ a linear transformation, then the *norm of* T, $|T|$, will be defined by

$$|T|^2 = \sum_{j,k=1}^{n} \langle Te_j, e_k \rangle^2,$$

where $\{e_1, \ldots, e_n\}$ is any orthonormal basis of V. One easily checks that $|T|$ is well defined, and that

$$|T|^2 \geq (\text{tr } T)^2/n, \tag{53}$$

with equality in (53) if and only if T is a scalar multiple of the identity.

If M is our given Riemannian manifold, with Levi–Civita connection ∇, and f is a C^2 function on M, then the *Hessian of*, $\text{Hess} f\colon TM \to TM$, is defined by

$$(\text{Hess} f)(\xi) = \nabla_\xi \, \text{grad} f,$$

for $\xi \in TM$. Of course, $(\text{Hess} f)(M_p) \subseteq M_p$ for all $p \in M$. Note that

$$\Delta f = \text{tr Hess} f. \tag{54}$$

THEOREM 9. Let M be an n-dimensional, compact Riemannian manifold with Ricci curvature satisfying, for some given constant $\kappa > 0$,

$$\text{Ric}(\xi, \xi) \geq \kappa(n-1)|\xi|^2 \tag{40}$$

for all $\xi \in TM$. Then $\lambda(M)$, satisfies

$$\lambda(M) \geq n\kappa, \tag{55}$$

with equality in (55) if and only if M is isometric to the n-sphere of constant sectional curvature κ.

Inequality (55) is due to Lichnerowicz [1], and the characterization of the case of equality is due to Obata [1].

GENERALIZED TOPONOGOV THEOREM (Cheng [2]). If M is as in the hypothesis of Theorem 9, and has maximal diameter, that is, (cf. the Bonnet–Myers theorem),

$$d(M) = \pi/\sqrt{\kappa},$$

then M is isometric to the n-sphere of constant sectional curvature κ.

We prove both theorems together. Inequality (55) follows the Bochner–Lichnerowicz formula (cf. Berger–Gauduchon–Mazet [1, p. 131 ff.] for a proof):

$$\tfrac{1}{2}\Delta(|\operatorname{grad} f|^2) = |\operatorname{Hess} f|^2 + \langle \operatorname{grad} f, \operatorname{grad} \Delta f\rangle + \operatorname{Ric}(\operatorname{grad} f, \operatorname{grad} f) \tag{56}$$

for any $f \in C^\infty(M)$.

Now if f is an eigenfunction on M with eigenvalue λ, then (40), (53), (54), (56) combine to imply

$$\tfrac{1}{2}\Delta(|\operatorname{grad} f|^2) \geq \lambda^2 f^2/n + \{\kappa(n-1) - \lambda\}|\operatorname{grad} f|^2 \tag{57}$$

on all of M. Integrate the inequality (57) over M; then

$$0 \geq \{\lambda^2/n + [\kappa(n-1) - \lambda]\lambda\}\|f\|^2 = \lambda(n-1)\{n\kappa - \lambda\}\|f\|^2/n \tag{58}$$

and inequality (55) follows.

We now turn to the Toponogov theorem. Suppose, with the given hypotheses, that we have points $p_1, p_2 \in M$ satisfying

$$d(p_1, p_2) = \pi/\sqrt{\kappa}. \tag{59}$$

Then

$$n\kappa \leq \lambda(M) \leq \max_{j=1,2} \lambda(\mathsf{B}(p_j; \pi/2\sqrt{\kappa})) \leq n\kappa.$$

The first inequality is (55), the second inequality is the max–min method, and the third inequality is the result of Theorem 7.

Equality (59) therefore implies

$$\lambda(M) = n\kappa;$$

and equality in the max–min inequality implies that

$$M = \bigcup_{j=1}^{2} \overline{\mathsf{B}(p_j; \pi/2\sqrt{\kappa})},$$

and

$$\lambda(\mathsf{B}(p_j; \pi/2\sqrt{\kappa})) = n\kappa \tag{60}$$

for $j = 1, 2$. But (60) is the case of equality in Theorem 7; thus each $\mathsf{B}(p_j; \pi/2\sqrt{\kappa})$ is isometric to the hemisphere of $\mathbb{M}_\kappa$. Then M is isometric to $\mathbb{M}_\kappa$.

Finally, we return to the case of equality in (55). Well, equality in (55) implies equality in (57) on all of M. Thus there exists a C^∞ function ϕ for which

$$\operatorname{Hess} f = \phi I$$

on all of M. Then

$$-\lambda f = \Delta f = \operatorname{tr} \operatorname{Hess} f = n\phi;$$

so

$$\phi = -\kappa f,$$

and

$$(\operatorname{Hess} f)(\xi) = -\kappa f \xi$$

for all $\xi \in TM$.

Now note that for any $\omega: (\alpha, \beta) \to M$ we have

$$\begin{aligned} -\kappa f |\omega'|^2 &= \langle (\operatorname{Hess} f)(\omega'), \omega' \rangle \\ &= \langle \nabla_t \operatorname{grad} f, \omega' \rangle \\ &= \langle \operatorname{grad} f, \omega' \rangle' - \langle \operatorname{grad} f, \nabla_t \omega' \rangle \\ &= (f \circ \omega)'' - \langle \operatorname{grad} f, \nabla_t \omega' \rangle. \end{aligned}$$

Therefore, for any geodesic $\gamma: (\alpha, \beta) \to M$, $|\gamma'| = 1$, we have

$$(f \circ \gamma)'' + \kappa (f \circ \gamma) = 0;$$

so

$$(f \circ \gamma)(t) = A \cos\sqrt{\kappa} t + B \sin\sqrt{\kappa} t$$

for some constants A, B.

If $p \in M$ has the property that

$$f(p) = \max_M f,$$

then for every $\xi \in \mathfrak{S}_p$ we have

$$f(\gamma_\xi(t)) = f(p) \cos\sqrt{\kappa} t,$$

which implies that $\exp_p|\mathsf{B}(p; \pi/\sqrt{\kappa})$ is a diffeomorphism of $\mathsf{B}(p; \pi/\sqrt{\kappa})$ onto $\mathsf{B}(p; \pi/\sqrt{\kappa})$. In particular, $d(M) \geq \pi/\sqrt{\kappa}$. Then $d(M) = \pi/\sqrt{\kappa}$ by the Bonnet–Myers theorem and the result follows by the Generalized Toponogov Theorem.

CHAPTER IV

Isoperimetric Inequalities

In this chapter we consider relations between geometric isoperimetric inequalities and physical isoperimetric inequalities. In our context, given an n-dimensional Riemannian manifold M, a *geometric isoperimetric inequality* will be an inequality for the $(n-1)$-dimensional volume of $\partial\Omega$, where Ω is the union of a finite number of regular domains, with compact closure in M, once the n-dimensional volume of Ω is given. A *physical isoperimetric inequality* will be an inequality for the eigenvalues of an arbitrary domain Ω, with compact closure in M, given the n-dimensional volume of Ω. Our emphasis will not be on the derivation of the geometric isoperimetric inequalities; rather, we shall emphasize the derivation of the physical inequalities from the geometric ones. Surveys on isoperimetric inequalities can be found in Polya–Szegö [1], Payne [1], and Osserman [3, 4].

1. THE CO-AREA FORMULA

Let M be an n-dimensional Riemannian manifold. In *all* that follows, we let $V(\ )$ denote n-dimensional volume of submanifolds of M, and, $A(\ )$, $(n-1)$-dimensional volume.

Now let Ω be a domain in M with compact closure, and $f: \bar{\Omega} \to \mathbb{R} \in C^0(\bar{\Omega}) \cap C^\infty(\Omega)$, with $f \mid \partial\Omega = 0$. The critical values of $f \mid \Omega$ have measure zero in $F(\Omega) \subseteq \mathbb{R}$, by Sard's theorem (Narasimhan [1, p. 19 ff.]). The regular values of $f \mid \Omega$ are an open subset of $\mathbb{R}$, henceforth to be denoted by R_f. For any $t \in R_f$, $f^{-1}[t] \cap \Omega$ is an $(n-1)$-manifold, and $f^{-1}[t] \cap \bar{\Omega}$ is compact.

Let $(\alpha, \beta) \subseteq R_f$, $\mu \in (\alpha, \beta)$. Then one can construct a diffeomorphism

$$\Psi: f^{-1}[\mu] \times (\alpha, \beta) \to f^{-1}[(\alpha, \beta)]$$

for which

$$f(\Psi(q, t)) = t \tag{1}$$

for all $(q, t) \in f^{-1}[\mu] \times (\alpha, \beta)$. Indeed, let X be the restriction of $(\operatorname{grad} f)/|\operatorname{grad} f|^2$ to $f^{-1}[(\alpha, \beta)]$, Φ_t the flow determined by X, and

$$\Psi(q, t) = \Phi_{t-\mu}(q).$$

Then Ψ satisfies (1). Moreover,

$$|\partial\Psi/\partial t| = |\operatorname{grad} f|^{-1},$$

and $\partial\Psi/\partial t$ is always orthogonal to the level surface $f^{-1}[t]$, for all $t \in (\alpha, \beta)$. The Riemannian density dV on Ω is given on $f^{-1}[(\alpha, \beta)]$ by

$$dV = |\operatorname{grad} f|^{-1}\, dA_{f^{-1}[t]}\, dt, \tag{2}$$

where $dA_{f^{-1}[t]}$ is the $(n-1)$-dimensional Riemannian density on $f^{-1}[t]$.

We now let

$$\Omega(t) = \{p \in M : |f(p)| > t\},$$
$$V(t) = V(\Omega(t)),$$
$$\Gamma(t) = \{p \in M : |f(p)| = t\}.$$

When t is a regular value of $|f|$, we let dA_t denote the Riemannian $(n-1)$-density on $\Gamma(t)$, and

$$A(t) = A(\Gamma(t)).$$

One immediately has

THEOREM 1 (co-area formula). The function $V(t)$ is C^∞ on $R_{|f|}$, and its derivative is given by

$$V'(t) = -\int_{\Gamma(t)} |\operatorname{grad} f|^{-1}\, dA_t. \tag{3}$$

For any $h \in L^1(\Omega)$ we have

$$\iint_\Omega h|\operatorname{grad} f|\, dV = \int_0^\infty dt \int_{\Gamma(t)} h\, dA_t. \tag{4}$$

In particular,

$$\iint_\Omega |\operatorname{grad} f|\, dV = \int_0^\infty A(t)\, dt. \tag{5}$$

2. THE FABER–KRAHN INEQUALITY

We are given a complete n-dimensional, $n \geq 2$, Riemannian manifold M, and for a fixed $\kappa \in \mathbb{R}$, the complete, simply connected, n-dimensional space form $\mathbb{M}_\kappa$ of constant sectional curvature κ.

To each open set Ω, consisting of a finite disjoint union of regular domains in M, associate the geodesic disk D in $\mathbb{M}_\kappa$ satisfying

$$V(\Omega) = V(\mathsf{D}). \tag{6}$$

If $\kappa > 0$ then only consider those Ω for which $V(\Omega) < V(\mathbb{M}_\kappa)$.

THEOREM 2 (Faber [1]; Krahn [1]). If, for all such Ω in M, equality (6) implies the isoperimetric inequality

$$A(\partial\Omega) \geq A(\partial\mathsf{D}), \tag{7}$$

with equality in (7) if and only if Ω is isometric to D, then we also have, for every normal domain Ω in M, that equality (6) implies the inequality

$$\lambda(\Omega) \geq \lambda(\mathsf{D}), \tag{8}$$

with equality in (8) if and only if Ω is isometric to D.

PROOF: Let f be an eigenfunction of $\lambda(\Omega)$, that is,

$$\Delta f + \lambda(\Omega) f = 0$$

on Ω, with $f \mid \partial\Omega = 0$. Since f never vanishes on Ω, we may assume $f \mid \Omega > 0$.

We first note that the function $V = V(t)$ is in $C^0([0, \infty)) \cap C^\infty(R_f)$. Indeed, we need only consider what happens when we are given $p \in \Omega$ for which $(\operatorname{grad} f)(p) = 0$. Then there exists a chart (any Riemann normal coordinates about p will do (cf. Section XII.8)) $x\colon U(p) \to \mathbb{R}^n$ for which

$$(\partial f^2/\partial x^{j^2})(p) < 0$$

for at least one $j \in \{1, 2, \ldots, n\}$. Thus there exists a neighborhood W of p such that if $q \in W$ and $(\operatorname{grad} f)(q) = 0$, then q is in the locus of

$$\partial f/\partial x^j = 0, \qquad \partial^2 f/\partial x^{j^2} < 0.$$

In particular, q is contained in an $(n-1)$-manifold. Thus $V(\Gamma(t)) = 0$ for *all* t, and the continuity of $V = V(t)$ follows.

The idea of the proof of (8) is to associate to f a function $F\colon \bar{\mathsf{D}} \to \mathbb{R}$ for which

$$\iint_\Omega |\operatorname{grad} f|^2 \, dV \geq \iint_{\mathsf{D}} |\operatorname{grad} F|^2 \, dV_\kappa, \tag{9}$$

$$\iint_\Omega f^2 \, dV = \iint_{\mathsf{D}} F^2 \, dV_\kappa, \tag{10}$$

where by dV_κ we mean the Riemann density of $\mathbb{M}_\kappa$, and for which $F \mid \partial\mathsf{D} = 0$.

To this end, let $\mathsf{B}_\kappa(\delta)$ be the geodesic disk, of radius δ, in $\mathbb{M}_\kappa$, with n-volume $V_\kappa(\delta)$, and $\mathsf{S}_\kappa(\delta) = \partial\mathsf{B}_\kappa(\delta)$, with $(n-1)$-volume $A_\kappa(\delta)$. For $T = \max f\,|\,\Omega$, and $0 \le t \le T$, we let $\mathsf{D}(t)$ be the geodesic disk in $\mathbb{M}_\kappa$ with volume equal to $V(t)$, namely, if $r(t)$ is the radius of $\mathsf{D}(t)$, then

$$V(t) = V_\kappa(r(t)).$$

Also, set $r_0 = r(0)$; in particular $\mathsf{D} = \mathsf{B}_\kappa(r_0)$.

The function $r\colon [0, T] \to [0, r_0]$ is in $C^0([0, T]) \cap C^\infty(R_f \cap (0, T))$, and is strictly decreasing. We let $\psi\colon [0, r_0] \to [0, T]$ be the inverse function of r, and define $F\colon \bar{\mathsf{D}} \to \mathbb{R}$ by

$$F = \psi \circ r,$$

where r denotes distance from the center of D.

To verify (10) we note that

$$\begin{aligned} V'(t) &= V_\kappa'(r(t))r'(t) = A_\kappa(r(t))r'(t),\\ 1 &= \psi'(r(t))r'(t), \end{aligned}$$

for all $t \in R_f$. Therefore

$$\begin{aligned} \iint_{\mathsf{D}} F^2\, dV_\kappa &= \int_0^{r_0} \psi^2(r)A_\kappa(r)\, dr\\ &= -\int_0^T \psi^2(r(t))A_\kappa(r(t))r'(t)\, dt\\ &= -\int_0^T t^2 V'(t)\, dt = \iint_\Omega f^2\, dV, \end{aligned}$$

which is (10).

To prove (9) note that, for $t \in R_f$, we have

$$\begin{aligned} r'(t) &= V'(t)/A_\kappa(r(t))\\ &= \frac{-1}{A_\kappa(r(t))} \int_{\Gamma(t)} \frac{dA_t}{|\operatorname{grad} f|}. \end{aligned}$$

Set

$$p(t) = \begin{cases} \displaystyle\int_{\Gamma(t)} |\operatorname{grad} f|\, dA_t, & t \in R_f,\\ 0, & \text{otherwise}. \end{cases}$$

The Cauchy–Schwarz inequality and (7) then imply

$$-r'(t)p(t) \ge A^2(t)/A_\kappa(r(t)) \ge A_\kappa(r(t)).$$

From the co-area formula we now have

$$\begin{aligned}\iint_\Omega |\mathrm{grad}\, f|^2\, dV &= \int_0^T p(t)\, dt\\ &\geq -\int_0^T \{A_\kappa(r(t))/r'(t)\}\, dt\\ &= -\int_0^T \psi'^2(r(t))A_\kappa(r(t))r'(t)\, dt\\ &= \iint_{\mathsf{D}} |\mathrm{grad}\, F|^2\, dV_\kappa,\end{aligned}$$

which implies (9).

If $\lambda(\Omega) = \lambda(\mathsf{D})$, then

$$A(t) = A_\kappa(r(t))$$

for all $t \in R_f$. But then for every $t \in R_f$, $\Omega(t)$ is isometric to D(t), and Sard's theorem implies that $\Omega = \Omega(0)$ is isometric to $\mathsf{D}(0) = \mathsf{D}$.

So Theorem 2 states that the geometric isoperimetric inequality

$$(*) \qquad V(\Omega) = V(\mathsf{D}) \;\Rightarrow\; A(\partial\Omega) \geq A(\partial\mathsf{D})$$

for all Ω consisting of a finite disjoint union of regular domains implies the physical isoperimetric inequality

$$(**) \qquad V(\Omega) = V(\mathsf{D}) \;\Rightarrow\; \lambda(\Omega) \geq \lambda(\mathsf{D})$$

for all normal domains Ω. (For normal domains Ω in $\mathbb{R}^2$, $\kappa = 0$, the result (**) was first conjectured by Lord Rayleigh [1, Section 210].) We now discuss the validity of (*).

The geometric isoperimetric inequality (*) is valid when $M = \mathbb{M}_\kappa$, for all $\kappa \in \mathbb{R}$ and $n \geq 2$. Compare, for example, Schmidt [1] and the survey by Osserman [3].

If M is 2-dimensional with Gaussian curvature less than or equal to κ, then (*) is valid whenever Ω is a finite disjoint union of regular domains, each of which is diffeomorphic to a 2-disk. Compare Fiala [1], Bol [1], Chavel–Feldman [5], and Osserman [4]. The Faber-Krahn inequality (**) is then valid for normal domains Ω, in M, which are diffeomorphic to a 2-disk. Indeed, all one has to note is that if Ω is diffeomorphic to a 2-disk, then, for any $t \in R_f$, $\Omega(t)$ has a finite number of components each of which is diffeomorphic to a 2-disk. The argument for (**) is, then, as presented here.

Recently, C. B. Croke [5] has shown that (*) is valid when M has nonpositive sectional curvature, dim $M = 4$, and $\mathbb{M}_\kappa = \mathbb{R}^4$. Again, we have the associated Faber–Krahn inequality (**).

Remark 1: We note that the hypothesis that Ω is a normal domain was only used to guarantee the existence of an eigenfunction for $\lambda^*(\Omega) = \lambda(\Omega)$. Therefore, for example, if Ω is a nodal domain of an eigenvalue λ of M, where M is either a compact manifold (with closed eigenvalue problem) or a regular domain (with Dirichlet eigenvalue problem), then the argument (assuming the geometric isoperimetric inequality holds for finite disjoint unions of regular domains) remains valid, and the theorem holds for $\lambda^*(\Omega)$ replacing $\lambda(\Omega)$ in (8).

Now the fact that f is an eigenfunction, is used only to guarantee the continuity of $V(t)$ even when grad f is zero. In particular, if Ω is any domain with compact closure in M, then for any $f \in \mathscr{H}_0(\Omega)$ for which $V(t)$ is continuous, one has the symmetrization F of f as described, for which the Rayleigh quotient associated to F is less than or equal to the Rayleigh quotient associated to f.

Therefore, if M has the property that the geometric isoperimetric inequality (relative to $\mathbb{M}_\kappa$) holds for finite unions of regular domains in M, then the physical isoperimetric inequality

$$\lambda^*(\Omega) \geq \lambda(D) \tag{8'}$$

holds for all domains Ω with compact closure in M satisfying (6). Indeed, one can construct a sequence f_k of continuous functions in $\mathscr{H}_0(\Omega)$, with compact support E_k in Ω, having the properties that E_k is a C^∞ $(n-1)$-submanifold of Ω, $f_k \in C^\infty(E_k)$, f has only nondegenerate critical points on E_k, and finally,

$$D[f_k, f_k]/\|f_k\|^2 \to \lambda^*(\Omega).$$

(Compare Aubin [2, p. 40].) By applying the above symmetrization argument, one easily shows (8′). Note, however, that in this situation the characterization of equality in (8′) is not at all obvious.

Remark 2: Also, it has recently been noted that one can obtain the Obata theorem (Section III.4) via the Faber–Krahn argument! namely, let M be a compact, n-dimensional, Riemannian manifold whose Ricci curvature satisfies, for a given constant $\kappa > 0$,

$$\mathrm{Ric}(\xi, \xi) \geq \kappa(n-1)|\xi|^2$$

for all $\xi \in TM$. Let

$$\beta = V(M)/V(\mathbb{M}_\kappa).$$

(Of course, $\beta \leq 1$.) Given any Ω consisting of a disjoint union of a finite number of normal domains in M, let D be the disk in $\mathbb{M}_\kappa$ for which

$$V(\Omega) = \beta V(\mathsf{D}). \tag{11}$$

Then Gromov [1] proved (cf. our sketch of the argument in Section XII.8) that

$$A(\partial\Omega) \geq \beta A(\partial\mathsf{D}),$$

with equality if and only if M is isometric to $\mathbb{M}_\kappa$, and Ω is isometric to D. P. Berard and D. Meyer [1] then gave the appropriate variation of the argument of Theorem 2 (viz., $r(t)$ is defined by $V(t) = \beta V_\kappa(r(t))$) to show that for any normal domain Ω in M, equality (11) implies

$$\lambda(\Omega) \geq \lambda(\mathsf{D}),$$

with equality if and only if M is isometric to $\mathbb{M}_\kappa$, and Ω is isometric to D. The same result holds for Ω a nodal domain of an eigenvalue of M with $\lambda^*(\Omega)$ replacing $\lambda(\Omega)$.

To obtain the Obata theorem, let ϕ be an eigenfunction of $\lambda(M)$. Then ϕ has precisely two nodal domains Ω_1, Ω_2. We assume that $V(\Omega_1) \leq V(\Omega_2)$. If D is the disk in $\mathbb{M}_\kappa$ for which $V(\Omega_1) = \beta V(D)$, then D is contained in a hemisphere of $\mathbb{M}_\kappa$, and

$$\lambda(M) = \lambda(\Omega_1) \geq \lambda(\mathsf{D}) \geq n\kappa$$

(cf. (II.43)), which is the Lichnerowicz inequality. If $\lambda(M) = n\kappa$, then M must be isometric to $\mathbb{M}_\kappa$, which is Obata's result. Thus, the Obata theorem is a consequence of the Levy–Gromov isoperimetric inequality and the Faber–Krahn method.

Remark 3: We note that C. Croke [3] has sharpened the Levy–Gromov inequality as follows: Let M, $\kappa > 0$, β, Ω, and D be as in Remark 2. If the diameter $d(M) = \pi/\sqrt{\kappa}$, then $M = \mathbb{M}_\kappa$; so assume that we are given a constant α for which $\sqrt{\kappa}d(M) \leq \alpha$. Then there exists a positive constant $c(n, \alpha) > 1$ such that (11) implies

$$A(\partial\Omega) \geq c(n, \alpha)\beta A(\partial D).$$

The Bérard–Meyer variant of the Faber–Krahn argument then yields

$$\lambda(M) \geq c(n, \alpha)n\kappa.$$

Croke then notes that if we are given the hypotheses of the Obata theorem, and we are given $\lambda(M) \leq c(n, \pi/2)n\kappa$ then we conclude that $d(M) > \pi/2\sqrt{\kappa}$, which implies, by a theorem of Grove–Shiohama [1], that

M is homeomorphic to a sphere. Thus one obtains an *eigenvalue pinching theorem* (Croke [3]): Given an n-dimensional, $n \geq 2$, compact Riemannian manifold satisfying (III.40) for a positive constant κ, on all of TM, then there exists a constant $c(n) > 1$ such that

$$n\kappa \leq \lambda(M) \leq c(n) n\kappa$$

implies that M is homeomorphic to a sphere (cf. Croke's paper for references to earlier versions of the result).

Remark 4: We return to the Faber–Krahn inequality for $M = \mathbb{M}_\kappa = \mathbb{R}^n$. Recall that D is a disk of radius r_0, and

$$V(\Omega) = V(\mathsf{D}) = \omega_n r_0^n,$$

$$\lambda(\Omega) \geq \lambda(\mathsf{D}) = (j_{n/2-1})^2/r_0^2,$$

where $j_{n/2-1}$ is the first zero of the Bessel function $J_{n/2-1}$ (cf. Theorem II.4 ff). So the Faber–Krahn inequality may be written analytically, in $\mathbb{R}^n$, as

$$\{\lambda(\Omega)\}^{n/2} V(\Omega) \geq \{j_{n/2-1}\}^n \omega_n \tag{12}$$

for all normal domains Ω, in $\mathbb{R}^n$, with equality if and only if Ω is a disk.

An immediate consequence of (12) and Pleijel's theorem (Section I.5) is that if Ω is a regular domain in $\mathbb{R}^n$, $n \geq 2$, with Dirichlet eigenvalues $0 < \lambda_1(\Omega) < \lambda_2(\Omega) \leq \lambda_3(\Omega) \leq \cdots$, $\{\phi_1, \phi_2, \phi_3, \ldots\}$ a complete orthonormal basis of $L^2(\Omega)$ with ϕ_k an eigenfunction of λ_k for all $k = 1, 2, \ldots$, and n_k is the number of nodal domains of ϕ_k, then

$$\limsup_{k \to \infty} \frac{n_k}{k} \leq \frac{\{2\pi/j_{n/2-1}\}^n}{\omega_n^2}. \tag{13}$$

But $\{2\pi/j_{n/2-1}\}^n/\omega_n^2$ is less than 1 (cf. Bérard–Meyer [1, Lemma 9]). Therefore, for the Dirichlet eigenvalue problem on regular domains in $\mathbb{R}^n$, we achieve equality in Courant's nodal theorem (Section I.5) for at most a finite number of eigenvalues of any given regular domain (Pleijel [1]).

Of course, if $n = 1$ then $n_k = k$ for all k, by (I.53), (I.54).

To return to $n \geq 2$, Bérard–Meyer [1] have extended the validity of (13) to any regular domain Ω in any n-dimensional, $n \geq 2$, Riemannian manifold M. First they establish the following geometric isoperimetric inequality: Given M with compact closure, and $\varepsilon > 0$, there exist $V_0 > 0$, depending on M, ε, such that for any Ω consisting of a disjoint finite union of regular domains satisfying

$$V(\Omega) \leq V_0,$$

then for the disk D, in $\mathbb{R}^n$, satisfying (6) we have (instead of (7)),

$$A(\partial\Omega) \geq (1 - \varepsilon)A(\partial\mathsf{D}).$$

The Faber–Krahn argument yields

$$\lambda(\Omega) \geq (1 - \varepsilon)^2\lambda(\mathsf{D}),$$

that is,

$$\{\lambda(\Omega)\}^{n/2}V(\Omega) \geq (1 - \varepsilon)^n\{j_{n/2-1}\}^n\omega_n$$

when Ω is any normal domain in M or a nodal domain of a Dirichlet eigenvalue or a regular domain in M containing Ω.

Now let Ω be any regular domain in M. Given $\varepsilon > 0$ (if M does not have compact closure, fix a domain, with compact closure, containing $\bar{\Omega}$), determine V_0, and let $l = [V(\Omega)/V_0] + 1$. If $\bar{n}_k$ are the number of nodal domains, of any eigenfunction of $\lambda_k(\Omega)$, with volume less than or equal to V_0, then

$$n_k - l \leq \bar{n}_k \leq n_k.$$

If $\Omega_1, \ldots, \Omega_{\bar{n}_k}$ is a list of the nodal domains with volume less than or equal to V_0, then, as in the proof of Pleijel's theorem (Section I.5),

$$\begin{aligned}\{\lambda_k(\Omega)\}^{n/2}V(\Omega) &\geq \{\lambda_k(\Omega)\}^{n/2}\sum_{j=1}^{\bar{n}_k} V(\Omega_j)\\ &= \sum_{j=1}^{\bar{n}_k}\{\lambda(\Omega_j)\}^{n/2}V(\Omega_j)\\ &\geq (n_k - l)(1 - \varepsilon)^n\omega_n\{j_{n/2-1}\}^n.\end{aligned}$$

Weyl's formula (I.50) implies

$$\limsup_{k\to\infty}(n_k/k) \leq (1 - \varepsilon)^{-n}\{2\pi/j_{n/2-1}\}/\omega_n^2$$

for all $\varepsilon > 0$—and (13) follows.

Thus, for M, with $\dim M \geq 2$, and for the Dirichlet eigenvalue problem on any given normal domain in M, equality can occur in Courant's nodal domain theorem (Section I.5) for at most a finite number of eigenvalues.

Remark 5: The Faber–Krahn inequality for $M = \mathbb{M}_\kappa = \mathbb{S}^2$ was applied by Barbosa–do Carmo [1] to the study of stability of minimal surfaces in $\mathbb{R}^3$. Compare Osserman [5] for a survey of this and related results.

Remark 6: We note that for a domain Ω in $\mathbb{R}^n$, the Neumann isoperimetric inequality goes the other way, namely,

$$(***) \qquad V(\Omega) = V(D) \Rightarrow \mu(\Omega) \leq \mu(D),$$

with equality if and only if Ω is isometric to D. The result (***) was first conjectured and investigated in Kornhauser–Stakgold [1], and proved for simply connected domains in $\mathbb{R}^2$ by Szegö [1]. A proof of the result for arbitrary domains in $\mathbb{R}^n$, $n \geq 2$, was first given in Weinberger [1]. His argument is easily adaptable to $\mathbb{M}_\kappa$, $\kappa < 0$, but the result in $\mathbb{S}^n$ (even for $n = 2$) has yet to be forthcoming. Compare some additional discussion in Chavel [3].

Remark 7: For compact M, isoperimetric results are available in the 2-dimensional case alone, namely, for any 2-dimensional Riemannian manifold M homeomorphic to $\mathbb{S}^2$ we have (Hersch [1])

$$\lambda(M)A(M) \leq 8\pi$$

(where $A(M)$ is the 2-dimensional area of M), with equality if and only if M is isometric to a constant curvature metric on $\mathbb{S}^2$. P. C. Yang and S. T. Yau [1] then proved that for any compact orientable Riemannian two-manifold of genus $g(M)$ we have

$$\lambda(M)A(M) \leq 8\pi(1 + g(M)).$$

(Both of the above results are actually consequences of lower bounds on $\sum_{j=1}^{3} \lambda_j^{-1}$.) In Li–Yau [2], another (and potentially more fruitful) argument was given for the above results, which also yielded

$$\lambda(M)A(M) \leq 12\pi$$

for M homeomorphic to $\mathbb{P}^2$, the real projective plane (a result obtained via a different argument by J. P. Bourguignon).

A consequence of the original Hersch estimate is the following result of S. Y. Cheng [3]: Suppose M is homeomorphic to $\mathbb{S}^2$, and ϕ_1, ϕ_2, ϕ_3 are three first eigenfunctions such that the sum of their squares is a constant function on M. Then M is actually isometric to a sphere of constant Gauss curvature. (Compare Cheng's paper for the nifty argument.)

For $n \geq 2$, one would naturally seek an upper bound on $\lambda^{n/2}V$ depending on, at most, the topology of M. H. Urakawa [1] has shown that no such upper bound exists when M is diffeomorphic to $\mathbb{S}^3$ (cf. also Bérard Bergery–Bourguignon [1]).

3. THE CHEEGER, SOBOLEV, AND ISOPERIMETRIC CONSTANTS

We let M be a noncompact Riemannian manifold, of dimension $n \geq 2$, possibly having nonempty boundary, and possibly having compact closure.

DEFINITION 1. The *Cheeger constant of* M, $\mathfrak{h}(M)$, is defined by

$$\mathfrak{h}(M) = \inf_{\Omega} \frac{A(\partial\Omega)}{V(\Omega)},$$

where Ω ranges over all open submanifolds of M, with compact closure in M, and smooth boundary.

THEOREM 3 (Cheeger [1]). For any normal domain Ω in M, we have

$$\lambda(\Omega) \geq \mathfrak{h}^2(\Omega)/4. \tag{14}$$

PROOF: Let u be an eigenfunction of $\lambda(\Omega)$, that is,

$$\Delta u + \lambda(\Omega)u = 0, \qquad u \mid \partial\Omega = 0.$$

Then

$$\operatorname{grad} u^2 = 2u(\operatorname{grad} u),$$

which implies, by the Cauchy–Schwarz inequality,

$$\lambda(\Omega) \geq \|\operatorname{grad} u\|^2/\|u\|^2 \geq \tfrac{1}{4}\left\{\iint_\Omega |\operatorname{grad} u^2|\,dV \Big/ \iint_\Omega u^2\,dV\right\}^2.$$

To estimate the last quotient from below, we use the co-area formula (with its notation) for $f = u^2$. Therefore,

$$\begin{aligned}\iint_\Omega |\operatorname{grad} u^2|\,dV &= \int_0^\infty A(t)\,dt \\ &\geq \mathfrak{h}(\Omega)\int_0^\infty V(t)\,dt \\ &= -\mathfrak{h}(\Omega)\int_0^\infty tV'(t)\,dt \\ &= \mathfrak{h}(\Omega)\iint_\Omega u^2\,dV.\end{aligned}$$

(One uses integration-by-parts to pass from the second line to the third), and the theorem follows.

An easy consequence of Cheeger's inequality is McKean's inequality (Theorem III.4), namely, if M is a complete, simply connected, n-dimensional Riemannian manifold, all of whose sectional curvatures are less than or equal to $\kappa < 0$, then for all normal domains Ω in M we have

$$\lambda(\Omega) \geq -(n-1)^2\kappa/4.$$

Indeed, given Ω, fix $p \notin \Omega$, and construct geodesic spherical coordinates (t, ξ) on M, based at p, as described at the end of Section III.1. Let $\sqrt{\mathfrak{g}}(t; \xi)$ be as in Definition III.6. We think of t as a function on M, namely,

$$t(q) = d(p, q).$$

Then explicit calculation, and Bishop's comparison theorem (I), imply

$$\Delta t = (\partial_t\sqrt{\mathfrak{g}})/\sqrt{\mathfrak{g}} \geq (n-1)\sqrt{-\kappa}\coth\sqrt{-\kappa}\,t \geq (n-1)\sqrt{-\kappa}.$$

Therefore, for all Ω', with $\overline{\Omega'} \subseteq \Omega$, we have

$$(n-1)\sqrt{-\kappa}\,V(\Omega') \leq \iint_{\Omega'} \Delta t\, dV = \int_{\partial\Omega'} \langle \operatorname{grad} t, \nu\rangle\, dA \leq A(\partial\Omega'),$$

which implies

$$\mathfrak{h}(\Omega) \geq (n-1)\sqrt{-\kappa}. \tag{15}$$

Cheeger's inequality then implies McKean's inequality (Yau [1]).

Inequality (15) actually implies that Cheeger's inequality is sharp. If M, just considered, has constant sectional curvature equal to -1, with geodesic disk $\mathsf{B}_{-1}(\delta)$, of radius δ, having lowest Dirichlet eigenvalue $\lambda_{-1}(\delta)$, then

$$(n-1)^2/4 \leq \mathfrak{h}^2(\mathsf{B}_{-1}(\delta))/4 \leq \lambda_{-1}(\delta) \leq (n-1)^2/4 + o(\delta)$$

as $\delta \to +\infty$, by (II.46). Our claim follows (Osserman [4]).

Interesting applications of Cheeger's inequality can be found in Brooks [1–4], Dodziuk [3, 4], Donnelly [2], and Osserman [2].

DEFINITION 2. The *Sobolev constant of* M, $\mathfrak{s}(M)$, is defined by

$$\mathfrak{s}(M) = \inf_f\left(\left\{\int_M |\operatorname{grad} f|\, dV\right\}^n \Big/ \left\{\int_M |f|^{n/(n-1)}\, dV\right\}^{n-1}\right),$$

where f varies over non-identically vanishing functions in $C_c^\infty(M)$, (henceforth), the compactly supported C^∞ functions on M.

DEFINITION 3. The *isoperimetric constant* of M, $\mathfrak{I}(M)$, is defined by

$$\mathfrak{I}(M) = \inf_\Omega \frac{\{A(\partial\Omega)\}^n}{\{V(\Omega)\}^{n-1}},$$

where Ω ranges over all open submanifolds of M, with compact closure in M, and smooth boundary.

THEOREM 4 (Federer–Fleming [1]). We always have

$$\mathfrak{s}(M) = \mathfrak{I}(M). \tag{16}$$

PROOF: Let Ω be an open submanifold of M, with compact closure in M, and smooth boundary; and for sufficiently small $\varepsilon > 0$, define

$$f_\varepsilon(p) = \begin{cases} 1, & p \in \Omega, \quad d(p, \partial\Omega) \geq \varepsilon, \\ (1/\varepsilon) d(p, \partial\Omega), & p \in \Omega, \quad d(p, \partial\Omega) < \varepsilon, \\ 0, & p \in M \backslash \Omega. \end{cases}$$

Then for each ε we have (by approximating f_ε with suitable functions in $C_c^\infty(M)$),

$$\mathfrak{s}(M) \leq \left\{\int_M |\operatorname{grad} f_\varepsilon| \, dV\right\}^n \bigg/ \left\{\int_M |f_\varepsilon|^{n/(n-1)} \, dV\right\}^{n-1}.$$

One, now, easily has

$$\lim_{\varepsilon \to 0} \int_M |f_\varepsilon|^{n/(n-1)} \, dV = V(\Omega).$$

Also,

$$|\operatorname{grad} f_\varepsilon| = \begin{cases} 1/\varepsilon, & p \in \Omega, \quad d(p, \partial\Omega) < \varepsilon, \\ 0, & \text{otherwise,} \end{cases}$$

which implies

$$\lim_{\varepsilon \to 0} \int_M |\operatorname{grad} f_\varepsilon| \, dV = \lim_{\varepsilon \to 0} (1/\varepsilon) V(\Omega \cap \{p : d(p, \partial\Omega) < \varepsilon\}) = A(\partial\Omega).$$

Thus,

$$\mathfrak{s}(M) \leq \{A(\partial\Omega)\}^n / \{V(\Omega)\}^{n-1}$$

for all such Ω, from which we conclude

$$\mathfrak{s}(M) \leq \mathfrak{I}(M).$$

So it remains to show that

$$\left\{\int_M |\operatorname{grad} f| \, dV\right\}^n \geq \mathfrak{I}(M) \left\{\int_M |f|^{n/(n-1)} \, dV\right\}^{n-1} \tag{17}$$

for all $f \in C_c^\infty(M)$.

We fix $f \in C_c^\infty(M)$, and apply the co-area formula using the notation of Section 1. Then

$$\int_M |\operatorname{grad} f|\, dV = \int_0^\infty A(t)\, dt \geq \{\mathfrak{I}(M)\}^{1/n} \int_0^\infty \{V(t)\}^{(n-1)/n}\, dt,$$

and

$$\begin{aligned} \int_M |f|^{n/(n-1)}\, dV &= \int_M dV \int_0^{|f|} \{n/(n-1)\} t^{1/(n-1)}\, dt \\ &= \{n/(n-1)\} \int_0^\infty t^{1/(n-1)}\, dt \int_{\Omega(t)} dV \\ &= \{n/(n-1)\} \int_0^\infty t^{1/(n-1)} V(t)\, dt. \end{aligned}$$

So, to prove (17) it suffices to establish

$$\left\{ \int_0^\infty [V(t)]^{(n-1)/n}\, dt \right\}^{n/(n-1)} \geq \frac{n}{n-1} \int_0^\infty t^{1/(n-1)} V(t)\, dt. \tag{18}$$

To establish (18), set

$$F(x) = \left\{ \int_0^x [V(t)]^{(n-1)/n}\, dt \right\}^{n/(n-1)},$$

$$G(x) = \frac{n}{n-1} \int_0^x t^{1/(n-1)} V(t)\, dt,$$

and note that

$$F(0) = G(0);$$

also, since $V(x)$ is a decreasing function of x, we have

$$\begin{aligned} F'(x) &= \frac{n}{n-1} \left\{ \int_0^x \{V(t)\}^{(n-1)/n}\, dt \right\}^{1/(n-1)} \{V(x)\}^{(n-1)/n} \\ &\geq \frac{n}{n-1} x^{1/(n-1)} V(x) = G'(x). \end{aligned}$$

Thus (18) follows, and, with it, the theorem.

THEOREM 5 (Yau [1]). In the definitions of $\mathfrak{h}(M)$ (Definition 1), and $\mathfrak{I}(M)$ (Definition 3), it suffices to let Ω range over open submanifolds of M which are connected.

PROOF: We present the proof for $\mathfrak{I}(M)$. Let

$$\gamma_1 = \inf_{\Omega} \frac{\{A(\partial\Omega)\}^n}{\{V(\Omega)\}^{n-1}},$$

where Ω ranges over connected, open submanifolds of M with compact closure in M. Obviously, $\mathfrak{I} \leq \gamma_1$.

Now let Ω be any open submanifold of M, with compact closure in M. We wish to show

$$\{A(\partial\Omega)\}^n \geq \gamma_1 \{V(\Omega)\}^{n-1}. \tag{19}$$

To this end, we write

$$\partial\Omega = \bigcup_{j=1}^{k} S_j, \tag{20}$$

where $S_1, \ldots, S_k$ are compact, connected, $(n-1)$-dimensional submanifolds in M, and verify (19) via induction on k.

If $k = 1$, then Ω is connected and (19) is valid.

Assume (19) is valid for all $k \leq k_0$, and suppose $\partial\Omega$ is given by (20) with $k = k_0 + 1$. If Ω is connected, then (19) is already valid. If not we may assume that Ω can be written as the disjoint union of open sets Ω_1, Ω_2, that is, $\Omega = \Omega_1 \cup \Omega_2$. We number the components of $\partial\Omega$ so that

$$\partial\Omega_1 = S_1 \cup \cdots \cup S_l, \qquad \partial\Omega_2 = S_{l+1} \cup \cdots \cup S_{k_0+1}.$$

Then the induction hypothesis implies

$$\left\{\sum_{j=1}^{l} A(S_j)\right\}^n \geq \gamma_1 \{V(\Omega_1)\}^{n-1}, \qquad \left\{\sum_{j=l+1}^{k_0+1} A(S_j)\right\}^n \geq \gamma_1 \{V(\Omega_2)\}^{n-1},$$

and, using Minkowski's inequality, we have

$$\begin{aligned}\{A(\partial\Omega)\}^n &\geq \gamma_1 \{[V(\Omega_1)]^{(n-1)/n} + [V(\Omega_2)]^{(n-1)/n}\}^n \\ &\geq \gamma_1 \{V(\Omega_1) + V(\Omega_2)\}^{n-1} \\ &= \gamma_1 \{V(\Omega)\}^{n-1},\end{aligned}$$

which is (19).

4. THE SOBOLEV CONSTANT AND EIGENVALUE, EIGENFUNCTION, ESTIMATES

The work described in this section is that of P. Li [1]. Although his paper, and the details contained therein, are devoted to the closed eigenvalue problem on a compact Riemannian manifold, we shall find it more convenient to present here the estimates for the Dirichlet eigenvalue problem on some n-dimensional, $n \geq 2$, normal domain M (cf. Section I.5). All functions under consideration will be in the class of admissible functions $\mathfrak{H}(M)$ of the Dirichlet eigenvalue problem, namely, they are in the completion of $C_c^\infty(M)$ relative to the inner product

$$(f, h)_1 = (f, h) + (\operatorname{grad} f, \operatorname{grad} h),$$

and associated norm

$$\|f\|_1^2 = \|f\|^2 + \|\operatorname{grad} f\|^2.$$

For these functions we have, when ϕ is some eigenfunction,

$$(21) \qquad (\Delta\phi, f) = -D[\phi, f],$$

where $D[\ ,\ ]$ denotes the Dirichlet integral.

For any $p > 0$, we use the Riemannian density dV to define the L^p-space, by declaring a measurable function f to be an element of L^p if the integral

$$\left\{\int_M |f|^p \, dV\right\}^{1/p}$$

is finite. The *L^p-norm* of f, $\|f\|_p$, is given by the above expression, when finite. As we shall only consider the case $p > 1$, here, no confusion will result in the notation for the various norms.

Hölder's inequality states that for $p, q > 1$ satisfying

$$1/p + 1/q = 1,$$

we have for $\phi \in L^p$, $\psi \in L^q$,

$$\int |\phi\psi| \leq \|\phi\|_p \|\psi\|_q$$

(we drop the M and dV from our integral expressions for the rest of this section).

Note that since any admissible function is contained in L^2, it is contained in L^p for all $p \in [1, 2]$. Since any eigenfunction ϕ is continuous on all of $\overline{M}$, it is in L^p for all $p > 1$. Its L^∞-norm $\|\phi\|_\infty$, given by

$$\|\phi\|_\infty = \sup_M |\phi|,$$

is known to satisfy

$$(22) \qquad \lim_{p \to \infty} \|\phi\|_p / V^{1/p} = \|\phi\|_\infty,$$

where $V = V(M)$.

For all $f \in \mathfrak{H}(M)$, we still have

$$\left\{\int |\operatorname{grad} f|\right\}^n \geq \mathfrak{s}\left\{\int |f|^{n/(n-1)}\right\}^{n-1}, \tag{23}$$

that is,

$$\int |\operatorname{grad} f| \geq \mathfrak{s}^{1/n}\|f\|_{n/(n-1)}, \tag{24}$$

where $\mathfrak{s} = \mathfrak{s}(M)$.

THEOREM 6. If $n = 2$, then $\mathfrak{H}(M) \subseteq L^4$, and

$$\int |\operatorname{grad} f|^2 \geq \{\mathfrak{s}/4V^{1/2}\}\|f\|_4^2 \tag{25}$$

for all $f \in \mathfrak{H}(M)$.

PROOF: Given $f \in C_c^\infty(M)$, apply (23) to the function $h = f^2$. Then

$$\begin{aligned} \mathfrak{s}\int f^4 &\leq \left\{\int |\operatorname{grad} f^2|\right\}^2 \\ &= 4\left\{\int |f|\,|\operatorname{grad} f|\right\}^2 \\ &\leq 4\int f^2 \int |\operatorname{grad} f|^2 \\ &\leq 4V^{1/2}\left\{\int f^4\right\}^{1/2}\left\{\int |\operatorname{grad} f|^2\right\}, \end{aligned}$$

and the result follows easily.

THEOREM 7. If $n > 2$, then $\mathfrak{H}(M) \subseteq L^{2n/(n-2)}$, and

$$\int |\operatorname{grad} f|^2 \geq \mathfrak{c}\|f\|_{2n/(n-2)}^2 \tag{26}$$

for all $f \in \mathfrak{H}(M)$, where

$$\mathfrak{c} = \mathfrak{s}^{2/n}\{(n-2)/2(n-1)\}^2. \tag{27}$$

Proof: If we apply (23) to $h = |f|^{2(n-1)/(n-2)}$, then we obtain

$$\mathfrak{s}\left\{\int |f|^{2n/(n-2)}\right\}^{n-1} \le \left\{\frac{2(n-1)}{n-2}\int |f|^{n/(n-2)}|\operatorname{grad} f|\right\}^{n}$$
$$\le \{2(n-1)/(n-2)\}^{n}\left\{\int |f|^{2n/(n-2)}\right\}^{n/2}\left\{\int |\operatorname{grad} f|^{2}\right\}^{n/2},$$

from which one easily obtains (26).

THEOREM 8. There exists a positive constant $c(n)$ depending only on n, such that if ϕ is an eigenfunction of the eigenvalue λ then

$$\|\phi\|_{\infty}^{2} \le c(n)\|\phi\|_{2}^{2}\begin{cases}\{\lambda/\mathfrak{c}\}^{n/2}, & n > 2,\\ \{4\lambda/\mathfrak{s}\}^{2}V, & n = 2.\end{cases} \tag{28}$$

Proof: We first note that for any $\psi \in \mathfrak{H}(M) \cap C^{2}(M)$, and $\alpha > \frac{1}{2}$, we have

$$-\int |\psi|^{2\alpha-2}\psi\Delta\psi = \int \langle \operatorname{grad}|\psi|^{2\alpha-2}\psi, \operatorname{grad}\psi\rangle$$
$$= (2\alpha - 1)\int |\psi|^{2\alpha-2}|\operatorname{grad}\psi|^{2}$$
$$= \frac{2\alpha-1}{\alpha^{2}}\int |\operatorname{grad}|\psi|^{\alpha}|^{2}.$$

So for the eigenfunction ϕ of the eigenvalue λ, we have

$$\lambda\int |\phi|^{2\alpha} = \frac{2\alpha-1}{\alpha^{2}}\int |\operatorname{grad}|\phi|^{\alpha}|^{2}. \tag{29}$$

First we consider the case $n > 2$. Then (26) and (29) imply

$$\lambda\int |\phi|^{2\alpha} \ge \frac{2\alpha-1}{\alpha^{2}}\mathfrak{c}\|\,|\phi|^{\alpha}\|_{2n/(n-2)}^{2}. \tag{30}$$

If we set

$$\beta = n/(n-2), \tag{31}$$

and let α range over the numbers β^{k}, where $k = 0, 1, 2, \ldots$, then we obtain

$$\|\phi\|_{2\beta^{k+1}} \le \|\phi\|_{2}\prod_{l=0}^{k}\left\{\frac{\lambda}{\mathfrak{c}}\frac{\beta^{2l}}{2\beta^{l}-1}\right\}^{1/2\beta^{l}}$$

for all $k = 0, 1, 2, \ldots$. By (22) we have

$$\begin{aligned}\|\phi\|_\infty &\le \|\phi\|_2 \prod_{l=0}^{\infty} \left\{\frac{\lambda}{\mathfrak{c}} \frac{\beta^{2l}}{2\beta^l - 1}\right\}^{1/2\beta^l} \\ &= \|\phi\|_2 \{\lambda/\mathfrak{c}\}^{n/4} \prod_{l=0}^{\infty} \{\beta^{2l}/(2\beta^l - 1)\}^{1/2\beta^l} \\ &\equiv: c(n)\|\phi\|_2\{\lambda/\mathfrak{c}\}^{n/4},\end{aligned}$$

which is the claim for $n > 2$.

If $n = 2$, then (25) and (29) imply

$$\lambda \int |\phi|^{2\alpha} \ge \frac{2\alpha - 1}{\alpha^2} \frac{\mathfrak{s}}{4V^{1/2}} \left\{\int |\phi|^{4\alpha}\right\}^{1/2},$$

which we rewrite as

$$\|\phi\|_{4\alpha} \le \|\phi\|_{2\alpha} \left\{\frac{4\lambda V^{1/2}}{\mathfrak{s}} \frac{\alpha^2}{2\alpha - 1}\right\}^{1/2\alpha},$$

for all $\alpha > \frac{1}{2}$. One argues as above, and obtains

$$\begin{aligned}\|\phi\|_\infty &\le \frac{\|\phi\|_2}{V^{1/2}} \prod_{l=1}^{\infty} \left\{\frac{4\lambda V}{\mathfrak{s}} \frac{4^{l-1}}{2^l - 1}\right\}^{1/2^l} \\ &= \|\phi\|_2 \{4\lambda V^{1/2}/\mathfrak{s}\} \prod_{l=1}^{\infty} \{4^{l-1}/(2^l - 1)\}^{1/2^l} \\ &\equiv: c(n)\|\phi\|_2\{4\lambda V^{1/2}/\mathfrak{s}\},\end{aligned}$$

and the result follows.

PROPOSITION 1. Let E be a finite-dimensional subspace of $L^2 \cap C^0$. Then there exists $\psi \in E$ such that

(32) $$(\dim E)\|\psi\|_2^2 \le \|\psi\|_\infty^2 V.$$

PROOF: Let $\{\psi_1, \ldots, \psi_r\}$ be an orthonormal basis of E, and set

$$F(x) = \sum_{j=1}^{r} \psi_j^2(x).$$

One easily sees that F is independent of the choice of the orthonormal basis. Also we may assume that F is not identically zero.

Let

$$\|F\|_\infty = F(x_0)$$

for some $x_0 \in M$. Define the subspace E_0 of E to consist of those functions in E vanishing at x_0. Then $\dim E_0 = \dim E - 1$, and we may assume that $\psi_1(x_0) \neq 0$. Then

$$\dim E = \sum_j \int \psi_j^2 = \int F \leq \|F\|_\infty V = \psi_1^2(x_0) V \leq \|\psi_1\|_\infty^2 V;$$

so $\psi = \psi_1$ is desired function.

COROLLARY 1. If λ is an eigenvalue, with multiplicity m_λ, then

$$m_\lambda \leq c(n) \begin{cases} \{\lambda/\mathfrak{c}\}^{n/2} V, & n > 2, \\ \{4\lambda V/\mathfrak{s}\}^2, & n = 2. \end{cases} \tag{33}$$

One might describe (33) as providing a lower bound of λ in terms of its multiplicity m_λ. We now give an adjusted version of the previous argument which will produce a lower bound for λ_k in terms of k.

LEMMA 1. If $f_1, \ldots, f_k$ are linearly independent on M, $q \geq 2$, and $0 < \lambda_1 \leq \cdots \leq \lambda_k$, then there exists a subset $S \subseteq \{1, \ldots, k\}$ such that

$$\int \left| \sum_{j=1}^k \lambda_j f_j \right|^q \leq \lambda_k^q \int \left| \sum_{s \in S} f_s \right|^q. \tag{34}$$

PROOF: Let

$$F(\lambda_1, \ldots, \lambda_k) = \int \left| \sum_{k=1}^k \lambda_j f_j \right|^q,$$

and note that

$$\frac{\partial^2 F}{\partial \lambda_j^2} = q(q-1) \int f_j^2 \left| \sum_j \lambda_j f_j \right|^{q-2} \geq 0.$$

Thus F is convex in each variable. Therefore F may be majorized by, appropriately, replacing each λ_j by 0 or λ_k. But that is the claim.

Now let $0 < \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_k$ be the first k eigenvalues of M, $\{\phi_1, \ldots, \phi_k\}$ orthonormal eigenfunctions such that ϕ_j is an eigenfunction of λ_j for each $j = 1, \ldots, k$, and let $E = \operatorname{span}\{\phi_1, \ldots, \phi_k\}$.

We assume that $n > 2$, and let $\mathfrak{c}$, β, be as defined in (27), (31), respectively. As in the beginning of the proof of Theorem 8, we have for any $\psi \in E$, $\alpha > \frac{1}{2}$,

$$-\int |\psi|^{2\alpha-2} \psi \Delta \psi \geq \frac{2\alpha - 1}{\alpha^2} \mathfrak{c} \left\{ \int |\psi|^{2\alpha n/(n-2)} \right\}^{(n-2)/n}. \tag{35}$$

Let $\alpha = 1$; then (35) implies

$$\|\psi\|^2_{2n/(n-2)} \le c^{-1}\int -\psi\Delta\psi \le \{\lambda_k/c\}\|\psi\|_2^2,$$

that is,

$$\|\psi\|_{2n/(n-2)} \le \{\lambda_k/c\}^{1/2}\|\psi\|_2. \tag{36}$$

For any fixed l in $\{0, 1, 2, \ldots\}$, let $h_l \in E$ have the property that

$$\|f\|_{2\beta^{l+1}}/\|f\|_2 \le \|h_l\|_{2\beta^{l+1}}/\|h_l\|_2 \tag{37}$$

for all $f \in E$. The function h_l can be written as

$$h_l = \sum_{j=1}^{k} \sigma_j\phi_j.$$

By setting $\alpha = \beta^l$ in (35), and using Lemma 1, we obtain

$$\begin{aligned}\left\{\int |h_l|^{2\beta^{l+1}}\right\}^{1/\beta} &\le \frac{\beta^{2l}}{2\beta^l - 1}\frac{1}{c}\int -|h_l|^{2\beta^l-2}h_l\Delta h_l \\ &\le \frac{\beta^{2l}}{2\beta^l - 1}\frac{1}{c}\left\{\int |h_l|^{2\beta^l}\right\}^{(2\beta^l-1)/2\beta^l}\left\{|\Delta h_l|^{2\beta^l}\right\}^{1/2\beta^l},\end{aligned}$$

and

$$\begin{aligned}\left\{\int |\Delta h_l|^{2\beta^l}\right\}^{1/2\beta^l} &\le \left\{\int |\Delta h_l|^{2\beta^{l+1}}\right\}^{1/2\beta^{l+1}} V^{(\beta-1)/2\beta^{l+1}} \\ &= \left\{\int \left|\sum_j \lambda_j\sigma_j\phi_j\right|^{2\beta^{l+1}}\right\}^{1/2\beta^{l+1}} V^{(\beta-1)/2\beta^{l+1}} \\ &\le \lambda_k\left\{\int \left|\sum_{s\in S} \sigma_s\phi_s\right|^{2\beta^{l+1}}\right\}^{1/2\beta^{l+1}} V^{(\beta-1)/2\beta^{l+1}} \\ &\le \lambda_k\{\|h_l\|_{2\beta^{l+1}}/\|h_l\|_2\}\left\|\sum_{s\in S}\sigma_s\phi_s\right\|_2 V^{(\beta-1)/2\beta^{l+1}} \\ &\le \lambda_k\|h_l\|_{2\beta^{l+1}}V^{(\beta-1)/2\beta^{l+1}}.\end{aligned}$$

Therefore,

$$\{\|h_l\|_{2\beta^{l+1}}\}^{2\beta^l-1} \le \frac{\lambda_k}{c}\frac{\beta^{2l}}{2\beta^l - 1}V^{(\beta-1)/2\beta^{l+1}}\{\|h_l\|_{2\beta^l}\}^{2\beta^l-1},$$

which implies

$$\frac{\|h_l\|_{2\beta^{l+1}}}{V^{1/2\beta^{l+1}}} \le \left\{\frac{\lambda_k V^{2/n}}{c}\frac{\beta^{2l}}{2\beta^l - 1}\right\}^{1/(2\beta^l-1)}\frac{\|h_l\|_{2\beta^l}}{V^{1/2\beta^l}}. \tag{38}$$

Since h_l satisfies (36), one obtains via induction

$$\frac{\|\psi\|_{2\beta^{l+1}}}{\|\psi\|_2 V^{1/2\beta^{l+1}}} \le \frac{\|h_1\|_{2\beta}}{\|h_1\|_2} V^{-1/2\beta} \prod_{r=1}^{l} \left\{\frac{\lambda_k V^{2/n}}{c} \frac{\beta^{2r}}{2\beta^r - 1}\right\}^{1/(2\beta^r-1)}$$

for all $\psi \in E$, $l = 0, 1, 2, \ldots$, which implies

$$\frac{\|\psi\|_\infty}{\|\psi\|_2} \le \frac{\|h_1\|_{2\beta}}{\|h_1\|_2} V^{-1/2\beta} \prod_{r=1}^{\infty} \left\{\frac{\lambda_k V^{2/n}}{c} \frac{\beta^{2r}}{2\beta^r - 1}\right\}^{1/(2\beta^r-1)} \tag{39}$$

for all $\psi \in E$.

Now $r \ge 1$, $\beta > 1$ imply

$$1/2\beta^r \le 1/(2\beta^r - 1) \le 1/\beta^r,$$

from which one obtains

$$1/2(\beta - 1) \le \sum_{r=1}^{\infty} 1/(2\beta^r - 1) \le 1/(\beta - 1).$$

Set

$$C = \prod_{r=1}^{\infty} \left\{\frac{\beta^{2r}}{2\beta^r - 1}\right\}^{1/(2\beta^r-1)}. \tag{40}$$

If

$$\lambda_k V^{2/n}/c \ge 1, \tag{41}$$

then (36) and (39) imply

$$\|\psi\|_\infty/\|\psi\|_2 \le C\{\lambda_k/c\}^{(n-1)/2} V^{(n-2)/2n}$$

for all $\psi \in E$, and Proposition 1 implies

$$k \le C^2\{\lambda_k V^{2/n}/c\}^{n-1}$$

for all k for which (41) is valid.

If, on the other hand,

$$\lambda_k V^{2/n}/c \le 1,$$

then (38), for $l = 0$, and (39) imply

$$\frac{\|\psi\|_\infty}{\|\psi\|_2} \le \frac{C}{V^{1/2}} \left\{\frac{\lambda_k V}{c}\right\}^{(n+2)/4};$$

and, by Proposition 1, we have

$$k \le C^2\{\lambda_k V/c\}^{(n/2)+1}.$$

When $n = 2$, (35) is replaced by

$$-\int |\psi|^{2\alpha-2}\psi\Delta\psi \geq \frac{2\alpha-1}{\alpha^2}\frac{\mathfrak{s}}{4V^{1/2}}\left\{\int |\psi|^{4\alpha}\right\}^{1/2}$$

for all $\alpha > \frac{1}{2}$. Then $\psi \in E$ implies (with $\alpha = 1$)

$$\|\psi\|_4^2 \leq \{4\lambda_k V^{1/2}/\mathfrak{s}\}\|\psi\|_2^2.$$

For all $l = 0, 1, 2, \ldots$ we let h_l be the function in E for which

$$\|f\|_{2^{l+2}}/\|f\|_2 \leq \|h_l\|_{2^{l+2}}/\|h_l\|_2$$

for all $f \in E$. An argument similar to the one given above shows that

$$\frac{\|h_l\|_{2^{l+2}}}{\|h_l\|_2 V^{1/2^{l+2}}} \leq \left\{\frac{4\lambda_k V}{\mathfrak{s}}\frac{4^l}{2^{l+1}-1}\right\}^{1/(2^{l+1}-1)} \frac{\|h_l\|_{2^{l+1}}}{\|h_l\|_2 V^{1/2^{l+1}}}$$

for every $l = 0, 1, 2, \ldots$, and

$$\frac{\|\psi\|_\infty}{\|\psi\|_2} \leq C\left\{\frac{\|h_0\|_4}{\|h_0\|_2}\frac{1}{V^{1/4}}\right\}\prod_{l=1}^{\infty}\left\{\frac{4\lambda_k V}{\mathfrak{s}}\right\}^{1/(2^{l+1}-1)}$$

for all $\psi \in E$, where, for $n = 2$,

(42) $$C = \prod_{l=1}^{\infty}\left\{\frac{4^l}{2^{l+1}-1}\right\}^{1/(2^{l+1}-1)}$$

One then concludes, using Proposition 1, and arguing as before, that

$$k \leq C^2\{4\lambda_k V/\mathfrak{s}\}^3$$

for all $k = 1, 2, 3, \ldots$.

We summarize the discussion as follows:

THEOREM 9. There exists a positive constant $C(n)$, depending only on $n = \dim M$, such that

(43) $$k \leq C(n)\begin{cases}[\lambda_k V^{2/n}/\mathfrak{c}]^{n-1}, & n > 2,\\ [4\lambda_k V/\mathfrak{s}]^3, & n = 2,\end{cases}$$

for all $k \geq C(n)$.

If $k \leq C(n)$ then (43) remains valid if $\lambda_k V^{2/n}/\mathfrak{c} > 1$ (resp., $4\lambda_k V/\mathfrak{s} > 1$) and $n > 2$ (resp., $n = 2$). Otherwise, we have

$$k \leq C(n)\begin{cases}[\lambda_k V^{2/n}/\mathfrak{c}]^{n/2+1}, & n > 2,\\ [4\lambda_k V/\mathfrak{s}]^3, & n = 2,\end{cases}$$

when $\lambda_k V^{2/n}/\mathfrak{c} < 1$ (resp., $4\lambda_k V/\mathfrak{s} < 1$), and $n > 2$ (resp., $n = 2$).

From Theorems 8 and 9 one easily derives:

THEOREM 10 (Chavel–Feldman [6]). Given positive constants τ, μ, γ, then the formal heat kernel, $q\colon \bar{M} \times \bar{M} \times (0, \infty) \to \mathbb{R}$, associated to the Dirichlet eigenvalue problem (cf. Section I.4), and given by

$$q(x, y, t) = \sum_{j=1}^{\infty} e^{-\lambda_j t}\phi_j(x)\phi_j(y)$$

(where $\phi_1, \phi_2, \ldots$ is a complete orthonormal basis of $L^2(M)$, with each ϕ_j an eigenfunction of λ_j, $j = 1, 2, \ldots$) converges uniformly on $\bar{M} \times \bar{M} \times [\tau, \infty)$, and for $V(M) \in (0, \mu)$, $\mathfrak{s}(M) \in [\gamma, \infty)$. That is, an upper bound on $V(M)$, and a lower bound on $\mathfrak{s}(M) = \mathfrak{I}(M)$ provide a uniform upper bound on the heat kernel, for times bounded away from zero.

PROOF: We give the proof for $n > 2$. For $n = 2$ the argument is similar.

Once we are given that $V(M)$ is bounded from above, and $\mathfrak{s}(M)$ from below, we have for all $t \geq T > 0$, by Theorem 8,

$$|p(x, y, t)| \leq \text{const} \sum_{k=1}^{\infty} \lambda_k^{n/2} e^{-\lambda_k T}.$$

Next, one determines $\alpha_0 > 0$ for which

$$\alpha^{5n/2-2} e^{-\alpha T} \leq 1$$

for all $\alpha \geq \alpha_0$. From Theorem 9 we have

$$\lambda_k \geq \text{const} \cdot k^{1/(n-1)}$$

for all $k \geq C(n)$. So there exists a positive integer N such that

$$\lambda_k \geq \alpha_0$$

for all $k > N$, which implies

$$\sum_{k=N+1}^{\infty} \lambda_k^{n/2} e^{-\lambda_k T} \leq \text{const} \sum_{k=N+1}^{\infty} \lambda_k^{-2(n-1)} \leq \text{const} \sum_{k=N+1}^{\infty} k^{-2}.$$

If we set

$$\alpha_1 = \sup_{\alpha > 0} \alpha^{n/2} e^{-\alpha T},$$

then

$$\sum_{k=1}^{N} \lambda_k^{n/2} e^{-\lambda_k T} \leq \alpha_1 N;$$

and the theorem is proved.

5. THE CONSTANTS AND ESTIMATES FOR THE CLOSED EIGENVALUE PROBLEM

In this section M will be a compact Riemannian manifold of dimension $n \geq 2$, and all discussion of eigenvalues and eigenfunctions will be directed to the closed eigenvalue problem on M. Recall that we list the eigenvalues by $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$.

DEFINITION 4. The *Cheeger constant of* M, $\mathfrak{h}(M)$, is defined by

$$\mathfrak{h}(M) = \inf_S \frac{A(S)}{\min\{V(M_1), V(M_2)\}},$$

where S ranges over all compact $(n-1)$-dimensional submanifolds, S, of M, which divide M into 2 open submanifolds M_1, M_2 satisfying $\partial M_1 = \partial M_2 = S$.

THEOREM 11 (Cheeger [1]). We always have

$$\lambda_1 \geq \mathfrak{h}^2/4.$$

PROOF: λ_1 has two nodal domains Ω_1, Ω_2 with, say, Ω_1 having the smaller volume. Since λ_1 is the lowest Dirichlet eigenvalue of Ω_1, we have

$$\lambda_1 \geq \mathfrak{h}^2(\Omega_1)/4 \geq \mathfrak{h}^2(M)/4,$$

which is the claim.

A detailed study of Cheeger's inequality, including its sharpness, can be found in Buser [4]. The influence of Cheeger's constant $\mathfrak{h}(M)$ on determining an upper bound for $\lambda(M)$ is studied in Buser [5], the main result being: If the Ricci curvature of the compact manifold M is bounded below by $-(n-1)\kappa^2$, $\kappa > 0$, then there exists a constant $c(n)$, depending only on n, such that

$$\lambda(M) \leq c(n)\{\mathfrak{h}(M)\kappa + \mathfrak{h}^2(M)\}.$$

Buser also notes therein, that one cannot dispense with the hypothesis on the lower bound for the curvature, for it is possible to construct examples in which $\mathfrak{h}$ is arbitrarily small with λ hardly affected, if we make no hypothesis on the behavior of the Ricci curvature. Compare Buser [3] and the examples constructed in Chapter IX.

DEFINITION 4. The *Sobolev constant of M*, $\mathfrak{s}(M)$, is defined to be the supremum of all $C > 0$ for which

$$\left\{\int_M |\mathrm{grad}\, f|\, dV\right\}^n \geq C \inf_{\alpha\in\mathbb{R}} \left\{\int_M |f-\alpha|^{n/(n-1)}\, dV\right\}^{n-1} \tag{44}$$

for all $f \in C^\infty(M)$.

In order to characterize the infimum on the right-hand side of (44), we note that, for $\alpha_0, s > 0$,

$$\int \{\operatorname{sgn}(f-\alpha_0)\}|f-\alpha_0|^s = 0 \tag{45}$$

implies

$$\int |f-\alpha_0|^{s+1} \leq \int |f-\alpha|^{s+1} \tag{46}$$

for all $\alpha \in \mathbb{R}$. Indeed, write

$$\begin{aligned} |f-\alpha_0|^{s+1} &= \{\operatorname{sgn}(f-\alpha_0)\}(f-\alpha_0)|f-\alpha_0|^s \\ &= \{\operatorname{sgn}(f-\alpha_0)\}(f-\alpha+\alpha-\alpha_0)|f-\alpha_0|^s; \end{aligned}$$

then (45) implies

$$\begin{aligned} \int |f-\alpha_0|^{s+1} &= \int \{\operatorname{sgn}(f-\alpha_0)\}(f-\alpha)|f-\alpha_0|^s \\ &\leq \int |f-\alpha|\,|f-\alpha_0|^s. \end{aligned}$$

Now use Holder's inequality with $p = s+1$, and $q = (s+1)/s$. Then

$$\int |f-\alpha_0|^{s+1} \leq \left\{\int |f-\alpha|^{s+1}\right\}^{1/(s+1)} \left\{\int |f-\alpha_0|^{s+1}\right\}^{s/(s+1)},$$

and (46) follows easily. (The argument was communicated to us by R. Sacksteder.)

DEFINITION 5. The *isoperimetric constant of M*, $\mathfrak{I}(M)$, is defined by

$$\mathfrak{I}(M) = \inf_S \frac{\{A(S)\}^n}{\{\min(V(M_1), V(M_2))\}^{n-1}},$$

where S ranges over all compact $(n-1)$-dimensional submanifolds S, of M, which divide M into 2 open submanifolds M_1, M_2 satisfying $\partial M_1 = \partial M_2 = S$.

THEOREM 12. We always have

$$\mathfrak{I}(M) \le \mathfrak{s}(M) \le 2\mathfrak{I}(M).$$

PROOF: Assume we are given S with $V(M_1) \le V(M_2)$. Let $f_\varepsilon\colon M \to \mathbb{R}$ be defined as in the beginning of the proof of Theorem 3, except that M_1, here, will replace Ω, there. Then

$$\inf_{\alpha\in\mathbb{R}} \left\{\int |f_\varepsilon - \alpha|^{n/(n-1)}\right\}^{n-1} \mathfrak{s}(M) \le \left\{\int |\operatorname{grad} f_\varepsilon|\right\}^n$$

for all $\varepsilon > 0$. If we let $\varepsilon \downarrow 0$, then we obtain

$$\begin{aligned} \{A(S)\}^n &\ge \mathfrak{s}(M) \inf_{\alpha\in\mathbb{R}} \{|1-\alpha|^{n/(n-1)}V(M_1) + |\alpha|^{n/(n-1)}V(M_2)\}^{n-1} \\ &\ge \mathfrak{s}(M)\{V(M_1)\}^{n-1} \inf_{\alpha\in\mathbb{R}} \{|1-\alpha|^{n/(n-1)} + |\alpha|^{n/(n-1)}\}^{n-1} \\ &\ge \mathfrak{s}(M)\{V(M_1)\}^{n-1}/2, \end{aligned}$$

which implies $\mathfrak{s}(M) \le 2\mathfrak{I}(M)$.

To prove $\mathfrak{I}(M) \le \mathfrak{s}(M)$, we first note that given $f \in C^\infty$, there exists $\alpha_0 \in \mathbb{R}$ such that for

$$M_1 = \{x : f(x) < \alpha_0\}, \qquad M_2 = \{x : f(x) > \alpha_0\},$$

we have, for $j = 1, 2$,

$$V(M_j) \le V(M)/2.$$

Apply the second half of the proof of Theorem 3 to $h = f - \alpha_0$. Then

$$\begin{aligned} \int_{M_j} |\operatorname{grad} f| &\ge \{\mathfrak{I}(M_j)\}^{1/n}\left\{\int_{M_j} |f-\alpha_0|^{n/(n-1)}\right\}^{(n-1)/n} \\ &\ge \{\mathfrak{I}(M)\}^{1/n}\left\{\int_{M_j} |f-\alpha_0|^{n/(n-1)}\right\}^{(n-1)/n} \end{aligned}$$

for $j = 1, 2$—remember: $\mathfrak{I}(M_j)$ is given by Definition 3, and $\mathfrak{I}(M)$ by Definition 5. Thus if we let χ_j denote the characteristic function of M_j, $j = 1, 2$, then

$$\begin{aligned} \int_M |\operatorname{grad} f| &= \sum_j \int_M |\operatorname{grad} \chi_j f| \\ &\ge \{\mathfrak{I}(M)\}^{1/n} \sum_j \left\{\int_M \chi_j |f-\alpha_0|^{n/(n-1)}\right\}^{(n-1)/n} \\ &\ge \{\mathfrak{I}(M)\}^{1/n}\left\{\int_M |f-\alpha_0|^{n/(n-1)}\right\}^{(n-1)/n} \\ &\ge \{\mathfrak{I}(M)\}^{1/n} \inf_{\alpha\in\mathbb{R}}\left\{\int_M |f-\alpha_0|^{n/(n-1)}\right\}^{(n-1)/n}; \end{aligned}$$

one goes from the third line to the fourth via Minkowski's inequality. Then $\mathfrak{s}(M) \ge \mathfrak{I}(M)$ follows immediately.

THEOREM 13 (Yau [1]). In the definitions of $\mathfrak{h}(M)$ and $\mathfrak{I}(M)$ for the closed eigenvalue problem, it suffices to let S range over compact, $(n-1)$-dimensional submanifolds of M for which M_1, M_2 are connected.

THEOREM 14 (Li [1]). Let M be a compact n-dimensional Riemannian manifold, with volume V, and constants $\mathfrak{s}, \mathfrak{I}, \mathfrak{c}$ given by Definition 4, Definition 5, and (27), respectively. Then there exists a constant $C(n)$, depending only on n, such that

$$\|\phi\|_\infty^2 \le C(n)\|\phi\|_2^2 \begin{cases} (\lambda/\mathfrak{c})^{n/2}, & n > 2, \\ (4\lambda/\mathfrak{s})^2 V, & n = 2, \end{cases}$$

for any eigenfunction on M having eigenvalue equal to λ.

Also there exists a constant $C(n) > 0$, depending only on n, such that

$$k \le C(n) \begin{cases} (\lambda_k V^{2/n}/\mathfrak{c})^{n-1}, & n > 2, \\ (4\lambda_k V/\mathfrak{s})^3, & n = 3, \end{cases}$$

for all eigenvalues $\lambda_1, \lambda_2, \ldots$.

THEOREM 15 (Chavel–Feldman [6]). Given positive constants τ, μ, γ, then the formal heat kernel $p\colon M \times M \times (0, \infty) \to \mathbb{R}$ associated to the closed eigenvalue problem of a compact Riemannian manifold M (cf. Section I.4), and given by

$$p(x, y, t) = \sum_{j=0} e^{-\lambda_j t}\phi_j(x)\phi_j(y)$$

(where ϕ_1, $\phi_2, \ldots$ is a complete orthonormal basis of $L^2(M)$, with each ϕ_j an eigenfunction of λ_j, $j = 0, 1, 2, \ldots$) converges uniformly on $\bar{M} \times \bar{M} \times [\tau, \infty)$, and for $V(M) \in (0, \mu]$, $\mathfrak{s}(M) \in [\gamma, \infty)$. That is, an upper bound on $V(M)$, and a lower bound on $\mathfrak{s}(M)$ (equivalent to the isopermetric constant, $\mathfrak{I}(M)$) provide a uniform upper bound on the heat kernel for times bounded away from zero.

The arguments of Theorems 13, 14, 15 are similar to those of Theorems 5, 8, 9, and 10, respectively.

CHAPTER V

Eigenvalues and the Kinematic Measure

In this chapter we present the methods of C. B. Croke [1] in estimating the Cheeger and Sobolev constants from below, via the kinematic measure. His arguments are based, in turn, on those of M. Berger and J. Kazdan in their contributions to solving the Blaschke conjecture in higher dimensions (compare Besse [1] and Berger–Kazdan [1]).

Applications of Croke's estimates, to estimating the heat kernel of noncompact Riemannian manifolds, can be found in Cheng–Li–Yau [1] and Cheeger–Gromov–Taylor [1] (cf. our discussion in Section VIII.4). We apply the Croke arguments in our discussion of eigenvalues of compact manifolds with small handles—cf. Section IX.2.

1. THE ANALYTIC INEQUALITY

We start with an N-dimensional vector space V, and a continuous map $R(t)$ of the interval $[0, \pi]$ into the space of self-adjoint linear transformations of V. The associated Jacobi equation will be

$$A'' + RA = 0, \tag{1}$$

where $A(t)$ is a linear transformation of V, for each $t \in [0, \pi]$. Of course, for each $\xi \in V$, the vector function

$$\eta(t) = A(t)\xi$$

is a solution of the vector Jacobi equation

$$\eta'' + R(t)\eta = 0.$$

In what follows, we let $A(t)$ denote the solution of (1) satisfying

$$A(0) = 0, \qquad A'(0) = I,$$

when $I: V \to V$ is the identity transformation; for every $s \in [0, \pi]$, we let $C(t; s)$ be the solution of (1) satisfying

$$C(s; s) = 0, \qquad C'(s; s) = I,$$

where the prime $'$ denotes differentiation with respect to the first variable.

The basic inequality is the following:

THEOREM 1 (Berger–Kazdan [1]). If $A(t)$ is invertible for all $t \in (0, \pi)$, then for any continuous function $m(t)$, on $[0, \pi]$, satisfying

$$m(t) > 0, \qquad m(\pi - t) = m(t) \tag{2}$$

on $(0, \pi)$, we have

$$\int_0^\pi ds \int_s^\pi m(t - s) \det C(t; s)\, dt \geq \int_0^\pi ds \int_s^\pi m(t - s) \sin^N(t - s)\, dt.$$

Equality in (3) is achieved if and only if $R(t) = I$ on $[0, \pi]$ (i.e., $C(t; s) = \{\sin(t - s)\}I$).

PROOF: First one establishes, with standard arguments,

$$A^*A' = A'^*A$$

on $(0, \pi)$, and

$$C(t; s) = A(t)\left\{\int_s^t (A^*A)^{-1}(\tau)\, d\tau\right\}A^*(s).$$

Next we apply the following version of *Jensen's inequality*: If $F = F(B)$ is a strictly convex function defined on the convex set of positive definite, self-adjoint linear transformations of V, and ν is any positive measure on $\mathbb{R}$, then

$$F\left\{\frac{1}{\nu((\alpha, \beta))} \int_\alpha^\beta B(\tau)\, d\nu(\tau)\right\} \leq \frac{1}{\nu((\alpha, \beta))} \int_\alpha^\beta F(B(\tau))\, d\nu(\tau), \tag{4}$$

with equality in (4) if and only if $B(s)$ is a constant function.

To apply the Jensen inequality, one sets

$$F(B) = (\det B)^{-1}, \qquad \phi(t) = \{\det A(t)\}^{1/N},$$

$$B(\tau) = \phi^2(\tau)(A^*A)^{-1}(\tau), \qquad d\nu(\tau) = \phi^{-2}(\tau)\, d\tau.$$

The Jensen inequality then implies that

$$\det \int_s^t (A^*A)^{-1}(\tau)\, d\tau \geq \left\{\int_s^t \phi^{-2}(\tau)\, d\tau\right\}^N,$$

which, in turn, implies

$$\det C(t; s) \geq \left\{ \phi(t)\phi(s) \int_s^t \phi^{-2}(\tau)\, d\tau \right\}^N. \tag{5}$$

One checks that equality is attained in (5) if and only if $A = \phi I$ on $(0, \pi)$.

The next step is to apply Hölder's inequality to

$$f = \{\det C(t; s)\}^{1/N}, \qquad p = N,$$
$$h = \sin^{N-1}(t - s), \qquad q = N/(N - 1),$$

and measure ε, given by

$$d\varepsilon = m(t - s)\, dt\, ds.$$

Then the Hölder inequality

$$\int |fh|\, d\varepsilon \leq \left\{ \int |f|^p\, d\varepsilon \right\}^{1/p} \left\{ \int |h|^q\, d\varepsilon \right\}^{1/q}$$

when $1/p + 1/q = 1$ (with equality if and only if there exist constants α, β not both 0, such that $\alpha|f|^p = \beta|h|^q$ a.e. $[d\varepsilon]$), and (5), combine to imply

$$\begin{aligned} &\int_0^\pi ds \int_s^\pi m(t - s) \det C(t - s)\, dt \\ &\geq \{G(\phi)\}^N \left\{ \int_0^\pi ds \int_s^\pi \sin^N(t - s) m(t - s)\, dt \right\}^{1-N}, \end{aligned} \tag{6}$$

where

$$G(\phi) = \int_0^\pi \phi(s)\, ds \int_s^\pi m(t - s) \sin^{N-1}(t - s) \phi(t)\, dt \int_s^t \phi^{-2}(\tau)\, d\tau.$$

Equality is achieved in (6) if and only if

$$\sin^N(t - s) = \det C(t; s), \qquad A(t) = \phi(t) I,$$

that is, if and only if

$$A(t) = \{\sin t\} I \tag{7}$$

on $[0, \pi]$.

Thus we are led to study $G(\phi)$. Note that if $\phi(t) = \sin t$, then

$$\int_s^t \phi^{-2}(\tau)\, d\tau = \int_s^t \frac{d\tau}{\sin^2 \tau} = \frac{\sin(t - s)}{(\sin t)(\sin s)},$$

which implies

$$G(\sin) = \int_0^\pi ds \int_s^\pi m(t-s)\sin^N(t-s)\,dt.$$

Therefore, inequality (3) is a consequence of the inequality

$$G(\phi) \geq G(\sin). \tag{8}$$

We shall prove (8) under the hypothesis that $\phi(t)$ has the form

$$\phi(t) = t^\alpha(\pi - t)^\beta h(t),$$

where $0 \leq \alpha, \beta \leq 1$, and $h(t)$ is positive and continuous on all of $[0, \pi]$.

Let Ω be the subset of $\mathbb{R}^3$ consisting of those (τ, t, s) for which

$$s \leq \tau \leq t, \qquad s \leq t \leq \pi, \qquad 0 \leq s \leq \pi,$$

and, on Ω, define the measure σ by

$$d\sigma = \frac{(\sin s)(\sin t)}{\sin^2\tau} m(t-s)\,d\tau\,dt\,ds.$$

Note that

$$G(\sin) = \sigma(\Omega).$$

Now write

$$\phi(t) = e^{u(t)}\sin t,$$

where u is continuous on $(0, \pi)$. Then the usual form of Jensen's inequality implies

$$G(\phi) \geq G(\sin)\exp\left\{\frac{1}{G(\sin)}\int_\Omega \{u(s) + u(t) - 2u(\tau)\}\,d\sigma\right\}. \tag{9}$$

So (8) will be a consequence of

$$\int_\Omega \{u(s) + u(t) - 2u(\tau)\}\,d\sigma = 0 \tag{10}$$

for all u. Note that we have yet to use the symmetry hypothesis (cf. (2)) for $m(t)$. We shall show that (10) is valid for all u under consideration if and only if $m(t) = m(\pi - t)$ for all $t \in [0, \pi]$.

First one employs some manipulation to rewrite (10) as

$$\int_0^\pi u(t)f(t)(\sin t)^{-2}\,dt = 0 \tag{11}$$

for all u, where $f(t)$ is a C^2 function on $[0, \pi]$, satisfying $f(0) = 0$, and

$$\{(\sin t)^{-2}f'(t)\}' = (\sin t)^{-2}\{(\sin^3 t)[m(t) - m(\pi - t)]\}'.$$

If $m(t) = m(\pi - t)$ for all $t \in [0, \pi]$, then $f = 0$, and (11) is valid for all u. Conversely, if (11) is valid for all u, then $f = 0$ and $m(t) - m(\pi - t) = \text{const.}$ To evaluate the constant, set $t = \pi/2$. The constant is 0, and inequality (3) is proved.

Equality in (3) implies equality in (6), which implies (7); the theorem is proved.

2. M. BERGER'S ISOEMBOLIC INEQUALITY

M is our given complete Riemannian manifold, with tangent bundle TM. We let $\mathfrak{S}M$ denote the *unit tangent bundle* of M, that is, the (2(dim M) − 1)-submanifold of TM consisting of tangent vectors with length equal to 1. We denote the projection map, which associates to every unit vector in $\mathfrak{S}M$ the point in M in whose tangent space it belongs, by $\pi\colon \mathfrak{S}M \to M$. Thus

$$\pi^{-1}[p] = \mathfrak{S}_p,$$

for all $p \in M$.

DEFINITION 1. For all $n \geq 0$, we denote the *volume of* $\mathbb{S}^n$ by $\mathbf{c}_n$.

Recall, from Section III.1, that for any $p \in M$, we let μ_p denote the standard measure on $\mathfrak{S}_p$.

DEFINITION 2. The *kinematic measure* μ is defined to be the measure on $\mathfrak{S}M$ for which

$$\int_{\mathfrak{S}M} f\,d\mu = \int_M dV(p) \int_{\mathfrak{S}_p} (f \mid \mathfrak{S}_p)\,d\mu_p$$

for all compactly supported continuous functions on $\mathfrak{S}M$.

Note that if M is compact, n-dimensional, then $\mathfrak{S}M$ is also compact, and

$$\mu(\mathfrak{S}M) = \mathbf{c}_{n-1} V(M). \tag{12}$$

LIOUVILLE'S THEOREM. Let $\{\Phi_t\}\colon \mathfrak{S}M \to \mathfrak{S}M$ be the *geodesic flow* on $\mathfrak{S}M$, that is, $\{\Phi_t\}$ is the one-parameter group of diffeomorphisms of $\mathfrak{S}M$ defined by

$$\Phi_t \xi = \gamma'_\xi(t)$$

for all $(t, \xi) \in \mathbb{R} \times \mathfrak{S}M$. Then the kinematic measure is invariant with respect to the geodesic flow on $\mathfrak{S}M$, that is, for any $f \in L^1(\mu)$ and domain $G \subseteq \mathfrak{S}M$, we have

$$\int_G \Phi_t^* f \, d\mu = \int_{\Phi_t(G)} f \, d\mu.$$

Recall that geodesics always have constant speed, so Φ_t, indeed, always maps $\mathfrak{S}M$ into $\mathfrak{S}M$. We also have

$$(\pi \circ \Phi_t)(\xi) = \gamma_\xi(t)$$

for all $(t, \xi) \in \mathbb{R} \times \mathfrak{S}M$. For a proof of Liouville's theorem, compare A. Besse [1, Chap. 1].

We recall some notations and facts from our discussion of geodesic polar coordinates in Section III.1. For a fixed $p \in M$, the function $\mathrm{c} = \mathrm{c}(\xi)$, $\xi \in \mathfrak{S}_p$ is the distance to the cut point of p along γ_ξ, that is, $t = d(p, \gamma_\xi(t))$ for all $t \in [0, \mathrm{c}(\xi))$ (include $\mathrm{c}(\xi)$ itself, if $\mathrm{c}(\xi) < +\infty$), and $t > d(p, \gamma_\xi(t))$ for all $t > \mathrm{c}(\xi)$. Also, γ_ξ is the only minimizing geodesic joining p to $\gamma_\xi(t)$ for $t \in [0, \mathrm{c}(\xi))$. The function $\mathrm{c}: \mathfrak{S}M \to (0, +\infty]$ is known to be continuous, and

$$\mathrm{inj}(M) \equiv: \inf_{\mathfrak{S}M} \mathrm{c}$$

is called the *injectivity radius of M*. The domain in M_p of the geodesic spherical coordinates $\mathfrak{D}_p$ is given by

$$\mathfrak{D}_p = \{(t, \xi) \in [0, \infty) \times \mathfrak{S}_p : 0 \le t < \mathrm{c}(\xi)\},$$

and is mapped diffeomorphically onto its image D_p in M, by exp. Thus,

$$\mathrm{inj}(M) = \sup\{\rho \ge 0\}$$

such that $\mathsf{B}(p; \rho) \subseteq \mathsf{D}_p$ for all $p \in M$.

Fix $\xi \in M_p$. Then $\tau_t: M_p \to M_{\gamma_\xi(t)}$ denotes parallel translation along γ_ξ, $\xi^\perp$ is the orthogonal complement of $\mathbb{R}\xi$ in M_p, $\mathscr{R}(t): \xi^\perp \to \xi^\perp$ is the self-adjoint linear transformation defined by

$$\mathscr{R}(t)\eta = (\tau_t)^{-1} R(\phi_t \xi, \tau_t \eta) \phi_t \xi,$$

and $\mathscr{A}(t;\ )$ is the solution to

$$\mathscr{A}'' + \mathscr{R}\mathscr{A} = 0 \tag{13}$$

satisfying

$$\mathscr{A}(0; \xi) = 0, \qquad \mathscr{A}'(0; \xi) = I.$$

We write

$$ds^2(\exp t\xi) = (dt)^2 + |\mathscr{A}(t;\xi)\,d\xi|^2$$

on D_p, and for

$$\sqrt{\mathfrak{g}}(t;\xi) = \det \mathscr{A}(t;\xi)$$

we have that $\sqrt{\mathfrak{g}} \in C^\infty([0,\infty) \times \mathfrak{S}M$, and

$$dV(\exp t\xi) = \sqrt{\mathfrak{g}}(t;\xi)\,dt\,d\mu_p(\xi).$$

Recall that M has constant sectional curvature κ if and only if

$$\mathscr{A}(t;\xi) = \mathbf{S}_\kappa(t)I$$

on $[0,\infty) \times \mathfrak{S}M$.

PROPOSITION 1. For all $n \geq 1$, we have

$$\mathbf{c}_n = \mathbf{c}_{n-1}\int_0^\pi \sin^{n-1} t\,dt, \tag{14}$$

and

$$\pi\mathbf{c}_n/2\mathbf{c}_{n-1} = \int_0^\pi ds \int_0^{\pi-s} \sin^{n-1} t\,dt. \tag{15}$$

PROOF: Equation (14) is the result of direct calculation in geodesic spherical coordinates in $\mathbb{S}^n$. To prove (15), we have

$$\begin{aligned}\pi\mathbf{c}_n/\mathbf{c}_{n-1} &= \int_0^\pi ds \int_0^\pi \sin^{n-1} t\,dt\\ &= \int_0^\pi ds \int_0^{\pi-s} \sin^{n-1} t\,dt + \int_0^\pi ds \int_{\pi-s}^\pi \sin^{n-1} t\,dt,\end{aligned}$$

and

$$\begin{aligned}\int_0^\pi ds \int_{\pi-s}^\pi \sin^{n-1} t\,dt &= \int_0^\pi ds \int_{\pi-s}^\pi \sin^{n-1}(\pi - t)\,dt\\ &= \int_0^\pi ds \int_0^s \sin^{n-1} t\,dt\\ &= \int_0^\pi ds \int_0^{\pi-s} \sin^{n-1} t\,dt,\end{aligned}$$

which implies the claim.

THEOREM 2 (Berger's isoembolic inequality [6]). If M is a compact Riemannian manifold of dimension $n \geq 1$, then

$$V(M) \geq \mathbf{c}_n\{\mathrm{inj}(M)/\pi\}^n, \tag{16}$$

with equality in (16) if and only if M is isometric to $\mathbb{M}_\kappa$ with $\kappa = \{\pi/\mathrm{inj}(M)\}^2$.

PROOF: Normalize the Riemannian metric on M, so that $\mathrm{inj}(M) = \pi$ (viz., change the Riemannian metric on M by multiplying the length of every element of TM by $\pi/\mathrm{inj}(M)$).

Our first step in the proof is to note that for all $r \in [0, \pi/2]$, and all $\xi \in \mathfrak{S}M$, we have

$$V(M) \geq V(\mathsf{B}(\gamma_\xi(0); r)) + V(\mathsf{B}(\gamma_\xi(\pi); \pi - r)). \tag{17}$$

Indeed, the disks $\mathsf{B}(\gamma_\xi(0); r)$ and $\mathsf{B}(\gamma_\xi(\pi); \pi - r)$ are pairwise disjoint.

We now note that

$$\{V(M)\}^2 \geq \int_M \{V(\mathsf{B}(p; r)) + V(\mathsf{B}(p; \pi - r))\}\, dV(p) \tag{18}$$

for all $r \in [0, \pi/2]$. To verify (18), integrate (17) over $\mathfrak{S}M$. We then obtain

$$\begin{aligned}\{V(M)\}^2\mathbf{c}_{n-1} &= V(M)\mu(\mathfrak{S}M)\\ &\geq \int_{\mathfrak{S}M} \{V(\mathsf{B}(\gamma_\xi(0); r)) + V(\mathsf{B}(\gamma_\xi(\pi); \pi - r))\}\, d\mu(\xi),\end{aligned}$$

and

$$\int_{\mathfrak{S}M} V(\mathsf{B}(\gamma_\xi(0); r))\, d\mu(\xi) = \mathbf{c}_{n-1}\int_M V(\mathsf{B}(p; r))\, dV(p).$$

Also, by Liouville's theorem,

$$\begin{aligned}\int_{\mathfrak{S}M} V(\mathsf{B}(\gamma_\xi(\pi); \pi - r))\, d\mu(\xi) &= \int_{\mathfrak{S}M} V(\mathsf{B}(\pi \circ \Phi_\pi(\xi); \pi - r))\, d\mu(\xi)\\ &= \int_{\mathfrak{S}M} V(\mathsf{B}(\pi(\xi); \pi - r))\, d\mu(\xi)\\ &= \mathbf{c}_{n-1}\int_M \mathsf{B}(p; \pi - r)\, dV(p),\end{aligned}$$

and inequality (18) follows immediately.

Our next comment is that one easily verifies

$$\int_M V(\mathsf{B}(p; r))\, dV(p) = \int_0^r dt \int_{\mathfrak{S}M} \sqrt{\mathsf{g}}(t; \xi)\, d\mu(\xi), \tag{19}$$

for all $r \in [0, \pi]$. Also one checks that

$$\tag{20} \begin{aligned} &\int_0^{\pi/2} dr \int_0^r \sqrt{\mathfrak{g}}(t;\xi)\,dt + \int_0^{\pi/2} dr \int_0^{\pi - r} \sqrt{\mathfrak{g}}(t;\xi)\,dt \\ &\qquad = \int_0^{\pi/2} \{(\pi - t)\sqrt{\mathfrak{g}}(t;\xi) + t\sqrt{\mathfrak{g}}(\pi - t;\xi)\}\,dt \end{aligned}$$

for all $\xi \in \mathfrak{S}M$.

If we now integrate (18) over $r \in [0, \pi/2]$, then we have, via (19) and (20),

$$\begin{aligned} &(\pi/2)\{V(M)\}^2 \\ &\geq \int_0^{\pi/2} dr \left\{ \int_0^r dt \int_{\mathfrak{S}M} \sqrt{\mathfrak{g}}(t;\xi)\,d\mu(\xi) + \int_0^{\pi - r} dt \int_{\mathfrak{S}M} \sqrt{\mathfrak{g}}(t;\xi)\,d\mu(\xi) \right\} \\ &= \int_0^{\pi/2} dt \int_{\mathfrak{S}M} \{(\pi - t)\sqrt{\mathfrak{g}}(t;\xi) + t\sqrt{\mathfrak{g}}(\pi - t;\xi)\}\,d\mu(\xi). \end{aligned}$$

But

$$\begin{aligned} (\pi - t)\int_{\mathfrak{S}M} \sqrt{\mathfrak{g}}(t;\xi)\,d\mu(\xi) &= \int_0^{\pi - t} dr \int_{\mathfrak{S}M} \sqrt{\mathfrak{g}}(t;\xi)\,d\mu(\xi) \\ &= \int_0^{\pi - t} dr \int_{\mathfrak{S}M} \sqrt{\mathfrak{g}}(t;\Phi_r\xi)\,d\mu(\xi), \end{aligned}$$

by Liouville's theorem, and, similarly,

$$t\int_{\mathfrak{S}M} \sqrt{\mathfrak{g}}(\pi - t;\xi)\,d\mu(\xi) = \int_0^t dr \int_{\mathfrak{S}M} \sqrt{\mathfrak{g}}(\pi - t;\Phi_r\xi)\,d\mu(\xi).$$

Thus we have

$$\tag{21} \begin{aligned} &(\pi/2)\{V(M)\}^2 \\ &\quad \geq \int_0^{\pi/2} dt \int_{\mathfrak{S}M} d\mu(\xi) \left\{ \int_0^{\pi - t} \sqrt{\mathfrak{g}}(t;\Phi_r\xi)\,dr + \int_0^t \sqrt{\mathfrak{g}}(\pi - t;\Phi_r\xi)\,dr \right\}. \end{aligned}$$

We are now ready to apply the Berger–Kazdan inequality (3), for $N = n - 1$, and

$$C(t;s) = \mathscr{A}(t - s;\Phi_s\xi),$$

$$\det C(t;s) = \sqrt{\mathfrak{g}}(t - s;\Phi_s\xi).$$

Inequality (3) can then be written as

$$\tag{22} \int_0^\pi ds \int_0^{\pi - s} m(r)\sqrt{\mathfrak{g}}(r;\Phi_s\xi)\,dr \geq \int_0^\pi ds \int_0^{\pi - s} m(r)\sin^{n-1} r\,dr.$$

For the function m we pick

$$m = \delta_t + \delta_{\pi - t}, \tag{23}$$

the sum of the Dirac distribution at t and the Dirac distribution at $\pi - t$. Inequality (3) is valid for this choice of m since (3) is valid for all positive continuous functions $m = m(t)$ on $[0, \pi]$ satisfying the symmetry condition in (2). An easy change of order of integration leads to

$$\int_0^{\pi - t} \sqrt{\mathbf{g}}(t; \Phi_r \xi)\, dr = \int_0^{\pi} dr \int_0^{\pi - r} \delta_t(s) \sqrt{\mathbf{g}}(s; \Phi_r \xi)\, ds,$$

and

$$\int_0^t \sqrt{\mathbf{g}}(\pi - t; \Phi_r \xi)\, dr = \int_0^{\pi} dr \int_0^{\pi - r} \delta_{\pi - t}(s) \sqrt{\mathbf{g}}(s; \Phi_r \xi)\, ds.$$

Therefore, (21) and (22) imply

$$(\pi/2)\{V(M)\}^2 \geq \int_0^{\pi/2} dt \int_{\mathfrak{S}M} d\mu(\xi) \int_0^{\pi} ds \int_0^{\pi - s} (\delta_t + \delta_{\pi - t})(r) \sin^{n-1} \mathrm{r}\, \mathrm{dr}. \tag{24}$$

We claim that (24) implies the theorem. First, if M is, in fact, isometric to $\mathbb{S}^n$, then one easily checks that all the inequalities, in the whole argument, are equalities. Thus the right-hand side of (24) is equal to $(\pi/2)\{\mathbf{c}_n\}^2$, and then (24) is to be written as

$$(\pi/2)\{V(M)\}^2 \geq (\pi/2)\{\mathbf{c}_n\}^2,$$

and (16) follows.

If we have equality in (22), with m given by (23), for all $\xi \in \mathfrak{S}M$, then (cf. (6) ff.)

$$\mathscr{A}(t; \xi) = \sqrt{\mathbf{g}}(t; \xi) I$$

for all $(t; \xi) \in [0, \pi] \times \mathfrak{S}M$, and

$$\sqrt{\mathbf{g}}(r - s; \xi) = \sin(r - s)$$

almost everywhere, on $\{(r, s) : 0 \leq s \leq r, 0 \leq r \leq \pi\}$, with respect to the measure

$$d\varepsilon = (\delta_t + \delta_{\pi - t})(r - s)\, dr\, ds,$$

for all $\xi \in \mathfrak{S}M$. Thus

$$\sqrt{\mathbf{g}}(t; \xi) = \sin t, \qquad \sqrt{\mathbf{g}}(\pi - t; \xi) = \sin(\pi - t)$$

for all $t \in [0, \pi/2]$, that is,

$$\sqrt{\mathfrak{g}}(t; \xi) = \sin t$$

for all $(t; \xi) \in [0, \pi] \times \mathfrak{S}M$, which implies that M has constant sectional curvature equal to 1.

Then the universal Riemannian covering of M is isometric to $\mathbb{S}^n$. Since $V(M) = \mathbf{c}_n$, the covering must be an isometry, and the theorem is proved.

We conclude this section by highlighting the role played by isoembolic inequality in the proof of the Blaschke conjecture.

DEFINITION 3. Let M be a complete Riemannian manifold. We say that M is a *wiedersehnsraum* if there exists a constant $l > 0$ such that

$$\mathscr{A}(l; \xi) = 0 \tag{25}$$

for every $\xi \in \mathfrak{S}M$.

M is called a wiedersehnsraum since, by Proposition III.1, the validity of (25) for all $\xi \in \mathfrak{S}M$ is equivalent with the fact that, for every $p \in M$, the image, under the exponential map, of the sphere $\mathfrak{S}(p; l)$, in M_p, consists of a single point—we denote it by $Q(p)$. So when M is a wiedersehnsraum, all unit-speed geodesics leaving p meet at the same time l at $Q(p)$—hence the name.

The *Blaschke conjecture* is that a wiedersehnsraum M is isometric to the standard sphere, or real projective space, of the same dimension as M, and having constant sectional curvature equal to $(\pi/l)^2$.

The basic steps of the proof of the Blaschke conjecture are: (i) The map $p \to Q(p)$ is an involutive isometry of M, and M is diffeomorphic to the standard sphere, or real projective space, with dimension equal to that of M. When M is diffeomorphic to real projective space, then its universal Riemannian covering is a wiedersehnsraum diffeomorphic to the sphere. So it suffices to consider the case of a wiedersehnsraum M, diffeomorphic to the sphere, in which case $c(\xi) = l$ for all $\xi \in \mathfrak{S}M$ (Green [1]). (ii) Berger's isoembolic inequality applies. (iii) One has the case of equality in the isoembolic inequality (Weinstein [1]; Yang [1]).

For more material on these matters we refer the reader to Berger [6], Berger–Kazdan [1], and Besse [1].

Remark 1: The isoembolic inequality is a "universal" inequality for compact Riemannian manifolds, as it has no assumptions on curvature. Other inequalities can be found in Berger [6] and Croke [5].

3. CHEEGER AND ISOPERIMETRIC CONSTANTS, AND THE KINEMATIC MEASURE

The results of this section are the work of C. B. Croke [1].

We let Ω be a Riemannian normal domain with smooth boundary. For any $\xi \in \mathfrak{S}\Omega$ we let

$$\tau(\xi) = \sup\{\tau > 0 : \gamma_\xi(t) \in \Omega \; \forall t \in [0, \tau]\}.$$

We note that if G is an open subset of $\mathfrak{S}\Omega$, and $0 < \tau < \inf\{\tau(\xi) : \xi \in G\}$, then $\Phi_t(G) \subseteq \mathfrak{S}\Omega$ for all $t \in [0, \tau]$, and Liouville's theorem is valid for G, and $t \in [0, \tau]$.

For any $\xi \in \mathfrak{S}\Omega$, we also define

$$l(\xi) = \inf(\mathrm{c}(\xi), \tau(\xi)),$$

where $\mathrm{c}(\xi)$ is, as usual, the distance to the cut point of $\gamma_\xi(0)$ along γ_ξ. Since we are, at the moment, not thinking of Ω as an imbedded domain in a larger Riemannian manifold, when $\tau(\xi) < +\infty$, $\gamma_\xi(\tau(\xi))$ may be the last point at which the geodesic is defined. In this case, if γ_ξ minimizes arc length from $\gamma_\xi(0)$ to $\gamma_\xi(\tau(\xi))$, then $\mathrm{c}(\xi) = \tau(\xi)$ by definition. So, when Ω is not an imbedded domain in a larger Riemannian manifold, we will always have $\mathrm{c}(\xi) \leq \tau(\xi)$. (Of course, when Ω is an imbedded domain in a larger Riemannian manifold, then it will be possible that $\mathrm{c}(\xi) > \tau(\xi)$.)

The version of the Berger–Kazdan inequality that we shall apply here is

THEOREM 3. For all $l \in [0, \mathrm{c}(\xi)]$, $\xi \in \mathfrak{S}\Omega$, we have

$$\int_0^l ds \int_0^{l-s} \sqrt{\mathbf{g}}(r; \Phi_s \xi)\, dr \geq (\pi \mathbf{c}_n / 2\mathbf{c}_{n-1})(l/\pi)^{n+1}, \tag{26}$$

with equality in (26) if and only if

$$\mathscr{A}(t; \xi) = (\sin t\pi/l)I \tag{27}$$

for $t \in [0, l]$.

PROOF: Indeed, we have normalized the interval in (3) from $[0, \pi]$ to $[0, l]$, set $N = n - 1$ and $m = 1$, and used (15).

We define

$$\mathfrak{U}\Omega = \{\xi \in \mathfrak{S}\Omega : \mathrm{c}(\xi) = \tau(\xi)\}$$

(if Ω is an imbedded domain in a larger Riemannian manifold, then the defining property should read as $\mathbf{c}(\xi) \geq \tau(\xi)$),

$$\mathfrak{U}_p = (\pi\,|\,\mathfrak{U}\Omega)^{-1}[p], \qquad \omega_p = \mu_p(\mathfrak{U}_p)/\mathbf{c}_{n-1},$$
$$\omega = \inf\{\omega_p : p \in \Omega\}.$$

We call ω_p the *visibility angle of* $\partial\Omega$ *from* p.

We now let $\bar{\nu}$ be the inward pointing normal unit vector field along $\partial\Omega$, dA the Riemannian density of $\partial\Omega$,

$$\mathfrak{S}^+\partial\Omega = \{\xi \in \mathfrak{S}\bar{\Omega}\,|\,\partial\Omega : \langle \xi, \bar{\nu}\rangle > 0\}, \qquad \mathfrak{S}_q^+ = (\pi\,|\,\mathfrak{S}^+\partial\Omega)^{-1}[q],$$

for $q \in \partial\Omega$, and $d\sigma$ the induced density on $\mathfrak{S}^+\partial\Omega$, that is, $d\sigma$ is locally given by

$$d\sigma(\xi) = d\mu_{\pi(\xi)}(\xi)\, dA(\pi(\xi))$$

for $\xi \in \mathfrak{S}^+\partial\Omega$.

An important tool will be

SANTALO'S FORMULA (Santalo [1, p. 336ff.]). For any $f \in L^1(\mu)$ we have

$$\iint_{\mathfrak{U}\Omega} f\, d\mu = \int_{\mathfrak{S}^+\partial\Omega} \langle \xi, \bar{\nu}_{r(\xi)}\rangle\, d\sigma(\xi) \int_0^{l(\xi)} f(\Phi_t\,\xi)\, dt. \tag{28}$$

In particular, for $f = 1$, we have

$$\mu(\mathfrak{U}\Omega) = \int_\Omega \omega_p\, dV(p) = \int_{\mathfrak{S}^+\partial\Omega} l(\xi)\langle \xi, \bar{\nu}_{\pi(\xi)}\rangle\, d\sigma(\xi) \tag{29}$$

It will be convenient to note

PROPOSITION 2. For any $e \in \mathbb{S}^n$, we have

$$\int_{H_e} \langle \xi, e\rangle\, dV(\xi) = \mathbf{c}_{n-1}/n, \tag{30}$$

where H_e is the hemisphere of $\mathbb{S}^n$ centered at e. (The proof is an easy exercise.)

LEMMA 1. Let $d(\Omega)$ denote the diameter of Ω. Then

$$A(\partial\Omega)/V(\Omega) \geq (n-1)\mathbf{c}_{n-1}\omega/\mathbf{c}_{n-2}\,d(\Omega), \tag{31}$$

with equality in (31) if M is a hemisphere of constant positive curvature, in which case $\omega = 1$, and $d(\Omega)$ is the diameter of the hemisphere.

PROOF: We have

$$\begin{aligned}\omega \mathbf{c}_{n-1} V(\Omega) &\le \mu(\mathfrak{U}\Omega)\\ &= \int_{\mathfrak{S}^+\partial\Omega} l(\xi)\langle \xi, \bar{\nu}_{\pi(\xi)}\rangle \, d\sigma(\xi)\\ &= \int_{\partial\Omega} dA(q) \int_{\mathfrak{S}_q^+} l(\xi)\langle \xi, \bar{\nu}_q\rangle \, d\mu_q(\xi)\\ &\le d(\Omega)\mathbf{c}_{n-2}A(\partial\Omega)/(n-1),\end{aligned}$$

that is,

$$\omega \mathbf{c}_{n-1} V(\Omega) \le d(\Omega)\mathbf{c}_{n-2}A(\partial\Omega)/(n-1),$$

which implies (31).

Equality in (31), when Ω is a hemisphere of constant positive curvature, is easily verified.

LEMMA 2. Let M be a compact n-dimensional Riemannian manifold, all of whose Ricci curvatures satisfy

$$\operatorname{Ric}(\xi, \xi) \ge (n-1)\kappa \tag{32}$$

for a fixed constant κ, and for all $\xi \in \mathfrak{S}M$. Let d denote the diameter of M. If S is any $(n-1)$-dimensional compact submanifold of M dividing M into open submanifolds M_1, M_2, satisfying $\partial M_1 = \partial M_2 = S$, then, setting $\Omega = M_1$, we have

$$\omega_p \ge V(M_2)/\mathbf{c}_{n-1}\int_0^d \mathbf{S}_\kappa \tag{33}$$

for all $p \in M_1$.

PROOF: Note that the discussion through Santalo's formula is valid even when Ω is a finite disjoint union of regular domains.

To prove (33), for $p \in M_1$, let

$$\Theta_p = \{q \in M : q = \gamma_\xi(t),\ \xi \in \mathfrak{U}_p,\ t \le \mathrm{c}(\xi)\}.$$

Then

$$M_2 \subseteq \Theta_p \subseteq \mathsf{D}_p,$$

which implies

$$\begin{aligned} V(M_2) &\le \int_{\mathfrak{U}_p} d\mu_p(\xi) \int_0^{\mathbf{c}(\xi)} \sqrt{\mathfrak{g}}(t;\xi)\,dt \\ &\le \int_{\mathfrak{U}_p} d\mu_p(\xi) \int_0^{\mathbf{c}(\xi)} \mathbf{S}_\kappa \le \omega_p \mathbf{c}_{n-1} \int_0^d \mathbf{S}_\kappa; \end{aligned}$$

note that the right-hand side of the first line is precisely equal to $V(\Theta_p)$; and that the second line is a consequence of the Bishop comparison theorem (II) (cf. Section III.3).

Using Cheeger's inequality (Theorem IV.10), Theorem IV.12, and our two lemmata here, one immediately obtains

THEOREM 4. Let M be as in Lemma 2. Then

$$\lambda_1 \ge \left\{(n-1)V(M)/4\mathbf{c}_{n-2}d\int_0^d \mathbf{S}_\kappa\right\}^2.$$

Thus, λ_1, of a compact Riemannian manifold, may be bounded from below in terms of the diameter, volume, and a lower bound for the Ricci curvatures, of M—a result first proved by Yau [1]. Since then, Li and Yau [1] have shown, via the maximum principle applied to appropriate gradient estimates, that one can bound λ_1 from below in terms of the diameter, and lower bound on the Ricci curvature, alone (cf. a different argument in Section XII.8). Thus, together with Cheng's theorem (Corollary III.3) one currently has both upper and lower bounds on λ_1, of a compact Riemannian manifold in terms of the diameter, and lower bound on the Ricci curvature.

THEOREM 5. Let Ω be our normal domain with smooth boundary. Then

$$\frac{\{A(\partial\Omega)\}^n}{\{V(\Omega)\}^{n-1}} \ge \frac{\{\mathbf{c}_{n-1}\}^n}{\{\mathbf{c}_n/2\}^{n-1}}\,\omega^{n+1}, \tag{34}$$

with equality in (34) if and only if Ω is a hemisphere of a constant sectional curvature sphere.

PROOF: For every $p \in \Omega$, we have

$$V(\Omega) = \int_{\mathfrak{S}_p} d\mu_p(\xi) \int_0^{l(\xi)} \sqrt{\mathfrak{g}}(t;\xi)\,dt. \tag{35}$$

Now integrate (35) over all of Ω. Then

$$
\begin{align}
\{V(\Omega)\}^2 &= \int_\Omega dV(p)\int_{\mathfrak{S}_p} d\mu_p(\xi)\int_0^{l(\xi)} \sqrt{\mathbf{g}}(t;\,\xi)\,dt \notag\\
&= \int_{\mathfrak{S}\Omega} d\mu(\xi)\int_0^{l(\xi)} \sqrt{\mathbf{g}}(t;\,\xi)\,dt \tag{36}\\
&\ge \int_{\mathfrak{U}\Omega} d\mu(\xi)\int_0^{l(\xi)} \sqrt{\mathbf{g}}(t;\,\xi)\,dt \notag\\
&= \int_{\mathfrak{S}^+\partial\Omega} \langle \xi, \bar{v}_{\pi(\xi)}\rangle\,d\sigma(\xi)\int_0^{l(\xi)} ds\int_0^{l(\phi_s\xi)} \sqrt{\mathbf{g}}(t;\,\phi_s\xi)\,dt \notag\\
&\ge \int_{\mathfrak{S}^+\partial\Omega} \langle \xi, \bar{v}_{\pi(\xi)}\rangle\,d\sigma(\xi)\int_0^{l(\xi)} ds\int_0^{l(\xi)-s} \sqrt{\mathbf{g}}(t;\,\phi_s\xi)\,dt \notag\\
&\ge \{\mathbf{c}_n/2\pi^n\mathbf{c}_{n-1}\}\int_{\mathfrak{S}^+\partial\Omega} \{l(\xi)\}^{n+1}\langle \xi, \bar{v}_{\pi(\xi)}\rangle\,d\sigma(\xi) \tag{37}\\
&\ge \frac{\mathbf{c}_n}{2\pi^n\mathbf{c}_{n-1}}\,\frac{\{\int_{\mathfrak{S}^+\partial\Omega} l(\xi)\langle \xi, \bar{v}_{\pi(\xi)}\rangle\,d\sigma(\xi)\}^{n+1}}{\{\int_{\mathfrak{S}^+\partial\Omega} \langle \xi, \bar{v}_{\pi(\xi)}\rangle\,d\sigma(\xi)\}^{n}} \tag{38}\\
&= \frac{\mathbf{c}_n}{2\pi^n\mathbf{c}_{n-1}}\left\{\frac{n-1}{\mathbf{c}_{n-2}}\right\}^n \frac{\{\mu(\mathfrak{U}\Omega)\}^{n+1}}{\{A(\partial\Omega)\}^n} \notag\\
&\ge \frac{\omega\mathbf{c}_n}{2}\left\{\frac{(n-1)\mathbf{c}_{n-1}}{\pi\mathbf{c}_{n-2}}\right\}^n \frac{\{V(\Omega)\}^{n+1}}{\{A(\partial\Omega)\}^n}. \notag
\end{align}
$$

One goes from the third to the fourth line using (28), with

$$f(\xi) = \int_0^{l(\xi)} \sqrt{\mathbf{g}}(t;\,\xi)\,dt;$$

one goes from the fourth to the fifth line using

$$l(\Phi_s\xi) \ge l(\xi) - s;$$

one goes from the fifth to the sixth line using (26); one goes from the sixth to the seventh line using (29) (for the numerator) and (30) (for the denominator).

Thus,

$$\frac{\{A(\partial\Omega)\}^n}{\{V(\Omega)\}^{n-1}} \ge \frac{\omega\mathbf{c}_n}{2}\left\{\frac{(n-1)\mathbf{c}_{n-1}}{\pi\mathbf{c}_{n-2}}\right\}^n.$$

Note that if Ω is a hemisphere in a constant sectional curvature sphere, then $\omega = 1$ and we have equality in every step of the argument. In particular,

$$\frac{\mathrm{c}_n}{2}\left\{\frac{(n-1)\mathrm{c}_{n-1}}{\pi \mathrm{c}_{n-2}}\right\}^n = \frac{\{\mathrm{c}_{n-1}\}^n}{\{\mathrm{c}_n/2\}^{n-1}}.$$

Thus (34) is valid, with equality if Ω is a hemisphere in a constant sectional curvature sphere.

It therefore remains to show that equality in (34) implies that Ω is a hemisphere.

First note that $\mathrm{c} = \mathrm{c}(\xi)$ is continuous on $\mathfrak{S}\Omega \cup \mathfrak{S}^+\partial\Omega$, $\tau = \tau(\xi)$ is lower semicontinuous on $\mathfrak{S}\Omega \cup \mathfrak{S}^+\partial\Omega$, and $\mathrm{c}(\xi) \leq \tau(\xi)$ on all of $\mathfrak{S}\Omega \cup \mathfrak{S}^+\partial\Omega$. Therefore, if we are given ξ_0 in $\mathfrak{S}\Omega \cup \mathfrak{S}^+\partial\Omega$ for which $\mathrm{c}(\xi_0) < \tau(\xi_0)$, then there exists a neighborhood G of ξ_0, in $\mathfrak{S}\Omega \cup \mathfrak{S}^+\partial\Omega$, on which $\mathrm{c} < \tau$.

Now assume that we are given equality in (34). Equality in (36) implies that $\mathfrak{U}M = \mathfrak{S}M$, that is, $\mathrm{c}(\xi) = \tau(\xi)$ for all $\xi \in \mathfrak{S}\Omega \cup \mathfrak{S}^+\partial\Omega$. Equality in (38) implies $l(\xi)$ is constant, say $l(\xi) = l$ on all of $\mathfrak{S}^+\partial\Omega$. Thus

$$\{\gamma_\xi(t) : \xi \in \mathfrak{S}^+\partial\Omega, 0 \leq t \leq l\}$$

covers all of Ω, and equality in (37) then implies that Ω has constant sectional curvature equal to $(\pi/l)^2$. Then

$$\mathrm{c}(\xi) = \tau(\xi) = l$$

for all $\xi \in \mathfrak{S}^+\partial\Omega$ implies that Ω is a hemisphere.

THEOREM 6. Let $\mathbb{S}_l^+$ denote the n-dimensional hemisphere of the constant sectional curvature sphere having (Riemannian) diameter equal to l. Let Ω be a Riemannian normal domain, with smooth boundary, such that

$$\tau(\xi) \leq l$$

for all $\xi \in \mathfrak{S}\Omega$, that is, every geodesic starting in Ω hits $\partial\Omega$ not later than having traveled distance l. Then

$$\lambda(\Omega) \geq \lambda(\mathbb{S}_l^+). \tag{39}$$

If $\omega = 1$, that is,

$$\mathrm{c}(\xi) = \tau(\xi)$$

for all $\xi \in \mathfrak{S}\Omega$, then equality holds in (39) if and only if Ω is isometric to S_l^+.

LEMMA 3. Let $f \in C^\infty([0, l])$, with $f(0) = f(l) = 0$. Then

$$\int_0^l f'^2 \geq (\pi/l)^2 \int_0^l f^2, \tag{40}$$

with equality in (40) if and only if

$$f(t) = \text{const} \cdot \sin \pi t/l$$

on $[0, l]$.

Proof: The lemma is merely a restatement of Rayleigh's theorem (Section I.5) for the Dirichlet eigenvalue problem on $[0, l]$. (The lemma is known as *Wirtinger's inequality*.)

LEMMA 4 (Santalo [1, p. 336ff.]). If $\tau(\xi)$ is finite for all $\xi \in \mathfrak{S}\Omega$, then for $f \in L^1(\mu)$ we have

$$\iint_{\mathfrak{S}\Omega} f \, d\mu = \int_{\mathfrak{S}^+\partial\Omega} \langle \xi, \bar{\nu}_{\pi(\xi)} \rangle \, d\sigma(\xi) \int_0^{\tau(\xi)} f(\Phi_t \xi) \, dt. \tag{41}$$

Proof of Theorem 6: Let $f \in C^\infty(\bar{\Omega})$, with $f \mid \partial\Omega = 0$. We wish to show that

$$\int_\Omega |\text{grad} \, f|^2 \geq \lambda(\mathbb{S}_l^+) \int_\Omega f^2. \tag{42}$$

Well, it is standard that for every $p \in \Omega$ we have

$$|\text{grad} \, f|^2(p) = \frac{n}{\mathbf{c}_{n-1}} \int_{\mathfrak{S}_p} \langle \text{grad} \, f, \xi \rangle^2 \, d\mu_p(\xi),$$

which implies

$$\begin{aligned}
&\int_\Omega |\text{grad} \, f|^2 \, dV \\
&= \left\{\frac{n}{\mathbf{c}_{n-1}}\right\} \int_{\mathfrak{S}\Omega} \langle \text{grad} \, f \circ \pi, \xi \rangle^2 \, d\mu(\xi) \\
&= \left\{\frac{n}{\mathbf{c}_{n-1}}\right\} \int_{\mathfrak{S}^+\partial\Omega} \langle \xi, \bar{\nu}_{\pi(\xi)} \rangle \, d\sigma(\xi) \int_0^{\tau(\xi)} \langle \text{grad} \, f \circ \pi, \phi_t \xi \rangle^2 \, dt \\
&= \left\{\frac{n}{\mathbf{c}_{n-1}}\right\} \int_{\mathfrak{S}^+\partial\Omega} \langle \xi, \bar{\nu}_{\pi(\xi)} \rangle \, d\sigma(\xi) \int_0^{\tau(\xi)} \{(f \circ \gamma_\xi)'(t)\}^2 \, dt \\
&\geq \left\{\frac{n\pi^2}{l^2 \mathbf{c}_{n-1}}\right\} \int_{\mathfrak{S}^+\partial\Omega} \langle \xi, \bar{\nu}_{\pi(\xi)} \rangle \, d\sigma(\xi) \int_0^{\tau(\xi)} \{(f \circ \gamma_\xi)(t)\}^2 \, dt \\
&= \left\{\frac{n\pi^2}{l^2 \mathbf{c}_{n-1}}\right\} \int_{\mathfrak{S}\Omega} (f \circ \pi)^2 \, d\mu \\
&= \left(\frac{n\pi^2}{l^2}\right) \int_\Omega f^2 \, dV = \lambda(\mathbb{S}_l^+) \int_\Omega f^2 \, dV,
\end{aligned}$$

which is (42).

If we are given equality in (39), then

$$\tau(\xi) = l \qquad \forall \xi \in \mathfrak{S}^+\partial\Omega,$$

and

$$(f \circ \gamma_\xi)(t) = \alpha(\xi) \sin \pi t/l \qquad \forall \xi \in \mathfrak{S}^+\partial\Omega.$$

For our convenience, we assume that $p \in \Omega$ such that

$$f(p) = \max_\Omega f = 1.$$

We first note that if $\xi \in \mathfrak{S}^+\partial\Omega$ and γ_ξ passes through p, then, since $f \circ \gamma_\xi$ has precisely one maximum, it must occur when $\gamma_\xi(t) = p$. Thus $p = \gamma_\xi(l/2)$. On the other hand, every geodesic leaving p must hit $\partial\Omega$, and by the remark, just made,

$$\tau(\xi) = l/2 \qquad \forall \xi \in \mathfrak{S}_p.$$

Thus,

$$\Omega = \exp \mathfrak{B}(p; l/2), \qquad \partial\Omega = \exp \mathfrak{S}(p; l/2),$$

which implies

$$(f \circ \gamma_\xi)(t) = \cos \pi t/l \qquad \forall \xi \in \mathfrak{S}_p;$$

therefore

$$f(q) = \cos \pi d(p, q)/l$$

for all $q \in \Omega$, and $\exp | \mathfrak{B}(p; l/2)$ is a diffeomorphism of $\mathfrak{B}(p; l/2)$ onto Ω, with

$$\partial\Omega = S(p; l/2).$$

Furthermore, since $\operatorname{grad} f \,|\, \partial\Omega$ is orthogonal to $\partial\Omega$, we also have

$$\alpha(\xi) = \langle \xi, \bar{v}_{\pi(\xi)} \rangle$$

for all $\xi \in \mathfrak{S}^+\partial\Omega$.

Now let $q \in \partial\Omega$, $\eta \in (\partial\Omega)_q$, $|\eta| = 1$. By continuity, $\gamma_\eta(t) \in \bar{\Omega}$ for all $t \in [0, l]$, and

$$(f \circ \gamma_\eta)(t) = \langle \eta, \bar{v}_q \rangle \sin \pi t/l = 0.$$

Thus, $\gamma_\eta([0, l]) \subseteq \partial\Omega$. Therefore, $\partial\Omega$ is *totally geodesic*, that is, all geodesics, with initial velocity vector tangent to $\partial\Omega$, remain in $\partial\Omega$.

We now *add* the assumption $\omega = 1$, that is,

$$c(\xi) = \tau(\xi)$$

for all $\xi \in \mathfrak{S}\Omega \cup \mathfrak{S}^+\partial\Omega$.

Given $q \in \partial\Omega$, let

$$\bar{q} = \gamma_{\bar{\nu}_q}(l) \in \partial\Omega.$$

Then

$$\gamma_{\bar{\nu}_q}(l/2) = p$$

which implies

$$d(q, q^*) < l$$

for all $q^* \in \Omega \backslash \{\bar{q}\}$. Since

$$\mathrm{c}(\xi) = \tau(\xi) = l \qquad \forall \xi \in \mathfrak{S}^+\partial\Omega,$$

we have

$$\gamma_\xi(l) = \bar{q} \qquad \forall \xi \in \mathfrak{S}_q^+,$$

which implies that $\partial\Omega$ is a wiedersehnsraum, which, in turn, implies that $\partial\Omega$ is isometric to $\partial\mathbb{S}_l^+$.

Now (41) implies

$$\begin{aligned} \mathbf{c}_{n-1} V(\Omega) &= \mu(\mathfrak{S}\Omega) \\ &= l \int_{\mathfrak{S}^+\partial\Omega} \langle \xi, \bar{\nu}_{\pi(\xi)} \rangle \, d\sigma(\xi) \\ &= l A(\partial\Omega)\mathbf{c}_{n-1}/n \\ &= l A(\partial\mathbb{S}_l^+)\mathbf{c}_{n-1}/n \\ &= \mathbf{c}_{n-1} V(\mathbb{S}_l^+); \end{aligned}$$

the third line is a consequence of (30); the fourth line, a consequence of the isometry of $\partial\Omega$ and $\partial\mathbb{S}_l^+$; and the last line, a consequence of the same argument connecting the first and fourth lines.

Thus, $A(\partial\mathbb{S}_l^+) = A(\partial\Omega)$, and $V(\mathbb{S}_l^+) = V(\Omega)$, which implies, with $\omega = 1$, equality in (34), which implies Ω is isometric to a hemisphere of a constant sectional curvature sphere. Theorem 6 is proven.

Remark 2: Theorem 6 provides for a sharp universal lower bound for $\lambda(\Omega)$. Universal upper bounds for lowest Dirichlet eigenvalues of geodesic disks are also discussed in Croke's paper, and in Berger [5], Bérard–Besson [2], and Hebda [1].

THEOREM 7. Let M be a compact n-dimensional Riemannian manifold all of whose Ricci curvatures satisfy

$$\text{Ric}(\xi, \xi) \geq (n-1)\kappa \tag{32}$$

for a fixed constant κ, and for all $\xi \in \mathfrak{S}M$. Let d denote the diameter of M. Then

$$\mathfrak{I}(M) \geq [1/4(\mathbf{c}_n)^{n-1}\mathbf{c}_{n-1}]\left\{V(M)\Big/\int_0^d \mathbf{S}_\kappa\right\}^{n+1}. \tag{43}$$

PROOF: One obtains (43) by combining Lemma 2 with Theorem 5.

Thus, combining Theorem 7 with Li's results (Section IV.5), we have that upper bounds on the volume and diameter of a compact Riemannian manifold, and a lower bound on the Ricci curvature, imply uniform lower bounds on the eigenvalues of the manifold and a uniform upper bound on the formal heat kernel.

CHAPTER VI

The Heat Kernel for Compact Manifolds

We are given a fixed Riemannian manifold M of dimension $n \geq 1$, with associated Laplace–Beltrami operator Δ. The *heat operator L on M* will be defined to be the differential operator

$$L = \Delta - \partial/\partial t$$

acting on functions in $C^0(M \times (0, \infty))$, which are C^2 in the "space" variable x, varying over M, and C^1 in the "time" variable t, varying over $(0, \infty)$. The *homogeneous heat equation*, or *heat equation*, for short, is given by

$$Lu = 0 \tag{1}$$

that is,

$$\Delta u = \partial u/\partial t.$$

The heat equation is interpreted by considering the Riemannian manifold M as a homogeneous isotropic medium (with all the physical constants suitably normalized), and $u(x, t)$ as the temperature of $x \in M$ at time t. The equation, then, determines the conduction of heat through the medium, assuming no sources of heat, or refrigeration, are present.

If heat is generated in M so that at $x \in M$ heat is supplied, or withdrawn, so that the temperature changes at the instantaneous rate $F(x, t)$, with respect to time, then the temperature distribution $u(x, t)$ satisfies the *inhomogeneous heat equation*

$$Lu = -F. \tag{2}$$

Equations (1) and (2) are derived as follows: Given any regular domain Ω in M, the instantaneous change in the total temperature in Ω, with respect to time, is equal to the total change due to the supply of heat to, or withdrawal from, Ω, added to the spatial change of the heat distribution across $\partial\Omega$, that is,

$$\iint_{\Omega} \frac{\partial u}{\partial t}(x, t)\, dV(x) = \iint_{\Omega} F(x, t)\, dV(x) + \int_{\partial\Omega} \frac{\partial u}{\partial \nu}(w, t)\, dA(w).$$

By Green's formula (I.39) we have

$$\iint_{\Omega} \frac{\partial u}{\partial t}\, dV = \iint_{\Omega} \{F + \Delta u\}\, dV,$$

that is,

$$\iint_{\Omega} \{Lu + F\}\, dV = 0,$$

for all such domains Ω in M. One immediately obtains (2). Case (1) corresponds to $F = 0$.

Our primary interest is in the *fundamental solution of the heat equation on* M, or the *heat kernel of* M. Its definition is as follows:

DEFINITION 1. A *fundamental solution of the heat equation on* M is a continuous function $p = p(x, y, t)$, defined on $M \times M \times (0, \infty)$, which is C^2 with respect to x, C^1 with respect to t, and which satisfies

$$L_x p = 0, \qquad \lim_{t \downarrow 0} p(\ , y, t) = \delta_y,$$

where δ_y is the Dirac delta function, that is, for all bounded continuous functions f on M we have, for every $y \in M$,

$$\lim_{t \downarrow 0} \int_M p(x, y, t) f(x)\, dV(x) = f(y). \tag{3}$$

Thus, $p(\ , y,\ )$ is the solution of the heat equation (1) resulting from an initial temperature distribution (i.e., at time $t = 0$), having total temperature 1, and completely concentrated at y.

If at time $t = 0$ we have an initial temperature distribution having total temperature α, and completely concentrated at y, then we expect the corresponding solution to the heat equation to be given by $\alpha p(\ , y,\ )$.

If at time $t = 0$ we are given the initial temperature distribution $f(y)$, then we expect the solution to the heat equation (1) to be obtained by "summing spatially" the contribution of each point to the initial data, namely,

$$u(x, t) = \int_M p(x, y, t) f(y)\, dV(y), \tag{4}$$

with

$$f(x) = \lim_{t \downarrow 0} \int_M p(x, y, t) f(y)\, dV(y). \tag{5}$$

If, in addition, heat, or refrigeration, is supplied to M, as described by $F(x, t)$ above, then the contribution of $F(\ , \tau)$, $\tau \in (0, t)$, to the temperature distribution at time t is given by

$$\int_M p(x, y, t - \tau) F(y, \tau)\, dV(y)$$

—so the total contribution to the temperature distribution, at time t, by F is given by

$$\int_0^t d\tau \int_M p(x, y, t - \tau) F(y, \tau)\, dV(y).$$

In sum, we expect the solution of the inhomogeneous heat equation (2), with initial data given by f, to be given by

$$u(x, t) = \int_M p(x, y, t) f(y)\, dV(y) + \int_0^t d\tau \int_M p(x, y, t - \tau) F(y, \tau)\, dV(y). \tag{6}$$

In Section 1, we derive formulas (4)–(6) as consequences of the existence and uniqueness of the heat kernel on M, in the case where M is compact. We also obtain the basic facts about the existence of eigenvalues, and their eigenfunctions, summarized in Theorem I.1 (and the ensuing discussion there) for the closed eigenvalue problem. Our treatment is, essentially, that of Milgram–Rosenbloom [1]. In Section 2 we consider the heat kernel on Euclidean space, and, in the subsequent sections, the existence of the heat kernel on compact Riemannian manifolds. In this latter topic, we follow the construction of Minakshisundaram [2]. The reader is also referred to Berger–Gauduchon–Mazet [1] and McKean–Singer [1].

1. DUHAMEL'S PRINCIPLE AND ITS CONSEQUENCES

In this section, we will assume that M is a given compact Riemannian manifold.

PROPOSITION 1. Let $u(x, t)$ be a solution to the heat equation (1) on M. Then

$$\int_M u(x, t)\, dV(x)$$

is a constant function of t, and

$$\int_M u^2(x, t)\, dV(x)$$

a decreasing function of t.

PROOF: For both cases, one differentiates with respect to t under the integral sign, applies the heat equation (1), and Green's formulas. Either result follows easily.

COROLLARY 1. Given continuous functions $f: M \to \mathbb{R}$ and $F: M \times (0, \infty) \to \mathbb{R}$, there exists at most one function $u: M \times [0, \infty) \to \mathbb{R} \in C^0$ satisfying (2), with initial data

$$u(\ ,0) = f. \tag{7}$$

Indeed, if u_1, u_2 were two solutions to the initial-value problem (2) : (7), then their difference, $v = u_2 - u_1$, would satisfy the heat equation (1), with identically vanishing initial data. The second claim of Proposition 1 then implies that v vanishes identically on $M \times [0, \infty)$. So $u_2 = u_1$.

Duhamel's principle (I): Let u, v be continuous functions on $M \times (0, t)$, both C^2 in the space variable and C^1 in the time variable. Then for all $[\alpha, \beta] \subseteq (0, t)$ we have

$$\begin{aligned} &\int_M \{u(z, t - \beta)v(z, \beta) - u(z, t - \alpha)v(z, \alpha)\}\, dV(z) \\ &= \int_\alpha^\beta d\tau \int_M \{(Lu)(z, t - \tau)v(z, \tau) - u(z, t - \tau)(Lv)(z, \tau)\}\, dV(z). \end{aligned} \tag{8}$$

PROOF: One has

$$\begin{aligned} (Lu)(z, t - \tau)v(z, \tau) &- u(z, t - \tau)(Lv)(z, \tau) \\ &= (\Delta u)(z, t - \tau)v(z, \tau) - u(z, t - \tau)(\Delta v)(z, \tau) \\ &\quad + (\partial/\partial\tau)\{u(z, t - \tau)v(z, \tau)\}. \end{aligned}$$

Then (8) is obtained by applying Green's formula, followed by integrating with respect to t.

THEOREM 1. Any fundamental solution to the heat equation on M is symmetric in the two space variables; also the fundamental solution is unique.

Assume M has the fundamental solution p. Then given bounded continuous functions $f\colon M \to \mathbb{R}$, $F\colon M \times (0, \infty) \to \mathbb{R}$, the solution to the initial-value problem (2) : (7), should it exist, is given by

$$(6) \qquad \begin{aligned} u(x, t) &= \int_M p(x, y, t) f(y)\, dV(y) \\ &\quad + \int_0^t d\tau \int_M p(x, y, t - \tau) F(y, \tau)\, dV(y). \end{aligned}$$

In particular,

$$(9) \qquad \int_M p(x, y, t)\, dV(y) = 1$$

on all of $M \times (0, \infty)$.

PROOF: Let p_1, p_2 be fundamental solutions of the heat equation on M; set

$$u(z, \tau) = p_1(z, x, \tau), \qquad v(z, \tau) = p_2(z, y, \tau),$$

and apply Duhamel's principle. Let $\alpha \downarrow 0$ and $\beta \uparrow t$. Then one obtains

$$(10) \qquad p_2(x, y, t) = p_1(y, x, t).$$

If we apply the argument to any given fundamental solution p, by setting $p_1 = p_2 = p$, we obtain

$$p(x, y, t) = p(y, x, t)$$

for all $x, y \in M$. So any fundamental solution is symmetric in the space variables. But then, when given arbitrary p_1, p_2, we may apply the symmetry to p_1, and (10) will then imply $p_1 = p_2$.

Suppose we are given the fundamental solution p, and the initial-value problem (2) : (7). Pick

$$v(z, \tau) = p(x, z, \tau),$$

substitute into Duhamel's principle, and let $a \downarrow 0$, $\beta \uparrow t$, as above. Then (6) will follow.

Thus, once one has the existence of a fundamental solution for the heat operator, one also has another proof for uniqueness of solutions to the initial-value problem (2) : (7).

We note that a third proof of uniqueness of solutions of the initial-value problem (2) : (7) will be given in Section VIII.1, via the maximum principle.

We now assume, for the rest of this section, the existence of a fundamental solution on M.

Given a continuous function f defined on M, then the formula

$$u(x, t) = \int_M p(x, y, t) f(y)\, dV(y) \tag{4}$$

does, indeed, provide a solution to the initial-value problem (1) : (7), since differentiation of $u(x, t)$ can be carried out under the integral sign. The point of view we shall take here is that for each fixed $t > 0$ formula (4) defines an integral operator $\mathfrak{P}_t$ acting on functions f on M. One immediately verifies

$$\mathfrak{P}_t \mathfrak{P}_{t_1} = \mathfrak{P}_{t+t_1}, \tag{11}$$

or equivalently,

$$p(x, z, t + t_1) = \int_M p(x, y, t) p(y, z, t_1)\, dV(y). \tag{12}$$

For each $t > 0$, the continuity and symmetry of $p(x, y, t)$ implies that $\mathfrak{P}_t$ is a self-adjoint, compact operator on $L^2(M)$. Also,

$$(\mathfrak{P}_t f, f) = \|\mathfrak{P}_{t/2} f\|^2 \geq 0,$$

which implies that $\mathfrak{P}_t$ is a positive operator for all $t > 0$. In particular,

$$p \geq 0$$

on $M \times M \times (0, \infty)$ [that $p > 0$ on $M \times M \times (0, \infty)$ requires more, for example, the maximum principle (Section VIII.1)].

If $f \in L^2(M)$, then $(\mathfrak{P}_t f)(x)$, given by (4), is a solution to the heat equation, and has whatever smoothness is possessed by $p(x, y, t)$. As $t \downarrow 0$ we have

$$\lim_{t \downarrow 0} \mathfrak{P}_t f = f$$

uniformly, when $f \in C^0(M)$; if we are only given that $f \in L^2(M)$, then the limit is valid in $L^2(M)$. This last claim is a consequence of the second claim of Proposition 1 (viz., $\|\mathfrak{P}_t f\|^2$ is a decreasing of t), and the L^2-density of continuous functions in $L^2(M)$.

The Sturm–Liouville decomposition: For M compact, there exists a complete orthonormal basis $\{\phi_0, \phi_1, \phi_2, \ldots\}$ of $L^2(M)$, consisting of eigenfunctions of Δ, with ϕ_j having eigenvalue λ_j satisfying

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \uparrow +\infty.$$

In particular, each eigenvalue has finite multiplicity. Each ϕ_j is as smooth as p; in particular, $p \in C^\infty$ implies $\phi_j \in C^\infty$ for every $j = 0, 1, 2, \ldots$.

Finally,

$$p(x, y, t) = \sum_{j=0}^{\infty} e^{-\lambda_j t}\phi_j(x)\phi_j(y) \tag{13}$$

with convergence absolute, and uniform, for each $t > 0$. In particular

$$\int_M p(x, x, t)\, dV(x) = \sum_{j=0}^{\infty} e^{-\lambda_j t}. \tag{14}$$

PROOF: We refer the reader to Riesz–Nagy [1, Chap. VI] for the basic facts, on self-adjoint, compact (or, completely continuous) operators, used in what follows.

For each $t > 0$, the self-adjoint compact operator $\mathfrak{P}_t$ has eigenvalues

$$\lambda_0(t) \geq \lambda_1(t) \geq \lambda_2(t) \geq \cdots \downarrow 0,$$

with respective eigenfunctions

$$\phi_0(t), \quad \phi_1(t), \quad \phi_2(t), \ldots,$$

forming a complete orthonormal basis of $L^2(M)$. The semigroup property of $\mathfrak{P}_t$, (10), then implies that for each $t > 0$, and positive integer k, we have

$$\lambda_j(kt) = (\lambda_j(t))^k, \qquad \phi_j(kt) = \phi_j(t), \tag{15}$$

from which we have the validity of (15) even when k is positive and rational. The continuity of $\mathfrak{P}_t$, with respect to t, then implies that (15) is valid for all $t, k > 0$. So

$$\lambda_j(t) = (\lambda_j(1))^t = e^{t(\ln \lambda_j(1))}, \qquad \phi_j(t) = \phi_j(1),$$

for all $t > 0$. Since $\|\mathfrak{P}_t f\|$ is decreasing in t, we have

$$\ln \lambda_j(1) \leq 0.$$

Set

$$\lambda_j =: -\ln \lambda_j(1), \qquad \phi_j =: \phi_j(1);$$

then

$$\mathfrak{P}_t\phi_j = e^{-\lambda_j t}\phi_j$$

for all $j = 0, 1, 2, \ldots$. In particular, ϕ_j is as smooth as p.

Next, we have

$$\begin{aligned} 0 = L(\mathfrak{P}_t\phi_j) &= e^{-\lambda_j t}\Delta\phi_j - \phi_j(d/dt)e^{-\lambda_j t} \\ &= e^{-\lambda_j t}\{\Delta\phi_j + \lambda_j\phi_j\}, \end{aligned}$$

which implies that ϕ_j is an eigenfunction of Δ with eigenvalue λ_j. Note that since $\lambda_j(1)$ is decreasing with respect to j, we have λ_j increasing, and since $\lambda_j(1) \downarrow 0$ as $j \to \infty$, we have $\lambda_j \uparrow +\infty$ as $j \to \infty$.

The Hilbert–Schmidt theory implies that (13) is valid in $L^2(M \times M)$, and the continuity of p, with the positivity of $\mathfrak{P}_t$, imply, via Mercer's theorem (Riesz–Nagy [1, p. 245]) that (13) is valid with respect to absolute and uniform convergence. Equation (14) follows immediately.

We now wish to show that for M compact, the "heat is evenly distributed over M as the time becomes large."

THEOREM 2. For any $f \in L^2(M)$, the function $\mathfrak{P}_t f$ converges uniformly, as $t \uparrow +\infty$, to a harmonic function on M. Since M is compact, the limit function is a constant.

PROOF: By Proposition 1, $\|\mathfrak{P}_t f\|^2$ is a decreasing function of t, with nonnegative limit as $t \uparrow +\infty$, say, $\alpha \geq 0$. But

$$\|\mathfrak{P}_t f - \mathfrak{P}_\tau f\|^2 = \|\mathfrak{P}_t f\|^2 - 2\|\mathfrak{P}_{(t+\tau)/2} f\|^2 + \|\mathfrak{P}_\tau f\|^2$$

then implies that $\mathfrak{P}_t f$ converges in $L^2(M)$.

Now for fixed $T > 0$, we have for all $x \in M$, $t, \tau \in (0, \infty)$,

$$\begin{aligned} |\mathfrak{P}_{t+T} f - \mathfrak{P}_{\tau+T} f|(x) &= |\mathfrak{P}_T(\mathfrak{P}_t f - \mathfrak{P}_\tau f)|(x) \\ &= \left| \int_M p(x, y, T)(\mathfrak{P}_t f - \mathfrak{P}_\tau f)(y)\, dV(y) \right| \\ &\leq \text{const} \cdot \|\mathfrak{P}_t f - \mathfrak{P}_\tau f\|, \end{aligned}$$

by the Cauchy–Schwarz inequality and the continuity of p. So $\mathfrak{P}_t f$ converges uniformly on M, as $t \to +\infty$, to a continuous function which we call Hf.

The argument just given also implies

$$\begin{aligned} |\mathfrak{P}_{t+h} f - \mathfrak{P}_t Hf|(x) &\leq \|\mathfrak{P}_h f - Hf\| \sup_x \|p(x,\ , t)\| \\ &\leq \|\mathfrak{P}_h f - Hf\| \sup_x \|p(x,\ , T)\| \end{aligned}$$

for all $t > T$, by Proposition 1. So

$$0 = \lim_{h \to \infty} \mathfrak{P}_{t+h} f - \mathfrak{P}_t Hf = Hf - \mathfrak{P}_t Hf,$$

that is,

$$\mathfrak{P}_t Hf = Hf$$

for all $t > 0$, from which one has that $Hf \in C^\infty$ and is harmonic.

2. THE HEAT EQUATION ON $\mathbb{R}^n$

DEFINITION 2. We set, for $(x, y, t) \in \mathbb{R}^n \times \mathbb{R}^n \times (0, \infty)$,

$$\mathrm{e}(x, y, t) = (4\pi t)^{-n/2} e^{-|x-y|^2/4t}.$$

THEOREM 3. The function $\mathrm{e}(x, y, t)$ is a fundamental solution to the heat equation on $\mathbb{R}^n$.

PROOF: The discovery of $\mathrm{e}(x, y, t)$, as a candidate for a fundamental solution, is easily motivated by taking the Fourier transform (with respect to the space variable) of a solution to the heat equation, solving the resulting ordinary differential equation with respect to the time variable, and then taking the inverse transform (cf. Section XII.4).

One easily verifies that $\mathrm{e}(\ , y,\ )$ is a solution to the heat equation. So we concentrate on the limit of

$$u(x, t) = \int_{\mathbb{R}^n} \mathrm{e}(x, y, t) f(y)\, dV(y) \tag{16}$$

as $t \downarrow 0$.

We shall actually consider a broader (than bounded and continuous) class of functions on $\mathbb{R}^n$.

DEFINITION 3. We say that a measurable function f on $\mathbb{R}^n$ is *locally in* $L^1(\mathbb{R}^n)$, written as $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, if to each $x \in \mathbb{R}^n$ there exists $\delta > 0$ so that

$$\int_{\mathbb{B}(x;\delta)} |f| < +\infty.$$

Now let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, such that, for every $\alpha > 0$, we have

$$f(x) = o(e^{\alpha|x|^2}) \tag{17}$$

as $|x| \to +\infty$. For any given $x \in \mathbb{R}^n$, set

$$y = x + r\xi, \qquad r = |y - x|, \quad \xi \in \mathbb{S}^{n-1}, \tag{18}$$

$$r^2 = 4ts^2. \tag{19}$$

Then

$$\begin{aligned}
&\int_{\mathbb{R}^n} \mathrm{e}(x, y, t) f(y)\, dV(y) \\
&\qquad = (4\pi t)^{-n/2} \int_0^\infty e^{-r^2/4t} r^{n-1}\, dr \int_{\mathbb{S}^{n-1}} f(x + r\xi)\, dA(\xi) \\
&\qquad = \pi^{-n/2} \int_0^\infty e^{-s^2} s^{n-1}\, ds \int_{\mathbb{S}^{n-1}} f(x + 2\sqrt{t} s\xi)\, dA(\xi).
\end{aligned}$$

So for (x, t) restricted to a bounded subset, K, of $\mathbb{R}^n \times (0, \infty)$, we have by (17),

$$f(x + 2\sqrt{t} s\xi) = o(e^{s^2/2})$$

as $s \to +\infty$, uniformly on $K \times \mathbb{S}^{n-1}$. Thus the integral $u(x, t)$ given by (16), converges absolutely, uniformly on K.

Note that the same argument will also establish that u, given by (16), is C^∞ on $\mathbb{R}^n \times (0, \infty)$, and that the derivatives of u may be calculated by differentiating the right-hand side of (16) under the integral sign. So u given by (16) is a solution to the heat equation.

Given the function $u(x, t)$ defined by (16), we wish to consider what happens when $t \downarrow 0$. If $f \equiv 1$ on $\mathbb{R}^n$, then $u \equiv 1$ on $\mathbb{R}^n \times (0, \infty)$ since

$$\int_{\mathbb{R}^n} \mathrm{e}(x, y, t)\, dV(y) = \pi^{-n/2} \mathrm{c}_{n-1} \int_0^\infty e^{-s^2} s^{n-1}\, ds = \pi^{-n/2} \int_{\mathbb{R}^n} e^{-|x|^2}\, dV(x) = 1$$

by the standard arguments (cf. Section XII.1).

If $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ and satisfies (17), then we have

$$\begin{aligned}
&f(x) - \int_{\mathbb{R}^n} \mathrm{e}(x, y, t) f(y)\, dV(y) \\
&\quad = \int_{\mathbb{R}^n} \mathrm{e}(x, y, t)\{f(x) - f(y)\}\, dV(y) \\
&\quad = \pi^{-n/2} \int_0^\infty e^{-s^2} s^{n-1}\, ds \int_{\mathbb{S}^{n-1}} \{f(x) - f(x + 2\sqrt{t} s\xi)\}\, dA(\xi) \\
&\quad \equiv \int_0^\delta (\cdots)\, ds + \int_\delta^\infty (\cdots)\, ds \\
&\quad \leq \left\{ \sup_{y \in \mathbb{B}(x; 2\sqrt{t}\delta)} |f(x) - f(y)| \right\} + \int_\delta^\infty o(e^{-s^2/2})\, ds.
\end{aligned}$$

One now easily concludes that if f is continuous at x, then

$$\lim_{t\downarrow 0}\int_{\mathbb{R}^n}\mathfrak{e}(x,y,t)f(y)\,dV(y)=f(x); \tag{20}$$

so $\mathfrak{e}(x,y,t)$ is a fundamental solution of the heat equation on $\mathbb{R}^n$.

COROLLARY 2. If f is continuous on $\mathbb{R}^n$, satisfying (17), for all $\alpha > 0$, as $|x| \to +\infty$, then (16) is a solution to the initial-value problem (1):(7).

Furthermore, if f is uniformly continuous on some set K, and

$$|f(x)-f(x+2\sqrt{ts}\xi)|=o(e^{s^2/2})$$

uniformly in $(x,t,\xi)\in K\times[0,T]\times\mathbb{S}^{n-1}$ for some $T>0$, then the convergence in (20) is uniform on K.

THEOREM 4. If f is a continuous function on $\mathbb{R}^n$ "vanishing at infinity", that is,

$$\lim_{|x|\to\infty}f(x)=0,$$

and u is the solution to the initial-value problem (1):(7), given by (16), then

$$u(x,t)\to 0,\qquad (\operatorname{grad}_x u)(x,t)\to 0$$

for each $t > 0$, as $|x| \to \infty$, that is, u and its gradient vanish at infinity for all $t > 0$.

PROOF: One easily has

$$\begin{aligned}|u(x,t)| &\le \int_{\mathbb{R}^n}\mathfrak{e}(x,y,t)|f(y)|\,dV(y)\\ &\le \left\{\max_{B(x;\delta)}|f|\right\}\int_{\mathbb{B}(x;\delta)}\mathfrak{e}(x,y,t)\,dV(y)\\ &\quad+\left\{\max_{\mathbb{R}^n}|f|\right\}\int_{\mathbb{R}^n\setminus\mathbb{B}(x;\delta)}\mathfrak{e}(x,y,t)\,dV(y)\\ &\le \max_{\mathbb{B}(x;\delta)}|f|+\left\{\max_{\mathbb{R}^n}|f|\right\}\frac{\mathfrak{c}_{n-1}}{(4_\pi t)^{n/2}}\int_\delta^\infty e^{-r^2/4t}r^{n-1}\,dr\end{aligned}$$

for all $(x,t)\in\mathbb{R}^n\times(0,\infty)$, and $\delta>0$. That $u(\ ,t)$ vanishes at ∞ follows easily.

Similarly, for each $j = 1, \ldots, n$,

$$\frac{\partial u}{\partial x^j}(x, t) = \int_{\mathbb{R}^n} \frac{\partial \mathbf{e}}{\partial x^j}(x, y, t) f(y)\, dV(y) \tag{21}$$

$$= \frac{\pi^{-n/2}}{\sqrt{t}} \int_0^\infty s^n e^{-s^2}\, ds \int_{\mathbb{S}^{n-1}} f(x + 2\sqrt{t}s\xi)\, dA(\xi),$$

via the substitutions (18), (19). Then one has, as for u,

$$\left|\frac{\partial u}{\partial x^j}(x, t)\right| \leq \left\{ \max_{\mathbb{B}(x;\, 2\sqrt{t}\delta)} |f| \right\} \frac{\pi^{-n/2}\mathbf{c}_{n-1}}{\sqrt{t}} \int_0^\infty s^n e^{-s^2}\, ds$$

$$+ \left\{ \max_{\mathbb{R}^n} |f| \right\} \frac{\pi^{-n/2}\mathbf{c}_{n-1}}{\sqrt{t}} \int_\delta^\infty s^n e^{-s^2}\, ds$$

from which we draw the desired conclusion.

We also note that if $f \in C^0(\mathbb{R}^n)$ is supported on the compact set K, then u satisfies the inequality

$$|u(x, t)| \leq (4\pi t)^{-n/2} e^{-d^2(x,K)/4t} \left\{ \iint_K |f|\, dV \right\}. \tag{22}$$

A similar inequality holds for $|\mathrm{grad}_x\, u|$.

Also, if $f \in L^1(\mathbb{R}^n)$ then u given by (16) converges and

$$\int_{\mathbb{R}^n} u(x, t)\, dV(x) = \int_{\mathbb{R}^n} f(x)\, dV(x) \tag{23}$$

for all $t > 0$—simply use Fubini's theorem.

To consider uniqueness questions, we give the Duhamel principle for normal domains.

Duhamel's principle (II): Let M be a Riemannian manifold with compact closure and smooth boundary, and let $u, v\colon \bar{M} \times (0, t) \to \mathbb{R} \in C^1$, and C^2 in the space variable varying over M. Then for any $[\alpha, \beta] \subseteq (0, t)$ we have

$$\iint_M \{u(z, t - \beta)v(z, \beta) - u(z, t - \alpha)v(z, \alpha)\}\, dV(z)$$

$$= \int_\alpha^\beta d\tau \iint_M \{(Lu)(z, t - \tau)v(z, \tau) - u(z, t - \tau)(Lv(z, \tau)\}\, dV(x)$$

$$+ \int_\alpha^\beta \int_{\partial M} \left\{u(w, t - \tau)\frac{\partial v}{\partial \nu}(w, \tau) - \frac{\partial u}{\partial \nu}(w, t - \tau)v(w, z)\right\} dA(w).$$

The proof is the same argument as the Duhamel principle in the compact case. One simply uses Green's formulas (II).

Now let u be a solution to the initial value problem (2) : (7) on $\mathbb{R}^n$. In the Duhamel principle, set

$$v(z, \tau) = \mathbf{e}(x, z, \tau)$$

and let $\alpha \downarrow 0$, $\beta \uparrow t$. Then we have for $M = \mathbb{B}(\delta)$,

$$\begin{aligned} u(x, t) = &\iint_{\mathbb{B}(\delta)} \mathbf{e}(x, y, t) f(y)\, dV(y) \\ &+ \int_0^t d\tau \iint_{\mathbb{B}(\delta)} \mathbf{e}(x, y, \tau) F(y, t - \tau)\, dV(y) \\ &+ \int_0^t d\tau \int_{\mathbb{S}(\delta)} \{u(w, t - \tau) \frac{\partial \mathbf{e}}{\partial r_w}(x, w, z) \\ &\qquad - \frac{\partial u}{\partial r_w}(w, t - \tau)\mathbf{e}(x, w, z)\}\, dA(w). \end{aligned}$$

So, if we know that u grows slowly enough so that the integral over $\mathbb{S}(\delta) \times (0, t)$ tends to 0 as $\delta \to +\infty$, then u is determined by the formula

$$\begin{aligned} u(x, t) = &\int_{\mathbb{R}^n} \mathbf{e}(x, y, t) f(y)\, dV(y) \\ &+ \int_0^t d\tau \int_{\mathbb{R}^n} \mathbf{e}(x, y, \tau) F(y, t - \tau)\, dV(y). \end{aligned} \tag{24}$$

One immediately has

THEOREM 5. If f is a bounded continuous function on $\mathbb{R}^n$, then $u(x, t)$ given by (16) is the unique solution to the initial-value problem for the heat equation (1) : (7) among all functions $v(x, t)$ on $\mathbb{R}^n \times (0, \infty)$ for which

$$|v(x, t)| \le \text{const}, \qquad |(\text{grad}_x\, u)(x, t)| \le \text{const}/\sqrt{t},$$

where the constants are independent of (x, t).

Finally, we consider a restricted case of the nonhomogeneous initial-value problem.

THEOREM 6. Given $F\colon \mathbb{R}^n \times [0, \infty) \to \mathbb{R} \in C^1$, such that supp $F \subseteq K \times [0, \infty)$, for some compact K in $\mathbb{R}^n$. Then a solution to the initial-value problem (2) : (7) is given by (24).

PROOF: By the linearity of the heat operator, and the fact that u given by (16) satisfies the homogeneous initial-value problem, namely, (1) : (7), it suffices to show that

$$u(x, t) = \int_0^t d\tau \int_{\mathbb{R}^n} \mathbf{e}(x, y, t)F(y, t - \tau)\, dV(y) \tag{25}$$

is a solution to the initial-value problem

$$LU = -F, \qquad \lim_{t \downarrow 0} U = 0. \tag{26}$$

Now u given by (25) satisfies, via the substitutions (18), (19),

$$u(x, t) = \int_0^t d\tau \int_0^\infty \pi^{-n/2} e^{-s^2} s^{n-1}\, ds \int_{\mathbb{S}^{n-1}} F(x + 2\sqrt{\tau} s\xi, t - \tau)\, dA(\xi), \tag{27}$$

from which we conclude that $u \in C^1$ for $t > 0$, and that differentiations of u with respect to the space variables may be obtained by differentiating under the integral sign in (27). So, using the chain rule, and integration by parts, we have

$$\begin{aligned}
\frac{\partial u}{\partial x^j}(x, t) &= \int_0^t d\tau \int_{\mathbb{R}^n} \mathbf{e}(x, y, t) \frac{\partial F}{\partial y^j}(y, t - \tau)\, dV(y) \\
&= -\int_0^t d\tau \int_{\mathbb{R}^n} \frac{\partial \mathbf{e}}{\partial y^j}(x, y, t)F(y, t - \tau)\, dV(y) \\
&= \int_0^t d\tau \int_{\mathbb{R}^n} \frac{\partial \mathbf{e}}{\partial x^j}(x, y, t)F(y, t - \tau)\, dV(y) \\
&= \int_0^t \tau^{-1/2}\, d\tau \int_0^\infty \pi^{-n/2} e^{-s^2} s^n\, ds \\
&\quad \cdot \int_{\mathbb{S}^{n-1}} F(x + 2\sqrt{\tau} s\xi, t - \tau)\xi^j\, dA(\xi).
\end{aligned}$$

Similarly, we have $\partial u/\partial x^j \in C^1$, and

$$\frac{\partial^2 u}{\partial x^{j2}}(x, t) = \int_0^t d\tau \int_{\mathbb{R}^n} \frac{\partial^2 \mathbf{e}}{\partial x^{j2}}(x, y, \tau)F(y, t - \tau)\, dV(y).$$

Finally, we have

$$\begin{aligned}\frac{\partial u}{\partial t}(x, t) &= \int_{\mathbb{R}^n} \mathbf{e}(x, y, t)F(y, 0)\, dV(y) \\ &\quad + \int_0^t d\tau \int_{\mathbb{R}^n} \mathbf{e}(x, y, \tau)\frac{\partial F}{\partial t}(y, t - \tau)\, dV(y) \\ &= \int_{\mathbb{R}^n} \mathbf{e}(x, y, t)F(y, 0)\, dV(y) \\ &\quad + \int_0^t d\tau \int_{\mathbb{R}^n} \mathbf{e}(x, y, t - \tau)\frac{\partial F}{\partial \tau}(y, \tau)\, dV(y)\end{aligned}$$

Now, for $0 < s < t$,

$$\begin{aligned}&\int_0^s d\tau \int_{\mathbb{R}^n} \mathbf{e}(x, y, t - \tau)\frac{\partial F}{\partial \tau}(y, \tau)\, dV(y) \\ &\quad = \int_{\mathbb{R}^n} \{\mathbf{e}(x, y, t - s)F(y, s) - \mathbf{e}(s, y, t)F(y, 0)\}\, dV(y) \\ &\quad\quad + \int_0^s d\tau \int_{\mathbb{R}^n} \frac{\partial \mathbf{e}}{\partial \tau}(x, y, t - \tau)F(y, \tau)\, dV(y),\end{aligned}$$

which implies, by letting $s \uparrow t$,

$$\frac{\partial u}{\partial t}(x, t) = F(x, t) + \int_0^t d\tau \int_{\mathbb{R}^n} \frac{\partial \mathbf{e}}{\partial t}(x, y, \tau)F(y, t - \tau)\, dV(y).$$

One now has that u, given by (25), satisfies (26).

3. THE MINAKSHISUNDARAM–PLEIJEL RECURSION FORMULAS

For any given Riemannian manifold M, with distance function $d(\ ,\ )$, we set

$$\mathscr{E}(x, y, t) = (4\pi t)^{-n/2}e^{-d^2(x,y)/4t}. \tag{28}$$

While there can be no pretense that $\mathscr{E}$, in general, will be the fundamental solution to the heat equation on M, it turns out that one can produce a fundamental solution by suitably tampering with $\mathscr{E}$. The first step is given in the following construction.

For convenience, we assume that M is complete, and that its injectivity radius is strictly positive. To each $y \in M$, we define the sequence of functions

$$u_j(\ ,y)\colon \mathsf{B}(y;\operatorname{inj}(M)) \to \mathbb{R},$$

by requiring that

$$u_0(y,y) = 0, \tag{29}$$

and that the functions

$$\mathsf{H}_k(x,y,t) = \mathscr{E}(x,y,t)\sum_{j=0}^{k} t^j u_j(x,y), \tag{30}$$

$k = 0, 1, 2, \ldots$, satisfy

$$L_x \mathsf{H}_k = \mathscr{E} t^k \Delta_x u_k. \tag{31}$$

To calculate the functions u_j, $j = 0, 1, 2, \ldots$, introduce geodesic spherical coordinates on $\mathsf{B}(y;\operatorname{inj}(M))$ by

$$x = \exp_y r\xi, \qquad r \in [0, \operatorname{inj}(M)), \qquad \xi \in \mathfrak{S}_y. \tag{32}$$

Then for functions $f, h\colon \mathsf{B}(y;\operatorname{inj}(M)) \to \mathbb{R}$ we have

$$\begin{aligned}&\langle \operatorname{grad} f, \operatorname{grad} h\rangle(\exp_y r\xi)\\ &\quad= \frac{\partial f}{\partial r}\frac{\partial h}{\partial r} + \langle \operatorname{grad}_{S(y;r)}(f \mid S(y;r)), \operatorname{grad}_{S(y;r)}(h \mid S(y;r))\rangle,\end{aligned}$$

where the right-hand side is evaluated at $\exp_y r\xi$. Also,

$$\Delta f = \frac{\partial_r(\sqrt{\mathfrak{g}}\,\partial_r f)}{\sqrt{\mathfrak{g}}} + \Delta_{S(y;r)}(f \mid S(y;r)),$$

where $\sqrt{\mathfrak{g}}$ is defined for the spherical coordinates, based at y, by Definition III.6.

Next, set

$$\varphi(x;y) = \sqrt{\mathfrak{g}}(r;\xi)/r^{n-1} \tag{33}$$

with y fixed and x given by (32). Then explicit calculation shows that (31) is equivalent to

$$(\partial_r u_0)/u_0 = -\tfrac{1}{2}(\partial_r \varphi)/\varphi,$$

and

$$\partial_r u_j + \left\{\frac{1}{2}\frac{\partial_r \varphi}{\varphi} + \frac{j}{r}\right\} u_j = \frac{\Delta u_j}{r}$$

for all $j = 1, 2, \ldots$, which implies, using (28), that

$$(34) \qquad \begin{aligned} u_0(x, y) &= \varphi^{-1/2}(x; y), \\ u_j(x, y) &= \varphi^{-1/2}(x; y) r^{-j} \int_0^r s^{j-1} (\varphi^{-1/2} \Delta u_{j-1})(\exp_y s\xi, y)\, ds \\ &= \varphi^{-1/2}(x; y) \int_0^1 \tau^{j-1} (\varphi^{-1/2} \Delta u_{j-1})(\exp_y \tau r \xi, y)\, d\tau, \end{aligned}$$

$j = 1, 2, \ldots$. Note that $u_j \in C^\infty$ on

$$\mathsf{B}_M =: \{(x, y) \in M \times M : d(x, y) < \operatorname{inj}(M)\}.$$

The above construction was first given in Minakshisundaram–Pleijel [1], and then applied by Minakshisundaram [2] to the construction of the fundamental solution of the heat equation on compact Riemannian manifolds. We leave it to the reader to verify, using Riemann normal coordinates (Section XII.8), that

$$(36) \qquad u_1(y, y) = \tfrac{1}{6} S(y)$$

where $S(y)$ is the scalar curvature of M at y. Calculation of the constants

$$a_j = \int_M u_j(y, y)\, dV(y)$$

have been considered in Berger [2], McKean–Singer [1], and Sakai [1]. Compare also Berger–Gauduchon–Mazet [1]. Of course, for $j = 1$ and $n = 2$, we have

$$(37) \qquad a_1 = 2\pi\chi(M)/3$$

by the Gauss–Bonnet theorem.

Next, we note that if M is the 3-dimensional space form $\mathbb{M}_\kappa$, of constant sectional curvature κ, then one has, directly,

$$u_j = \kappa^j \varphi^{-1/2}/j!,$$

from which one concludes that

$$(38) \qquad \mathsf{H}(x, y, t) =: \lim_{k \to \infty} \mathsf{H}_k(x, y, t) = \mathscr{E}(x, y, t) \varphi^{-1/2}(d(x, y)) e^{\kappa t}.$$

One easily checks that H is a fundamental solution of the heat equation on the 3-dimensional $\mathbb{M}_\kappa$ (Debiard–Gaveau–Mazet [1]). For $\kappa \geq 0$ we have already discussed its uniqueness. We consider the uniqueness question for $\kappa < 0$ in Section VIII.1.

It is an unpublished result of Millson (reported in Debiard–Gaveau–Mazet [1]) that if

$$p_n(x, y, t) = \varphi_n(d(x, y), t)$$

is the heat kernel of the n-dimensional $\mathbb{M}_\kappa$, then, with $\varphi = \varphi(r, t)$, we have

$$p_{n+2}(x, y, t) = -e^{n\kappa t}(\partial_r \varphi_n)(d(x, y), t)/2\pi\kappa \mathrm{S}_\kappa$$

is the heat kernel of the $(n + 2)$-dimensional $\mathbb{M}_\kappa$. We leave the direct verification of the result to the reader.

For explicit formulas of the heat kernel for 2-dimensional $\mathbb{M}_\kappa$, $\kappa < 0$, compare Section X.2.

4. EXISTENCE OF THE HEAT KERNEL

DEFINITION 4. Given a Riemannian manifold M, a *parametrix for L on M, H,* is a mapping $H: M \times M \times (0, \infty) \to \mathbb{R} \in C^\infty$ such that $L_x H$ extends to a continuous map of $M \times M \times [0, \infty)$ to $\mathbb{R}$, with

$$H(\ , y, t) \sim \delta_y, \qquad H(x, \ , t) \sim \delta_x$$

as $t \downarrow 0$.

In what follows, we let M be a compact Riemannian manifold, with injectivity radius $\mathrm{inj}(M) =: \varepsilon > 0$.

To construct a parametrix on M, fix a function $\rho: [0, \infty) \to [0, 1] \in C^\infty$, such that

$$\rho \,|[0, \varepsilon/4] = 1, \qquad \rho\,|[\varepsilon/2, \infty] = 0,$$

define $\eta: M \times M \to [0, 1] \in C^\infty$ by

$$\eta(x, y) = \rho(d(x, y)),$$

and, for each $k = 0, 1, 2, \ldots$, define

$$H_k = \eta \mathsf{H}_k,$$

where H_k is given by (30).

We now require two lemmata, whose proof we leave to the reader. A proof of Lemma 1 can be found in Berger–Gauduchon–Mazet [1, pp. 210–211], and a proof of Lemma 2 can be modeled on the argument of Theorem 6.

LEMMA 1. For every $k > n/2$, H_k is a parametrix for L on M.

Furthermore, if ∂_x^α denotes the differentiation in any chart $x: U \to \mathbb{R}^n$ on M,

$$\frac{\partial^{\alpha_1}}{(\partial x^1)^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{(\partial x^n)^{\alpha_n}},$$

with

$$|\alpha| = \alpha_1 + \cdots + \alpha_n,$$

in the first space variable, then

$$\partial_x^\alpha L_x H_k = t^{k-(n/2+|\alpha|)} e^{-d^2/4t} F_{k,\alpha}, \tag{39}$$

where $F_{k,\alpha} \in C^\infty(M \times M \times [0, \infty))$. Also, if ∂_t^l denotes the lth derivative in the time variable, then

$$\partial_t^l L_x H_k = t^{k-(n/2+2l)} e^{-d^2/4t} G_{k,l}, \tag{40}$$

where $G_{k,l} \in C^\infty(M \times M \times [0, \infty))$.

Finally, if N is any topological space, and f is any continuous function defined on $N \times M$, then for $k > n/2$ we have

$$\lim_{t \downarrow 0} \int_M f(z, x) H_k(x, y, t)\, dV(x) = f(z, y),$$

$$\lim_{t \downarrow 0} \int_M H_k(x, y, t) f(z, y)\, dV(y) = f(z, x);$$

and the above limits are $(N \times M)$-locally uniform. In particular, if $N \times M$ are both compact, then both limits are uniform on $N \times M$.

For any functions $F, G \in C^0(M \times M \times [0, \infty))$ define the *convolution*, $F * G$, by

$$(F * G)(x, y, t) = \int_0^t d\tau \int_M F(x, z, \tau) G(z, y, t - \tau)\, dV(z).$$

and note that

$$(F * G) * H = F * (G * H).$$

So we write, for $l = 1, 2, \ldots,$

$$(F*)^l = F * \cdots * F \qquad (l \text{ times}).$$

LEMMA 2. For $F \in C^0(M \times M \times [0, \infty))$, and $k > n/2$, we may form $H_k * F$ by the above formula. Then $H_k * F$ will be in the domain of L_x, and satisfy

$$L_x(H_k * F) = -F + (L_x H_k) * F. \tag{41}$$

Once one has Lemmas 1 and 2, the idea is to seek a fundamental solution $p(x, y, t)$ of the form

$$p = H_k + H_k * F. \tag{42}$$

Then (41), (42) imply that finding p, for which

$$L_x p = 0$$

is equivalent to finding F for which

$$L_x H_k = F - (L_x H_k) * F,$$

from which one has

$$F = \sum_{l=1}^{\infty} (L_x H_k)^{*l}. \tag{43}$$

To investigate the convergence of series (43), we note that

$$L_x H_k = t^{k-n/2} e^{-d^2/4t} \mathscr{G}_k,$$

where $\mathscr{G}_k \in C^\infty(M \times M \times [0, \infty])$. We also have

LEMMA 3. Let x, y, z belong to any metric space. Then, for any $\tau \in (0, t)$, one has

$$\frac{d^2(x, y)}{\tau} + \frac{d^2(y, z)}{t - \tau} \geq \frac{d^2(x, z)}{t}. \tag{44}$$

Indeed, one uses calculus to show that

$$\min_{\tau \in (0,t)} \left\{ \frac{d^2(x, y)}{\tau} + \frac{d^2(y, z)}{t - \tau} \right\} = \frac{\{d(x, y) + d(y, z)\}^2}{t},$$

and (44) follows.

Now fix $k > n/2$ and $T > 0$. Let $V = V(M)$,

$$A = \sup_{M \times M \times [0,T]} |\mathscr{G}_k|, \qquad B = A T^{k-n/2}.$$

Then for all $t \in [0, T]$ we have

$$\begin{aligned}|L_x H_k * L_x H_k|(x, y, t) &= \left|\int_0^t \tau^{k-n/2}(t-\tau)^{k-n/2}\,d\tau \int_M \mathscr{G}_k(x, z, \tau)\mathscr{G}_k(z, y, t-\tau)\right.\\ &\qquad \left.\times\, e^{-d^2(x,z)/4\tau} e^{-d^2(z,y)/4(t-\tau)}\,dV(z)\right|\\ &\le \frac{ABt^{k+1-n/2}V}{k+1-n/2}\,e^{-d^2(x,y)/4t}.\end{aligned}$$

An induction argument easily establishes that for all $l \ge 2$,

$$|(L_x H_k)^{*l}|(x, y, t) \le \frac{A(BV)^{l-1}t^{k+l-n/2-1}e^{-d^2(x,y)/4t}}{\{(k+l-n/2-1)\cdots(k+1-n/2)\}}.$$

So $e^{d^2(x,y)/4t}F$ converges absolutely and uniformly on $M \times M \times [0, T]$, for every $T > 0$. Also, since $k > n/2$, we have

$$\lim_{t \downarrow 0} F = 0$$

uniformly on $M \times M$.

Furthermore, a similar argument, using (39) and (40), shows that for $k > n/2 + 2l$, $F \in C^l(M \times M \times [0, \infty))$, and all differentiations may be carried out by differentiating the series term-by-term.

In summary, for $k > n/2 + 2$,

$$\begin{aligned}p(x, y, t) &= H_k(x, y, t) + ((L_x H_k) * F)(x, y, t)\\ &= \frac{e^{-d^2(x,y)/4t}}{(4\pi t)^{n/2}}\left\{\rho(d(x, y))\sum_{j=0}^{k} t^j u_j(x, y) + O(t^{k+1})\right\},\end{aligned} \tag{45}$$

with

$$u_0(x, x) = 0, \qquad u_1(x, x) = S(x)/6, \tag{46}$$

is a fundamental solution for L on M. Since the fundamental solution for L on M is unique, all fundamental solutions, obtained by varying $k > n/2 + 2$, coincide. The fundamental solution is, therefore, C^∞ on $M \times M \times (0, \infty)$. The Sturm–Liouville argument of Section 1 then implies that all the eigenfunctions of the Laplacian are C^∞.

Note that (14), (45) imply

$$\begin{aligned}\sum_{j=0}^{\infty} e^{-\lambda_j t} &= \int_M p(x, x, t)\,dV(x)\\ &= (4\pi t)^{-n/2}\left\{\sum_{j=0}^{k} t^j \int_M u_j(x, x)\,dV(x)\right\} + O(t^{k-n/2+1});\end{aligned}$$

in particular,

$$\sum_{j=0}^{\infty} e^{-\lambda_j t} = \frac{1}{(4\pi t)^{n/2}}\left\{V(M) + \frac{t}{6}\int_M S\,dV + O(t^2)\right\} \tag{47}$$

as $t \downarrow 0$. Thus

$$\sum_{j=0}^{\infty} e^{-\lambda_j t} \sim V(M)/(4\pi t)^{n/2} \tag{48}$$

as $t \downarrow 0$. The Karamata theorem (Feller [1, p. 446]) then implies

Weyl's asymptotic formula: Let M be a compact Riemannian manifold, with eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$, each distinct eigenvalue repeated according to its multiplicity. Then for

$$N(\lambda) =: \sum_{\lambda_j \leq \lambda} 1$$

we have

$$N(\lambda) \sim \omega_n V(M)\lambda^{n/2}/(2\pi)^n$$

as $\lambda \uparrow +\infty$. In particular,

$$(\lambda_k)^{n/2} \sim (2\pi)^n k/\omega_n V(M)$$

as $k \uparrow +\infty$.

Note also that (47) implies that knowledge of the spectrum of M determines the integral of the scalar curvature of M. In the two-dimensional case, (46) reads as

$$\sum_{j=0}^{\infty} e^{-\lambda_j t} \sim \frac{V(M)}{4\pi t} + \frac{\chi(M)}{6} \tag{49}$$

as $t \downarrow 0$, by the Gauss–Bonnet theorem (cf. Section III.1). Thus, in the 2-dimensional case, knowledge of the spectrum of M determines the topology of M.

Remark 1: In Donnelly [1] it is remarked that the argument of this section is also valid when M is noncompact, if M has a properly discontinuous group of isometries Γ acting on M such that the quotient $M_0 = \Gamma\backslash M$ is compact, for example, if M is a Riemannian cover of the compact Riemannian manifold M_0. In such a situation the heat kernels p, p_0 of M, M_0, respectively, are related by

$$p_0(x, y, t) = \sum_{\gamma\in\Gamma} p(x, \gamma\cdot y, t). \tag{50}$$

If M_0 is not a manifold, one still has the existence of p_0—in fact, defined by (50)—and spectrum, with asymptotic formula (47) (for M_0) with a correction to the coefficient of $t^{1-n/2}$.

Remark 2: Following Minakshisundaram [1, 2], and Minakshisundaram–Pleijel [1], one defines the zeta function

$$z(s) = \sum_{j=1}^{\infty} \lambda_j^{-s}$$

for complex s. Then the Weyl formula implies that for any $\varepsilon > 0$, the series $z(s)$ converges uniformly on $\{s : \operatorname{Re} s \geq n/2 + \varepsilon\}$, thereby defining an analytic function on $\{s : \operatorname{Re} s > n/2\}$.

To study the continuation of $z(s)$, we note that since the gamma function $\Gamma(s)$ (cf. Section XII.1) satisfies

$$\Gamma(s)a^{-s} = \int_0^{\infty} e^{-at} t^{s-1}\, dt,$$

we have, for $\operatorname{Re} s > n/2 + \varepsilon$,

$$\begin{aligned}\Gamma(s) \sum_{j=1}^{\infty} \lambda_j^{-s} &= \sum_{j=1}^{\infty} \Gamma(s)\lambda_j^{-s} = \sum_{j=1}^{\infty} \int_0^{\infty} e^{-\lambda_j t} t^{s-1}\, dt \\ &= \int_0^{\infty} \left(\sum_{j=1} e^{-\lambda_j t}\right) t^{s-1}\, dt = \int_0^1 + \int_1^{\infty}.\end{aligned}$$

It is easy to see that

$$\int_1^{\infty} \left(\sum_{j=1}^{\infty} e^{-\lambda_j t}\right) t^{s-1}\, dt$$

can be analytically continued, in s, to all of $\mathbb{C}$.

Given any nonnegative integer N, we have for $s \geq n/2 + 1 - N$,

$$\begin{aligned}&\int_0^1 \left(\sum_{j=1}^{\infty} e^{-\lambda_j t}\right) t^{s-1}\, dt \\ &\quad = \int_0^1 \left\{-1 + \int_M p(x, x, t)\, dV(x)\right\} t^{s-1}\, dt \\ &\quad = -1/s + \int_0^1 \left[\left\{\sum_{k=0}^{N} a_k t^{k+s-1-n/2}\right\} + O(t^{s+N-n/2})\right] dt \\ &\quad = -1/s + \sum_{k=0}^{N} a_k/(s - n/2 + k) + O(1).\end{aligned}$$

So $\Gamma(s)z(s)$ can be continued to a meromorphic function on the complete s plane, with simple zeros at $s = 0$, and at $s = n/2 - k$, $k = 0, 1, 2, \ldots$. Thus $z(s)$ also has an analytic continuation as a meromorphic function, with the location of guaranteed poles and zeros depending on whether n is even or odd.

One can also give a similar discussion for the zeta function

$$z(s; x, y) = \sum_{j=1}^{\infty} \varphi_j(x)\varphi_j(y)\lambda_j^{-s},$$

where $\{\varphi_0, \varphi_1, \varphi_2, \ldots\}$ is a complete orthonormal sequence in $L^2(M)$ consisting of eigenfunctions of Δ, with λ_j the eigenvalue of φ_j.

When M is a compact Riemann surface of constant negative curvature, a detailed study of $z(s)$ and $z(s; x, y)$ is possible, with consequences for the asymptotic growth of eigenvalues, and can be found in Randol [2, 8]. For the general Riemannian case, we refer the reader to Duistermaat–Guillemin [1].

Remark 3: We refer the reader to Donnelly–Li [1, 2] wherein it is shown how to go from an upper bound on

$$\sum_{j=0}^{\infty} e^{-\lambda_j t}$$

valid for all $t > 0$, to a lower bound on $\lambda_k k^{-2/n}$ for large k. The result is that if there exist constants $c_1 > 0$, $c_2, c_3 \geq 0$ such that

$$\sum_{j=0}^{\infty} e^{-\lambda_j t} \leq e^{c_3 t}(c_1 t^{-n/2} + c_2 t) \tag{51}$$

for all $t > 0$, then given any $c_4 > 0$, there exists $c_5 = c_5(c_1, c_2, c_3, c_4) > 0$ such that

$$\lambda_k \geq c_4 > 0 \quad \Rightarrow \quad \lambda_k \geq c_5 k^{-2/n}.$$

If $c_2, c_3 = 0$ in (51), then one easily has

$$\lambda_k \geq \{c_1 e\}^{-2/n} k^{2/n}$$

for all $k > 0$. Indeed, we would then have

$$ke^{-\lambda_k t} \leq \sum_{j=0}^{\infty} e^{-\lambda_j t} \leq c_1 t^{-n/2}$$

for all $t > 0$. Now set $t = 1/\lambda_k$.

Also discussed in these papers are geometric conditions, in the spirit of Section VIII.3, for which (51) is valid. Also, compare Cheng–Li [1].

CHAPTER VII

The Dirichlet Heat Kernel for Regular Domains

In this chapter we carry out the program of Chapter VI for the Dirichlet eigenvalue problem on a regular domain. Given a regular domain M, and a fundamental solution to the heat equation on M, $q = q(x, y, t)$, we say that q is a *Dirichlet heat kernel of* M if q is extendable to a continuous function on $\bar{M} \times \bar{M} \times (0, \infty)$ such that

$$q(\ , y, t) \,|\, \partial\Omega = 0$$

for all $(y, t) \in M \times (0, \infty)$. In the context of the physical interpretation of the fundamental solution, discussed in the introduction to Chapter VI, $q(x, y, t)$ is the heat distribution at x at time t resulting from one unit of heat concentrated at y at time $t = 0$, *and* subject to absolute refrigeration of ∂M for all time $t > 0$.

Nearly all that we do, here, has a corresponding formulation for the Neumann eigenvalue problem. We prefer the Dirichlet eigenvalue problem because of the powerful applications, of the domain monotonicity of the Dirichlet heat kernel, that are developed in the following chapters.

Another qualification of our treatment, here, is in order. We will not show that the Dirichlet heat kernel q of M can be extended to a C^1 function, in each of the space variables, on $\bar{M}$. Rather, we shall only show that the gradient vector field of q, with respect to each of the space variables, can be extended, in that variable, to a continuous vector field on $\bar{M}$. This less ambitious result suffices for most of the applications already presented, in earlier chapters, and those contemplated in the coming chapters.

Our construction of the Dirichlet heat kernel follows that of Minakshisundaram [1, Sec. 4], and, the application to the existence of Dirichlet eigenvalues, that of Milgram–Rosenbloom [1]. We also refer the reader to Dodziuk [1].

1. PRELIMINARIES

We start with noting that for all $x > 0$, $\alpha > 0$,

$$x^\alpha e^{-x} \leq \alpha^\alpha e^{-\alpha}.$$

In particular, for any $\mu > 0$, $n \geq 1$, we have the existence of a constant, depending on μ, n, such that for any $r > 0$,

$$r^{-(n/2+1)} e^{-r^2/4t} \leq \text{const} \cdot t^{-\mu} r^{-(n+2-2\mu)}. \tag{1}$$

DEFINITION 1. Let M be a Riemannian manifold with heat kernel $p(x, y, t)$. We say that p is *almost Euclidean* if p and $\mathscr{E}$ [given by (IV.28)] are of the same order, locally uniformly in (x, y), as $t \downarrow 0$, and if a similar statement holds for the first derivatives of p and $\mathscr{E}$.

Certainly, the heat kernel of any compact Riemannian manifold is almost Euclidean.

If M is an n-dimensional Riemannian manifold with almost Euclidean heat kernel, Ω a regular domain in M with boundary Γ carrying the outward unit normal vector field ν, then given any $\mu > 0$, and $T > 0$, there exists a constant, depending on n, μ, T, Γ, M, such that

$$\left|\frac{\partial p}{\partial \nu_w}(w_0, w, t)\right| \leq \text{const} \cdot t^{-\mu} d^{-n+2\mu}(w_0, w) \tag{2}$$

for all $w_0, w \in \Gamma$, and $t \in (0, T]$. The improvement in the exponent of $d(w_0, w)$ in (2) over the exponent of r in (1) is due to the fact that the derivative, here in (2), is in the direction normal to Γ.

We also leave to the reader to verify that if M, Ω, Γ are as above, then there exists a constant such that, for every $\alpha, \beta \in (0, n-1)$, we have

$$\int_\Gamma d^{-\alpha}(w_1, w) d^{-\beta}(w, w_2)\, dA(w) \leq \text{const} \cdot d^{n-1-(\alpha+\beta)}(w_1, w_2) \tag{3}$$

for all $w_1, w_2 \in \Gamma$, $w_1 \neq w_2$.

THEOREM 1 (the jump relation). Let M be a Riemannian manifold, Ω a regular domain in M, with boundary Γ carrying the outward unit normal vector field ν. Let $p(x, y, t)$ be a C^∞ almost Euclidean heat kernel on M. Let $\psi(w, t)$ be a continuous function on $\Gamma \times [0, T]$, and $u(x, t)$ defined on $\Omega \times [0, T]$ by

$$u(x, t) = -\int_0^t d\tau \int_\Gamma \frac{\partial p}{\partial \nu_w}(x, w, \tau)\psi(w, t-\tau)\, dA(w). \tag{4}$$

Then $u \in C^\infty(\Omega \times [0, T])$, with

$$u(\ , 0) = 0;$$

and for $w_0 \in \Gamma$ we have

$$(5)\qquad \lim_{x \to w_0} u(x, t) = \frac{\psi(w_0, t)}{2} - \int_0^t d\tau \int_\Gamma \frac{\partial p}{\partial \nu_w}(w_0, w, \tau)\psi(w, t - \tau)\, dA(w).$$

PROOF: We only give the proof for $M = \mathbb{R}^n$, $\psi = 1$, and Ω having the property that $\Gamma \cap \mathbb{B}(w_0; \rho)$ is a flat $(n-1)$-disk, for some $\rho > 0$. All other cases are technical generalizations of this one.

In the case under consideration, the heat kernel is given by

$$\mathbf{e}(x, y, t) = (4\pi t)^{-n/2} e^{-|x-y|^2/4t}$$

and its gradient by

$$(\operatorname{grad}_y \mathbf{e})(x, y, t) = (4\pi t)^{-n/2} e^{-|x-y|^2/4t}(x - y)/2t.$$

The convergence of the integral in (4) presents no difficulty, by virtue of the estimates (2).

Now let $\Gamma_1 = \Gamma \cap \mathbb{B}(w_0; \rho_1)$, where $\rho_1 \in (0, \rho)$, and let $\Gamma_2 = \Gamma \backslash \Gamma_1$. Also, set $\nu_0 = \nu(w_0)$.

For $x = w_0 - \alpha\nu_0$ we certainly have

$$\lim_{\alpha \downarrow 0} -\int_0^t d\tau \int_{\Gamma_2} \frac{\partial \mathbf{e}}{\partial \nu_w}(w_0 - \alpha\nu_0, w, \tau)\, dA(w) = -\int_0^t d\tau \int_{\Gamma_2} \frac{\partial \mathbf{e}}{\partial \nu_w}(w_0, w, \tau)\, dA(w).$$

For $x = w_0 - \alpha\nu_0$, $w \in \Gamma_1$, we have

$$(w - w_0) \cdot \nu_0 = 0,$$

which implies

$$|w - x|^2 = |w - w_0|^2 + \alpha^2.$$

Therefore,

$$\begin{aligned}
-\int_0^t d\tau \int_{\Gamma_1} &\frac{\partial \mathbf{e}}{\partial \nu_w}(w_0 - \alpha\nu_0, w, \tau)\, dA(w) \\
&= \int_0^t d\tau \int_0^{\rho_1} \mathbf{c}_{n-2}\, e^{-\alpha^2/4\tau - r^2/4\tau} \alpha 2^{-(n+1)} \pi^{-n/2} \tau^{-(1+n/2)} r^{n-2}\, dr \\
&= \alpha \mathbf{c}_{n-2}\, \pi^{-n/2} 2^{-(n+1)} \int_0^t \tau^{-(1+n/2)} e^{-\alpha^2/4\tau}\, d\tau \int_0^{\rho_1} e^{-r^2/4\tau} r^{n-2}\, dr \\
&= (\alpha \mathbf{c}_{n-2}\, \pi^{-n/2}/4) \int_0^t \tau^{-3/2} e^{-\alpha^2/4\tau}\, d\tau \int_0^{\rho_1/2\tau^{1/2}} e^{-s^2} s^{n-2}\, ds \\
&= \frac{\mathbf{c}_{n-2}}{2\pi^{n/2}} \int_{\alpha^2/4t}^{\infty} \mu^{-1/2} e^{-\mu}\, d\mu \int_0^{\rho_1 \mu^{1/2}/\alpha} e^{-s^2} s^{n-2}\, ds \\
&\to \frac{\mathbf{c}_{n-2}}{2\pi^{n/2}} \int_0^{\infty} \mu^{-1/2} e^{-\mu}\, d\mu \int_0^{\infty} e^{-s^2} s^{n-2}\, ds = \tfrac{1}{2}
\end{aligned}$$

as $\alpha \downarrow 0$, for every $\rho_1 > 0$. One, indeed, has to check that one can pass to the limit, as indicated. However, that is not difficult. Evaluation of the limit is done via Section XII.1.

The jump relation (5) is called the *jump relation for the double layer potential* (4). The *jump relation for the single-layer potential* is as follows:

THEOREM 2. Let $M, p, \Omega, \Gamma, \nu, \psi$ be given as in Theorem 1, and define $v: \Omega \times [0, T] \to \mathbb{R}$ by

$$v(x, t) = \int_0^t d\tau \int_\Gamma p(x, w, z)\psi(w, t - \tau)\, dA(w). \tag{6}$$

Then $v \in C^\infty(\Omega \times [0, T])$ extends to a continuous function on $\bar{\Omega} \times [0, T]$, and satisfies

$$v(\ , 0) = 0.$$

Furthermore, if for $w_0 \in \Gamma$ we let

$$\gamma(s) = \gamma_{\nu(w_0)}(s),$$

then

$$\begin{aligned} &\lim_{s \uparrow 0} \langle (\operatorname{grad}_x v)(\gamma(s), t), \gamma'(s) \rangle \\ &= \frac{\psi(w_0, t)}{2} + \int_0^t d\tau \int_\Gamma \frac{\partial p}{\partial \nu_{w_0}}(w_0, w, \tau)\psi(w, t - \tau)\, dA(w). \end{aligned} \tag{7}$$

2. THE DIRICHLET HEAT KERNEL FOR REGULAR DOMAINS

We start the discussion by considering the following initial-boundary-value problem: we are given a regular domain Ω, with boundary Γ, in a Riemannian manifold M. We are also given a continuous function $\varphi: \Gamma \times (0, \infty) \to \mathbb{R}$; what we seek is a solution $u: \bar{\Omega} \times [0, \infty) \to \mathbb{R} \in C^0$ to the heat equation on $\Omega \times (0, \infty)$ satisfying

$$u \mid \Omega \times \{0\} = 0, \qquad u \mid \Gamma \times (0, \infty) = \varphi. \tag{8}$$

We assume that M possesses an almost Euclidean heat kernel. For if it does not, one may replace the complement of Ω with a Riemannian manifold, having boundary Γ, whose union with $\bar{\Omega}$ is a C^∞ compact Riemannian manifold. Simply pick Λ to be a relatively compact domain in M, having smooth boundary, and containing $\bar{\Omega}$; let M' be the double of Λ

with the metric smoothed out across the boundary of Λ, and unchanged in a neighborhood of $\bar{\Omega}$ (cf. Duff [1]). This new Riemannian manifold is compact, contains Ω, and possesses an almost Euclidean heat kernel.

To find a solution to the heat equation, satisfying the initial-boundary values given in (8), we attempt to determine a function $\psi(w, \tau)$ on $\Gamma \times [0, \infty)$ so that

$$u(x, t) = -\int_0^t d\tau \int_\Gamma \frac{\partial p}{\partial \nu_w}(x, w, t - \tau)\psi(w, \tau)\, dA(w). \tag{9}$$

Should such a ψ exist, the jump relation (5) would imply, for $w_0 \in \Gamma$,

$$\varphi(w_0, t) = \psi(w_0, t)/2 - \int_0^t d\tau \int_\Gamma \frac{\partial p}{\partial \nu_w}(w_0, w, t - \tau)\psi(w, \tau)\, dA(w),$$

that is,

$$\psi(w_0, t) = 2\varphi(w_0, t) + 2\int_0^t d\tau \int_\Gamma \frac{\partial p}{\partial \nu_w}(w_0, w, t - \tau)\psi(w, \tau)\, dA(w). \tag{10}$$

A formal solution to (10) is discovered, via iterating (10) itself, to be

$$\psi(w_0, t) = 2\varphi(w_0, t) + 2\int_0^t d\tau \int_\Gamma \mathscr{M}(w_0, w, t - \tau)\varphi(w, \tau)\, dA(w)$$

for all $(w_0, t) \in \Gamma \times [0, \infty)$, where

$$\mathscr{M}(w_0, w, t) = \sum_{k=1}^{\infty} M_k(w_0, w, t),$$

$$M_1(w_0, w, t) = 2\frac{\partial p}{\partial \nu_w}(w_0, w, t),$$

and

$$M_{l+1}(w_0, w, t) = 2\int_0^t d\tau \int_\Gamma M_l(w_0, \bar{w}, t - \tau)\frac{\partial p}{\partial \nu_w}(\bar{w}, w, \tau)\, dA(\bar{w})$$

for $l \geq 1$.

If we then set

$$\mathscr{F}(x, w, t) = 2\frac{\partial p}{\partial \nu_w}(x, w, t) + 2\int_0^t d\tau \int_\Gamma \frac{\partial p}{\partial \nu_w}(x, \bar{w}, t - \tau)\mathscr{M}(\bar{w}, w, \tau)\, dA(\bar{w}),$$

then we may rewrite (9) as

$$u(x, t) = -\int_0^t d\tau \int_\Gamma \mathscr{F}(x, w, t - \tau)\varphi(w, \tau)\, dA(w) \tag{11}$$

and we have a *formal* solution of the initial-boundary value problem (8) for the heat equation. It remains to verify whether the candidate is a good one.

The first thing to investigate is the convergence of $\mathscr{M}(w_0, w, t)$. Let, as usual $n = \dim M$, and fix a number μ in $(\frac{1}{2}, 1)$ and $T > 0$. By (2), we have

$$|M_1(w_0, w, t)| \leq \mathrm{const} \cdot t^{-\mu} d^{2\mu - n}(w_0, w)$$

for all distinct $w_0, w \in \Gamma$, $t \in (0, T]$. Now, if $\alpha, \beta > 0$, and

$$|M_l(w_0, w, t)| \leq \mathrm{const} \cdot t^{-\alpha} d^{-\beta}(w_0, w),$$

then one easily has

$$|M_{l+1}(w_0, w, t)| \leq \mathrm{const} \cdot t^{1-\mu-\alpha} d^{2\mu-1-\beta}(w_0, w).$$

So the power of t increases by $1 - \mu$, and the power of $d(w_0, w)$ by $2\mu - 1$. Ultimately, at least one will become nonnegative.

Next, note that

$$\int_0^t (t - \tau)^\alpha \tau^{-\mu} \, d\tau = \frac{\Gamma(1 + \alpha)\Gamma(1 - \mu)}{\Gamma(2 + \alpha - \mu)} t^{\alpha + (1 - \mu)}$$

for all $\alpha \geq 0$, where $\Gamma(x)$ is the classical gamma function (cf. Section XII.1); and note that if $\alpha \in \mathbb{R}$, then the validity of

$$|M_l(w_0, w, t)| \leq m_1 t^\alpha, \tag{12}$$

for some fixed constant m_1, on all of $\Gamma \times \Gamma \times (0, T]$, implies that

$$|M_{l+1}(w_0, w, t)| \leq m_1 m_2 \int_0^t (t - \tau)^\alpha \tau^{-\mu} \, d\tau,$$

where

$$m_2 = \sup_{w \in \Gamma} \int_\Gamma \mathrm{d}^{2\mu - n}(\overline{w}, w) \, dA(\overline{w}) < +\infty.$$

We can, therefore, conclude that there exists $l \geq 1$ and positive constants $m_0, \alpha > 0$, such that (12) is valid on $\Gamma \times \Gamma \times (0, T]$. Thus, for all $k \geq 1$, we have

$$|M_{l+k}(w_0, w, t)| \leq m_1(m_2) t^{\alpha + k(1-\mu)} \Gamma(\alpha + k)\Gamma(1 - \mu)/\Gamma(\alpha + k + 1 - \mu),$$

which implies that

$$\sum_{k=1}^{\infty} M_{l+k}(w_0, w, t)$$

converges absolutely (with exponential speed), and uniformly on $\Gamma \times \Gamma \times [0, T]$ for all $T > 0$.

Thus, the series

$$t^{\mu}d^{n-2\mu}(w_0, w)\mathscr{M}(w_0, w, t)$$

converges absolutely and uniformly on $\Gamma \times \Gamma \times [0, T]$ for all $T > 0$. Therefore, for all $T > 0$, *$\mathscr{F}(x, w, t)$ is continuous on $\Omega \times \Gamma \times [0, T]$, C^∞ in the x variable, and uniformly bounded on $K \times \Gamma \times [0, T]$ for compact $K \subseteq \Omega$.* One now verifies directly that u given by (11) is a solution to the initial-boundary-value problem.

To obtain the existence of a Dirichlet heat kernel on Ω, let Ω be contained in the Riemannian manifold M, having almost Euclidean heat kernel $p(x, y, t)$. For each $y \in \Omega$ solve the initial-boundary-value problem for the heat equation with initial-boundary data given by (8), with

$$\varphi(w, t) = -p(w, y, t).$$

The solution is given by, using (11),

$$h(x, y, t) = \int_0^t d\tau \int_\Gamma \mathscr{F}(x, w, t - \tau)p(w, y, \tau)\, dA(w).$$

One easily verifies that

$$Q(x, y, t) \equiv: p(x, y, t) + h(x, y, t)$$

is a Dirichlet heat kernel on Ω.

Note that given a compact set $K \subseteq \Omega$, and $T > 0$, we have a positive constant for which

$$|h(x, y, t)| \leq \text{const} \cdot t^{-n/2}e^{-d^2(y,\Gamma)/8t} \tag{13}$$

for all $(x, y, t) \in K \times \Omega \times (0, T]$.

Also note that the jump relation, for the single-layer potential (Theorem 2), implies that $(\text{grad}_y\, Q)(x, y, t)$ can be extended to a continuous vector field on $\bar{\Omega}$.

3. DUHAMEL'S PRINCIPLE

In what follows, Ω is a regular domain in a fixed Riemannian manifold M, and has boundary Γ carrying the outward unit normal vector field ν.

DEFINITION 2. If $v: \Omega \to \mathbb{R} \in C^1$, we say that $v \in \bar{C}^1(\Omega)$, if v extends to a continuous function on $\bar{\Omega}$, and grad v extends to a continuous vector field on $\bar{\Omega}$. If $v \in \bar{C}^1(\Omega)$, then for any $w \in \Gamma$, we let $\gamma_w(s)$ denote the geodesic emanating from w, with initial velocity vector $\nu(w)$, and define the *normal derivative of* v *at* $w \in \Gamma$, $(\partial v/\partial \nu)(w)$, by

$$\frac{\partial v}{\partial \nu}(w) = \lim_{s \uparrow 0} \langle (\operatorname{grad} v)(\gamma_w(s)), \gamma_w'(s) \rangle. \tag{14}$$

One now easily proves

THEOREM 3. Let $u \in C^0(\bar{\Omega}) \cap C^1(\Omega)$ and $v \in \bar{C}^1(\Omega) \cap C^2(\Omega)$. Then we have the extended Green's formula

$$\iint_\Omega \{u\Delta v + \langle \operatorname{grad} u, \operatorname{grad} v \rangle\}\, dV = \int_\Gamma u(\partial v/\partial \nu)\, dA. \tag{15}$$

If u is also in $\bar{C}^1(\Omega) \cap C^2(\Omega)$, then we also have

$$\iint_\Omega \{u\Delta v - v\Delta u\}\, dV = \int_\Gamma \{u(\partial v/\partial \nu) - v(\partial u/\partial \nu)\}\, dA. \tag{16}$$

For $v = v(x, t)$ defined on $\Omega \times (\alpha, \beta)$ and C^1 there, we say that $v \in \bar{C}^1$ on $\Omega \times (\alpha, \beta)$ if v extends to a continuous function on $\bar{\Omega} \times (\alpha, \beta)$, and $\operatorname{grad} v$ (taken with respect to x and t) to a continuous vector field on $\bar{\Omega} \times (\alpha, \beta)$. For $w \in \Gamma$ we, now, define $(\partial v/\partial \nu_w)(w, t)$ by

$$(\partial v/\partial \nu_w)(w, t) = \lim_{s \uparrow 0} \langle (\operatorname{grad}_x v)(\gamma_w(s), t), \gamma_w'(s) \rangle, \tag{17}$$

where $\gamma_w(s)$ is the geodesic, in M, emanating from w with initial velocity vector $\nu(w)$.

PROPOSITION 1. Let $u \in \bar{C}^1(\Omega \times (\alpha, \beta))$ be a solution of the heat equation

$$Lu = 0 \tag{18}$$

on $\Omega \times (\alpha, \beta)$. Then for the functions

$$U(t) = \iint_\Omega u(x, t)\, dV(x), \tag{19}$$

$$\mathfrak{U}(t) = \tfrac{1}{2} \iint_\Omega u^2(x, t)\, dV(x), \tag{20}$$

we have

$$U'(t) = \int_\Gamma (\partial u/\partial \nu_w)(w, t)\, dA(w) \tag{21}$$

$$\mathfrak{U}'(t) = -\iint_\Omega |\mathrm{grad}_x\, u|^2(x, t)\, dV(x) + \int_\Gamma (u\, \partial u/\partial \nu_w)(w, t)\, dA(w). \tag{22}$$

One, simply, differentiates each of the above, with respect to t, under the integral sign, applies the heat equation, and then, the Green's formula (15).

COROLLARY 1. Suppose we are given continuous functions $F\colon \Omega \times (0, \infty) \to \mathbb{R}$, $\varphi\colon \Gamma \times (0, \infty) \to \mathbb{R}$, $f\colon \Omega \to \mathbb{R}$. Then there exists at most one continuous function

$$u\colon (\bar{\Omega} \times [0, \infty))\backslash(\Gamma \times \{0\}) \to \mathbb{R} \in \bar{C}^1 \cap C^2(\Omega \times (0, \infty))$$

which satisfies the differential equation

$$Lu = -F, \tag{23}$$

with the initial-boundary conditions

$$u \,|\, \Gamma \times (0, \infty) = \varphi, \tag{24}$$

$$u \,|\, \Omega \times \{0\} = f. \tag{25}$$

PROOF: Given u_1, u_2 satisfying (23)–(25), one applies (22) to $u \equiv: u_1 - u_2$ to conclude $\mathfrak{U}'(t) \le 0$. But $\mathfrak{U}(0) = 0$. Therefore $u = 0$ on all of $\Omega \times [0, \infty)$, which implies the claim.

We note that by using the maximum principle (cf. Section VIII.1), one can obtain the uniqueness result of Corollary 1, without the assumption that $u \in \bar{C}^1(\Omega \times (0, \infty))$.

We recall the *Duhamel principle* (II) (for normal domains) (Section VI.2). Here $u, v \in (\bar{C}^1 \cap C^2)(\Omega \times (0, t))$, and $[\alpha, \beta] \subseteq (0, t)$. Then

$$\begin{aligned} &\int_\alpha^\beta d\tau \iint_\Omega \{(Lu)(x, t - \tau)v(z, \tau) - u(z, t - \tau)(Lv)(z, \tau)\}\, dV(z) \\ &= \int_\alpha^\beta d\tau \int_\Gamma \left\{\frac{\partial u}{\partial \nu_w}(w, t - \tau)v(w, \tau) - u(w, t - \tau)\frac{\partial v}{\partial \nu_w}(w, \tau)\right\} dA(w) \\ &\quad + \iint_\Omega \{u(z, t - \beta)v(z, \beta) - u(z, t - \alpha)v(z, \alpha)\}\, dV(z). \end{aligned} \tag{26}$$

THEOREM 4. Any Dirichlet heat kernel $q(x, y, t)$ on Ω, having the property that $q(\ , y,\ ) \in \bar{C}^1(\Omega \times (0, \infty))$ for all $y \in \Omega$, is symmetric in the space variables x and y.

Furthermore, Ω has precisely one Dirichlet heat kernel, $q(x, y, t)$, with the property that $q(\ , y,\ ) \in \bar{C}^1(\Omega \times (0, \infty))$ for all $y \in \Omega$.

PROOF: For u and v in (26) we simply take

$$u(z, t) = q(z, x, t), \qquad v(z, t) = q(z, y, t),$$

and then let $\alpha \downarrow 0$, $\beta \uparrow t$. One obtains, immediately, the desired symmetry.

Consider, now, the Dirichlet heat kernel of Ω, $Q(x, y, t)$, constructed in Section 2. Imagine the manifold M to be compact, so as to guarantee that its heat kernel $p(x, y, t)$ is symmetric. Then one can verify that Q is a Dirichlet heat kernel relative to the *second* space variable, namely, the function $q(x, y, t)$ given by

$$q(x, y, t) = Q(y, x, t)$$

is a Dirichlet heat kernel. The first claim of our theorem, here, shows that Q is symmetric in x and y—in particular $Q(\ , y,\ ) \in \bar{C}^1(\Omega \times (0, \infty))$ for all $y \in \Omega$. We now have to show that Q is unique.

Indeed, if q_2 is any Dirichlet heat kernel on Ω, satisfying the given hypotheses, then set

$$u(z, t) = Q(z, x, t), \qquad v(z, t) = q_2(z, y, t).$$

One easily deduces from (26) that

$$Q(y, x, t) = q_2(x, y, t).$$

But both heat kernels are symmetric; therefore $Q = q_2$.

COROLLARY 2. The Dirichlet heat kernel q on Ω is almost Euclidean. Furthermore, if Ω_1, Ω_2 are regular domains in a Riemannian manifold, with respective Dirichlet heat kernels $q_{\Omega_1}, q_{\Omega_2}$, then

$$q_{\Omega_1}(x, y, t) - q_{\Omega_2}(x, y, t) = O(t^N)$$

for all $N \geq 1$, $(x, y) \in (\Omega_1 \cap \Omega_2) \times (\Omega_1 \cap \Omega_2)$, as $t \downarrow 0$. The limit is uniform on compact subsets of $\Omega_1 \cap \Omega_2$.

We leave the proof of the corollary to the reader.

THEOREM 5. Let Ω be a regular domain with boundary Γ and Dirichlet heat kernel q. Suppose we are given bounded continuous functions $F\colon \Omega \times (0,\infty) \to \mathbb{R}$, $\varphi\colon \Gamma \times (0,\infty) \to \mathbb{R}$, and $f\colon \Omega \to \mathbb{R}$. Then the unique solution $u\colon \overline{\Omega} \times [0,\infty) \to \mathbb{R}$, which is C^0 on $(\overline{\Omega} \times [0,\infty))\backslash(\Gamma \times \{0\})$ and C^1 on $\Omega \times (0,\infty)$, to the initial-boundary-value problem (23)–(25) is given by

$$\begin{aligned} u(x,t) &= \iint_\Omega q(x,y,t)f(y)\,dV(y) \\ &\quad + \int_0^t d\tau \iint_\Omega q(x,y,t-\tau)F(y,\tau)\,dV(y) \qquad (27) \\ &\quad - \int_0^t d\tau \int_\Gamma \frac{\partial q}{\partial \nu_w}(x,w,t-\tau)\varphi(w,\tau)\,dA(w). \end{aligned}$$

PROOF: To derive (27), one substitutes

$$v(z,\tau) = q(x,z,\tau)$$

into (26), and lets $\alpha \downarrow 0$, $\beta \uparrow t$.

Once one has (27), one verifies that it, in fact, provides a solution to the initial-boundary-value problem (23)–(25). We note that $(\partial h/\partial \nu_w)(x,w,t)$ is given, via the jump relation for the single-layer potential (Theorem 2), by

$$\frac{\partial h}{\partial \nu_w}(x,w,t) = \frac{\mathscr{F}(x,w,t)}{2} + \int_0^t d\tau \int_\Gamma \mathscr{F}(x,\overline{w},t-\tau)\frac{\partial p}{\partial \nu_w}(\overline{w},w,\tau)\,dA(\overline{w}).$$

One now can easily show that $u \to \varphi$ as $x \to \Gamma$ (the only integral to really consider is the last one in (27)). Establishing $u \to f$, as $t \downarrow 0$, is easy. Finally, we note that, establishing that u satisfies (23), requires an extension of Theorem VI.6, and Lemma VI.6.

DEFINITION 3. For each $t > 0$, define the operator $\mathfrak{Q}_t$ on $L^2(\Omega)$ by

$$(\mathfrak{Q}_t f)(x) = \iint_\Omega q(x,y,t)f(y)\,dV(y). \qquad (28)$$

PROPOSITION 2. For each $t > 0$, $\mathfrak{Q}_t$ is a self-adjoint, compact (completely continuous), operator on $L^2(\Omega)$. For each $f \in L^2(\Omega)$,

$$u(x,t) =: (\mathfrak{Q}_t f)(x)$$

is a solution of the heat equation on $\Omega \times (0,\infty)$.

If we also set $\mathfrak{Q}_0 = I$, then

$$\mathfrak{Q}_t \circ \mathfrak{Q}_T = \mathfrak{Q}_{t+T} \qquad (29)$$

for all t, $T \geq 0$. For each $t \geq 0$, $\mathfrak{Q}_t \geq 0$, which implies

$$q \geq 0$$

on $\Omega \times \Omega \times (0, \infty)$ (the maximum principle, cf. Section VIII.1, implies $q > 0$). Also

$$\lim_{t \downarrow 0} \mathfrak{Q}_t f = f$$

for all $f \in L^2(\Omega)$. If $f \in L^2(\Omega)$ is also continuous on Ω, then $\mathfrak{Q}_t f$ converges to f, as $t \downarrow 0$, uniformly on compact subsets of Ω.

Finally, for each $x \in \Omega$,

$$\iint_\Omega q(x, y, t)\, dV(y)$$

is a decreasing function of t. In particular,

$$\iint_\Omega q(x, y, t)\, dV(y) \leq 1 \tag{30}$$

for all $(x, t) \in \Omega \times (0, \infty)$.

The proof of the proposition is straightforward. We only need note that the last claim is a consequence of the nonnegativity of q, and Eq. (21).

The Sturm–Liouville decomposition for the Dirichlet eigenvalue problem: Given the normal domain Ω, there exists a complete orthonormal basis, $\{\varphi_1, \varphi_2, \varphi_3, \ldots\}$ of $L^2(\Omega)$ consisting of Dirichlet eigenfunctions of Δ, with φ_j having eigenvalue λ_j satisfying

$$0 < \lambda_1 < \lambda_2 \leq \lambda_3 \leq \cdots \uparrow +\infty.$$

In particular, each eigenvalue has finite multiplicity. Each

$$\varphi_j \in C^\infty(\Omega) \cap \bar{C}^1(\Omega).$$

Finally,

$$q(x, y, t) = \sum_{j=1}^{\infty} e^{-\lambda_j t} \varphi_j(x) \varphi_j(y), \tag{31}$$

with convergence absolute, and uniform, for each $t > 0$. In particular

$$\iint_\Omega q(x, x, t)\, dV(x) = \sum_{j=1}^{\infty} e^{-\lambda_j t}. \tag{32}$$

The proof is the same as that for the closed eigenvalue problem (Section VI.1).

We now wish to consider what happens as $t \uparrow +\infty$. For $f \in L^2(\Omega) \cap C^0(\Omega)$, $(\mathfrak{Q}_t f)(x)$ is the solution of the heat equation on Ω, with initial heat distribution f. Of course, as $x \to \Gamma$, the boundary of Ω, $(\mathfrak{Q}_t f)(x) \to 0$ for each $t > 0$. So we are thinking of the temperature distribution in a room whose walls are absolutely frozen for all time $t > 0$. We expect the temperature of the room itself to, ultimately, be absolutely cold, namely, one proves, as in Theorem VI.2, that as $t \uparrow +\infty$, $\mathfrak{Q}_t f$ converges, uniformly on $\bar{\Omega}$, to a function which is harmonic on Ω. Since the boundary values of this limit function are identically 0, we have

$$\lim_{t \uparrow +\infty} \mathfrak{Q}_t f = 0 \tag{33}$$

uniformly on $\bar{\Omega}$.

Theorem 6 is a more ambitious version of this result.

THEOREM 6. Suppose we are given continuous functions $f: \bar{\Omega} \to \mathbb{R}$, and $\varphi: \Gamma \to \mathbb{R}$, and $u: (\bar{\Omega} \times [0,\infty)) \backslash (\Gamma \times \{0\}) \to \mathbb{R} \in C^0$ a solution to the initial-boundary-value problem:

$$Lu = 0,$$

$$u(\ ,0) = f, \qquad u(\ ,t) \,|\, \Gamma = \varphi.$$

Then $u(x,t)$ converges, uniformly on $\bar{\Omega}$, to a function harmonic on Ω and taking boundary values φ on Γ.

PROOF: From (27), we have

$$u(x,t) = \iint_{\Omega} q(x,y,t) f(y)\, dV(y) - \int_0^t d\tau \int_{\Gamma} \frac{\partial q}{\partial \nu_w}(x,w,\tau)\varphi(w)\, dA(w).$$

By (33) we may assume $f = 0$ on all of Ω. So we assume that u is given by

$$u(x,t) = -\int_0^t d\tau \int_{\Gamma} \frac{\partial q}{\partial \nu_w}(x,w,\tau)\varphi(w)\, dA(w).$$

Note that the nonnegativity of q, on Ω, implies

$$\frac{\partial q}{\partial \nu_w}(x,w,t) \le 0$$

on all of $\Omega \times \Gamma \times (0,t)$.

Now $t > T$ implies

$$\begin{aligned}
|u(x,t) - u(x,T)| &= \left|\int_T^t d\tau \int_\Gamma \frac{\partial q}{\partial v_w}(x,w,\tau)\varphi(w)\,dA(w)\right| \\
&\leq \{\max|\varphi|\}\int_T^t d\tau \int_\Gamma -\frac{\partial q}{\partial v_w}(x,w,\tau)\,dA(w) \\
&= -\text{const}\cdot\int_T^t d\tau \iint_\Omega (\Delta_y q)(x,y,\tau)\,dV(y) \\
&= \text{const}\cdot\iint_\Omega \{q(x,y,T) - q(x,y,t)\}\,dV(y) \\
&= \text{const}\cdot\{(\mathfrak{Q}_T - \mathfrak{Q}_t)1\}(x),
\end{aligned}$$

which implies that $u(x,t)$ converges, uniformly on $\bar{\Omega}$, as $t \uparrow +\infty$, to a continuous function, H, on Ω. Of course

$$H\,|\,\Gamma = \varphi, \tag{34}$$

and

$$H(x) = -\int_0^\infty d\tau \int_\Gamma \frac{\partial q}{\partial v_w}(x,w,\tau)\varphi(w)\,dA(w). \tag{35}$$

We wish to show that H is harmonic. It suffices to show that

$$H(x) = \iint_\Omega q(x,y,t)H(y)\,dV(y) - \int_0^t d\tau \int_\Gamma \frac{\partial q}{\partial v_w}(x,w,\tau)\varphi(w)\,dA(w) \tag{36}$$

for all $t > 0$. For the right-hand side of (36) is a solution of the heat equation, which, by the left-hand side, will be independent of t. This function must then be harmonic.

To prove (36), one has by (35)

$$\begin{aligned}
&\iint_\Omega q(x,y,t)H(y)\,dV(y) \\
&\quad= -\iint_\Omega q(x,y,t)\,dV(y)\int_0^\infty d\tau \int_\Gamma \frac{\partial q}{\partial v_w}(y,w,\tau)\varphi(w)\,dA(w) \\
&\quad= -\int_0^\infty d\tau \int_\Gamma \varphi(w)\,dA(w)\iint_\Omega q(x,y,t)\frac{\partial q}{\partial v_w}(y,w,\tau)\,dV(y) \\
&\quad= -\int_t^\infty d\tau \int_\Gamma \frac{\partial q}{\partial v_w}(x,w,t+\tau)\varphi(w)\,dA(w) \\
&\quad= -\int_t^\infty d\tau \int_\Gamma \frac{\partial q}{\partial v_w}(x,w,\tau)\varphi(w)\,dA(w).
\end{aligned}$$

This reduces (36) to (35), and the claim is proven.

Note that, by (27), the right-hand side of (36) is the heat distribution determined by the initial distribution H and boundary distribution φ. Equation (36) implied that H is harmonic. But the identity also suggests

THEOREM 7. Let $H \in C^0(\bar{\Omega}) \cap (\bar{C}^1 \cap C^2)(\Omega)$ be harmonic in Ω, $\varphi = H \mid \partial\Omega$. Let $u(x, t)$ be the solution to the heat equation on Ω with initial heat distribution H on Ω, and boundary distribution φ for all time. Then u is a *steady-state solution*, that is,

$$u(\ , t) = H$$

for all $t > 0$.

We leave the proof to the reader, as an exercise.

Weyl's asymptotic formula for Dirichlet eigenvalues: Let Ω be a normal domain with Dirichlet eigenvalues: $0 < \lambda_1 < \lambda_2 \leq \lambda_3 \leq \cdots$, each distinct eigenvalue repeated according to its multiplicity. Then, for

$$N(\lambda) =: \sum_{\lambda_l \leq \lambda} 1,$$

we have

$$N(\lambda) \sim \omega_n V(\Omega) \lambda^{n/2}/(2\pi)^n$$

as $\lambda \uparrow +\infty$. In particular,

$$(\lambda_k)^{n/2} \sim (2\pi)^n k/\omega_n V(\Omega)$$

as $k \uparrow +\infty$.

PROOF: As in the proof of the compact case (Section VI.4), we wish to estimate

$$\sum_{j=1}^{\infty} e^{-\lambda_j t} = \iint_{\Omega} q(x, x, t)\, dV(x),$$

where q is the Dirichlet heat kernel of Ω (cf. (32)).

To estimate the integral, fix any $\varepsilon > 0$, set

$$D_\varepsilon = \{x \in \Omega : d(x, \Gamma) \geq \varepsilon\},$$

where Γ is the boundary of Ω, and write

$$q(x, y, t) = p(x, y, t) + h(x, y, t),$$

where p is the heat kernel of a compact Riemannian manifold containing Ω. Then, by (13), for fixed $T > 0$, there exists a positive constant, depending on ε, T, so that

$$|h(x, x, t)| \leq \text{const} \cdot t^{-n/2} e^{-d^2(x,\Gamma)/8t}$$

for all $x \in D_\varepsilon$, $t \in (0, T]$. Then (VI.44), (VI.45) imply

$$(4\pi t)^{n/2} \iint_{D_\varepsilon} q(x, x, t)\, dV(x) \to V(D_\varepsilon)$$

as $t \downarrow 0$.

To estimate the integral in question over $\Omega \backslash D_\varepsilon$, we note that by the maximum principle (Section VIII.1), we have

$$0 \leq q \leq p$$

on all of Ω. Therefore,

$$\begin{aligned}(4\pi t)^{n/2} \iint_{\Omega \backslash D_\varepsilon} q(x, x, t)\, dV(x) &\leq \{\sup_{\bar{\Omega} \times [0,T]} (4\pi t)^{n/2} p(x, x, t)\} V(\Omega \backslash D_\varepsilon) \\ &\leq \text{const} \cdot \varepsilon.\end{aligned}$$

Thus

$$\sum_{k=1}^{\infty} e^{-\lambda_j t} \sim V(\Omega)/(4\pi t)^{n/2}, \tag{37}$$

which implies the desired result by the Karamata theorem (Feller [1, p. 446]).

We now discuss the Green's function of the regular domain Ω. For convenience, we let Ω_D denote the *diagonal of* Ω, that is,

$$\Omega_D = \{(x, y) \in \Omega \times \Omega : x = y\}.$$

DEFINITION 4. A continuous function $G: (\bar{\Omega} \times \bar{\Omega}) \backslash \Omega_D \to \mathbb{R}$ is a *Green's function of* Ω, if $G \mid \{(\Omega \times \Omega) \backslash \Omega_D\} \in C^2$,

$$(\Delta_y G)(x, y) = 0$$

for all $x, y \in \Omega$,

$$G \mid \Omega \times \Gamma = 0$$

(where, as usual, $\Gamma = \partial\Omega$), and, near Ω_D, G is given by

$$G(x, y) = \psi(x, y) + h(x, y), \tag{38}$$

where $h \in C^0(\bar{\Omega} \times \bar{\Omega}) \cap C^2(\Omega \times \Omega)$, and

$$\psi(x, y) = \frac{1}{c_{n-1}} \begin{cases} d^{2-n}(x, y)/(n-2), & n > 2, \\ -\ln d(x, y), & n = 2. \end{cases} \tag{39}$$

The existence of G is derived in Duff [1]. We shall also assume here that for each $x \in \Omega$, $(\operatorname{grad}_y G)(x, \ )$ extends to a continuous vector field on $\bar{\Omega}\backslash\{x\}$.

One easily sees that the maximum principle (Section XII.11) implies that G is strictly positive on $(\Omega \times \Omega)\backslash\Omega_D$, and that there is at most one Green's function for Ω.

THEOREM 8. Let $u \in \bar{C}^1(\Omega) \cap C^2(\Omega)$. Then

$$-u(x) = \iint_\Omega G(x, z)(\Delta u)(z)\, dV(z) + \int_\Gamma \frac{\partial G}{\partial \nu_w}(x, w) u(w)\, dA(w). \tag{40}$$

In particular,

$$-u(x) = \iint_\Omega G(x, z)(\Delta u)(z)\, dV(z) \tag{41}$$

if $u | \Gamma = 0$; and

$$u(x) = -\int_\Gamma \frac{\partial G}{\partial \nu_w}(x, w) u(w)\, dA(w) \tag{42}$$

if u is harmonic on Ω.

PROOF: We apply Green's formula to the given u, and to v given by

$$v(z) = G(x, z);$$

but, because of the singularity of v at x, we must apply Green's formula to the given u and v on the domain

$$\Omega_\varepsilon =: \Omega\backslash B(x; \varepsilon),$$

for sufficiently small $\varepsilon > 0$. We then obtain

$$\begin{aligned} -\iint_{\Omega_\varepsilon} & G(x, z)\Delta u(z)\, dV(z) \\ &= \int_{\partial\Omega} \frac{\partial G}{\partial \nu_w}(x, w) u(w)\, dA(w) \\ &\quad - \int_{S(x;\varepsilon)} \left\{ \frac{\partial G}{\partial r_w}(x, w) u(w) - G(x, w) \frac{\partial u}{\partial r}(w) \right\} dA(w), \end{aligned} \tag{43}$$

where $\partial/\partial r$ is the directional derivative in the radial directions emanating from x.

Note that for all dimensions $n \geq 2$, we have

$$-\frac{\partial\psi}{\partial r_w}(x, w) = d^{1-n}(x, w)/\mathbf{c}_{n-1}.$$

Introduce geodesic spherical coordinates

$$y = q(r, \xi) = \exp_x r\xi,$$

$r \geq 0$, $\xi \in \mathfrak{S}_x$, about x, as described in Section III.1. As discussed there, the volume element dV is given by

$$dV(q(r, \xi)) = \sqrt{\mathbf{g}}(r; \xi)\, dr\, d\mu_x(\xi),$$

where $d\mu_x$ is the standard $(n-1)$-measure on $\mathfrak{S}_x$; and the $(n-1)$-dimensional area element of $\mathsf{S}(x; r)$ is given by

$$dA(q(r, \xi)) = \sqrt{\mathbf{g}}(r; \xi)\, d\mu_x(\xi).$$

The discussion there (cf. Sections III.1 and XII.8), shows that

$$\lim_{r \to 0} \frac{\sqrt{\mathbf{g}}(r; \xi)}{r^{n-1}} = 1. \tag{44}$$

One now easily lets $\varepsilon \downarrow 0$ in (43), and obtains (40).

Note that if $x, y \in \Omega$, and we pick

$$u(z) = G(y, z), \qquad v(z) = G(x, z)$$

and

$$\Omega_\varepsilon = \Omega \backslash \{\mathsf{B}(x; \varepsilon) \cup \mathsf{B}(y; \varepsilon)\},$$

then the argument just given will yield the symmetry of G, namely,

$$G(x, y) = G(y, x).$$

Recall that the *Dirichlet problem* for Ω is: given $\varphi: \Gamma \to \mathbb{R} \in C^0$, find $u: \bar{\Omega} \to \mathbb{R} \in C^0(\bar{\Omega}) \cap C^2(\Omega)$ such that u is harmonic on Ω and $u | \Gamma = \varphi$. Then (42) says that once we have a Green's function G for Ω, and we know that $u \in \bar{C}^1(\Omega)$, then we must have

$$u(x) = -\int_\Gamma \frac{\partial G}{\partial \nu_w}(x, w)\varphi(w)\, dA(w). \tag{45}$$

However, to verify that, given φ, the function u proposed by (45) is, indeed, the solution to the Dirichlet problem for boundary data φ, we require more delicate information of the boundary behavior of G. (After all, given φ, there is no a priori way of predicting that $u \in \bar{C}^1(\Omega)$.)

What one requires is that given $w_0 \in \Gamma$, $\delta > 0$, then

$$\lim_{x \to w_0} \frac{\partial G}{\partial v_w}(x, w) = 0 \tag{46}$$

uniformly on

$$\Gamma_2(w_0; \delta) \mathrel{=:} \{w \in \Gamma : d(w, w_0) > \delta\}.$$

Once one has this information (and we shall assume that we, indeed, have it) the argument that (45) supplies a solution to the Dirichlet problem, for boundary data φ, is as follows:

First, we note that

$$1 = -\int_\Gamma \frac{\partial G}{\partial v}(x, w)\, dA(w) \tag{47}$$

for all $x \in \Omega$. Indeed, the argument given to derive (42), for $u \equiv 1$, is perfectly rigorous.

Second, we note that

$$\frac{\partial G}{\partial v}(x, w) \leq 0 \tag{48}$$

since $G > 0$ on $\Omega \times \Omega$, and $G \mid \Omega \times \Gamma = 0$.

Finally, set

$$\Gamma_1(w_0; \delta) = \Gamma \cap \mathsf{B}(w_0; \delta).$$

Then, by (47),

$$\begin{aligned} u(x) - \varphi(w_0) &= \int_\Gamma -\frac{\partial G}{\partial v_w}(x, w)\{\varphi(w) - \varphi(w_0)\}\, dA(w) \\ &= \int_{\Gamma_1(w_0;\delta)} (\cdots) + \int_{\Gamma_2(w_0;\delta)} (\cdots). \end{aligned}$$

To estimate the second integral, we have, since φ is bounded and $(\partial G/\partial v_w)(x, \ )$ is nonpositive,

$$\begin{aligned} &\left| \int_{\Gamma_2(w_0;\delta)} -\frac{\partial G}{\partial v_w}(x, w)\{\varphi(w) - \varphi(w_0)\}\, dA(w) \right| \\ &\qquad \leq 2\{\sup|\varphi|\} \int_{\Gamma_2(w_0;\delta)} -\frac{\partial G}{\partial v_w}(x, w)\, dA(w). \end{aligned}$$

For the first integral we have, using (47) and (48),

$$\left|\int_{\Gamma_1(w_0;\delta)} -\frac{\partial G}{\partial \nu_w}(x, w)\{\varphi(w) - \varphi(w_0)\}\, dA(w)\right|$$

$$\leq \{\sup_{\Gamma_1(w_0;\delta)} |\varphi(w) - \varphi(w_0)|\} \int_{\Gamma_1(w_0;\delta)} -\frac{\partial G}{\partial \nu_w}(x, w)\, dA(w)$$

$$\leq \sup_{\Gamma_1(w_0;\delta)} |\varphi(w) - \varphi(w_0)|.$$

One now uses (46), (47), and the continuity of φ at w_0 to obtain

$$\lim_{x \to w_0} u(x) = \varphi(w_0),$$

which is the claim.

Recall the problem of heat conduction considered in Theorem 6. We are given Ω with initial temperature distribution $f \in C^0(\Omega)$, and allow the heat to diffuse in Ω, subject to the condition that the temperature at the boundary be given by the function $\varphi\colon \Gamma \to \mathbb{R} \in C^0$, independent of time. The statement of Theorem 6, in virtue of (35) and (45), is equivalent to saying

$$\lim_{t \uparrow +\infty} \int_0^t d\tau \int_\Gamma \frac{\partial q}{\partial \nu_w}(x, w, t - \tau)\varphi(w)\, dA(w) = \int_\Gamma \frac{\partial G}{\partial \nu_w}(x, w)\varphi(w)\, dA(w),$$

where q is the Dirichlet heat kernel of Ω. This suggests the formula

$$G(x, y) = \int_0^\infty q(x, y, t)\, dt, \tag{49}$$

the validity of which we now investigate.

Recall that $\mathfrak{Q}_t$ denotes the operator

$$(\mathfrak{Q}_t f)(x) = \iint_\Omega q(x, y, t) f(y)\, dV(y); \tag{50}$$

let $\mathfrak{G}$ denote the operator given by

$$(\mathfrak{G} f)(x) = \iint_\Omega G(x, y) f(y)\, dV(y). \tag{51}$$

Note that

$$(\mathfrak{G}1)(x) < +\infty, \tag{52}$$

by (38), (39), and (44). Also note that for $u \in \bar{C}^1(\Omega) \cap C^2(\Omega)$, satisfying $u \mid \Gamma = 0$, (41) reads as

$$\mathfrak{G}\Delta u = -u. \tag{53}$$

Now for $f \in C^0(\bar{\Omega})$ we have

$$\begin{aligned}\Delta_x \int_0^T dt \iint_\Omega q(x,y,t) f(y)\, dV(y) &= \int_0^T dt \iint_\Omega (\Delta_x q)(x,y,t) f(y)\, dV(y) \\ &= \int_0^T dt \iint_\Omega (\partial q/\partial t)(x,y,t) f(y)\, dV(y) \\ &= \iint_\Omega q(x,y,T) f(y)\, dV(y) - f(x),\end{aligned}$$

that is,

$$\Delta\left(\int_0^T \mathfrak{Q}_t\, dt\right) = \mathfrak{Q}_T - I \tag{54}$$

on $C^0(\bar{\Omega})$. Thus

$$\mathfrak{G}(\mathfrak{Q}_T - I) = \mathfrak{G}\Delta\left(\int_0^T \mathfrak{Q}_t\, dt\right) = -\int_0^T \mathfrak{Q}_t\, dt.$$

By (52) and the uniformity of (33) on Ω, we have

$$\mathfrak{G} = +\int_0^\infty \mathfrak{Q}_t\, dt$$

on $C^0(\bar{\Omega})$, and (49) follows.

It is important to note that one also has

$$\Delta\mathfrak{G} = -I \tag{55}$$

on $C_c^1(\Omega)$—a proof can be found in Duff [1, Sec. V.3]. More generally, (55) is valid on bounded, locally Hölder continuous functions on Ω—cf. Gilbarg–Trudinger [1, Sec. 4.2]. Still more generally, (55) can be interpreted in the sense of acting on distributions—cf. Wermer [1, Sec. 4].

CHAPTER VIII

The Heat Kernel for Noncompact Manifolds

In this chapter we extend the theory of the Dirichlet heat kernel for regular domains to arbitrary noncompact manifolds, namely, given a noncompact Riemannian manifold M, one can exhaust M by a sequence of regular domains, thereby creating an increasing sequence of Dirichlet heat kernels, the limit of which is the minimal positive heat kernel of M (Section 2). While this construction is always valid, it is not necessary that this heat kernel be the only one on the manifold, and that it satisfy the conservation law (VI.9) (cf. Azencott [1, Sec. 7]). The situation we consider here is when M is complete with Ricci curvature bounded from below. The key tools are maximum principles, and the extension of differential inequalities past the cut locus of a point in M—this last technique developed by Calabi [1]. The treatment here is that of Dodziuk [2]—almost word-for-word. Other constructions of the heat kernel, and uniqueness and conservation of heat considerations, can be found in Azencott [1], Cheeger–Yau [1], Karp [4], Karp–Li [1], Seeley [1], Vauthier [1], and Yau [4].

In Section 3 we then present the comparison theorems for the heat kernel, the first when all sectional curvatures are bounded from above, and the second when the Ricci curvatures are bounded from below. These theorems were first proved in Debiard–Gaveau–Mazet [1] as being valid inside the cut locus, and the second one was extended by Cheeger–Yau [1] past the cut locus. The arguments of Sections 1 and 2 yield an easy proof. (We note that Cheeger–Yau [1] also considers generalized Neumann boundary conditions, in which case the results are unavailable by the methods of this chapter, or by those of Debiard–Gaveau–Mazet [1].)

As the bound for the heat kernel of Section 3, when the Ricci curvature is bounded from below, is a lower bound, we then turn our attention to an upper bound for the heat kernel valid past the cut locus. This was first considered in Cheng–Li–Yau [1], and was then treated in Cheeger–Gromov–Taylor [1]. Our treatment follows the latter, and is based on the lectures of J. Dodziuk to the seminar in differential geometry (Fall 1981) at the Graduate School of the City University of New York. We note that these upper bounds do not directly require bounds on curvature; rather, they involve lower bounds on the isoperimetric constant of geodesic disks in the manifold. We use the upper bounds, in this form, in our discussion of manifolds with small handles (Chapter IX) where one has no a priori bounds on the curvature of the small handles. For general theorems, however, one can bound the isoperimetric constants below in terms of bounds on the curvature, which, in turn, supply upper bounds for the heat kernel in terms of bounds on the curvature. Compare Cheng–Li–Yau [1, Sec. 1] and Cheeger–Gromov–Taylor [1, Sec. 4] (this last paper determining the bounds in terms of lower bounds on the Ricci curvature).

Recently, Li and Yau [4] have given upper bounds for the heat kernel depending directly on lower bounds for the Ricci curvature. Compare their paper for the details.

1. THE MAXIMUM PRINCIPLE, AND UNIQUENESS THEOREMS, FOR THE HEAT OPERATOR

We are given a fixed Riemannian manifold M with Laplacian Δ, and associated heat operator $L = \Delta - \partial/\partial t$.

STRONG MAXIMUM PRINCIPLE. Let u be a bounded continuous function on $M \times [0, T]$, which is C^2 on $M \times (0, T)$, and which satisfies

$$Lu \geq 0 \tag{1}$$

on $M \times (0, T)$. If there exists (x_0, t_0) in $M \times (0, T]$ such that

$$u(x_0, t_0) = \sup_{M \times [0,T]} u,$$

then

$$u \mid M \times [0, t_0] = u(x_0, t_0).$$

We refer the reader to Protter–Weinberger [1, Sec. III.3] for a proof of the theorem when M is diffeomorphic to a domain in Euclidean space. A standard continuation argument then extends the theorem to arbitrary M.

Of course, if one is given

$$Lu \leq 0, \tag{2}$$

instead of (1), on $M \times (0, T)$, then one has a corresponding minimum principle. For solutions to the heat equation, both principles are valid.

Easy applications of the maximum principle are the following:

THEOREM 1. If M is compact with heat kernel p, then p is strictly positive on $M \times M \times (0, \infty)$.

Similarly, if M is a regular domain with Dirichlet heat kernel q, then q is strictly positive on $M \times M \times (0, \infty)$.

More generally, if Ω is a regular domain in M, q is the Dirichlet heat kernel of Ω, and p is the heat kernel (resp., Dirichlet heat kernel) of M, for M compact (resp., for M a regular domain) then

$$0 < q < p \tag{3}$$

on $\bar{\Omega} \times \bar{\Omega} \times (0, \infty)$.

In particular, for a regular domain Ω,

$$\iint_{\Omega} q(x, y, t)\, dV(y) < 1 \tag{4}$$

for all $t > 0$.

We now wish to consider existence and uniqueness theorems for the heat kernel on arbitrary noncompact M. The uniqueness results require only a weaker form of the maximum principle.

DEFINITION 1. Let $u: M \times (\alpha, \beta) \to \mathbb{R} \in C^0$, and ρ be a real number. We say that u *satisfies the weak differential inequality*.

$$Lu \succ \rho$$

at $(x, t) \in M \times (\alpha, \beta)$ if for every $\varepsilon > 0$ there exists a neighborhood $V = V_\varepsilon(x, t)$, of (x, t), in $M \times (\alpha, \beta)$, and a function $u_\varepsilon: V \to \mathbb{R} \in C^0$ which is C^2 in the space variable, and C^1 in the time variable, such that

$$\min_V (u - u_\varepsilon) = (u - u_\varepsilon)(x, t),$$

and

$$(Lu_\varepsilon)(x, t) \geq \rho - \varepsilon.$$

We say

$$Lu \prec \rho$$

if

$$L(-u) \succ -\rho.$$

PROPOSITION 1. If $u \in C^0(M \times (\alpha, \beta))$, and u is C^2 in the space variable, and C^1 in the time variable, then for any $x \in M$, $\rho \in \mathbb{R}$, we have

$$(Lu)(x) \geq \rho$$

if and only if

$$Lu \succ \rho$$

at x.

For the proof, see the proof of Proposition 2, below, which is basically the same argument.

THEOREM 2. Let Ω be a relatively compact domain in M, $u\colon \bar{\Omega} \times [\alpha, \beta] \to \mathbb{R} \in C^0$, such that

$$Lu \succ 0$$

on all of $\Omega \times (\alpha, \beta)$. Then

$$\sup_{\bar{\Omega} \times [\alpha, \beta]} u = \sup_{(\bar{\Omega} \times \{\alpha\}) \cup (\partial\Omega \times [\alpha, \beta])} u.$$

PROOF: Given any $\delta > 0$, $\tau > 0$, set

$$v^\delta = u - \delta(t - \alpha),$$

$$\Lambda_\tau = \bar{\Omega} \times [\alpha, \beta - \tau], \qquad \Lambda_\tau^* = (\bar{\Omega} \times \{\alpha\}) \cup (\partial\Omega \times [\alpha, \beta - \tau]).$$

We will show that

$$\sup_{\Lambda_\tau} v^\delta = \sup_{\Lambda_\tau^*} v^\delta \tag{5}$$

for all δ, τ. The desired result will then follow easily.

To prove (5) first note that

$$Lv^\delta \succ \delta > 0$$

on M.

Now assume there exists $(x, t) \in \Omega \times (\alpha, \beta - \tau)$ for which

$$v^\delta(x, t) = \max_{\Lambda_\tau} v^\delta.$$

Consider, at (x, t), the function $u_{\delta/2}$, defined on a neighborhood of (x, t), whose existence is postulated in Definition 1, for $\rho = 0$, $\varepsilon = \delta/2$. Then

$$(Lu_{\delta/2})(x, t) \geq \delta/2 > 0; \tag{6}$$

furthermore,

$$v^\delta - u_{\delta/2} \geq (v^\delta - u_{\delta/2})(x, t)$$

on the neighborhood in question. Therefore,

$$u_{\delta/2}(x, t) \geq u_{\delta/2} + v^\delta(x, t) - v^\delta \geq u_{\delta/2},$$

since v^δ has its local maximum at (x, t). Thus

$$(Lu_{\delta/2})(x, t) \leq 0$$

in contradiction to (6). So v^δ can have no maximum in $\Omega \times (\alpha, \beta - \tau)$; and (5) is proven.

We now consider, on M, the initial-value problem for the heat equation, that is, given $\varphi: M \to \mathbb{R} \in C^0$, we seek a function $u: M \times [0, \infty) \to \mathbb{R} \in C^0$ which is C^2 in the space variable, C^1 in the time variable, satisfies the heat equation, and satisfies

$$u \,|\, M \times \{0\} = \varphi.$$

THEOREM 3 (Dodziuk [2]). If M is complete, with Ricci curvature bounded from below, then bounded solutions to the initial-value problem are uniquely determined by their initial data.

To prove the theorem, we require a differential inequality, which is not only valid within the cut locus of a given point, but extends past the cut locus.

Recall, from Section III.1, the construction of geodesic spherical coordinates on the domain D_x, about a given $x \in M$. The cut locus of x is the boundary of D_x.

LEMMA 1. Assume the Ricci curvature of M is bounded below by $(n - 1)\kappa$, $n = \dim M$. Let $\psi: (0, \infty) \to \mathbb{R} \in C^2$, and for fixed $x \in M$, set

$$f(y) = \psi(d(x, y)). \tag{7}$$

If ψ' is nonnegative on $(0, \infty)$, then

$$\Delta f \leq \{\psi'' + (n - 1)(\mathbf{C}_\kappa/\mathbf{S}_\kappa)\psi'\} \circ d(x,\ ) \tag{8}$$

on $\mathsf{D}_x \backslash \{x\}$.

Proof: The inequality is a direct consequence of the Bishop comparison theorem (II) (cf. Section III.3).

DEFINITION 2. Let $u: M \to \mathbb{R} \in C^0$, $x \in M$, and ρ a real number. We say that u *satisfies the weak differential inequality*

$$\Delta u \succ \rho$$

at $x \in M$, if, for every $\varepsilon > 0$, there exists a neighborhood $V = V_\varepsilon(x)$ of x, and a function $u_\varepsilon: V \to \mathbb{R} \in C^2$ such that

$$\min_V (u - u_\varepsilon) = (u - u_\varepsilon)(x),$$

and

$$(\Delta u_\varepsilon)(x) \geq \rho - \varepsilon.$$

We say that

$$\Delta u \prec \rho$$

if

$$\Delta(-u) \succ -\rho.$$

PROPOSITION 2. If $u \in C^2(M)$, then at $x \in M$ we have

$$(\Delta u)(x) \geq \rho$$

if and only if

$$\Delta u \succ \rho$$

at x.

PROOF: If $(\Delta u)(x) \geq \rho$, then for any $\varepsilon > 0$, pick $V_\varepsilon = M$, $u_\varepsilon = u$. One easily has $\Delta u \succ \rho$ at x.

Now assume $\Delta u \succ \rho$ at x. Given $\varepsilon > 0$, one picks u_ε so that $u - u_\varepsilon$ has a minimum at x, and

$$(\Delta u_\varepsilon)(x) \geq \rho - \varepsilon.$$

But since $u - u_\varepsilon$ has a minimum at x, we have

$$(\Delta(u - u_\varepsilon))(x) \geq 0.$$

Therefore, at x, we have

$$\Delta u = \Delta u_\varepsilon + \Delta(u - u_\varepsilon) \geq \Delta u_\varepsilon \geq \rho - \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the claim follows.

LEMMA 2 (Calabi [1]). Given the assumptions of Lemma 1, and the assumption that

$$\mathbf{S}_\kappa \circ d(x,\ \ ) > 0 \tag{9}$$

on all of $M\backslash\{x\}$, we have

$$\Delta f \prec \{\psi'' + (n-1)(\mathbf{C}_\kappa/\mathbf{S}_\kappa)\psi\} \circ d(x,\ \ ) \tag{10}$$

on all of $M\backslash\{x\}$.

PROOF: Of course, we need only consider $y \in M\backslash \mathsf{D}_x$, that is, y is in the cut locus of x.

For $y \in M\backslash D_x$, let $\delta \in (0, d(x,y))$, and let $\gamma\colon [0, d(x,y)] \to M$, $|\gamma'| = 1$, be a geodesic from x to y, that is, $\gamma(0) = x$, $\gamma(d(x,y)) = y$. Set $x_\delta = \gamma(\delta)$, $f_\delta = \psi \circ d(x_\delta,\ \ )$. Then $y \in \mathsf{D}_{x_\delta}$, and

$$\Delta f_\delta \le \{\psi'' + (n-1)(\mathbf{C}_\kappa/\mathbf{S}_\kappa)\psi'\} \circ d(x_\delta,\ \ ),$$

by Lemma 1.

Now $f - f_\delta$ achieves a maximum when $d(x,\ \ ) - d(x_\delta,\ \ )$ achieves a maximum (since $\psi' \ge 0$). In general, the triangle inequality implies

$$d(x,\ \ ) - d(x_\delta,\ \ ) \le \delta,$$

but

$$d(x,y) - d(x_\delta, y) = \delta.$$

Given $\varepsilon > 0$, there certainly exists $\delta > 0$ such that

$$\Delta f_\delta \le \{\psi'' + (n-1)(\mathbf{C}_\kappa/\mathbf{S}_\kappa)\psi'\} \circ d(x,\ \ ) + \varepsilon.$$

Then this f_δ satisfies the requirements of Definition 2 to imply (10).

Remark 1: Note that $\mathbf{S}_\kappa(d(x,y)) = 0$ if and only if $\kappa > 0$ and $d(x,y) = \pi/\sqrt{\kappa}$. By the Bonnet–Myers theorem (Section III.3), M is compact, and, by the generalized Toponogov theorem (Section III.4), M is isometric to the standard n-sphere of radius $\pi/\sqrt{\kappa}$. But for M compact, we already have the desired uniqueness theorem (cf. Section VI.1 or Theorem 1 of this chapter).

PROOF OF THEOREM 3: We fix a function $\psi\colon (0,\infty) \to \mathbb{R} \in C^2$, for which $\psi' \ge 0$ on all of $(0,\infty)$, and

$$\psi\,|\,(0,\tfrac{1}{2}) = 0, \qquad \psi\,|\,[1,\infty) = id.$$

Let $f\colon M \to \mathbb{R}$ be given by (7), for a fixed $x \in M$. Then by Lemma 2, and the fact that f vanishes on a neighborhood of x, we have the existence of a constant δ for which

$$\Delta f \prec \delta$$

on all of M.

For $u\colon M \times [0, \infty) \to \mathbb{R} \in C^0$ a solution of the heat equation on $M \times (0, \infty)$, fix $T > 0$, let

$$N_0 = \sup_M |u(\ ,0)|, \qquad N = \sup_{M \times [0,T)} |u|,$$

and define v by

$$v = u - N_0 - (N/R)(f + \delta t),$$

where R is a positive constant.

Restrict v to $\overline{\mathsf{B}(x;R)} \times [0, T]$. One immediately has

$$Lv = (N/R)(\delta - \Delta f) \succ 0$$

and

$$v \,|\, (\overline{\mathsf{B}(x;R)} \times \{0\}) \cup (\mathsf{S}(x;R) \times [0,T)) \leq 0,$$

which implies, by Theorem 2, that

$$v \,|\, \overline{\mathsf{B}(x;R)} \times [0,T] \leq 0,$$

that is,

$$u \leq N_0 + (N/R)(f + \delta t)$$

on $\overline{\mathsf{B}(x;R)} \times [0,T)$. Thus

$$u \leq N_0 = \sup_M |u(\ ,0)|$$

on all of $M \times [0, T)$, for all $T > 0$. Since u satisfies the heat *equation*, we may apply the foregoing argument to $-u$. We obtain

$$|u| \leq \sup_M |u(\ ,0)|$$

on all of $M \times [0, \infty)$ (in particular, $N = N_0$ for every $T > 0$).

COROLLARY 1. If for $R > 0$, we set

$$N_1 = \sup_{\overline{\mathsf{B}(x;R)}} |u(\ ,0)|,$$

then

$$|u| \leq N_1 + (N_0/R)(f + \delta t) \tag{11}$$

on $\overline{\mathsf{B}(x;R)} \times [0, \infty)$.

2. THE HEAT KERNEL FOR NONCOMPACT MANIFOLDS

We now turn to existence questions, but first we require

LEMMA 3. Let M be an arbitrary Riemmanian manifold, and $u_l\colon M \times (\alpha, \beta) \to \mathbb{R} \in C^\infty$, $l = 1, 2, \ldots$, such that u_l is a nondecreasing sequence of solutions to the heat equation on $M \times (\alpha, \beta)$, satisfying

$$\int_M |u_l(x, t)| \, dV(x) \leq \text{const} \tag{12}$$

for all $l = 1, 2, \ldots, t \in (\alpha, \beta)$, where the constant is independent of l and t. Then the sequence u_l converges to a C^∞ solution u to the heat equation on $M \times (\alpha, \beta)$. All derivatives, of all orders, of u_l converge to those of u. All convergence, referred to above, is uniform on compact subsets of M.

PROOF: Let Ω be a regular domain in M, with boundary Γ, and Dirichlet heat kernel q_Ω; let $[t_1, t] \subseteq (\alpha, \beta)$, $t_1 < t$. Fix $h \in C_c^\infty(\Omega)$ such that $h = 1$ on some $V \subseteq \Omega$. Then for $x \in V$ and any solution v of the heat equation on $M \times (\alpha, \beta)$, we have, by Duhamel's principle (II), that is, (VII.26),

$$\begin{aligned}
v(x, t) &= \int_\Omega v(y, t_1) h(y) q_\Omega(x, y, t - t_1) \, dV(y) \\
&\qquad - \int_{t_1}^{t} d\tau \int_\Omega \{2\langle \operatorname{grad}_y v, \operatorname{grad} h\rangle + v\, \Delta h\}(y, \tau) q_\Omega(x, y, t - \tau) \, dV(y) \\
&= \int_\Omega v(y, t_1) h(y) q_\Omega(x, y, t - t_1) \, dV(y) \\
&\qquad + \int_{t_1}^{t} d\tau \int_\Omega v(y, \tau) \{2\langle (\operatorname{grad} h)(y), (\operatorname{grad}_y q_\Omega)(x, y, t - \tau)\rangle \\
&\qquad\qquad + (\Delta h)(y) q_\Omega(x, y, t - \tau)\} \, dV(y) \\
&= \int_\Omega v(y, t_1) h(y) q_\Omega(x, y, t - t_1) \, dV(y) \\
&\qquad + \int_{t_1}^{t} d\tau \int_{\Omega \setminus V} v(y, \tau) \{2\langle (\operatorname{grad} h)(y), (\operatorname{grad}_y q_\Omega)(x, y, t - \tau)\rangle \\
&\qquad\qquad + (\Delta h)(y) q_\Omega(x, y, t - \tau)\} \, dV(y).
\end{aligned}$$

One obtains the second equality via Green's formula; and the third equality is a consequence of $h \,|\, V = 1$.

Thus $v(x,t)$ is expressed as integrals, over $\operatorname{supp} h$ and $\{\operatorname{supp} h\}\backslash V$, respectively, against C^∞ kernels. Space differentiations of v may be carried out by differentiating the kernels under the integral sign. The result is that locally uniform upper bounds of $|v|$, and its space derivatives, may be expressed as constant multipliers of the L^1-norm of v on Ω. The time derivatives are estimated via $\Delta v = \partial v/\partial t$.

We now consider the nondecreasing sequence u_l of solutions to the heat equation satisfying (12). Both $\{u_l\}$, $\{\operatorname{grad} u_l\}$ are locally uniformly bounded, from which one concludes that u_l converge to a continuous function u. By Dini's theorem, the convergence is uniform on compacta. By using the above integral expression for $v = u_l$, $l = 1, 2, \ldots$, one obtains the locally uniform convergence of derivatives of u_l, the existence of the corresponding derivatives of u, and the locally uniform convergence of derivatives of u_l to those of u. This proves the lemma.

We are now given a noncompact Riemannian manifold M. It need not be complete. To construct a fundamental solution of the heat equation on M, pick an exhaustion $\Omega_1, \Omega_2, \ldots$, of M by regular domains, namely,

$$\bar{\Omega}_j \subseteq \Omega_{j+1}$$

for all $j = 1, 2, \ldots$, and

$$\bigcup_{j=1}^{\infty} \Omega_j = M.$$

For each $j = 1, 2, \ldots$, let q_j be the Dirichlet heat kernel of Ω_j; we think of q_j as defined on all of $M \times M \times (0, \infty)$, vanishing identically whenever at least one of the space variables is in $M\backslash\Omega_j$. By (3), the sequence $\{q_j\}$ is an increasing sequence, so we may define

$$p = \lim_{j\to\infty} q_j.$$

THEOREM 4. The limit p is finite everywhere on $M \times M \times (0, \infty)$, and is a strictly positive C^∞ function on $M \times M \times (0, \infty)$. Also p is symmetric in the space variables, and satisfies

$$(L_x p)(x, y, t) = 0$$

on $M \times M \times (0, \infty)$. For all $t, s > 0$ we have

$$\int_M p(x, z, t)p(z, y, s)\, dV(z) = p(x, y, t + s) \tag{13}$$

on $M \times M$. More generally,

$$p = \sup_{\Omega} q_\Omega, \tag{14}$$

where Ω ranges over regular domains in M, and q_Ω denotes the Dirichlet heat kernel of Ω.

Finally, p is a fundamental solution to the heat equation on M; in fact, it is the minimal positive fundamental solution. If M is complete, with Ricci curvature bounded from below, then p is the *unique* heat kernel on M. In this last case, p is the unique continuous function on $M \times M \times (0, \infty)$ for which

$$u(x, t) = \int_M p(x, y, t)\varphi(y)\, dV(y) \tag{15}$$

always gives a solution to the initial value problem for the heat equation on M, with initial values φ, where φ is a bounded continuous on M.

PROOF: By (4), Lemma 3, and the symmetry of q_j, we have that for every $y \in M$, the sequence of functions $u_l(x, t) = q_l(x, y, t)$ converge locally uniformly, with derivatives, to a solution of the heat equation. So p is finite; it is positive since $p \geq q_j$, and $q_j > 0$ on Ω_j, for all j; and p is a solution of the heat equation in (x, t), for every fixed $y \in M$. The same argument also yields the symmetry of p in the space variables.

Equation (13) is a consequence of (VII.29) via Lebesgue's monotone convergence theorem; and (14) is obvious.

To show that $p \in C^\infty(M \times M(0, \infty))$ we consider the "heat operator" $\bar{L}$ on $(M \times M) \times (0, \infty)$ given by

$$\bar{L} = \Delta_x + \Delta_y - 2\,\partial/\partial t.$$

We have for every j,

$$\bar{L}q_j = 0,$$

and

$$0 < \iint_{\Omega \times \Omega} q_j(x, y, t)\, dV(x)\, dV(y) \leq V(\Omega)$$

for all j, and regular domains Ω in M. Then Lemma 3 implies that $p \in C^\infty(M \times M \times (0, \infty))$.

Next we note that (4) implies

$$\int_M p(x, y, t)\, dV(y) \leq 1 \tag{16}$$

for all $(x, t) \in M \times (0, \infty)$. Therefore, given a continuous bounded function φ on M, the function u, given by (15), is well defined and continuous on $M \times (0, \infty)$. To show that u satisfies the heat equation, it suffices (since φ is bounded) to consider the case where φ is nonnegative. In this case, u is the increasing limit of solutions to the heat equation, given by

$$u_j(x, t) = \int_M q_j(x, y, t)\varphi(y)\, dV(y),$$

with

$$\int_\Omega u_j(x, t)\, dV(x) \leq \{\sup \varphi\} V(\Omega),$$

for every bounded domain Ω in M. One now uses Lemma 3.

It remains to consider the behavior of p as $t \downarrow 0$. From (16) we have

$$\limsup_{t \downarrow 0} \int_M p(x, y, t)\, dV(y) \leq 1$$

for all $x \in M$. Next, we have for any open set U in M, and any regular domain, Ω, with compact closure in U,

$$\int_\Omega q_\Omega(x, y, t)\, dV(y) \leq \int_U p(x, y, t)\, dV(y)$$

for all $(x, t) \in M \times (0, \infty)$, from which one has

$$\liminf_{t \downarrow 0} \int_U p(x, y, t)\, dV(y) \geq 1,$$

for all $x \in M$, U open in M. Thus

$$\lim_{t \downarrow 0} \int_U p(x, y, t)\, dV(y) = 1 \tag{17}$$

for all $x \in M$, U open in M. But one now easily concludes that for bounded, measurable φ, and u given by (15), that

$$\lim_{t \downarrow 0} u(x, t) = \varphi(x)$$

when φ is continuous at x.

Therefore, p is a fundamental solution of the heat equation on M, and via (15), is the heat kernel for the initial value problem with bounded initial data. That p is the *minimal* positive fundamental solution, follows from (14), and the fact that for any fundamental solution P on M, and regular domain Ω in M, we have

$$q_\Omega \leq P.$$

The uniqueness of p, when M is complete with Ricci curvature bounded below, follows directly from Theorem 3.

THEOREM 5. Let M be complete, with Ricci curvature bounded from below. Then

$$\int_M p(x, y, t)\, dV(y) = 1 \tag{18}$$

for all $t > 0$. If $\varphi \in C_c^\infty(M)$, and u is given by (15), then

$$\int_M u(x, t)\, dV(x) = \int_M \varphi(x)\, dV(x) \tag{19}$$

for all $t > 0$ (Gaffney [1]).

If φ is continuous, bounded, and

$$\lim_{x \to \infty} \varphi(x) = 0,$$

then, for u given by (15), we have

$$\lim_{x \to \infty} u(x, t) = 0$$

for all $t > 0$ (Yau [4]).

PROOF: Equation (18) is the direct consequence of Theorem 3, (15), and (17). Equation (19) is a consequence of (18) via Fubini's theorem.

It remains to consider the last claim. Given $\varepsilon > 0$ there exists a compact set $K = K_\varepsilon$, in M, such that

$$|\varphi| \,|\, M \backslash K < \varepsilon.$$

For $x \in M\backslash K$, let $R = d(x, K)$, f the function in Theorem 3, centered at x (even though the function ψ, there, is independent of M, the function f depends on the choice of x). Then by Corollary 1,

$$|u(x, t)| \leq \varepsilon + N_0 \delta t / d(x, K).$$

One now easily has the desired result.

3. COMPARISON THEOREMS FOR HEAT KERNELS

We start by recalling from Section II.1 that if M is a Riemannian manifold, and $\Phi: M \to M$ is an isometry of M, then for any function $f: M \to \mathbb{R} \in C^2$, and induced action of Φ on f, given by

$$(\Phi^* f) = f \circ \Phi,$$

we have

$$\Phi^*(\Delta f) = \Delta(\Phi^* f).$$

Now let $\mathbb{M}_\kappa$ be the n-dimensional simply connected space form of constant sectional curvature κ, fix $x \in \mathbb{M}_\kappa$, and introduce geodesic spherical coordinates, about x, given by

$$y(t, \xi) = \exp t\xi,$$

where $0 \le t < \pi/\sqrt{\kappa}$ (we write $\pi/\sqrt{\kappa}$ for $+\infty$ when $\kappa \le 0$), $\xi \in \mathfrak{S}_x$. Fix $\delta \in (0, \pi/\sqrt{\kappa})$ and let q_κ denote the Dirichlet heat kernel of $\mathsf{B}_\kappa(x; \delta)$.

If $\Phi: \mathbb{M}_\kappa \to \mathbb{M}_\kappa$ is an isometry of $\mathbb{M}_\kappa$ for which

$$\Phi(x) = x,$$

then one can easily verify that

$$q_\kappa(x, y, t) = q_\kappa(x, \Phi(y), t).$$

Since the action of isometries of M, which leave x fixed, is transitive on $\mathfrak{S}_x$, we obtain the existence of a function $\mathscr{E}_\kappa: [0, \delta) \times (0, \infty) \to \mathbb{R}$ for which

$$q_\kappa(x, y, t) = \mathscr{E}_\kappa(d(x, y), t).$$

LEMMA 4. Write $\mathscr{E}_\kappa = \mathscr{E}_\kappa(r, t)$. Then

$$\partial \mathscr{E}_\kappa / \partial r < 0 \tag{20}$$

for $r, t > 0$.

PROOF: For sufficiently small $\varepsilon > 0$, $0 < r, t < \varepsilon$, we have, by Sections 3 and 4 of Chapter VI,

$$\mathscr{E}_\kappa(r, t) \sim \frac{e^{-r^2/4t}}{(4\pi t)^{n/2}} \{a_0(r) + a_1(r)t\},$$

and

$$\frac{\partial \mathscr{E}_\kappa}{\partial r} \sim \frac{e^{-r^2/4t}}{(4\pi t)^{n/2}}\left\{\frac{-r}{2t}\right\}\{a_0(r) + a_1(r)t\} + \frac{e^{-r^2/4t}}{(4\pi t)^{n/2}}\{a_0'(r) + a_1'(r)t\},$$

with $a_0(0) = 1$, and $a_0'(0) = a_1'(0) = 0$. So (20) is certainly valid in this case.

We now write

$$Q(y, t) = \mathscr{E}_\kappa(d(x, y), t), \tag{21}$$

$$(\partial Q/\partial r)(y, t) = (\partial \mathscr{E}_\kappa/\partial r)(d(x, y), t), \tag{22}$$

and let G be the subset of $\{\overline{\mathsf{B}_\kappa(x; \delta)} \times (0, t)\}$ consisting of those points at which $\partial Q/\partial r$ is strictly positive. For each $\tau \in (0, t)$ let

$$G_\tau = \{y \in \mathsf{B}_\kappa(x; \delta) : (y, \tau) \in G\};$$

then for any s in $(0, t)$ we have

$$\begin{aligned}
\tfrac{1}{2}\iint_{G_t} (\partial Q/\partial t)^2\, dV &- \tfrac{1}{2}\iint_{G_s} (\partial Q/\partial r)^2\, dV \\
&= \int_s^t d\tau \iint_{G_\tau} \frac{\partial^2 Q}{\partial \tau \partial r}\frac{\partial Q}{\partial r}\, dV \\
&= \int_s^t d\tau \iint_{G_\tau} \frac{\partial^2 Q}{\partial r \partial \tau}\frac{\partial Q}{\partial r}\, dV \\
&= \int_s^t d\tau \iint_{G_\tau} \left(\frac{\partial}{\partial r}\Delta Q\right)\left(\frac{\partial Q}{\partial r}\right) V \\
&= \int_s^t d\tau \iint_{G_\tau} \langle \operatorname{grad} \Delta Q, \operatorname{grad} Q\rangle\, dV \\
&= -\int_s^t d\tau \iint_{G_\tau} (\Delta Q)^2\, dV + \int_s^t d\tau \int_{\partial G_\tau} (\Delta Q)(\partial Q/\partial r)\, dA \\
&= -\int_s^t d\tau \iint_{G_\tau} (\Delta Q)^2\, dV.
\end{aligned}$$

We note that the first equality requires an argument modeled on that of Lemma 1.1 in Cheeger–Yau [1]; that Green's formula may be applied in the fifth equality, since G_τ is an, at most countable, union of open geodesic annuli. The last equality results from the fact that $\partial Q/\partial r$ vanishes on ∂G_τ, for each $\tau \in (0, t)$. Indeed, since $Q > 0$ on $\mathsf{B}_\kappa(x; \delta) \times (0, t)$, and $Q = 0$ on $\mathsf{S}_\kappa(x; \delta) \times (0, t)$, it is impossible that $\overline{G}$ intersect $\mathsf{S}_\kappa(x; \delta) \times (0, t)$.

So for $0 < s < t$, we have

$$\frac{1}{2}\iint_{G_t} (\partial Q/\partial r)^2\,dV - \frac{1}{2}\iint_{G_s} (\partial Q/\partial r)^2\,dV = -\int_s^t d\tau \iint_{G_\tau} (\Delta Q)^2\,dV.$$

Since G is bounded away from $(x, 0)$, we have, by letting $s \downarrow 0$,

$$\frac{1}{2}\iint_{G_t} (\partial Q/\partial r)^2\,dV = -\int_0^t d\tau \iint_{G_\tau} (\Delta Q)^2\,dV$$

–a contradiction, unless G is empty.

Therefore $\partial \mathscr{E}_\kappa/\partial r \le 0$ on all of $[0, \delta] \times (0, \infty)$. But $U = \partial \mathscr{E}_\kappa/\partial r$ is a solution of

$$\frac{\partial^2 U}{\partial r^2} + (n-1)\frac{\mathbf{C}_\kappa}{\mathbf{S}_\kappa}\frac{\partial U}{\partial r} - \frac{(n-1)}{\mathbf{S}_\kappa^2}U = \frac{\partial U}{\partial t}$$

on $(0, \delta) \times (0, \infty)$. The maximum principle (Protter–Weinberger [1, pp. 173–175]), on which the strong maximum principle of Section 1 is based, states that if U is nonconstant, then U has no maximum on $(0, \delta) \times (0, \infty)$. We therefore have $\partial \mathscr{E}_\kappa/\partial r < 0$ on $(0, \delta) \times (0, \infty)$.

Finally, we have $(\partial \mathscr{E}_\kappa/\partial r)(\delta, t) < 0$ by the strong maximum principle (Protter–Weinberger [1, pp. 173–175]) applied to

$$\frac{\partial^2 \mathscr{E}_\kappa}{\partial r^2} + (n-1)\frac{\mathbf{C}_\kappa}{\mathbf{S}_\kappa}\frac{\partial \mathscr{E}_\kappa}{\partial r} = \frac{\partial \mathscr{E}_\kappa}{\partial t}, \tag{23}$$

with $\mathscr{E}_\kappa(\delta, t) = 0$. The lemma is now proven.

We now consider an arbitrary noncompact Riemannian manifold M, $x \in M$, and $\delta > 0$, such that $\overline{\mathsf{B}(x;\delta)}$ is in the domain of the exponential map. In particular, $\mathsf{B}(x;\delta)$ is relatively compact in M. We let q be the heat kernel of $\mathsf{B}(x;\delta)$, and $Q(y, t)$ be given by (21)—*except that*, now, *Q is defined on* $\mathsf{B}(x;\delta) \times (0, \infty)$, and not on $\mathsf{B}_\kappa(x;\delta) \times (0, \infty)$.

Note that

$$LQ = \frac{\partial^2 Q}{\partial r^2} + \frac{\partial r\sqrt{\mathfrak{g}}}{\sqrt{\mathfrak{g}}}\frac{\partial Q}{\partial r} - \frac{\partial Q}{\partial t}, \tag{24}$$

where $\sqrt{\mathfrak{g}}$ is given by Definition III.6, and $\mathscr{E}_\kappa$ satisfies (23).

THEOREM 6 (Debiard–Gaveau–Mazet [1]; Cheeger–Yau [1]. If

$$\overline{\mathsf{B}(x;\delta)} \subseteq \mathsf{D}_x, \tag{25}$$

that is, $\mathsf{B}(x;\delta)$ is contained inside the cut locus of x, and all the sectional curvatures of $\mathsf{B}(x;\delta)$ are $\leq\kappa$, then

$$q(x, y, t) \leq Q(y, t) \tag{26}$$

for all $(y, t) \in \mathsf{B}(x;\delta) \times (0, \infty)$, with equality at some

$$(y_0, t_0) \in \mathsf{B}(x;\delta) \times (0, \infty)$$

if and only if $\mathsf{B}(x;\delta)$ is isometric to the geodesic disk, of radius δ, in $\mathbb{M}_\kappa$.

If, on the other hand, we are given that all the Ricci curvatures of $\mathsf{B}(x;\delta)$ are $\geq(n-1)\kappa$, where $n = \dim M$, then

$$q(x, y, t) \geq Q(y, t) \tag{27}$$

for all $(y, t) \in \mathsf{B}(x;\delta) \times (0,\infty)$, with equality at some $(y_0, t_0) \in \mathsf{B}(x;\delta) \times (0,\infty)$ if and only if $\mathsf{B}(x;\delta)$ is isometric to the geodesic disk, of radius δ, in $\mathbb{M}_\kappa$. In this second case, where the Ricci curvatures of $\mathsf{B}(x;\delta)$ are bounded from below, we do not require assumption (25).

PROOF: We first assume that δ is picked sufficiently small so that (25) is valid.

If all the sectional curvatures of $\mathsf{B}(x;\delta)$ are $\leq\kappa$, then the Bishop comparison theorem (I) (cf. Section III.2) implies

$$\partial_r\sqrt{\mathfrak{g}}/\sqrt{\mathfrak{g}} \geq (n-1)\mathbf{C}_\kappa/\mathbf{S}_\kappa,$$

which implies, by using (20), (23), (24),

$$LQ \leq 0 = L(q(x, \ , \ )).$$

Thus

$$L(Q - q(x, \ , \ )) \leq 0, \tag{28}$$

and

$$\{Q - q(x, \ , \ )\} \,|\, (\mathsf{B}(x;\delta) \times \{0\}) \cup (\mathsf{S}(x;\delta) \times (0,\infty)) = 0. \tag{29}$$

The strong maximum principle now implies inequality (26), and the characterization of equality in (26).

If, on the other hand, all the Ricci curvatures on $\mathsf{B}(x;\delta)$ are bounded below by $(n-1)\kappa$, then

$$L(Q - q(x, \ , \ )) \geq 0, \tag{30}$$

with (29). The strong maximum principle will then imply (27), with the characterization of equality in (27).

If we are not given that (25) is valid, then, when all Ricci curvatures of $\mathsf{B}(x;\delta)$ are bounded below by $(n-1)\kappa$, the earlier arguments of Lemma 2 imply that (30) can be replaced by

$$L(Q - q(x,\ ,\ )) \succ 0. \tag{31}$$

By (29), and Theorem 2, we have (27).

The case of equality in (27), without assumption (25), requires a stronger version of Theorem 2. One can construct this stronger version by using Theorem 5 of Protter–Weinberger [1, Section III.3].

An immediate application of Theorem 6 is

THEOREM 7. Let M be complete with Ricci curvature bounded below by $(n-1)\kappa$, $n = \dim M$, on all of M, and let p_κ the heat kernel of $\mathbb{M}_\kappa$ be given by

$$p_\kappa(\underline{x}, \underline{y}, t) = \mathbf{E}_\kappa(d_\kappa(\underline{x}, \underline{y}), t)$$

for $\underline{x}$, $\underline{y} \in \mathbb{M}_\kappa$, where d_κ denotes distance in $\mathbb{M}_\kappa$ and

$$\mathbf{E}_\kappa\colon [0, \pi/\sqrt{\kappa}] \times (0, \infty) \to \mathbb{R}.$$

Then

$$p(x, y, t) \geq \mathbf{E}_\kappa(d(x, y), t) \tag{32}$$

for all $(x, y, t) \in M \times M \times (0, \infty)$.

Note that to give an upper bound of p, one will require that the injectivity radius of M be $\pi/\sqrt{\kappa}$—even then, the upper bound, when $\kappa > 0$, is only valid in $\mathsf{B}(x; \pi/\sqrt{\kappa})$.

4. UPPER BOUNDS FOR THE HEAT KERNEL

We now discuss an upper bound for the heat kernel that will be valid past the cut locus. M is still a noncompact Riemannian manifold. The starting point of this discussion is the following result of Moser [1, pp. 115–117]:

LEMMA 5. Let u be a nonnegative solution of the heat equation on $M \times [0, \infty)$. There exists a constant, depending only on $n = \dim M$, such that for all $r > 0$, for which $\mathsf{B}(x; r)$ is in the domain of $\exp_x$, we have

$$u(x, T) \leq \text{const}\{T^{-(n/2+1)} + r^{-(n+2)}\}^{1/2} \cdot (\mathfrak{I}(\mathsf{B}(x; r)))^{-1/2} \|u\|_{\mathsf{B}(x;r)\times[0,T]}, \tag{33}$$

where $\mathfrak{I}(\mathsf{B}(x;r))$ is the isoperimetric constant of $\mathsf{B}(x;r)$ (cf. Definition IV.3; recall that, by Theorem IV.4, $\mathfrak{I}(\mathsf{B}(x;r))$ is equal to the Sobolev constant $\mathfrak{s}(\mathsf{B}(x;r))$), and $\|u\|_{\mathsf{B}(x;r)\times[0,T]}$ is the L^2-norm of u over the manifold $\mathsf{B}(x;r)\times[0,T]$, endowed with the natural product metric.

We shall first discuss the applications of Lemma 5, and then its proof. For convenience we shall henceforth write

$$\mathfrak{I}(x;r)=\mathfrak{I}(\mathsf{B}(x;r)). \tag{34}$$

Let $u(x,t)$ be given by

$$u(x,t)=\int_M p(x,y,t)\varphi(y)\,dV(y), \tag{15}$$

where $\varphi\in C_c^\infty(M)$, $\varphi\geq 0$. Then

$$u^2(x,t)\leq\int_M p(x,y,t)\varphi^2(y)\,dV(y)$$

by the Cauchy–Schwarz inequality (applied to the measure $p(x,\ ,t)\,dV$, using (16)). Thus

$$\|u(\ ,t)\|\leq\|\varphi\| \tag{35}$$

which implies by (33),

$$u(x,T)\leq\text{const}\cdot\sqrt{T}\{T^{-(n/2+1)}+r^{-(n+2)}\}^{1/2}\mathfrak{I}^{-1/2}(x;r)\|\varphi\|. \tag{36}$$

If we think of $u(x,t)$ as

$$u(x,t)=\langle p(x,\ ,t),\varphi\rangle,$$

then (36) reads as

$$\langle p(x,\ ,T),\varphi\rangle\leq\text{const}\cdot\{T^{-n/2}+Tr^{-(n+2)}\}^{1/2}\mathfrak{I}^{-1/2}(x;r)\|\varphi\|,$$

which implies

$$\|p(x,\ ,T)\|_{\mathsf{B}(y;r)}\leq\text{const}\cdot\{T^{-n/2}+Tr^{-(n+2)}\}^{1/2}\mathfrak{I}^{-1/2}(x;r).$$

Therefore

$$\|p(x,\ ,T)\|_{\mathsf{B}(y;r)\times[T/2,T]}\leq\text{const}\cdot\sqrt{T}\{T^{-n/2}+Tr^{-(n+2)}\}^{1/2}\mathfrak{I}^{-1/2}(x;r). \tag{37}$$

Finally, set, for fixed x,

$$v(z,t)=p(x,z,t+T/2).$$

Then, applying (33) to v, and using (37), we obtain

THEOREM 8. For all $x, y \in M$, and $r > 0$, for which $\overline{\mathsf{B}(x; r)}$ and $\overline{\mathsf{B}(y; r)}$ are in the image of $\exp_x$ and $\exp_y$, respectively, we have the estimate

$$p(x, y, t) \leq \text{const} \cdot \{t^{-n/2} + tr^{-(n+2)}\}(\mathfrak{V}(x; r)\mathfrak{V}(y; r))^{-1/2}, \tag{38}$$

the constant depending only on $n = \dim M$.

We now give an improved estimate under the assumption that M is complete, and that x, y, r satisfy

$$d(x, y) > 2r. \tag{39}$$

THEOREM 9. If M is complete and x, y, r satisfy (39), then

$$p(x, y, t) \leq \text{const} \cdot \{t^{-n/2} + tr^{-(n+2)}\}(\mathfrak{V}(x; r)\mathfrak{V}(y; r))^{-1/2}e^{-(d-2r)^2/4t}, \tag{40}$$

where $d = d(x, y)$.

Surprisingly enough, the proof that we sketch is based on a fundamental contrast between the heat and wave equations, namely, the speed of propagation of the initial data. From (15), and the strict positivity of the heat kernel p on $M \times M \times (0, \infty)$, one has that the speed of the propagation of the initial data is infinite—in sharp contrast to physical understanding of heat diffusion via molecular collisions. Of course, the point of (40) is that, as in Euclidean space, the *effects* of the propagation are exceptionally small, for short times and across large distances.

On the other hand, light travels at finite speed, in fact, solutions to the wave equation

$$\Delta v = \partial^2 v/\partial t^2 \tag{41}$$

propagate their initial data at unit speed.

PROPOSITION 3. *Given* $\varphi \in C_c^\infty(M)$, *let* K *be the support of* φ, *and* $v(x, t)$ *the solution of the wave equation (41) satisfying the initial conditions*

$$v(\ , 0) = \varphi, \qquad (\partial_t v)(\ , 0) = 0.$$

Then

$$\operatorname{supp} v(\ , t) \subseteq \{x \in M : d(x, K) \leq t\}, \tag{42}$$

and

$$\|v(\ ,t)\| \leq \|\varphi\|, \tag{43}$$

for all t. The solution v always determines a solution u of the heat equation, with initial data φ, via

$$u(x,t) = \int_{-\infty}^{\infty} \frac{e^{-s^2/4t}}{\sqrt{4\pi t}} v(x,s)\, ds. \tag{44}$$

For u given by (44) we have, for all $t > 0$,

$$\|u(\ ,t)\| \leq \|\varphi\|, \tag{45}$$

$$\|u(\ ,t)\|_{M\setminus \mathsf{B}(x;r)} \leq \text{const} \cdot \|\varphi\| e^{-(r-\rho)^2/4t}, \tag{46}$$

where $K \subseteq \mathsf{B}(x;\rho)$.

Once one has this proposition, the proof of Theorem 4 is the same as the derivation of Theorem 8 from Lemma 5, except that, for x, y satisfying (39), one only allows φ to vary over $C_c^\infty(\mathsf{B}(y;r))$. Then (46), applied to $M\setminus\mathsf{B}(y;r)$, will imply

$$\|u(\ ,t)\|_{\mathsf{B}(x;r)} \leq \text{const} \cdot \|\varphi\| e^{-(d-2r)^2/4t}. \tag{47}$$

One now uses (47) instead of (35).

PROOF OF PROPOSITION 3: Certainly, to prove (42), it suffices to show that if

$$\varphi \mid \mathsf{B}(x;\rho) = 0 \tag{48}$$

for a given $\rho > 0$, then

$$v(x,\rho) = 0. \tag{49}$$

Assume φ satisfies (48) for some given $\rho > 0$. Set

$$\Omega = \{(y,t) : d(x,y) \leq \rho - t\},$$
$$\Omega_t = \Omega \cap (M \times \{t\}),$$

for $t \in [0,\rho]$, and

$$E(t) = \tfrac{1}{2} \iint_{\Omega_t} \{|\text{grad}\, v|^2 + |\partial_t v|^2\}\, dV,$$

where the gradient (henceforth) is with respect to the space variable only. Then

$$\begin{aligned} E'(t) &= \iint_{\Omega_t} \{\langle \operatorname{grad} \partial_t v, \operatorname{grad} v\rangle + \partial_t^2 v \partial_t v\}\, dV \\ &\quad - \tfrac{1}{2} \int_{\partial\Omega_t} \{|\operatorname{grad} v)^2 + |\partial_t v|^2\}\, dV \\ &= \int_{\partial\Omega_t} \left\{\partial_t v \frac{\partial v}{\partial \nu} - \tfrac{1}{2}[|\operatorname{grad} v|^2 + |\partial_t v|^2]\right\} dA, \end{aligned}$$

and

$$\partial_t v(\partial v/\partial \nu) \leq |\partial_t v|\,|\operatorname{grad} v| \leq \tfrac{1}{2}\{|\operatorname{grad} v|^2 + |\partial_t v|^2\}.$$

Thus $E'(t) \leq 0$ for all $t > 0$, and (49) follows.

To prove (45), let Ω be a regular domain in M with

$$K \subset \Omega, \qquad d(K, M \backslash \Omega) > t$$

for a given t. Then $v(\ ,t) \in C_c^\infty(\Omega)$. Let $\{\varphi_1, \varphi_2, \ldots\}$ be a complete orthonormal basis of $L^2(\Omega)$, consisting of Dirichlet eigenfunctions of Ω, so indexed that λ_j, the eigenvalue of φ_j, is nondecreasing with respect to j. One easily verifies that for each $l = 1, 2, 3, \ldots$, $\Delta^l\varphi$ has the Sturm–Liouville expansion

$$\Delta^l \varphi \sim \sum_{j=1}^{\infty} (-\lambda_j)^l \alpha_j \varphi_j,$$

which implies

$$\alpha_j = O(\lambda_j^{-l}) \tag{50}$$

as $j \to \infty$, for all $l = 1, 2, \ldots$. It is standard that $v(x, t)$ has the expansion

$$v(x, t) = \sum_{j=1}^{\infty} \alpha_j (\cos\sqrt{\lambda_j}\, t)\, \varphi_j(x),$$

and (45) follows by Parseval's theorem.

To study $u(x, t)$ given by (44), one requires an argument most neatly presented via spectral theory (and we shall omit it here, but compare Rauch–Taylor [1] and Kannai [1]) that the formal calculations of Lu, the heat operator applied to u, is legitimately carried out under the integral sign. The key is that

$$e(s, t) \equiv: e^{-s^2/4t}/\sqrt{4\pi t}$$

is the heat kernel of $\mathbb{R}^1$, with s playing the role of the space variable (cf. Hersh [1]). It is rather easy to establish that

$$u(x, t) \to \varphi(x)$$

as $t \downarrow 0$. And to prove (45), one simply has

$$\begin{aligned}\int_M u^2(x, t)\,dV(x) &= \int_M dV(x)\left\{\int_{-\infty}^{\infty} e(s, t)v(x, s)\,ds\right\}^2\\ &\le \int_M dV(x)\int_{-\infty}^{\infty} e(s, t)v^2(x, s)\,ds\\ &= \int_{-\infty}^{\infty} \|v(\ , s)\|^2 e(s, t)\,ds\\ &\le \|\varphi\|^2.\end{aligned}$$

To obtain (4.6), we have

$$\begin{aligned}\int_{M\setminus B(x;r)} u^2(y, t)\,dV(y) &\le \int_{M\setminus B(x;r)} dV(y)\int_{-\infty}^{\infty} e(s, t)v^2(y, s)\,ds\\ &= 2\int_{M\setminus B(x;r)} dV(y)\int_{r-\rho}^{\infty} e(s, t)v^2(y, s)\,ds\\ &\le 2\|\varphi\|^2\int_{r-\rho}^{\infty} e^{-s^2/4t}/\sqrt{4\pi t}\,ds\\ &\le \text{const}\cdot\|\varphi\|^2 e^{-(r-\rho)^2/4t}.\end{aligned}$$

The third line uses (42), and the fact that v is even with respect to s; and the last inequality is an easy argument. This concludes the discussion of the proposition.

THEOREM 10 (Varadhan [1]). If M is a complete Riemannian manifold, with Ricci curvature bounded from below, then

$$\lim_{t\downarrow 0} -(4t)\ln p(x, y, t) = d^2(x, y).$$

PROOF: From (40) we have

$$\liminf_{t\downarrow 0} -(4t)\ln p(x, y, t) \ge \{d(x, y) - 2r\}^2$$

for all sufficiently small $r > 0$. Therefore

$$\liminf_{t\downarrow 0} -(4t)\ln p(x,y,t) \ge d^2(x,y).$$

On the other hand, if $(n-1)\kappa$ is the lower bound of the Ricci curvatures, with $n = \dim M$, then (32) implies

$$\begin{aligned}\limsup_{t\downarrow 0} -(4t)\ln p(x,y,t) &\le \limsup_{t\downarrow 0} -(4t)\ln \mathrm{E}_\kappa(d(x,y),t)\\ &= d^2(x,y)\end{aligned}$$

—this last equality being explicitly verifiable (cf. Section VI.3).

We note that to use (38), (40) geometrically, one has to bound $\mathfrak{I}(x;r)$, from below, in terms of the geometry of the manifold. By Theorem V.5, it suffices to bound, from below, Croke's constant $\omega(\mathsf{B}(x;r))$. In Cheeger–Gromov–Taylor [1, Sec. 4] it is shown that knowledge of the volume of one geodesic disk, and a lower bound for the Ricci curvature, suffices to determine a lower bound (for "medium size" values of r), for $\omega(\mathsf{B}(x;r))$. Compare their discussion for this and other interesting results. Also, compare Cheng–Li–Yau [1, Sec. 1].

We now discuss the

PROOF OF LEMMA 5: We consider the case $n > 2$. Also we will find it convenient to think of the time variable as varying in $[-T, 0]$, and set

$$\mathsf{R}_{\rho,\tau} = \overline{\mathsf{B}(x;\rho)} \times [-\tau, 0]$$

for $0 < \rho < r$, $0 < \tau < T$. Also, we set

$$\mathfrak{I} = \mathfrak{I}(x;r)$$

for the rest of the argument. The main estimate is: for any $w \in C^\infty(\mathsf{R}_{\rho,\tau})$ satisfying

$$w \ge 0, \qquad Lw \ge 0,$$

we have

$$\int_{\mathsf{R}_{\rho',\tau'}} w^{2\alpha}\,dV\,dt \le c(n)\mathfrak{I}^{-2/n}\{(\rho-\rho')^{-2} + (\tau-\tau')^{-1}\}^{\alpha}\|w\|^{2\alpha}_{\mathsf{R}_{\rho,\tau}} \tag{51}$$

for all $\rho' \in (0,\rho)$, $\tau' \in (0,\tau)$, where $c(n)$ always denotes some constant depending only on n, and α, for the rest of the discussion, is given by

$$\alpha = 1 + 2/n.$$

For the moment, assume the validity of this last claim. We show how it implies the lemma. Fix ρ, τ, and set

$$\rho_\mu = (\tfrac{1}{2} + 2^{-\mu})\rho, \qquad \tau_\mu = (\tfrac{1}{2} + 2^{-\mu})\tau,$$
$$\mathsf{R}_\tau = \mathsf{R}_{\rho_\mu, \tau_\mu},$$

for $\mu = 1, 2, \ldots$. Then, employing the notation of Section IV.4, and using the fact that for any $\beta > 2$, we also have $Lw^\beta \geq 0$, we obtain, via (51),

$$\|w\|_{2\alpha^{\mu+1}, \mathsf{R}_{\mu+2}} \leq \{c(n)\mathfrak{I}^{-2/n}\}^{1/2\alpha^{\mu+1}} 2^{\mu/\alpha^\mu}\{\rho^{-2} + \tau^{-1}\}^{1/2\alpha^\mu} \|w\|_{2\alpha^\mu, \mathsf{R}_{\mu+1}},$$

which, in turns, implies

$$\sup_{\overline{\mathsf{B}(x;\rho)} \times [-\tau/2, 0]} w \leq c(n)\mathfrak{I}^{-1/2}\{\rho^{-2} + \tau^{-1}\}^{(n+2)/4} \|w\|_{\mathsf{R}_{\rho,\tau}},$$

which is the claim of the lemma.

We therefore wish to prove the estimate (51). To this end, fix, for the moment $\tau_1 > \tau_2 > 0$, and set

$$\mathsf{R} = \overline{\mathsf{B}(x_0; \rho)} \times [-\tau_1, -\tau_2].$$

Consider $\varphi \in C^\infty(\mathsf{R})$ such that φ is nonnegative, and vanishes on $\mathsf{S}(x_0; \rho) \times [-\tau_1, -\tau_2]$. Then $w \geq 0$, $Lw \geq 0$ imply, by Green's formula,

$$\iint_{\mathsf{R}} \{\varphi \partial_t w + \langle \operatorname{grad} \varphi, \operatorname{grad} w\rangle\}\, dV\, dt \leq 0.$$

If, in addition, φ is of the form

$$\varphi = \psi^2 w,$$

we then obtain

$$\begin{aligned}\iint_{\mathsf{R}} &\{\psi^2 w \partial_t w + \psi^2 |\operatorname{grad} w|^2\}\, dV\, dt \\ &\leq -2 \iint_{\mathsf{R}} \psi w \langle \operatorname{grad} \psi, \operatorname{grad} w\rangle\, dV\, dt \\ &\leq \iint_{\mathsf{R}} \{\tfrac{1}{2}\psi^2 |\operatorname{grad} w|^2 + 2w^2 |\operatorname{grad} \psi|^2\}\, dV\, dt.\end{aligned}$$

Thus

$$\tfrac{1}{2}\iint_{\mathsf{R}} \psi^2\{\partial_t(w^2) + |\text{grad } w|^2\}\, dV\, dt \le 2\iint_{\mathsf{R}} |\text{grad } \psi|^2 w^2\, dV\, dt,$$

which implies

$$\iint_{\mathsf{R}} \{\partial_t(\psi^2 w^2) + |\text{grad}(\psi w)|^2\}\, dV\, dt \le c \iint_{\mathsf{R}} \{\psi\partial_t\psi + |\text{grad } \psi|^2\}\, w^2\, dV\, dt,$$

(52)

where c is some universal constant.

Now pick ψ: $\mathsf{B}(x_0;\rho) \times [-\tau, 0] \to \mathbb{R}$ to be

$$\psi(x, t) = \psi_1(d(x, x_0))\psi_2(t),$$

where ψ_1, ψ_2 are given in Fig. 2 (p. 206).

Let $\tau_0 \ge 0$ have the property that

$$\int_{\mathsf{B}(x_0;\rho)\times\{t\}} \psi^2 w^2\, dV \le \int_{\mathsf{B}(x_0;\rho)\times\{-\tau_0\}} \psi^2 w^2\, dV$$

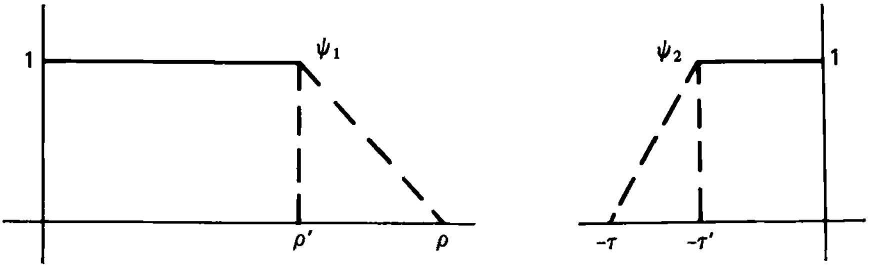
for all $t \in [-\tau, 0]$. Since $\psi \,|\, \mathsf{B}(x_0;\rho) \times \{-\tau\} = 0$, we have, by (52) for $\tau_1 = \tau$, $\tau_2 = \tau_0$,

$$\begin{aligned}\int_{\mathsf{B}(x_0;\rho)\times\{t\}} \psi^2 w^2\, dV &\le \int_{\mathsf{B}(x_0;\rho)\times\{-\tau_0\}} \psi^2 w^2\, dV \\ &= \iint_{\mathsf{B}(x_0;\rho)\times[-\tau,-\tau_0]} \partial_t(\psi^2 w^2)\, dV\, dt \\ &\le c\iint_{\mathsf{R}_{\rho,\tau}} \{\psi\, \partial_t\psi + |\text{grad } \psi|^2\}\, w^2\, dV\, dt \\ &\le c\{(\rho - \rho')^{-2} + (\tau - \tau')^{-1}\}\, \|w\|^2_{\mathsf{R}_{\rho,\tau}},\end{aligned}$$

that is,

$$\int_{\mathsf{B}(x_0;\rho)\times\{t\}} \psi^2 w^2\, dV \le c\{(\rho - \rho')^{-2} + (\tau - \tau')^{-1}\}\, \|w\|^2_{\mathsf{R}_{\rho,\tau}} \tag{53}$$

for all $t \in [-\tau, 0]$.



(a) (b)

Fig. 2

Again, because $\psi \mid \mathrm{B}(x_0;\rho) \times \{-\tau\} = 0$, we have, using (52) for $\tau_1 = \tau$, $\tau_2 = 0$,

$$\iint_{\mathrm{R}_{\rho,\tau}} |\mathrm{grad}(\psi w)|^2 \, dV\, dt \le c\{(\rho - \rho')^{-2} + (\tau - \tau')^{-1}\} \|w\|^2_{\mathrm{R}_{\rho,\tau}}. \tag{54}$$

Recall that $n > 2$. Then by Theorem IV.7 we have

$$D[f, f] \ge c(n)\mathfrak{I}^{-2/n}\|f\|^2_{2n/(n-2)}$$

for all $f \in C_c^\infty(\mathrm{B}(x; r))$, which implies, by Hölder's inequality,

$$\int w^{2\alpha}\, dV \le c(n)\mathfrak{I}^{-2/n}\|w\|^{4/n} D[w, w] \tag{55}$$

for $w \in C_c^\infty({}^{\mathrm{B}}(x; r))$.

From (53) and (55) we have, for all t,

$$\begin{aligned}
&\int_{\mathrm{B}(x_0;\rho') \times \{t\}} w^{2\alpha}\, dV \\
&\quad \le \int_{\mathrm{B}(x_0;\rho) \times \{t\}} (\psi w)^{2\alpha}\, dV \\
&\quad \le c(n)\mathfrak{I}^{-2/n}\left\{\int_{\mathrm{B}(x_0;\rho) \times \{t\}} |\mathrm{grad}(\psi w)|^2\, dV\right\} \|\psi w\|^{4/n}_{\mathrm{B}(x_0;\rho) \times \{t\}} \\
&\quad \le c(n)\mathfrak{I}^{-2/n}\{(\rho - \rho')^{-2} + (\tau - \tau')^{-1}\}^{2/n} \|w\|^{4/n}_{\mathrm{R}_{\rho,\tau}} \\
&\qquad \cdot \int_{\mathrm{B}(x_0;\rho) \times \{t\}} |\mathrm{grad}(\psi w)|^2\, dV,
\end{aligned}$$

and, using (54), we obtain

$$\begin{aligned}
&\iint_{\mathsf{R}_{\rho',\tau'}} w^{2\alpha}\,dV\,dt \\
&\qquad \le \iint_{\mathsf{R}_{\rho,\tau}} (\psi w)^{2\alpha}\,dV\,dt \\
&\qquad \le c(n)\mathfrak{J}^{-2/n}\{(\rho-\rho')^{-2}+(\tau-\tau')^{-1}\}^{2/n}\|w\|_{\mathsf{R}_{\rho,\tau}}^{4/n}\cdot\int_{\mathsf{R}_{\rho,\tau}} |\mathrm{grad}(\psi w)|^2\,dV\,dt \\
&\qquad \le c(n)\mathfrak{J}^{-2n}\{(\rho-\rho')^{-2}+(\tau-\tau')^{-1}\}^{1+2/n}\|w\|_{\mathsf{R}_{\rho,\tau}}^{2+4/n},
\end{aligned}$$

which is the claim.

The argument for $n = 2$ is similar.

CHAPTER IX

Topological Perturbations with Negligible Effect

In this chapter, we fix a Riemannian manifold M, with heat kernel $p(x, y, t)$. If M is compact, then p is the heat kernel constructed in Chapter VI; if M is a regular domain, then p is the Dirichlet heat kernel of M (constructed in Chapter VII); and if M is arbitrary noncompact, then p is the heat kernel constructed in Chapter VIII. (One also has results corresponding to those below, when M is a regular domain with Neumann heat kernel p.)

Our purpose in this chapter is to exhibit topological perturbations of M, supported on a compact subset of M, which have only negligible effect on the heat diffusion on, and the spectrum of, M. The precise meaning of "negligibility" is the content of the theorems stated below. The problems to be considered are as follows:

Fix a compact submanifold $\hat{M}$ of M; let B_ε be the tubular neighborhood of $\hat{M}$, in M, of radius $\varepsilon > 0$, that is,

$$\mathsf{B}_\varepsilon = \{x \in M : d(x, \hat{M}) < \varepsilon\};$$

and, finally, set

$$\Omega_\varepsilon = M \backslash \bar{\mathsf{B}}_\varepsilon, \qquad \Gamma_\varepsilon = \partial\Omega_\varepsilon = \partial\mathsf{B}_\varepsilon.$$

Assume that Ω_ε is connected and has minimal positive heat kernel q_ε. If M is compact, or a regular domain, then q_ε is the Dirichlet heat kernel of Ω_ε. We always think of q_ε as defined on all of M, vanishing whenever, at least, one of the space variables is in $\bar{\mathsf{B}}_\varepsilon$.

Our first interest will be in the validity of the limit

$$\lim_{\varepsilon \downarrow 0} q_\varepsilon = p. \tag{1}$$

If M is compact with eigenvalues

$$\{0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \cdots\},$$

and Ω_ε has Dirichlet eigenvalues

$$\{0 < \lambda_1(\varepsilon) < \lambda_2(\varepsilon) \le \lambda_3(\varepsilon) \le \cdots\},$$

then we are also interested in the validity of

$$\lim_{\varepsilon \downarrow 0} \lambda_j(\varepsilon) = \lambda_{j-1} \tag{2}$$

for all $j = 1, 2, \ldots$. (In both of the above lists of eigenvalues, the eigenvalues are listed according to their multiplicity.) Our basic result is that if

$$\mu =: \dim M - \dim \hat{M},$$

the codimension of $\hat{M}$ in M, is greater than or equal to 2, then the above limits (1), (2) are valid. Compare Theorems 1 and 2 below.

Our second interest is: assume $\hat{M}$ consists of 2 distinct points x_1, x_2. For sufficiently small $\varepsilon > 0$, we replace B_ε by a Riemannian manifold C_ε, which is diffeomorphic to $(-1, 1) \times \mathbb{S}^{n-1}$, where $n = \dim M$, in such a fashion that Ω_ε is isometrically imbedded in the new manifold

$$M_\varepsilon =: \Omega_\varepsilon \cup \Gamma_\varepsilon \cup \mathsf{C}_\varepsilon,$$

and that M_ε will be a C^∞ Riemannian manifold.

Let p_ε denote the heat kernel associated to the Laplace–Beltrami operator, acting on functions, on M_ε; and when M_ε is compact, let the eigenvalues of M_ε be given by

$$\{0 = \sigma_0(\varepsilon) < \sigma_1(\varepsilon) \le \sigma_2(\varepsilon) \le \cdots\},$$

with eigenvalues repeated according to their multiplicity. Then we seek conditions on the choice of Riemannian metrics on C_ε which will guarantee

$$\lim_{\varepsilon \downarrow 0} p_\varepsilon = p \tag{3}$$

on $M \backslash \hat{M}$, and, for compact M,

$$\lim_{\varepsilon \downarrow 0} \sigma_j(\varepsilon) = \lambda_j \tag{4}$$

for all $j = 0, 1, 2, \ldots$. Our basic result will be that a sufficient condition, in addition to the codimension of $\hat{M}$ in M being greater than or equal to 2, for the validity of (3) and (4), is that the isoperimetric constant of C_ε (cf. Definition IV.3) be bounded away from 0 as $\varepsilon \downarrow 0$. We then show how to construct examples satisfying the sufficient condition.

The results for (1) and (2) were first carried out, for domains in euclidean space, by Rauch–Taylor [2], J. Rauch [1], and then, for manifolds, by Chavel–Feldman [4]. The methods of this last paper are presented in Section 3. The "crushed-ice" arguments of Section 4 are from Rauch–Taylor

[2], J. Rauch [1], whose interest in the problem was inspired by Kac [1, 2] (compare, also, Papanicolaou–Varadhan [1]). Subsequently, (2) was treated in greater detail by Ozawa [1–3] (cf. further references in the bibliography of these papers), using variation of Green's functions as developed in Schiffer–Spencer [1]. This approach has been considered earlier in Swanson [1].

The results for (3) and (4) were first derived in Chavel–Feldman [6, 7], and extended and improved in Chavel–Feldman [8]. In Sections 1 and 2, we follow this last paper—however, we give a more elementary proof of Theorem 3 than the one given in Chavel–Feldman [8]. A protoversion of (4) was considered in Swanson [2] from the Schiffer–Spencer viewpoint.

1. STATEMENT AND DISCUSSION OF THE RESULTS

THEOREM 1. Let M, $\hat{M}$, p, q_ε, μ be as above. If

$$\mu \geq 2, \tag{5}$$

then (1) is valid uniformly for x, y bounded away from $\hat{M}$, and for t in bounded sets in $[0, \infty)$.

An immediate application of Theorem 1 is

THEOREM 2. If M is compact and μ satisfies (5), then (2) is valid for all $j = 1, 2, \ldots$.

We start our discussion by noting that if $\mu = 1$, then Theorem 2 is not necessarily true. For example, let M be $\mathbb{S}^1$, and $\hat{M}$ the zero-dimensional submanifold of M consisting of one point in M. For convenience, we think of $\mathbb{S}^1$ as $[-\pi/2, \pi/2]$, with $x = -\pi/2$ identified with $x = \pi/2$, and $\hat{M}$ the point corresponding to $x = -\pi/2$ and $x = \pi/2$. Then $\Omega_\varepsilon = (-\pi/2 + \varepsilon, \pi/2 - \varepsilon)$, and by (I.54),

$$\lambda_1(\varepsilon) = \{\pi/(\pi - 2\varepsilon)\}^2$$

which goes to 1 as $\varepsilon \downarrow 0$.

More generally, one might let $M = \mathbb{P}^n$, n-dimensional real projective space of constant sectional curvature 1, as described in Section II.4. Given $x_0 \in \mathbb{P}^n$, let $\hat{M}$ be the hyperplane at infinity dual to x_0. To realize the problem more concretely, visualize $\mathbb{P}^n$ as the hemisphere in $\mathbb{S}^n$ about x_0 (i.e., the geodesic disk about x_0 with radius $\pi/2$), with identification of antipodal points on the boundary, that is, the "equator" of x_0 is viewed as $\mathbb{P}^{n-1}$. We choose $\hat{M}$ to be this $\mathbb{P}^{n-1}$; so $\Omega_\varepsilon = \mathsf{B}(x_0; \pi/2 - \varepsilon)$, this geodesic disk considered as being in $\mathbb{S}^n$ as well as in $\mathbb{P}^n$. By (II.43),

$$\lambda_1(\varepsilon) \geq n$$

for all $\varepsilon > 0$. It is therefore impossible that $\lambda_1(\varepsilon) \to 0$ as $\varepsilon \downarrow 0$.

An intuitive way to view the matter is via heat diffusion, namely, Theorem 1. Recall, from the introduction to Chapter VI, that $p(x, y, t)$ is the temperature at x, at time t, resulting from the initial temperature distribution, at time $t = 0$, having total temperature 1, and completely concentrated at y. On the other hand, $q_\varepsilon(x, y, t)$ is the temperature at x, at time t, resulting from the same initial distribution as p, but subject to absolute refrigeration of B_ε for all time $t > 0$. If $\mu = 1$, then $\hat{M}$ locally separates M; so if x and y are close to each other but separated by $\hat{M}$, then $p(x, y, t)$ will reflect the fact that they are close to one another, but $q_\varepsilon(x, y, t)$ will view x and y as far from each other since q_ε is discounting the diffusion of any heat across $\hat{M}$.

If $\mu \geq 2$, then $\hat{M}$ does not even locally disconnect M, and $q_\varepsilon(x, y, t)$ represents, for x and y near $\hat{M}$, the diffusion *around* B_ε. A helpful analogy would be to replace the compact Riemannian manifold M by a regular domain, with the λ_j's now denoting Neumann eigenvalues (i.e., the regular domain is insulated at the boundary). On Ω_ε the boundary data are (i) vanishing Neumann data on ∂M, and (ii) vanishing Dirichlet data on Γ_ε. Then one might be considering the *fireman's pole problem* (Rauch–Taylor [2, p. 27]) (where $\dim M = 3$, and $\dim \hat{M} = 1$), in which case Theorem 1 is stating that heat diffusion in a firehouse is not seriously affected by the presence of the fireman's pole, even when it has an internal mechanism keeping it frozen. Or one might be considering a thermos with a finite number of spherical pieces of ice (this corresponds to $\dim \hat{M} = 0$) (Rauch–Taylor [2, p. 28]), cf. Section 4.

We now return to our Riemannian manifold M. If we think of M as filled with a monatonic gas, and the heat diffusion as effected via the excitation of an atom with a unit of thermal energy, with the energy propagated through M via the collision of thermally excited atoms with their neighbors, then (here we switch the roles of x and y, which is permissible by the symmetry of p in the space variables) $p(x, y, t)\, dV(y)$ may be interpreted as the proportion of continuous paths which emanate from x at time $t = 0$, that are near y at time t, relative to all continuous paths emanating from x at time $t = 0$.

A more precise formulation, of this proportion, is given by *Brownian motion on M*, described as follows: adjoin to M, via 1-point compactification, the point ∞ at infinity. (Thus, if M is compact, then ∞ is an isolated point of

$M \cup \{\infty\}$; and if M is a regular domain, then the complete boundary of M has been identified with ∞, and $M \cup \{\infty\}$ is connected.) For any continuous ω: $[0, +\infty) \to M \cup \{\infty\}$ with $\omega(0) \in M$, we let

$$\zeta(\omega) = \inf\{t > 0 : \omega(t) = \infty\}.$$

(If M is compact, then $\zeta \equiv +\infty$.) When $\zeta(\omega) < +\infty$, we redefine ω so that $\omega(t) = \infty$ for all $t \geq \zeta(\omega)$.

So the class of paths under consideration, $\mathscr{W}$, is the class of continuous mappings ω: $[0, +\infty) \to M \cup \{\infty\}$ such that $\omega(0) \in M$, and such that $\omega(t) = \infty$ for all $t \geq \zeta(\omega)$. For this space of paths, consider the collection of *cylinder subsets* of $\mathscr{W}$, namely, subsets $A \subseteq \mathscr{W}$ of the form

$$A = \{\omega \in \mathscr{W} : (\omega(t_1), \ldots, \omega(t_l)) \in B\},$$

where (i) l is any positive integer, (ii) B is any Borel set in M^l (the l-fold Cartesian product of M with itself), and (iii) the t_j's are any choice of times satisfying $0 \leq t_1 < \cdots < t_l$. On the σ-algebra generated by the cylinder sets we define, for each $x \in M$, a probability measure P_x such that (i) for

$$\mathscr{W}_x =: \{\omega \in \mathscr{W} : \omega(0) = x\}$$

we have

$$P_x(\mathscr{W}_x) = 1,$$

and (ii) for the cylinder set A, above, we have

$$P_x(A) = \int_B p(x, y_1, t_1) p(y_1, y_2, t_2 - t_1) \cdots \cdot p(y_{l-1}, y_l, t_l - t_{l-1})\, dV_l((y_1, \ldots, y_l)), \tag{6}$$

where dV_l is the Riemannian measure on M^l. It is known that the properties of the heat kernel (most notably, (VI.11) or (VII.29), (VII.48)) imply that P_x has a unique extension to a probability measure on the σ-algebra generated by the cylinder sets.

The easiest way to appreciate (6) is to consider the case $B = B_1 \times B_2 \subseteq M \times M$. The set A then consists of all paths which are in B_1 at time t_1, and in B_2 at the later time t_2. Formula (6) expresses the fact that once the path is in B_1 at time t_1, the probability that the path be in B_2 at time t_2 is independent of the history of the path prior to time t_1.

Note that the probability measures $\{P_x : x \in M\}$ possess the following consistency. Given $B_1, B_2 \subseteq M$, $t_1 < t_2$, as above, let t be any fixed number in (t_1, t_2). Then by (VI.11) and (6), we have

$$\begin{aligned} &P_x(\{\omega : (\omega(t_1), \omega(t), \omega(t_2)) \in B_1 \times M \times B_2\}) \\ &\quad = P_x(\{\omega : (\omega(t_1), \omega(t_2)) \in B_1 \times B_2\}), \end{aligned}$$

as it ought to be.

Note that

$$P_x(\{\omega : \zeta(\omega) > t\}) = \int_M p(x, y, t)\, dV(y). \tag{7}$$

For p the heat kernel of a compact Riemannian manifold, or the heat kernel of a complete Riemannian manifold with Ricci curvature bounded below, then the number given in (7) is equal to 1, by earlier theorems.

For $B \subseteq M$, one defines the *first hitting time of* B, T_B: $\mathscr{W} \to [0, \infty)$, by

$$T_B(\omega) = \inf\{t > 0 : \omega(t) \in B\}.$$

If Ω is a regular domain in M, with Dirichlet heat kernel, q_Ω, then $q_\Omega(x, y, t)$ is the probability density that a path starting at x at time $t = 0$, will be at y at time t, without ever having left Ω during the time interval $[0, t]$. More precisely, for any Borel set B in M,

$$P_x(\{\omega\colon \omega(t) \in B, t < \min(T_{M\setminus\Omega}(\omega), \zeta(\omega))\}) = \int_B q_\Omega(x, y, t)\, dV(y). \tag{8}$$

From these considerations one might expect that Theorem 1 is related to

THEOREM 3. If μ satisfies (5) then

$$\lim_{\varepsilon \downarrow 0} P_x(\{\omega : T_{\mathsf{B}_\varepsilon}(\omega) \le t\}) = 0. \tag{9}$$

uniformly for x bounded away from $\hat{M}$ and t in bounded subsets of $[0, +\infty)$; that is, the probability that a path, which starts at x at time $t = 0$, hits B_ε prior to time t, is negligible when ε is sufficiently small.

Remark 1: We note that if $\hat{M}$ consists of a single point $y \in M$, so that B_ε is now $\mathsf{B}(y; \varepsilon)$, we have the precise formula

$$\lim_{\varepsilon \downarrow 0} P_x(T_{\mathsf{B}(y;\varepsilon)} \le t)/\varepsilon^{n-2} = (n - 2)\mathbf{c}_{n-1} \int_0^t p(x, y, \tau)\, d\tau$$

when $n > 2$ (a corresponding formula, with $|\ln \varepsilon|^{-1}$ replacing ε^{n-2}, holds for $n = 2$). Compare Chavel–Feldman [9] for details and applications.

We shall show in Section 2 that Theorem 1, and all that follows, is a consequence of Theorem 3.

Note that, when $\mu = 1$, one can easily visualize why Theorem 3 would be false.

We assume $\hat{M}$ consists of 2 distinct points x_1, x_2 and consider the attachment of the cylinder C_ε across the geodesic spheres Γ_ε. Before stating the theorems we comment that the attachment of the handle is realized by a family of diffeomorphisms

$$\Psi_\varepsilon : M_\varepsilon \backslash \Omega_{2\varepsilon} \to [-2, 2] \times \mathbb{S}^{n-1}$$

for which

$$\mathbf{C}_\varepsilon = M_\varepsilon \backslash \bar{\Omega}_\varepsilon = \Psi_\varepsilon^{-1}[(-\tfrac{1}{2}, \tfrac{1}{2}) \times \mathbb{S}^{n-1}].$$

We also note that there can be no a priori expectation that the limit (4) is uniform in $j = 0, 1, 2, \ldots$. Indeed, when $n = 2$, we have by (IV.48)

$$\sum_{j=0}^{\infty} e^{-\lambda_j t} \sim \frac{V(M)}{4\pi t} + \frac{\chi(M)}{6},$$

$$\sum_{j=0}^{\infty} e^{-\sigma_j(\varepsilon) t} \sim \frac{V(M_\varepsilon)}{4\pi t} + \frac{\chi(M_\varepsilon)}{6},$$

as $t \downarrow 0$. Now

$$\begin{aligned} \left| \sum_{j=0}^{\infty} (e^{-\lambda_j t} - e^{-\sigma_j(\varepsilon) t}) \right| &= \left| \sum_{j=0}^{\infty} e^{-\lambda_j t} (1 - e^{-(\sigma_j(\varepsilon) - \lambda_j) t}) \right| \\ &\le \sum_{j=0}^{\infty} t e^{-\lambda_j t} |\sigma_j(\varepsilon) - \lambda_j| e^{|\sigma_j(\varepsilon) - \lambda_j| t} \end{aligned}$$

since $|e^x - 1| \le |x| e^{|x|}$ for all x. Assume that there exist $\rho, \varepsilon_0 > 0$ such that

$$(10) \qquad |\sigma_j(\varepsilon) - \lambda_j| < \rho$$

for all $j = 1, 2, \ldots$, and $0 < \varepsilon < \varepsilon_0$. Then

$$(11) \qquad \left| \sum_{j=0}^{\infty} (e^{-\lambda_j t} - e^{-\sigma_j(\varepsilon) t}) \right| \le \rho e^{\rho t} \sum_{j=0}^{\infty} t e^{-\lambda_j t}$$

is bounded for sufficiently small $t > 0$. Thus

$$4\pi t \left| \sum_{j=0}^{\infty} (e^{-\lambda_j t} - e^{-\sigma_j(\varepsilon) t}) \right| \le \text{const} \cdot t$$

as $t \downarrow 0$, from which we obtain

$$(12) \qquad V(M) = V(M_\varepsilon)$$

for all ε in $(0, \varepsilon_0)$.

Now make a stronger assumption than (10), namely, that (4) is valid uniformly in j. Fix t_0 so that

$$\sum_{j=0}^{\infty} t e^{-\lambda_j t} \le \text{const}$$

on $(0, t_0]$. Given $\rho > 0$, there exists $\varepsilon_0 > 0$ such that (10) is valid for all $j = 1, 2, \ldots$, and ε in $(0, \varepsilon_0)$. Then (11) and (12) imply that

$$|\chi(M) - \chi(M_\varepsilon) + o(t^0)| \leq \text{const} \cdot \rho e^{\rho t_0}$$

for all $t \in (0, t_0]$. But $\chi(M) - \chi(M_\varepsilon)$ is an integer. Therefore by picking ρ sufficiently small, we can guarantee the existence of $\varepsilon_0 > 0$ for which $\chi(M) = \chi(M_\varepsilon)$ for all ε in $(0, \varepsilon_0)$. But since M_ε differs from M by a handle, we have $\chi(M_\varepsilon) = \chi(M) - 2$, a contradiction. So in the 2-dimensional case, we are guaranteed that the convergence is not uniform.

On the other hand, we always have

PROPOSITION 1. For all $j = 1, 2, \ldots,$

$$\limsup_{\varepsilon \downarrow 0} \sigma_j(\varepsilon) \leq \lambda_j. \tag{13}$$

PROOF: For all $j = 1, 2, \ldots$ we have

$$\lambda_{j+1}(\varepsilon) > \lambda_j. \qquad \lambda_{j+1}(\varepsilon) > \sigma_j(\varepsilon)$$

by max–min methods (Section I.5), and

$$\lim_{\varepsilon \downarrow 0} \lambda_{j+1}(\varepsilon) = \lambda_j$$

by Theorem 2. Inequality (13) follows immediately.

THEOREM 4. Let $\nu(\varepsilon)$ be the lowest Dirichlet eigenvalue of $\mathbf{C}_\varepsilon$. Then a necessary condition for (4) to be valid for all $j = 0, 1, 2, \ldots$ is that

$$\liminf_{\varepsilon \downarrow 0} \nu(\varepsilon) = +\infty.$$

COROLLARY 1. If for some fixed $l > 0$, we have $(-l/2, l/2) \times \mathbb{S}^{n-1}(\varepsilon/4)$ (the "long-thin" handle) isometrically imbedded in $\mathbf{C}_\varepsilon$, for all ε, then

$$\nu(\varepsilon) \leq \pi^2/l^2$$

for all ε, and (4) cannot be satisfied for all j.

The sufficient condition for (4) will be given in terms of the isoperimetric constant (cf. Definition IV.3) of $\mathbf{C}_\varepsilon$, $\mathfrak{I}(\mathbf{C}_\varepsilon)$.

THEOREM 5. If there exists a positive constant $c > 0$ such that $\mathfrak{I}(\mathbf{C}_\varepsilon)$ satisfies

$$\mathfrak{I}(\mathbf{C}_\varepsilon) \geq c > 0 \tag{14}$$

for all sufficiently small $\varepsilon > 0$, then (4) is valid for all $j = 1, 2, \ldots$.

LEMMA 1. If the sufficient condition (14) is satisfied for all sufficiently small $\varepsilon > 0$, then there exist positive constants so that

$$V(\mathbf{C}_\varepsilon) \leq \text{const} \cdot \varepsilon^n, \tag{15}$$

$$\nu(\varepsilon) \geq \text{const}/\varepsilon^2 \tag{16}$$

for all sufficiently small $\varepsilon > 0$.

PROOF: From Definition IV.3 and (14) we have

$$c\{V(\mathbf{C}_\varepsilon)\}^{n-1} \leq \{A(\Gamma_\varepsilon)\}^n \leq \text{const } \varepsilon^{n(n-1)}$$

for sufficiently small $\varepsilon > 0$, which implies (15).

To prove (16), we note that, by (15), for any domain D, with compact closure in $\mathbf{C}_\varepsilon$, and with smooth boundary ∂D, we have

$$c \leq \{A(\partial D)/V(D)\}^n V(D) \leq \text{const} \cdot \varepsilon^n \{A(\partial D)/V(D)\}^n.$$

Therefore, the Cheeger constant of $\mathbf{C}_\varepsilon$, $\mathfrak{h}(\mathbf{C}_\varepsilon)$, given by Definition IV.1, satisfies

$$\mathfrak{h}(\mathbf{C}_\varepsilon) \geq \text{const} \cdot \varepsilon^{-1}$$

for all sufficiently small $\varepsilon > 0$, and (16) is then an immediate consequence of Cheeger's inequality (Theorem IV.3).

Note that in our example of the "long-thin" handle, we have, in contrast to (15),

$$V(\mathbf{C}_\varepsilon) \geq \text{const} \cdot \varepsilon^{n-1}$$

as $\varepsilon \downarrow 0$.

THEOREM 6. It is always possible to construct the family $\mathbf{C}_\varepsilon$ to satisfy (14) for all sufficiently small $\varepsilon > 0$.

2. PROOF OF THEOREMS 1–6

We start with Theorems 1, 2, 4, and 5, assuming that Theorem 3 has been proven, and then we go back to Theorem 3.

PROOF OF THEOREM 1: We assume that Theorem 3 has, in fact, been proved.

By Theorem VIII.1 we have

$$0 \le q_\varepsilon < p. \tag{17}$$

So for any bounded measurable function f on M, we have

$$\begin{aligned}&\left|\int_M \{p(x,y,t) - q_\varepsilon(x,y,t)\} f(y)\,dV(y)\right| \\ &\qquad \le \{\sup|f|\} \int_M \{p(x,y,t) - q_\varepsilon(x,y,t)\}\,dV(y).\end{aligned}$$

But from (7) we have

$$P_x(\{\omega : T_{\mathbf{B}_\varepsilon}(\omega) \le t\} \ge \int_M \{p(x,y,t) - q_\varepsilon(x,y,t)\}\,dV(y); \tag{18}$$

indeed,

$$\begin{aligned}P_x(\{\omega : T_{\mathbf{B}_\varepsilon}(\omega) \le t\}) &\ge P_x(\{\omega : T_{\mathbf{B}_\varepsilon}(\omega) \le t \le \zeta(\omega)\}) \\ &= \int_M \{p(x,y,t) - q_\varepsilon(x,y,t)\}\,dV(y)\end{aligned}$$

by (7), (8). Therefore

$$\begin{aligned}&\left|\int_M \{p(x,y,t) - q_\varepsilon(x,y,t)\} f(y)\,dV(y)\right. \\ &\qquad \le \{\sup|f|\} P_x(\{\omega : T_{\mathbf{B}_\varepsilon}(\omega) \le t\}).\end{aligned}$$

One can now easily obtain Theorem 1 from Theorem 3.

PROOF OF THEOREM 2: We use the fact that Theorem VIII.1 not only implies (17), but that it also implies that q_ε is strictly increasing with respect to ε. If (1) is valid on all of $(M\backslash\hat{M}) \times (M\backslash\hat{M}) \times (0, +\infty)$, then Lebesgue's monotone convergence theorem implies

$$\lim_{\varepsilon \downarrow 0} \int_M q_\varepsilon(x,x,t)\,dV(x) = \int_M p(x,x,t)\,dV(x) \tag{19}$$

for any $t > 0$. By the Sturm–Liouville expansions for the closed eigenvalue problem (Section VI.1) and the Dirichlet eigenvalue problem (Section VII.3), (19) can be rewritten as

$$\lim_{\varepsilon \downarrow 0} \sum_{j=1}^{\infty} e^{-\lambda_j(\varepsilon)t} = \sum_{j=1}^{\infty} e^{-\lambda_{j-1}t}. \tag{20}$$

Moreover, the standard max–min arguments (Section I.5) imply that $\lambda_j(\varepsilon)$ is an increasing function of ε, with

$$\lambda_j(\varepsilon) > \lambda_{j-1} \tag{21}$$

for all $\varepsilon > 0, j = 1, 2, \ldots$. One now easily deduces (2) from (20) and (21).

Proof of theorem 4: Let

$$\{0 < \mu_0(\varepsilon) \le \mu_1(\varepsilon) \le \cdots\}$$

be the collection of all Dirichlet eigenvalues of C_ε and Ω_ε, relisted in nondecreasing order, and repeated, as usual, according to multiplicity. Then the max–min arguments of Section I.5 imply that

$$\sigma_j(\varepsilon) \le \mu_j(\varepsilon)$$

for all $j = 1, 2, \ldots$, and $\varepsilon > 0$. Set

$$\alpha = \liminf_{\varepsilon \downarrow 0} \nu(\varepsilon),$$

assume that α is finite, and let λ_k be the first eigenvalue of M strictly greater than α. In particular,

$$\lambda_{k-1} \le \alpha < \lambda_k.$$

Now for any ε for which

$$\nu(\varepsilon) < \lambda_k,$$

we also have

$$\nu(\varepsilon) < \lambda_{k+1}(\varepsilon),$$

which implies

$$\nu(\varepsilon) \in \{\mu_0(\varepsilon), \ldots, \mu_k(\varepsilon)\},$$

which, in turn, implies

$$\sigma_k(\varepsilon) \le \mu_k(\varepsilon) \le \max\{\nu(\varepsilon), \lambda_k(\varepsilon)\}.$$

Therefore

$$\liminf_{\varepsilon \downarrow 0} \sigma_k(\varepsilon) \le \max\{\alpha, \lambda_{k-1}\} < \lambda_k,$$

which implies the claim.

Proof of Theorem 5: We first note that (3) is valid uniformly for x, y bounded away from $\hat{M}$, and t bounded in $[0, +\infty)$. Indeed, for $x \in M\backslash\hat{M}$, and sufficiently small $\varepsilon > 0$, we have

$$P_x(\{\omega : T_{\mathsf{B}_\varepsilon}(\omega) \leq t\}) = P_{\varepsilon,x}(\{\omega_\varepsilon : T_{\mathsf{C}_\varepsilon}(\omega_\varepsilon) \leq t\})$$
$$\geq \int_{M_\varepsilon} \{p_\varepsilon(x, y, t) - q_\varepsilon(x, y, t)\}\, dV(y),$$

where $P_{\varepsilon,x}$ is the probability measure on Brownian paths $\{\omega_\varepsilon\}$ in M_ε, starting at x. The point is that for $x \in \Omega_\varepsilon$ the Brownian particle starting from x is unaware of the change in the manifold until the particle hits the boundary of Ω_ε. One now shows, using the argument of Theorem 1, that (3) is valid uniformly for x, y bounded away from $\hat{M}$, and t bounded in $[0, +\infty)$.

LEMMA 2. To prove Theorem 5, it suffices to show the existence of a positive constant α, such that for

$$\mathsf{H}_{\varepsilon,\alpha} \equiv: M_\varepsilon\backslash\Omega_\alpha,$$

$0 < \varepsilon < \alpha$, and $t > 0$, we have $p_\varepsilon(\ ,\ , t)$ uniformly bounded on $\mathsf{H}_{\varepsilon,\alpha} \times \mathsf{H}_{\varepsilon,\alpha}$, independently of ε.

Proof: Assume the uniform boundedness of $p_\varepsilon(\ ,\ , t)$ as described in the lemma. We first claim that

$$\lim_{\varepsilon\downarrow 0} \int_{M_\varepsilon} p_\varepsilon(z, z, t)\, dV(z) = \int_M p(x, x, t)\, dV(x). \tag{22}$$

Indeed, we have

$$\int_{M_\varepsilon} p_\varepsilon(z, z, t)\, dV(z) - \int_M p(x, x, t)\, dV(x)$$
$$= \int_{\Omega_\varepsilon} \{p_\varepsilon(x, x, t) - p(x, x, t)\}\, dV(x)$$
$$+ \int_{\mathsf{C}_\varepsilon} p_\varepsilon(z, z, t)\, dV(z) - \int_{\mathsf{B}_\varepsilon} p(x, x, t)\, dV(x).$$

Now the last two integrals tend to 0 as $\varepsilon \downarrow 0$, since their integrands are bounded independently of ε, and their volumes tend to 0 as $\varepsilon \downarrow 0$. The uniform convergence of p_ε to p, for x, y bounded away from $\hat{M}$, and the uniform boundedness of p_ε on $\mathsf{H}_{\varepsilon,\alpha}$, combine to imply that $p_\varepsilon(z, z, t)$ is uniformly bounded in $z \in M_\varepsilon$, independent of ε. The first integral now converges to 0 as $\varepsilon \downarrow 0$, by Lebesgue's dominated convergence theorem.

We now note that (20), Fatou's lemma, and (13) imply

$$\begin{aligned}\sum_{j=0}^{\infty} e^{-\lambda_j t} &= \lim_{\varepsilon\downarrow 0} \sum_{j=0}^{\infty} e^{-\sigma_j(\varepsilon)t}\\ &\geq \sum_{j=0}^{\infty} \liminf e^{-\sigma_j(\varepsilon)t}\\ &= \sum_{j=0}^{\infty} e^{-\{\limsup \sigma_j(\varepsilon)\}t}\\ &\geq \sum_{j=0}^{\infty} e^{-\lambda_j t}.\end{aligned}$$

But if there exists $k \geq 1$ for which

$$\sigma_k =: \liminf_{\varepsilon\downarrow 0} \sigma_k(\varepsilon) < \lambda_k,$$

then there exists a sequence $\varepsilon_l \downarrow 0$, as $l \uparrow +\infty$, such that

$$\lim_{l\to\infty} \sigma_k(\varepsilon_l) = \sigma_k.$$

The above argument then yields

$$\begin{aligned}\sum_{j=0}^{\infty} e^{-\lambda_j t} &\geq \lim_{l\to\infty} \sum_{j=0}^{\infty} e^{-\sigma_j(\varepsilon_l)t}\\ &> e^{-\lambda_k t} + \sum_{j\neq k} e^{-\{\limsup \sigma_j(\varepsilon_l)\}t}\\ &\geq e^{-\lambda_k t} + \sum_{j\neq k} e^{-\lambda_j t}\\ &= \sum_{j=0}^{\infty} e^{-\lambda_j t}\end{aligned}$$

—a contradiction. Thus

$$\liminf_{\varepsilon\downarrow 0} \sigma_j(\varepsilon) \geq \lambda_j \tag{23}$$

for all $j = 0, 1, 2, \ldots$, and the lemma follows from (13) and (23).

Conclusion of the proof of Theorem 5: It therefore remains to exhibit an $\alpha > 0$ for which $p_\varepsilon(\ ,\ ,t)$ is uniformly bounded on $\mathsf{H}_{\varepsilon,\alpha} \times \mathsf{H}_{\varepsilon,\alpha}$, independently of ε.

First fix $\beta > 0$ such that

$$\beta < \inf_{x\in \mathsf{B}_\beta} \operatorname{inj}(x) \tag{24}$$

and set

$$\alpha = \tfrac{1}{3}\min\{\beta, d(x_1, x_2)/2\}. \tag{25}$$

To show that $p_\varepsilon(\ ,\ ,t)$ is bounded on $\mathsf{H}_{\varepsilon,\alpha} \times \mathsf{H}_{\varepsilon,\alpha}$ we note that by (VIII.38) we have for any $z, w \in M_\varepsilon$, $t > 0$,

$$p_\varepsilon(z, w, t) \leq \text{const}\cdot\left\{\frac{1}{t^{n/2}} + \frac{t}{\alpha^{n+2}}\right\}\{\mathfrak{I}(z;\alpha)\mathfrak{I}(w;\alpha)\}^{-1/2} \tag{26}$$

(where $\mathfrak{I}(z;\alpha)$ is the isoperimetric constant of $\mathfrak{B}(z;\alpha)$) with the constant depending only on $n = \dim M$. So we wish to bound $\mathfrak{I}(z;\alpha)$ from below, for $z \in \mathsf{H}_{\varepsilon,\alpha}$.

Let D be a regular domain in $\mathsf{B}(z;\alpha) \subseteq M_\varepsilon$, $z \in \mathsf{H}_{\varepsilon,\alpha}$ (therefore, $\mathsf{B}(z;\alpha) \subseteq \mathsf{H}_{\varepsilon,2\alpha}$), and assume that ∂D intersects Γ_ε finitely many times, all the intersections being transversal. Set

$$D_1 = D \cap \mathsf{C}_\varepsilon, \qquad D_2 = D \cap \Omega_\varepsilon$$

$$\gamma_1 = \partial D \cap \mathsf{C}_\varepsilon, \qquad \gamma_2 = \partial D \cap \Omega_\varepsilon, \qquad \gamma_\varepsilon = D \cap \Gamma_\varepsilon.$$

Then

$$\partial D_1 = \gamma_1 \cup \gamma_\varepsilon, \qquad \partial D_2 = \gamma_2 \cup \gamma_\varepsilon.$$

Now if

$$V(D_1) \geq V(D_2)$$

then standard calculations, based on geodesic spherical coordinates, imply the existence of positive constants, independent of ε for which

$$\begin{aligned} A(\partial D) &= A(\gamma_1) + A(\gamma_2) \\ &\geq A(\gamma_1) + \text{const}\cdot A(\gamma_\varepsilon) \\ &\geq \text{const}\cdot\{A(\gamma_1) + A(\gamma_\varepsilon)\} \\ &= \text{const}\cdot A(\partial D_1) \\ &\geq \text{const}\cdot\{I(\mathsf{C}_\varepsilon)\}^{1/n}\{V(D_1)\}^{(n-1)/n} \\ &\geq \text{const}\cdot\{I(\mathsf{C}_\varepsilon)/2^{n-1}\}^{1/n}\{V(D)\}^{(n-1)/n} \\ &\geq \text{const}\cdot\{V(D)\}^{(n-1)/n}. \end{aligned}$$

We now consider what happens when

$$V(D_1) \leq V(D_2).$$

In this case we estimate $\{A(\partial D)\}^n/\{V(D)\}^{(n-1)}$ using the lower bounds of C. B. Croke—Section V.3. Recall, from there, that for $x \in D$, the visibility angle of ∂D from x, ω_x, is the normalized $(n-1)$-measure of those unit vectors based at x whose geodesic intersects ∂D not later than the cut point of x along the geodesic. If one stops the chain of inequalities in the proof of Theorem IV.5, one line short of the last, then the argument provides the existence of positive constants, depending only on n, for which

$$\begin{aligned}
\frac{\{A(\partial D)\}^n}{\{V(D)\}^{n-1}} &\geq \text{const}\left\{\frac{1}{V(D)}\int_D \omega_x \, dV(x)\right\}^{n+1} \\
&\geq \text{const}\left\{\frac{1}{V(D)}\int_{D_2} \omega_x \, dV(x)\right\}^{n+1} \\
&\geq \text{const}\left\{\frac{1}{V(D_2)}\int_{D_2} \omega_x \, dV(x)\right\}^{n+1} \\
&\geq \text{const}\left\{\inf_{x \in D_2} \omega_x\right\}^{n+1}.
\end{aligned}$$

To estimate this last quantity from below, let ω_x' (resp., ω_x'', ω_x''') denote the visibility angle at $x \in D_2$ of ∂D_2 (resp., γ_2, Γ_ε). Now (24), (25) imply

$$\omega_x' = 1,$$

and that there exists a constant σ, independent of ε, such that

$$\tfrac{1}{2} \leq \sigma \leq 1,$$

and

$$\omega_x''' \leq \sigma$$

for all $x \in \mathsf{H}_{\varepsilon,2\alpha}$. We then have

$$\omega_x \geq \omega_x'' \geq \omega_x' - \omega_x''' \geq 1 - \sigma > 0$$

for all $x \in D_2$. This, then, gives a lower bound for $\{A(\partial D)\}^n/\{V(D)\}^{n-1}$, when $V(D_1) \leq V(D_2)$, independent of ε.

So $\mathfrak{I}(z;\alpha)$, $z \in \mathsf{H}_{\varepsilon,\alpha}$, is bounded from below independently of ε. From (26) we obtain a uniform upper bound of $p_\varepsilon(\; , \; , t)$ on $\mathsf{H}_{\varepsilon,\alpha} \times \mathsf{H}_{\varepsilon,\alpha}$ which is independent of ε. This concludes the proof of Theorem 5.

PROOF OF THEOREM 3: We first give a number of shorthand notations: given $x \in M$, the collection of Brownian paths $\mathscr{W}_x$, it is common to denote the arbitrary path in $\mathscr{W}_x$ by $X(t)$. If P_x is the probability measure on $\mathscr{W}_x$, it is common to write for a measurable function $f\colon \mathscr{W}_x \to \mathbb{R}$, and $A \subset \mathscr{W}_x$,

$$E_x(f; A) = \int_A f \, dP_x;$$

and for

$$A = \{\omega \in \mathscr{W}_x : T_{\mathsf{B}_\varepsilon}(\omega) \le t\}$$

it is common to simply write

$$A = (T_{\mathsf{B}_\varepsilon} \le t).$$

LEMMA 3. To prove Theorem 3, it suffices to show that there exists a $\delta_1 > 0$ such that for any δ in $(0, \delta_1)$, and $T > 0$, (9) is valid uniformly on $(\bar{\mathsf{B}}_{\delta_1} \setminus \mathsf{B}_\delta) \times [0, T]$.

The lemma is saying that one only has to consider (9) for x close to $\hat{M}$—indeed, a path starting far away from $\hat{M}$ cannot hit $\hat{M}$ without first "coming close" to $\hat{M}$. When the Brownian particle first arrives near $\hat{M}$ it has less time to find $\hat{M}$ than had it originally started close to M. And the history of the particle prior to its arrival near $\hat{M}$ has no effect on the ability of the particle to find $\hat{M}$ in the future. The formal argument is as follows:

PROOF OF LEMMA 3: Fix $\delta_1 > 0$. Given δ in $(0, \delta_1)$, let σ be the average of δ and δ_1. Then for any ε in $(0, \delta)$, and $x \in M \setminus \mathsf{B}_{\delta_1}$, we have, by the strong Markov property (Port–Stone [1, p. 12])

$$\begin{aligned} P_x(T_{\mathsf{B}_\varepsilon} \le t) &= E_x(P_{X(T_{\mathsf{B}_\sigma})}(T_{\mathsf{B}_\varepsilon} \le t - T_{\mathsf{B}_\sigma}); T_{\mathsf{B}_\sigma} < t) \\ &\le \sup_{y \in \mathsf{B}_{\delta_1} \setminus \mathsf{B}_\delta} P_y(T_{\mathsf{B}_\varepsilon} \le t), \end{aligned}$$

which implies the lemma.

The argument we now give, for Theorem 3, is based on that of Port–Stone [1, p. 21].

Fix $\delta, \delta_1 > 0$ satisfying

$$0 < \delta < \delta_1 < \alpha,$$

where α is given by (25), and also fix $T > 0$. Then there exists a positive constant, depending only on δ, δ_1, and T, such that for all $x \in \bar{\mathsf{B}}_{\delta_1} \setminus \mathsf{B}_\delta$, $z \in B_{\delta/2}$, and $t \in [0, T]$, we have

$$p(x, z, t) \le \text{const.} \tag{27}$$

LEMMA 4. For all $x \in \bar{\mathsf{B}}_{\delta_1} \setminus \mathsf{B}_\delta$, $t \in [0, T]$, and $0 < \varepsilon < r < \delta/2$, we have

$$E_x\left(\int_0^T d\tau \int_{\mathsf{B}_r} p(X(T_{\mathsf{B}_\varepsilon}), z, \tau)\, dV(z); T_{\mathsf{B}_\varepsilon} \le t\right) \le \text{const} \cdot r^\mu. \tag{28}$$

PROOF: We first note that for any Borel set $A \subseteq M$, $t > 0$, we have

$$\begin{aligned}\int_0^t d\tau \int_A p(x, y, \tau)\, dV(y) &= \int_0^t P_x(X(\tau) \in A)\, d\tau\\ &= \int_0^t E_x(\chi_A(X(\tau)))\, d\tau\\ &= E_x\left(\int_0^t \chi_A(X(\tau))\, d\tau\right);\end{aligned}$$

that is, $\int_0^t d\tau \int_A p(x, y, \tau)\, dV(y)$ is the average time the Brownian particle, starting from x at time 0, spends in A during the time interval $[0, t]$.

Then by the strong Markov property, and (27), we have

$$\begin{aligned}&E_x\left(\int_0^T d\tau \int_{\mathsf{B}_r} p(X(T_{\mathsf{B}_\varepsilon}), z, \tau)\, dV(z);\ T_{\mathsf{B}_\varepsilon} \le t\right)\\ &= E_x\left(\int_{T_{\mathsf{B}_\varepsilon}}^{T+T_{\mathsf{B}_\varepsilon}} \chi_{\mathsf{B}_r}(X(t))\, dt;\ T_{\mathsf{B}_\varepsilon} \le t\right)\\ &\le E_x\left(\int_0^{2T} \chi_{\mathsf{B}_r}(X(t))\, dt\right)\\ &= \int_0^{2T} d\tau \int_{\mathsf{B}_r} p(x, z, \tau)\, dV(z)\\ &\le \text{const} \cdot V(\mathsf{B}_r) \le \text{const} \cdot r^\mu,\end{aligned}$$

where χ_{B_r} is the characteristic function of B_r. The first four lines of the above chain of inequalities state that the average amount of time spent by the Brownian particle in B_r, during the first T time-units *after* hitting B_ε (if the Brownian particle does not hit B_ε prior to time t, then the contribution to the average is 0), is less than or equal to the average amount of time spent by the Brownian particle in B_r during the time interval $[0, 2T]$. But this is the claim of the lemma.

The next step is to estimate

$$\int_0^T d\tau \int_{\mathsf{B}_r} p(y, z, \tau)\, dV(z) \tag{29}$$

for $y \in \hat{M}$, from below, and then use the continuity of (29), with respect to $y \in M$, to estimate

$$\int_0^T d\tau \int_{\mathsf{B}_r} p(X(T_{\mathsf{B}_\varepsilon}), z, \tau)\, dV(z) \tag{30}$$

from below. By (28), an estimate of (30) from below will produce an upper bound on $P_x(T_{\mathsf{B}_\varepsilon} \leq t)$.

LEMMA 5. There exists $\delta_2 \in (0, \delta/2)$ such that

$$\int_0^T d\tau \int_{\mathsf{B}_r} p(y, z, \tau)\, dV(z) \geq \text{const} \int_0^T \tau^{-\mu/2}\, d\tau \int_0^r e^{-\text{const}\cdot\rho^2/\tau} \rho^{\mu-1}\, d\rho \tag{31}$$

for all $y \in \hat{M}$, $r \in (0, \delta_2)$—the positive constant depending only on δ_2, T.

PROOF: For $y \in \hat{M}$, $z \in \mathsf{B}(y; \delta/2)$, let $\bar{z}_y$ denote the lift of z to M_y, via $\exp_y|\mathsf{B}(y; \delta/2)$. Then it is possible to pick $\delta_2 > 0$ so that

$$\tfrac{1}{2} \leq d(z_1, z_2)/|(\bar{z}_1)_y - (\bar{z}_2)_y| \leq 2 \tag{32}$$

for all $y \in \hat{M}$, and $z_1, z_2 \in \mathsf{B}(y; \delta/2)$.

Next, we may pick δ_2 so that, in addition to the above, we have, for any $y \in \hat{M}$, $w \in \hat{M} \cap \mathsf{B}(y; \delta/2)$, and unit vector v in M_y, orthogonal to $\hat{M}_y$,

$$|\langle \bar{w}_y, v\rangle| \leq \text{const}\cdot|\bar{w}_y|^2. \tag{33}$$

We now think of $\hat{M}$ as a Riemannian manifold (should $\dim M > 0$). Let $\hat{d}(\ ,\ )$ denote the $\hat{M}$-distance function; $\hat{\mathsf{B}}(w; r)$ the $\hat{M}$-metric disk in $\hat{M}$, about $w \in \hat{M}$, of radius r; and $d\hat{V}$ the density of the Riemannian measure of $\hat{M}$.

Pick δ_2 sufficiently small so that, in addition to the above, δ_2 is less than $\frac{1}{2}$ the injectivity radius of $\hat{M}$, and that

$$\hat{\mathsf{B}}(w; r/2) \subseteq \hat{M} \cap \mathsf{B}(w; r) \tag{34}$$

for all $w \in \hat{M}$, $r \in (0, \delta_2)$.

Next, let $\mathfrak{N}\hat{M}$ denote the normal bundle of $\hat{M}$, in TM (i.e., $\mathfrak{N}M$ consists of all those vectors in TM, based at points of $\hat{M}$, which are orthogonal to $\hat{M}$), $\pi: \mathfrak{N}\hat{M} \to \hat{M}$ the natural projection, and Exp the exponential map of $\mathfrak{N}\hat{M}$ to M (i.e., $\text{Exp} = \exp|\mathfrak{N}\hat{M}$). Then δ_2 can, in addition to the above, be picked to be less than $\frac{1}{2}$ the injectivity radius of Exp.

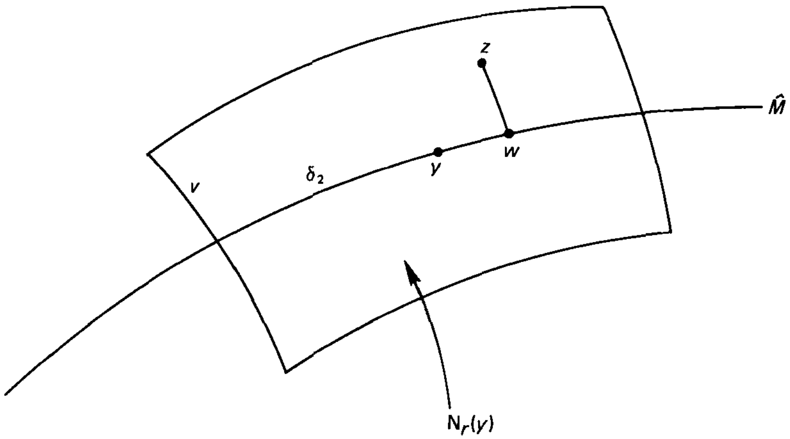


Fig. 3

Fix $y \in M$, and for all $r \in (0, \delta_2)$, set

$$\mathsf{N}_r(y) = \operatorname{Exp}(\{\xi \in \pi^{-1}[\hat{M} \cap \mathsf{B}(y; \delta_2)] : |\xi| < r\}.$$

For $z \in \mathsf{N}_r(y)$ set

$$\bar{z} = (\operatorname{Exp}|\mathsf{N}_{\delta_2}(y))^{-1}(z), \qquad w = \pi(\bar{z}).$$

Then we have

$$d^2(y, z) \leq \text{const} \cdot \hat{d}^2(y, w) + \text{const}|z|^2, \tag{35}$$

and

$$dV(z) \geq \text{const} \cdot d\bar{z}\, d\hat{V}(w), \tag{36}$$

where $d\bar{z}$ is the density of Lebesgue measure in the orthogonal complement of $\hat{M}_w$ in M_w.

Finally, δ_2 may be chosen so that, in addition to the above,

$$p(z_1, z_2, t) \geq \text{const} \cdot t^{-n/2} e^{-d^2(z_1, z_2)/4t} \tag{37}$$

for all $z_1, z_2 \in \mathsf{B}_{\delta_2}$, and $t \in (0, T]$.

From all the above we have, for any fixed $y \in \hat{M}$ and $r \in (0, \delta_2)$,

$$\begin{aligned}
&\int_0^T d\tau \int_{B_r} p(y, z, \tau)\, dV(z) \\
&\quad \geq \int_0^T d\tau \int_{N_r(y)} p(y, z, \tau)\, dV(z) \\
&\quad \geq \text{const} \int_0^T \tau^{-n/2}\, d\tau \int_{N_r(y)} e^{-d^2(y,z)/4\tau}\, dV(z) \\
&\quad \geq \text{const} \int_0^T \tau^{-n/2}\, d\tau \int_{\hat{M} \cap B(y;\delta_2/2)} e^{-\text{const}\cdot \hat{d}^2(y,w)/\tau}\, d\hat{V}(w) \cdot \int_0^r e^{-\text{const}\cdot \rho^2/\tau} \rho^{\mu-1}\, d\rho \\
&\quad \geq \text{const} \int_0^T \tau^{-n/2}\, d\tau \int_0^{\delta_2/2} e^{-\text{const}\cdot \eta^2/\tau} \eta^{n-\mu-1}\, d\eta \cdot \int_0^r e^{-\text{const}\cdot \rho^2/\tau} \rho^{\mu-1}\, d\rho \\
&\quad = \text{const} \int_0^T \tau^{-\mu/2}\, d\tau \int_0^{\text{const}/\sqrt{\tau}} e^{-s^2} s^{n-\mu-1}\, ds \cdot \int_0^r e^{-\text{const}\cdot \rho^2/\tau} \rho^{\mu-1}\, d\rho \\
&\quad \geq \text{const} \int_0^T \tau^{-\mu/2}\, d\tau \int_0^{\text{const}/\sqrt{T}} e^{-s^2} s^{n-\mu-1}\, ds \cdot \int_0^r e^{-\text{const}\cdot \rho^2/\tau} \rho^{\mu-1}\, d\rho \\
&\quad = \text{const} \int_0^T \tau^{-\mu/2}\, d\tau \int_0^r e^{-\text{const}\cdot \rho^2/\tau} \rho^{\mu-1}\, d\rho,
\end{aligned}$$

which implies the lemma.

To conclude the estimate of (30) from below, we note that, using the substitution

$$s = \text{const} \cdot \rho^2/\tau$$

(for fixed τ), followed by

$$\sigma = \text{const} \cdot r/\sqrt{\tau}$$

(for fixed r), one easily obtains

$$\begin{aligned}
\int_0^T \tau^{-\mu/2}\, d\tau \int_0^r e^{-\text{const}\cdot \rho^2/\tau} \rho^{\mu-1}\, d\rho &\geq \text{const} \cdot r^2 \int_{\text{const}\cdot r}^1 \sigma^{\mu-3}\, d\sigma \\
&\geq \text{const} \begin{cases} r^2, & \mu > 2, \\ r^2|\ln r|, & \mu = 2. \end{cases}
\end{aligned}$$

We therefore have positive constants so that for all $y \in \hat{M}$, $r \in (0, \delta_2)$,

$$\int_0^T d\tau \int_{B_r} p(y, z, \tau)\, dV(z) \geq \text{const} \begin{cases} r^2, & \mu > 2, \\ r^2|\ln r|, & \mu = 2. \end{cases} \tag{38}$$

The continuity of (30), with respect to y, implies that (38) is valid, with possibly lower constant, for all $y \in \mathsf{B}_{\varepsilon_0}$, for some $\varepsilon_0 > 0$.

From (28) and (38) we now have

$$P_x(T_{\mathsf{B}_\varepsilon} \leq t) \leq \text{const}\begin{cases} r^{\mu-2}, & \mu > 2, \\ |\ln r|^{-1}, & \mu = 2, \end{cases} \tag{39}$$

for all $x \in \bar{\mathsf{B}}_{\delta_1} - \mathsf{B}_{\delta/2}$, $t \in (0, T]$, $r \in (0, \delta_2)$, and ε in $(0, \varepsilon_0)$. But this then implies Theorem 3.

It, finally, remains to give a proof of Theorem 6. We first note that given any Riemannian manifold M, $x \in M$, and sufficiently small $\varepsilon > 0$, it is possible to change the Riemannian metric in $\mathsf{B}(x; \varepsilon)$ so that (i) the distance of any y, in $\mathsf{B}(x; \varepsilon)$, to x in the new metric is the same as in the old one, (ii) the new metric is euclidean on $\mathsf{B}(x; 3\varepsilon/5)$, and (iii) the new metric is the same as the old one on $\mathsf{B}(x; \varepsilon)\backslash\mathsf{B}(x; 4\varepsilon/5)$. Indeed, for each $\varepsilon > 0$, pick a C^∞ function $\psi_\varepsilon\colon [0, +\infty) \to [0, 1]$ such that $\psi_\varepsilon | [0, 3\varepsilon/5] = 0$ and $\psi_\varepsilon | [4\varepsilon/5, +\infty) = 1$. Let g denote the Riemannian metric on M; assume $\varepsilon < \text{inj}(x)$, and let G denote the pullback, via the local inverse of $\exp_x$, of the euclidean metric on $\mathfrak{B}(x; \varepsilon)$ to $\mathsf{B}(x; \varepsilon)$. One then defines for $y \in \mathsf{B}(x; \varepsilon)$,

$$g_\varepsilon(y) = \psi_\varepsilon(d(x, y))g(y) + \{1 - \psi_\varepsilon(d(x, y))\}G(y).$$

One easily sees that this will produce the desired result.

We also note that if ψ_ε is chosen so that $|\psi_\varepsilon'| \leq 10/\varepsilon$ on all of $[0, +\infty)$, and ε is sufficiently small, then a standard geodesic convexity argument proves that any unit-speed geodesic, with velocity vector at one point of $\mathsf{B}(x; \varepsilon)$ having positive projection onto the radial vector field determined by geodesic spherical coordinates based at x, must leave $\mathsf{B}(x; \varepsilon)$, in finite time, prior to that projection ever becoming 0.

To construct the example for Theorem 6, we first change the metric in B_ε by changing the metric about x_1 and x_2 as described above. We always use a function ψ_ε for which $|\psi_\varepsilon'| \leq 10/\varepsilon$ on all of $[0, +\infty)$. Next, let $\mathsf{C}_\varepsilon' = [-\varepsilon/2, \varepsilon/2] \times \mathbb{S}^{n-1}(\varepsilon/2)$, and attach C_ε' to $\mathsf{B}_\varepsilon\backslash\mathsf{B}_{\varepsilon/2}$ by identifying $\{-\varepsilon/2\} \times \mathbb{S}^{n-1}(\varepsilon/2)$ with the geodesic sphere about x_1 of radius $\varepsilon/2$, and identifying $\{\varepsilon/2\} \times \mathbb{S}^{n-1}(\varepsilon/2)$ with the geodesic sphere about x_2 of radius $\varepsilon/2$. We think of C_ε' as meeting two circular annuli at right angles in $\mathbb{R}^{n+1}$. The result of the identifications is to obtain the creased handle $\mathsf{C}_\varepsilon'' =: \mathsf{C}_\varepsilon' \cup \{\mathsf{B}_\varepsilon\backslash\mathsf{B}_{\varepsilon/2}\}$ smoothly attached, across Γ_ε, to $M\backslash\mathsf{B}_\varepsilon$.

Now in C_ε'', any geodesic hitting either of the creases transversally has a natural extension across the crease; so for any $x \in \mathsf{C}_\varepsilon''$ we may speak of the visibility angle ϑ_x of $\partial\mathsf{C}_\varepsilon'' = \Gamma_\varepsilon$ from x. Since the new B_ε has the geodesic convexity described above, we have for any $x \in \mathsf{B}_\varepsilon\backslash\mathsf{B}_{\varepsilon/2}$,

$$\vartheta_x \geq \tfrac{1}{2}.$$

For $x \in \mathsf{C}'_\varepsilon$, any geodesic emanating from x, which contributes to the visibility angle Θ_x of $\partial\mathsf{C}'_\varepsilon$ from x, must hit $\partial\mathsf{C}'_\varepsilon$ transversally, and, by geodesic convexity of the new B_ε, must contribute to ϑ_x. That is, for any $x \in \mathsf{C}'_\varepsilon$ we have

$$\vartheta_x = \Theta_x \geq \text{const}.$$

The constant depending only on n (since $[-\varepsilon/2, \varepsilon/2] \times \mathbb{S}^{n-1}(\varepsilon/2)$ is homothetic to $[-1, 1] \times \mathbb{S}^{n-1}$, for all ε, and homothetics leave visibility angles invariant). Thus the visibility angle of $\partial\mathsf{C}''_\varepsilon$ is bounded away from 0 as $\varepsilon \downarrow 0$.

Assume the above lower bound for the visibility angle of $\partial\mathsf{C}''_\varepsilon = \Gamma_\varepsilon$ is the constant $\delta > 0$. Then, for sufficiently small ε, one can smooth the creases of C''_ε in such a manner that the visibility angle of Γ_ε, associated to the uncreased handle C_ε, is bounded below by $\delta/2$ for all ε. Then Croke's original theorem (Theorem V.5) will imply that for the family of smooth handles C_ε we will have $\mathfrak{I}(\mathsf{C}_\varepsilon)$ bounded away from 0 independently of ε. This concludes the proof of Theorem 6.

3. USING RAYLEIGH'S CHARACTERIZATION OF EIGENVALUES

In this and the following section, we are mainly concerned with the first problem considered above, namely, the convergence of the eigenvalues of Ω_ε to those of M, as $\varepsilon \downarrow 0$. Here, we describe the arguments first given in Chavel–Feldman [4].

As in Section I.5, $\mathfrak{H}(M)$ is the Sobolev space of functions on M, which are in $L^2(M)$, possessing L^2-weak derivatives. $\mathfrak{H}(M)$, itself, is a Hilbert space relative to the inner product

$$(f, h)_1 = (f, h) + (\operatorname{Grad} f, \operatorname{Grad} h).$$

The induced norm is denoted by $\| \ \|_1$. (As mentioned in Section I.5, $C^\infty(M)$ is $\| \ \|_1$-dense in $\mathfrak{H}(M)$.) We let $\mathfrak{H}_\varepsilon(M)$ denote the subspace of $\mathfrak{H}(M)$ consisting of those functions in $\mathfrak{H}(M)$ vanishing identically on B_ε.

LEMMA 6. Given any $f \in C^\infty(M)$, we may associate to each sufficiently small $\varepsilon > 0$, a function $F_\varepsilon \in \mathfrak{H}_\varepsilon(M)$ such that

$$\lim_{\varepsilon \downarrow 0} \|F_\varepsilon - f\|_1 = 0.$$

PROOF OF THEOREM 2 FROM LEMMA 6: To each positive integer N, associate the hypothesis that there exists a sequence $\varepsilon_l \to 0$ as $l \to \infty$ such that

$$\lim_{l\to\infty} \lambda_j(\varepsilon_l) = \lambda_{j-1} \tag{40}$$

and

$$\lim_{l\to\infty} \|\varphi_j(\varepsilon_l) - \varphi_{j-1}\| = 0 \tag{41}$$

for all $j = 1, \ldots, N$, where $\{\varphi_1(\varepsilon_l), \ldots, \varphi_N(\varepsilon_l)\}$ is an orthonormal set in $L^2(\Omega_{\varepsilon_l})$, each $\varphi_j(\varepsilon_l)$ an eigenfunction of $\lambda_j(\varepsilon_l)$, and, similarly, $\{\varphi_0, \ldots, \varphi_{N-1}\}$ is an orthonormal set in $L^2(M)$, each φ_{j-1} an eigenfunction of λ_{j-1}.

Suppose we are given an N (for which this hypothesis is true). (For $N = 1$, (40) is valid by the lemma to $f = 1/\sqrt{V(M)}$. The validity of (41) will emerge from the ensuing argument.) We wish to show the hypothesis is, in fact, valid for $N + 1$.

To this end, let Φ be an eigenfunction of λ_N, with $\|\Phi\| = 1$, and Φ L^2-orthogonal to $\{\varphi_0, \ldots, \varphi_{N-1}\}$. To each ε_l in the hypothesis for N, associate

$$\mathscr{F}_l \equiv: F_{\varepsilon_l}$$

obtained by applying Lemma 6 to Φ and ε_l, and consider the Sturm–Liouville expansion

$$\mathscr{F}_l \sim \sum_{r=1}^{\infty} \alpha_r(\varepsilon_l)\varphi_r(\varepsilon_l), \tag{42}$$

where $\{\varphi_r(\varepsilon_l) : r = 1, 2, \ldots\}$ is a complete orthonormal basis of $L^2(\Omega_{\varepsilon_l})$, extending the orthonormal set $\{\varphi_1(\varepsilon_l), \ldots, \varphi_N(\varepsilon_l)\}$. Then

$$\lim_{l\to\infty} \alpha_r(\varepsilon_l) = \lim_{l\to\infty}(\mathscr{F}_l, \varphi_r(\varepsilon_l)) = (\Phi, \varphi_{r-1}) = 0$$

for $r = 1, \ldots, N$. From Rayleigh's principle (Section I.5) we have

$$\begin{aligned}\lambda_{N+1}(\varepsilon_l) &\leq \frac{\|\operatorname{grad}(\mathscr{F}_l - \sum_{r=1}^{N} \alpha_r(\varepsilon_l)\varphi_r(\varepsilon_l))\|^2}{\|\mathscr{F}_l - \sum_{r=1}^{N} \alpha_r(\varepsilon_l)\varphi_r(\varepsilon_l)\|^2}\\ &= \frac{\|\operatorname{grad} \mathscr{F}_l\|^2 - \sum_{r=1}^{N} \lambda_r(\varepsilon_l)\alpha_r^2(\varepsilon_l)}{\|\mathscr{F}_l\|^2 - \sum_{r=1}^{N} \alpha_r^2(\varepsilon_l)},\end{aligned}$$

which implies

$$\limsup_{l\to\infty} \lambda_{N+1}(\varepsilon_l) \leq \lambda_N \leq \liminf_{l\to\infty} \lambda_{N+1}(\varepsilon_l),$$

the last inequality resulting from the fact that $\lambda_{N+1}(\varepsilon_l)$ is always greater than λ_N. We therefore have

$$\lim_{l\to\infty} \lambda_{N+1}(\varepsilon_l) = \lambda_N \tag{43}$$

(this argument for $N = 0$ is even easier).

Now assume $\varphi_{N+1}(\varepsilon_l)$ in (42) is always an eigenfunction of $\lambda_{N+1}(\varepsilon_l)$. Then for $r = 1, \ldots, N$ we have

$$|(\varphi_{N+1}(\varepsilon_l), \varphi_{r-1})| = |(\varphi_{N+1}(\varepsilon_l), \varphi_{r-1} - \varphi_r(\varepsilon_l))| \leq \|\varphi_{r-1} - \varphi_r(\varepsilon_l)\|;$$

so if we expand $\varphi_{N+1}(\varepsilon_l)$ in $L^2(M)$ by

$$\varphi_{N+1}(\varepsilon_l) \sim \sum_{r=0}^{\infty} \beta_r(\varepsilon_l)\psi_r$$

where $\{\psi_r : r = 0, 1, 2, \ldots\}$ is a complete orthonormal basis of $L^2(M)$, extending the orthonormal set $\{\varphi_0, \ldots, \varphi_{N-1}\}$ (i.e., $\varphi_r = \psi_r$ for $r = 0, \ldots, N-1$), with each ψ_r an eigenfunction of λ_r, $r = 0, 1, 2, \ldots$, then

$$\lim_{l\to\infty} \beta_r(\varepsilon_l) = 0 \tag{44}$$

for all $r = 0, 1, \ldots, N-1$.

Assume

$$\lambda_N = \cdots = \lambda_{N+E-1}, \qquad \lambda_N < \lambda_{N+E}.$$

Then, by the argument of Rayleigh's principle, we have

$$\begin{aligned}\lambda_{N+1}(\varepsilon_l) = \|\text{grad}\, \varphi_{N+1}(\varepsilon_l)\|^2 &\geq \sum_{r=0}^{\infty} \lambda_r \beta_r^2(\varepsilon_l) \\ &\geq \sum_{r=0}^{N+E-1} \lambda_r \beta_r^2(\varepsilon_l) + \lambda_{N+E} \sum_{r=N+E}^{\infty} \beta_r^2(\varepsilon_l).\end{aligned}$$

From

$$\sum_{r=0}^{\infty} \beta_r^2(\varepsilon_l) = 1, \tag{45}$$

one easily has, now,

$$(\lambda_{N+E} - \lambda_N) \sum_{r=N}^{N+E-1} \beta_r^2(\varepsilon_l) \geq \lambda_{N+E} - \lambda_{N+1}(\varepsilon_l) + \sum_{r=0}^{N-1} (\lambda_r - \lambda_{N+E})\beta_r^2(\varepsilon_l),$$

which implies, via (43) and (44)

$$\liminf_{l\to\infty} \sum_{r=N}^{N+E-1} \beta_r^2(\varepsilon_l) \geq 1.$$

From (45) we obtain

$$\lim_{l\to\infty} \sum_{r=N}^{N+E-1} \beta_r^2(\varepsilon_l) = 1. \tag{46}$$

One therefore has a subsequence (ε_l') of (ε_l) for which the sequences $(\beta_r(\varepsilon_l'))$ converge, for each $r = N, \ldots, N + E - 1$. The desired φ_N is then given by

$$\varphi_N = \sum_{r=N}^{N+E-1} \left(\lim_{l\to\infty} \beta_r(\varepsilon_l')\right)\psi_r; \tag{47}$$

for from (44)–(46) we now have (41) for $j = N + 1$. (The argument can be written explicitly to verify (41) for $j = 1$.)

Note that if one assumes that all the eigenvalues of M have multiplicity 1 (and this is generically the case—cf. Uhlenbeck [1]) then this last argument shows that, up to possibly multiplying $\varphi_j(\varepsilon)$ by -1, we may also conclude in Theorem 2

$$\lim_{\varepsilon\downarrow 0} \varphi_j(\varepsilon) = \varphi_{j-1}$$

in $L^2(M)$ (it is *always* true for $j = 1$). Without the hypothesis of all eigenvalues having multiplicity 1, one has to even state the problem of convergence of eigenfunctions with care. Compare Chavel–Feldman [4], where this matter (also cf. Rauch–Taylor [2]), the convergence of eigenfunctions in the C^r-topology, $r = 1, 2, \ldots$, and the derivation of Theorem 1 from Theorem 2 are considered.

PROOF OF LEMMA 6: We shall assume here that

$$\mu > 2, \tag{5'}$$

instead of (5), and prove that F_ε can be chosen so that

$$\|F_\varepsilon - f\|_1 \le \text{const} \cdot \varepsilon^{\mu-2}, \tag{48}$$

where the constant depends only on M, $\hat{M}$, $\max|f|$, and $\max|\operatorname{grad} f|$. The case $\mu = 2$ has a similar argument, but is messier—one can find the argument in Chavel–Feldman [4].

Let $\mathfrak{N}\hat{M}$ denote the normal bundle of $\hat{M}$ in M, and Exp its exponential map, that is,

$$\text{Exp} = \exp|\mathfrak{N}\hat{M}.$$

Let $\delta > 0$ have the property that for each ε in $(0, \delta)$, Exp maps the collection of tangent vectors in $\mathfrak{N}\hat{M}$, having length less than ε, diffeomorphically onto B_ε. We only consider ε in $(0, \delta/2)$.

Given f, and ε in $(0, \delta/2)$, we associate F_ε to f and ε as follows: for any $x \in \Omega_{2\varepsilon}$ set

$$F_\varepsilon(x) = f(x).$$

To define F_ε on $\mathsf{B}_{2\varepsilon}$, define, for any unit tangent vector ξ in $\mathfrak{N}\hat{M}$,

$$F_\varepsilon(\operatorname{Exp} r\xi) = \begin{cases} f(\operatorname{Exp} 2\varepsilon\xi)(r - \varepsilon)/\varepsilon, & \varepsilon \le r \le 2\varepsilon, \\ 0, & 0 \le r \le \varepsilon. \end{cases}$$

Then F_ε is well defined, continuous on M, C^∞ on $M\backslash(\Gamma_{2\varepsilon} \cup \Gamma_\varepsilon)$, with grad F_ε having a jump discontinuity across Γ_ε, and possibly across $\Gamma_{2\varepsilon}$. So $F_\varepsilon \in \mathfrak{H}_c(M)$.

To verify (48), set

$$\alpha = \max_M |f|, \qquad \beta = \max_M |\operatorname{grad} f|.$$

Note that

$$\operatorname{supp}(f - F_\varepsilon) = \mathsf{B}_{2\varepsilon}.$$

We immediately have

$$\int_M |f - F_\varepsilon|^2 \, dV \le 4\alpha^2 V(\mathsf{B}_{2\varepsilon}) = O(\varepsilon^\mu).$$

To estimate $\|\operatorname{grad}(f - F_\varepsilon)\|^2$ from above we first derive the estimate

$$|\operatorname{grad}(f - F_\varepsilon)| \le \text{const}/\varepsilon^2 \tag{49}$$

on $\mathsf{B}_{2\varepsilon}$, where the constant depends only on M, $\hat{M}$, α, β. The estimate (49) will then imply

$$\|\operatorname{grad}(f - F_\varepsilon)\|^2 = \int_{B_{2\varepsilon}} |\operatorname{grad}(f - F_\varepsilon)|^2 \, dV = O(\varepsilon^{\mu - 2}),$$

which will complete the proof of the lemma.

To derive (49), we first write grad F_ε with more detail, namely, fix ε; and for any r in $(0, 2\varepsilon)$ map $h_r: \Gamma_r \to \Gamma_{2\varepsilon}$ by

$$h_r(\operatorname{Exp} r\xi) = \operatorname{Exp} 2\varepsilon\xi.$$

Then

$$F_\varepsilon \,|\, \Gamma_r = h_r^*(f \,|\, \Gamma_{2\varepsilon}) k_\varepsilon(r),$$

where

$$k_\varepsilon(r) = \begin{cases} (r - \varepsilon)/\varepsilon, & \varepsilon \le r \le 2\varepsilon, \\ 0, & 0 \le r \le \varepsilon, \end{cases}$$

which implies

$$\begin{aligned}(\operatorname{grad} F_\varepsilon) \mid \Gamma_r &= (\operatorname{grad}_{\Gamma_r}(h_r^*(f \mid \Gamma_{2\varepsilon})))k_\varepsilon(r) + (h_r^*(f \mid \Gamma_{2\varepsilon}))k_\varepsilon'(r)(\partial_r \mid \Gamma_r) \\ &= (h_r^{-1})_*(\operatorname{grad}_{\Gamma_{2\varepsilon}}(f \mid \Gamma_{2\varepsilon}))k_\varepsilon(r) + (h_r^*(f \mid \Gamma_{2\varepsilon}))k_\varepsilon'(r)(\partial_r \mid \Gamma_r),\end{aligned}$$

where ∂_r is the unit vector field on B_δ tangent to unit speed geodesics emanating orthogonally from $\hat{M}$.

Now a calculation in geodesic coordinates determined by Exp shows that $\mu \geq 2$ implies that there exists a positive constant, depending only on M, $\hat{M}$ such that

$$\begin{aligned}|(h_r^{-1})_*(\operatorname{grad}_{\Gamma_{2\varepsilon}}(f \mid \Gamma_{2\varepsilon}))|^2 &\leq \text{const} \cdot (\varepsilon^2/r^2)|\operatorname{grad}_{\Gamma_{2\varepsilon}}(f \mid \Gamma_{2\varepsilon})|^2 \\ &\leq \text{const} \cdot \beta^2\varepsilon^2/r^2\end{aligned}$$

One now easily obtains (49), using the fact that $|\partial_r| = 1$, and $\partial_r \mid \Gamma_r$ is orthogonal to Γ_r.

COROLLARY 2. If $\hat{M}$ consists of N distinct points, then

$$\lambda_1(\varepsilon) \leq \text{const} \cdot N\varepsilon^{n-2} \tag{50}$$

as $\varepsilon \downarrow 0$, where the constant depends only on M.

4. THE MATHEMATICS OF CRUSHED ICE

We now consider a fixed compact Riemannian manifold M of dimension $n \geq 2$. To each positive integer N we associate a positive number r_N so that $r_N \to 0$ as $N \to \infty$. Also, to each N we associate a collection of points $\{x_{1,N}, \ldots, x_{N,N}\}$ in M, set

$$\begin{aligned}\mathsf{B}_N &= \bigcup_{i=1}^{N} \mathsf{B}(x_{i,N}; r_N), \\ \Omega_N &= M \backslash \overline{\mathsf{B}_N}, \\ q_N &= \text{Dirichlet heat kernel of } \Omega_N.\end{aligned}$$

We seek results on the behavior of q_N as $N \to \infty$. The reader can find fascinating background, and similar phenomena in Kac [2] and J. Rauch [1]. Detailed results are in Rauch–Taylor [2].

THEOREM 7. Let D be a noncompact regular domain in M with relatively compact $D_1 \subseteq D$, such that $\mathsf{B}_N \subseteq D_1$ for all sufficiently large N. Furthermore, assume, $n > 2$, and

$$\lim_{N \to \infty} N r_N^{n-2} = 0. \tag{51}$$

Then for every $T > 0$, we have

$$\lim_{N \to \infty} q_N = p \tag{52}$$

uniformly on $(M \backslash D) \times (M \backslash D) \times [0, T]$. If $n = 2$, then the condition (51) is to be replaced by

$$\lim_{N \to \infty} N/|\ln r_N| = 0. \tag{51'}$$

PROOF: We first note that the proof of Theorem 1 shows that it suffices to prove

$$\lim_{N \to \infty} P_x(T_{\mathsf{B}_N} \leq t) = 0 \tag{53}$$

uniformly for $(x, t) \in (M \backslash D) \times [0, T]$, for each $T > 0$.

Next we note that

$$P_x(T_{\mathsf{B}_N} \leq t) \leq \sum_{i=1}^{N} P_x(T_{\mathsf{B}(x_{i,N}; r_N)} \leq t).$$

It therefore suffices to show that there exists a positive constant, depending only on M, D, D_1, and $T > 0$, for which (when $n > 2$)

$$P_x(T_{\mathsf{B}(w; r_N)} \leq t) \leq \text{const} \cdot r_N^{n-2} \tag{54}$$

for all sufficiently large N, for all $x \in M \backslash D$, $w \in D_1$, and $t \in [0, T]$. (When $n = 2$, the r_N^{n-2} in (54) should be replaced by $|\ln r_N|^{-1}$.) The proof is really contained in the argument of Theorem 3 so we will look at it closely.

Since M is compact, we may simply pick β in (24) to be less than $\frac{1}{3}$ the injectivity radius of M. Next, fix $\delta, \delta_1 > 0$ satisfying (26), and fix $T > 0$. Then the argument of Lemma 3 shows that it suffices to prove (54) uniformly for $x \in \overline{\mathsf{B}(w; \delta_1)} \backslash \mathsf{B}(w; \delta)$, $w \in D_1$, and $t \in [0, T]$.

Next, note that (27) is valid uniformly for all $x \in \overline{\mathsf{B}(w; \delta_1)} \backslash \mathsf{B}(w; \delta)$, $z \in \mathsf{B}(w; \delta/2)$, $w \in D_1$, and $t \in (0, T]$. If we pick N sufficiently large then Lemma 4 may be recast to state that: there exists a positive constant such that for all $x \in \overline{\mathsf{B}(w; \delta_1)} \backslash \mathsf{B}(w; \delta)$, $w \in D_1$, $t \in (0, T]$, and $r_N < \delta/6$, we have

$$E_x\left(\int_0^T d\tau \int_{\mathsf{B}(w; 3r_N)} p(X(T_{\mathsf{B}(w; r_N)}), z, \tau)\, dV(z);\ T_{\mathsf{B}(w; r_N)} \leq t\right) \leq \text{const} \cdot r_N^n.$$

We therefore wish to estimate

$$\int_0^T d\tau \int_{B(w;\,3r_N)} p(y, z, \tau)\, dV(z)$$

for $y \in B(w; r_N)$, from below, and thereby obtain the desired estimate for $P_x(T_{B(w;\,r_N)} \leq t)$.

Well, we have the existence of a positive constant such that for all $z_1, z_2 \in B(w; \delta_1)$, $w \in D_1$, $t \in [0, T]$, we have

$$p(z_1, z_2, t) \geq \text{const} \cdot t^{-n/2} e^{-d^2(z_1, z_2)/4t},$$

from which we have positive constants for which

$$\begin{aligned}\int_0^T d\tau \int_{B(w;\,3r_N)} p(y, z, \tau)\, dV(z) &\geq \int_0^T d\tau \int_{B(y;\,r_N)} p(y, z, \tau)\, dV(z)\\ &\geq \text{const} \int_0^T \tau^{-n/2}\, d\tau \int_0^{r_N} e^{-\rho^2/4\tau} \rho^{n-1}\, d\rho\\ &\geq \text{const} \begin{cases} r_N^2, & n > 2,\\ r_N^2 |\ln r_N|, & n = 2,\end{cases}\end{aligned}$$

for all $y \in B(w; r_N)$, $w \in D_1$, sufficiently large N. We now have (54) uniformly for $x \in \overline{B(w; \delta_1)} \backslash B(w; \delta)$, $w \in D_1$, $t \in [0, T]$, and sufficiently large N; and with it, the proof of the theorem.

We now consider what happens when $Nr_N^{n-2} \to +\infty$ as $N \to +\infty$ (for $n > 2$).

THEOREM 8. Suppose the sequences of points $\{x_{i,N} : i = 1, \ldots, N\}$, $N = 1, 2, \ldots$ are "evenly spaced," that is, suppose the points are chosen so that there exists a positive integer H, independent of N, and a sequence of numbers in $(0, \frac{1}{2}\,\text{inj}(M))$, δ_N, for which (i)

$$M = \bigcup_{i=1}^{N} B(x_{1,N}; \delta_N)$$

and (ii) any $y \in M$ belongs to at most H of the disks $\{B(x_{i,N}; \delta_N) : i = 1, \ldots, N\}$. Then there exists a positive constant, depending only on M, such that for $r_N \in (0, \delta_N)$ the lowest Dirichlet eigenvalue of Ω_N, $\lambda(\Omega_N)$ satisfies

$$\lambda(\Omega_N) \geq \text{const} \cdot Nr_N^{n-2}/H^2. \tag{55}$$

In particular, $\lambda(\Omega_N) \to +\infty$ if and only if $Nr_N^{n-2} \to \infty$ as $N \to +\infty$.

PROOF: First note that there exists positive constants such that, for all $r \in (0, \frac{1}{2}\operatorname{inj}(M))$ and $\xi \in \mathscr{S}M$, we have

$$\text{const} \leq \sqrt{\mathfrak{g}}(r; \xi)/r^{n-1} \leq \text{const} \tag{56}$$

(where $\sqrt{\mathfrak{g}}(r; \xi)$ is given in Definition III.6).

The first consequence of (56) is the existence of a positive constant such that, for all r in $(0, \frac{1}{2}\operatorname{inj}(M))$, and $y \in M$, we have

$$V(\mathsf{B}(y; r)) \geq \text{const} \cdot r^n,$$

from which we have

$$HV(M) \geq \sum_{i=1}^{N} V(\mathsf{B}(x_{1,N}; \delta_N)) \geq \text{const} \cdot N\delta_N^n,$$

that is,

$$HV(M) \geq \text{const} \cdot N\delta_N^n. \tag{57}$$

The next consequence of (56) is as follows: given any function $h\colon [r_N, \delta_N] \to \mathbb{R} \in C^1$, with $h(r_N) = 0$, we have, for all $\xi \in \mathfrak{S}M$,

$$h(\tau) = \int_{r_N}^{r} h' = \int_{r_N}^{r} h'\{\sqrt{\mathfrak{g}}(\ ; \xi)\}^{1/2}\{\sqrt{\mathfrak{g}}(\ ; \xi)\}^{-1/2},$$

which implies

$$h^2(\tau) \leq \left\{\int_{r_N}^{r} h'^2\sqrt{\mathfrak{g}}(\ ; \xi)\right\}\left\{\int_{r_N}^{r} [\sqrt{\mathfrak{g}}(\ ; \xi)]^{-1}\right\},$$

which, in turn, implies

$$\begin{aligned} \int_{\tau}^{\delta_N} h^2\sqrt{\mathfrak{g}}(\ ; \xi) &\leq \left\{\int_{\tau}^{\delta_N} h'^2\sqrt{\mathfrak{g}}\right\}\left\{\int_{\tau}^{\delta_N} [\sqrt{\mathfrak{g}}(\ ; \xi)]^{-1}\right\}\left\{\int_{\tau}^{\delta_N} \sqrt{\mathfrak{g}}(\ ; \xi)\right\} \\ &\leq \text{const} \cdot \tau^{2-n}\delta_N^n \int_{r_N}^{\delta} h'^2\sqrt{\mathfrak{g}}(\ ; \xi), \end{aligned}$$

that is,

$$\int_{r_N}^{\delta_N} h^2\sqrt{\mathfrak{g}}(\ ; \xi) \leq \text{const} \cdot r_N^{2-n}\delta_N^n \int_{r_N}^{\delta_N} h'^2\sqrt{\mathfrak{g}}(\ ; \xi). \tag{58}$$

We therefore have, for any $f \in C_c^\infty(\Omega_N)$,

$$\begin{aligned}
&\int_{\Omega_N} |\mathrm{grad}\, f|^2\, dV \\
&\qquad \geq H^{-1} \sum_{i=1}^{N} \int_{\mathsf{B}(x_{i,N};\delta_N)} |\mathrm{grad}\, f|^2\, dV \\
&\qquad \geq H^{-1} \sum_{i=1}^{N} \int_{\mathfrak{S}_{x_{i,N}}} d\mu_{x_{i,N}}(\xi) \int_{r_N}^{\delta_N} ((\partial_r f)^2 \circ \gamma_\xi)\sqrt{\mathbf{g}}(\ ;\xi) \\
&\qquad \geq \mathrm{const}\, \frac{r_N^{n-2}}{H\delta_N^n} \sum_{i=1}^{N} \int_{\mathfrak{S}_{x_{i,N}}} d\mu_{x_{i,N}}(\xi) \int_{r_N}^{\delta_N} (f^2 \circ \gamma_\xi)\sqrt{\mathbf{g}}(\ ;\xi) \\
&\qquad = \mathrm{const}\, \frac{r_N^{n-2}}{H\delta_N^n} \sum_{i=1}^{N} \int_{\mathsf{B}(x_{i,N};\delta_N)} f^2\, dV \\
&\qquad \geq \mathrm{const}\, \frac{N r_N^{n-2}}{H^2} \int_{\Omega_N} f^2\, dV,
\end{aligned}$$

which implies the claim, by Rayleigh's characterization of eigenvalues. (One goes from the first line to the second, using the hypotheses of the theorem; γ_ξ, in the third and fourth lines, denotes the geodesic determined by ξ; one goes from the third to the fourth line by (58); and one goes from the fifth to the sixth line by (57).)

For $n = 2$, one replaces $N r_N^{n-2}$ by $N/|\ln r_N|$ in (55).

The existence of "even spacing" is discussed (for a regular domain) in Baider–Feldman [1]. The paper also features an ingenious application, of the argument of Theorem 8, to the study of spectra of noncompact manifolds.

The statement $\lambda(\Omega_N) \to +\infty$ as $N \to +\infty$ implies that as N grows, the cooling of Ω_N, by B_N, becomes absolutely efficient, namely, let φ be any bounded, continuous function on M, and $u_N(x, t)$ the solution to the heat equation on Ω_N, vanishing on $\partial\Omega_N$ for all $t > 0$, and equal to φ at time $t = 0$. Then

$$\|u_N(\ , t)\|_{\Omega_N} \leq e^{-\lambda(\Omega_N)t}\|\varphi\| \tag{59}$$

for all $t > 0$. Indeed, for each $t > 0$,

$$\begin{aligned}
\lambda(\Omega_N)\|u_N(\ , t)\|_{\Omega_N}^2 &\leq \|(\mathrm{grad}\, u_N)(\ , t)\|_{\Omega_N}^2 \\
&= -(u(\ , t), (\Delta u)(\ , t))_{\Omega_N} \\
&= -\tfrac{1}{2}\partial_t(\|u(\ , t)\|_{\Omega_N}^2),
\end{aligned}$$

that is,

$$\partial_t(\|u_N(\ ,t)\|^2_{\Omega_N}) + \lambda(\Omega_N)\|u_N(\ ,t)\|^2_{\Omega_N} \le 0, \tag{60}$$

and (59) follows easily.

From (59) one has, by the Cauchy–Schwarz inequality,

$$\int_{\Omega_N} u_N(x, t)\, dV(x) \le e^{-\lambda(\Omega_N)t}\|\varphi\|\sqrt{V(\Omega_N)}.$$

Thus, for any $T > 0$, the total temperature at time $t \in [T, +\infty)$ tends to 0, uniformly in t, as $n \to +\infty$, if $\lambda(\Omega_N) \to +\infty$ as $N \to +\infty$.

Jeffrey Rauch [1, p. 356] has pointed out that Theorem 8 illustrates the well-known, but rarely admitted, dictum: formulas are smarter than people. For if B_N represents a block of ice of fixed volume chopped into spherical disks, in which case Nr_N^n is basically constant, then everyday experience suggests that as $N \to +\infty$ (i.e., the ice is being chopped up more finely) the cooling efficiency of the ice increases to absolute efficiency. (We are assuming internal mechanisms in the ice to keep them from melting.) While this suggestion conforms with the truth, it does not point to the truth. For everyday experience only suggests that one study the behavior of Nr_N^n (the volume of the ice), or Nr_N^{n-1} (the surface area of the ice), as $N \to +\infty$—not the behavior of Nr_N^{n-2}. Theorem 8 is saying that one can produce absolutely efficient cooling even when the surface area of the ice tends to 0 as $N \to +\infty$.

For the study of what happens when $Nr_N^{n-2} \to \alpha \in (0, +\infty)$ as $N \to +\infty$, cf. Kac [2], J. Rauch [1], Rauch–Taylor [2], Simon [1, Section 22], and, more recently, Chavel–Feldman [9].

CHAPTER X

Surfaces of Constant Negative Curvature

In this chapter we apply various techniques of earlier chapters, especially Chapter IV, to the study of eigenvalues on Riemann surfaces, that is, orientable 2-dimensional Riemannian manifolds of constant Gauss curvature -1. In recent years there has been a veritable explosion in the study of hyperbolic manifolds, and the analysis associated with them. Our study here can only be introductory. We present in Sections 3 and 4 the results of Buser [1], Randol [5], and Schoen–Wolpert–Yau [1], which give insight into the interaction of low eigenvalues with the geometry–topology of the underlying Riemann surface, the analytic methods consisting of max–min arguments, and Cheeger and isoperimetric inequalities. These stand in sharp contrast to the powerful and elegant theory of the Selberg trace formula, presented by B. Randol in Chapter XI.

In Section 2 we derive an explicit formula for the heat kernel on the hyperbolic plane, its existence and uniqueness being established in Chapter VIII. The method consists of generalizing the action of the Fourier transform on radial functions.

In the last section we derive the upper half-space model of hyperbolic space (the hyperbolic space being originally defined in Section II.5 via the ball model) in preparation for Chapter XI.

1. GEOMETRY OF THE HYPERBOLIC PLANE

We recall from Section II.5 that the simply connected space form of constant sectional curvature -1 is given by the manifold

$$\mathbb{B}^n = \{x \in \mathbb{R}^n : |x| < 1\}$$

carrying the Riemannian metric

$$ds = 2|dx|/\{1 - |x|^2\}.$$

In what follows we restrict ourselves to the case $n = 2$, and refer to the space form as $\mathbb{H}^2$, the *hyperbolic plane*. We change our notation to that of complex variables, and write $\mathbb{R}^2$ as $\mathbb{C}$, the complex numbers, with a typical $z \in \mathbb{C}$ written as

$$z = x + iy.$$

When viewing a point of $\mathbb{H}^2$ as a point of $\mathbb{C}$, we write it as an element in $\mathbb{C}$, namely, the underlying manifold of $\mathbb{H}^2$ is written as

$$D = \{z \in \mathbb{C} : |z| < 1\},$$

and the metric as

$$ds = 2|dz|/\{1 - |z|^2\}.$$

Also recall, from Section II.5, that geodesic polar coordinates about the origin are determined by setting

$$z = \rho e^{i\theta}, \qquad \rho = \tanh(r/2); \tag{1}$$

for then we have

$$ds^2 = dr^2 + \sinh^2 r\, d\theta^2.$$

Thus the straight Euclidean line segments in D, emanating from $z = 0$, are geodesics in $\mathbb{H}^2$, with associated arclength r given by (1).

To each $\sigma \in \mathbb{R}$ we let κ_σ denote the rotation of D about the origin, given by

$$\kappa_\sigma \cdot z = e^{i\sigma} z.$$

One easily sees that each κ_σ, $\sigma \in \mathbb{R}$, is an orientation preserving isometry of $\mathbb{H}^2$.

To each $\alpha \in D$, we let T_α denote the fractional linear transformation

$$T_\alpha \cdot z = (z + \alpha)/(1 + \bar{\alpha} z).$$

Since as a mapping of $\mathbb{C} \to \mathbb{C}$, T_α preserves $\mathbb{S}^1 = \partial D$, and takes 0 to α, we have that T_α is a conformal automorphism of $\bar{D}$. To see that $T_\alpha | D$ is an isometry of $\mathbb{H}^2$, one calculates, for $w = T_\alpha \cdot z$,

$$dw = (1 - |\alpha|^2)\, dz/(1 + \bar{\alpha} z)^2,$$

$$|dw| = (1 - |\alpha|^2)|dz|/|1 + \bar{\alpha} z|^2,$$

$$1 - |w|^2 = (1 - |\alpha|^2)(1 - |z|^2)/|1 + \alpha z|^2,$$

from which one easily obtains the claim.

Next, we note that

$$T_\beta \circ T_\alpha = T_{(\alpha+\beta)/(1+\bar{\alpha}\beta)},$$

so the collection of $\{T_\alpha : \alpha \in D\}$ form a group.

We then consider the group G of orientation preserving isometries of $\mathbb{H}^2$ generated by the groups $\{\kappa_\sigma : \sigma \in \mathbb{R}\}$, $\{T_\alpha : \alpha \in D\}$, and argue that G is, in fact, the complete group of orientation preserving isometries of $\mathbb{H}^2$. Indeed, given an orientation preserving isometry T of $\mathbb{H}^2$, let $\beta = T^{-1} \circ o$. Then the mapping $T_{-\beta} \cdot T^{-1}$ is an isometry of $\mathbb{H}^2$ preserving the origin of D. One easily shows that $T_{-\beta} \cdot T^{-1}$ is therefore a rotation of D, from which one has

$$T = \kappa_\sigma \circ T_{-\beta}$$

for some σ in $\mathbb{R}$.

Also note that this isometry group G acts transitively on the unit tangent bundle of $\mathbb{H}^2$.

To obtain the full group of isometries of $\mathbb{H}^2$, one takes the group generated by G, and the restriction to D of a euclidean reflection in a line through the origin of D.

We now consider the full collection of geodesics of $\mathbb{H}^2$. To do so, it will be convenient to refer to the full collection of lines and circles in $\mathbb{C}$ as *generalized circles.*

Recall that the geodesics emanating from the origin of D are the euclidean line segments through the origin with euclidean and hyperbolic distances related by (1). But the group G consists of fractional linear transformations of $\mathbb{C}$, leaving D invariant. This group, therefore, preserves the collection of generalized circles in $\mathbb{C}$ and their angles of mutual intersection. Thus the image of line segments through the origin of D, and intersecting ∂D, are generalized circle segments in D which intersect ∂D orthogonally. The transitivity of the action of the isometry group G guarantees that *all* generalized circle segments in D, intersecting ∂D orthogonally, are images, under G, of line segments through the origin of D which intersect ∂D. We therefore conclude that the full collection of geodesics of $\mathbb{H}^2$ coincides with the collection of generalized circle segments in D intersecting ∂D orthogonally.

One can easily check that, with the geodesics playing the role of lines, $\mathbb{H}^2$ is a hyperbolic geometry. Note that the classical area-excess formula for the angles of a geodesic triangle is, precisely, the Gauss–Bonnet formula (Section III.1)

$$-A = \left(\sum_{j=1}^{3} \alpha_j\right) - \pi,$$

where $\alpha_1, \alpha_2, \alpha_3$ are the interior angles of the triangle.

2. THE HEAT KERNEL OF THE HYPERBOLIC PLANE

We start by referring the reader to Section XII.4 in which the heat kernel of Euclidean space is obtained by (i) subjecting the initial value problem for the heat equation, on Euclidean space, to the Fourier transform, (ii) solving the resulting problem in the transform space, and (iii) subjecting the solution to the inverse transform. Whereas, however, the discussion is given in the usual rectangular coordinates on euclidean space, it could just as well be carried out in spherical coordinates, in which case the space variable consists of one real variable—distance from the origin. For examples, in the 2-dimensional case we write $\mathbb{R}^2$ as $\mathbb{C}$, and obtain, for any function φ depending only on distance from the origin,

$$\begin{aligned}\hat{\varphi}(w) &= (1/2\pi)\iint_{\mathbb{C}} \varphi(|z|)e^{i\operatorname{Re} z\bar{w}}\,dA(z)\\ &= (1/2\pi)\int_0^\infty \varphi(r)r\,dr\int_0^{2\pi} e^{ir\rho\cos\theta}\,d\theta\\ &= \int_0^\infty \varphi(r)J_0(r\rho)r\,dr,\end{aligned}$$

where $r = |z|$, $\rho = |w|$, and J_0 is the 0th-order Bessel function. Since $\hat{\varphi}$ depends only on ρ, the Fourier transform maps radial functions to radial functions. Since $r \to J_0(r\rho)$ is an eigenfunction of Δ on $\mathbb{R}^2$ (cf. Theorem III.4) with eigenvalue ρ^2, we have for radial functions with sufficiently rapid decrease at ∞, via integration-by-parts,

$$\widehat{\Delta\varphi} = -\rho^2\hat{\varphi}.$$

It is this equation that makes possible the easy solution of the transformed initial-value problem for the heat equation. The formula for the transform of convolutions then provides for an easy inversion of the solution, of the transformed problem, to the original radial variable.

This is the approach we take to determine the heat kernel on the hyperbolic plane. Introductory treatments to this material, and to the whole field of harmonic analysis on Lie groups and homogeneous spaces, can be found in Dym–McKean [1], Helgason [1], and Terras [1, 2]. These sources contain a plethora of references with which one can work his, or her, way back to the original literature.

Our search for the fundamental solution of the heat equation will focus on the calculations leading to the discovery of the appropriate candidate. Details of the arguments can be found in the above references.

To determine the radial eigenfunctions of $\mathbb{H}^2$, one has the ordinary differential equation

$$\frac{\partial^2 v}{\partial r^2} + \coth r \frac{\partial v}{\partial r} + \lambda v = 0 \tag{2}$$

which is solved by the Legendre function (Section XII.5)

$$v(r) = P_{-1/2+\sqrt{-\lambda+1/4}}(\cosh r),$$

normalized to be equal to 1 when $r = 0$. Guided by the fact that for any regular domain Ω in $\mathbb{H}^2$ we have $\lambda(\Omega) > \frac{1}{4}$, we only consider those eigenfunctions for which $\lambda \geq \frac{1}{4}$, namely, we set

$$i\rho = \sqrt{-\lambda + \tfrac{1}{4}}, \qquad \lambda = \tfrac{1}{4} + \rho^2, \qquad F_\rho(r) = P_{-1/2+i\rho}(\cosh r),$$

where $\rho \geq 0$.

So for any $\varphi \in C_c^\infty([0, +\infty))$ we define the *Mehler transform* (Mehler [1])

$$\hat{\varphi}(\rho) = \int_0^\infty \varphi(r) P_{-1/2+i\rho}(\cosh r) \sinh r \, dr \tag{3}$$

(note how $\sinh r$ replaces r in the Euclidean case), for which we have the *inversion formula* (Mehler [1]; Fock [1])

$$\varphi(r) = \int_0^\infty \hat{\varphi}(\rho) P_{-1/2+i\rho}(\cosh r) \rho \tanh \pi\rho \, d\rho. \tag{4}$$

Also, the usual integration-by-parts argument implies

$$\widehat{\Delta\varphi}(\rho) = -(\tfrac{1}{4} + \rho^2)\hat{\varphi}(\rho). \tag{5}$$

Therefore, if we are given the initial-value problem for the heat equation on $\mathbb{H}^2$

$$\Delta u = \partial u/\partial t, \qquad u(\cdot, 0) = f, \tag{6}$$

where f has compact support, and depends only on distance from a fixed point o in $\mathbb{H}^2$, then u will depend only on distance from o, and on the time t, and we obtain, applying the Mehler transform to the distance variable r of u and using (5),

$$\hat{u}(\rho, t) = \hat{f}(\rho) e^{-(1/4+\rho^2)t}. \tag{7}$$

We therefore seek a formula for taking the inverse transform of a product, namely, we seek some sort of convolution.

We work in the disk model of $\mathbb{H}^2$, and let o correspond to $z = 0$. To any $w \in D$, we associate the isometry of D, g_w, by

$$g_w = \kappa_\sigma T_R, \qquad w = Re^{i\sigma},$$

$R > 0$, $\sigma \in (-\infty, +\infty)$. Then for functions φ, ψ on $\mathbb{H}^2$ we define the *convolution* $\varphi * \psi$ *of* φ and ψ by

$$(\varphi * \rho)(z) = \int_{H^2} \varphi(w)\psi(g_w^{-1} \cdot z)\, dA(w)$$

(where dA is the hyperbolic area element). Then for φ a radial function, with respect to $z = 0$, we have for any rotation κ of D

$$\begin{aligned}(\varphi * \psi)(\kappa \cdot z) &= \int_{\mathbb{H}^2} \varphi(w)\psi(g_w^{-1}\kappa \cdot z)\, dA(w)\\ &= \int_{\mathbb{H}^2} \varphi(\kappa^{-1} \cdot w)\psi(g_w^{-1}\kappa \cdot z)\, dA(w)\\ &= \int_{\mathbb{H}^2} \varphi(w)\psi(g_w^{-1} \cdot z)\, dA(w)\\ &= (\varphi * \psi)(z),\end{aligned}$$

by the invariance of the area with respect to the isometries. Therefore, if φ is a radial function, then $\varphi * \psi$ is a radial function.

Now we study $\varphi * \psi$ with φ and ψ radial with respect to $z = 0$. First, we now write

$$F_\rho(z) = P_{-1/2+i\rho}(\cosh d(z, 0)).$$

Then $\hat{\varphi}$ is given by

$$\hat{\varphi}(\rho) = (1/2\pi)\int_{\mathbb{H}^2} \varphi(z)F_\rho(z)\, dA(z),$$

which implies

$$\begin{aligned}\widehat{\varphi * \psi}(\rho) &= (1/2\pi)\int_{\mathbb{H}^2} F_\rho(z)\, dA(z)\int_{\mathbb{H}^2} \varphi(w)\psi(g_w^{-1} \cdot z)\, dA(w)\\ &= (1/2\pi)\int_{\mathbb{H}^2} \varphi(w)\, dA(w)\int_{\mathbb{H}^2} \psi(g_w^{-1} \cdot z)F_\rho(z)\, dA(z)\\ &= (1/2\pi)\int_{\mathbb{H}^2} \varphi(w)\, dA(w)\int_{\mathbb{H}^2} \psi(z)F_\rho(g_w \cdot z)\, dA(z).\end{aligned}$$

Since ψ is radial with respect to $z = 0$, we have for any rotation κ of D,

$$\begin{aligned}\int_{\mathbb{H}^2} \psi(z)F_\rho(g_w \cdot z)\, dA(z) &= \int_{\mathbb{H}^2} \psi(\kappa^{-1} \cdot z)F_\rho(g_w \cdot z)\, dA(z)\\ &= \int_{\mathbb{H}^2} \psi(z)F_\rho(g_w\kappa \cdot z)\, dA(z),\end{aligned}$$

which implies

$$\begin{aligned}&\int_{\mathbb{H}^2}\psi(z)F_\rho(g_w\cdot z)\,dA(z)\\ (8)\qquad &=\int_{\mathbb{H}^2}\psi(z)\,dA(z)(1/2\pi)\int_0^{2\pi}F_\rho(g_w\kappa_\sigma\cdot z)\,d\sigma.\end{aligned}$$

LEMMA 1. We have

$$(9)\qquad (1/2\pi)\int_0^{2\pi}F_\rho(g_w\kappa_\sigma\cdot z)\,d\sigma=F_\rho(w)F_\rho(z).$$

PROOF: One approach to (9) is to reduce it, using (1), to the *classical addition formula*

$$\begin{aligned}&P_{-1/2+i\rho}(\cosh r)P_{-1/2+i\rho}(\cosh R)\\ (10)\qquad &=(1/2\pi)\int_0^{2\pi}P_{-1/2+i\rho}(\cosh r\cosh R+\sinh r\sinh R\cos\sigma)\,d\sigma.\end{aligned}$$

The other way is deduce (10) from (9), namely, the function

$$z\to(1/2\pi)\int_0^{2\pi}F_\rho(g_w\kappa_\sigma\cdot z)\,d\sigma$$

is a radial (with respect to $z=0$) eigenfunction on $\mathbb{H}^2$ with eigenvalue $\frac{1}{4}+\rho^2$. But for each λ, the collection of solutions of (2), which are bounded about $r=0$, is one-dimensional. So the left-hand side of (9) is a constant multiple of $F_\rho(z)$. The constant is evaluated by setting $z=0$.

Applying (9) to (8) we have

$$\int_{\mathbb{H}^2}\psi(z)F_\rho(g_w\cdot z)\,dA(z)=\int_{\mathbb{H}^2}\psi(z)F_\rho(w)F_\rho(z)\,dA(z)=2\pi F_\rho(w)\hat{\psi}(\rho),$$

which implies

$$(11)\qquad \widehat{\varphi*\psi}(\rho)=2\pi\hat{\varphi}(\rho)\hat{\psi}(\rho).$$

From (7) and (11) it follows that the solution to the initial-value problem for the heat equation (6) is $1/2\pi$ multiplied by the convolution of f with the inverse transform of $e^{-(1/4+\rho^2)t}$, namely,

$$u(z,t)=(1/2\pi)\int_{\mathbb{H}^2}f(w)\,dA(w)\int_0^\infty F_\rho(g_w^{-1}\cdot z)e^{-(1/4+\rho^2)t}\rho\tanh\pi\rho\,d\rho.$$

So the heat kernel on $\mathbb{H}^2$ will be given by

$$(12)\quad p(z, w, t) = (1/2\pi)\int_0^\infty e^{-(1/4+\rho^2)t}P_{-1/2+i\rho}(\cosh d(z, w))\rho \tanh \pi\rho \, d\rho.$$

To simplify (12), we note that by Eq. (7.47) of Lebedev [1, p. 173], we have

$$P_{-1/2+i\rho}(\cosh r) = \frac{\sqrt{2}}{\pi}\coth \pi\rho \int_r^\infty \frac{\sin \rho\beta}{\sqrt{\cosh \beta - \cosh r}}\, d\beta,$$

which implies

$$p(z, w, t) = \frac{1}{\sqrt{2}\pi^2}\int_{d(z,w)}^\infty \frac{d\beta}{\sqrt{\cosh \beta - \cosh d(z, w)}}\int_0^\infty e^{-(1/4+\rho^2)t}\rho \sin \rho\beta \, \mathrm{d}\rho.$$

For each nonnegative integer k, one has

$$\int_0^\infty \mathrm{e}^{-(1/4+\rho^2)t}\left\{\frac{\rho(-1)^k(\beta\rho)^{2k+1}}{(2k+1)!}\right\} d\rho = \frac{e^{-t/4}\beta\sqrt{\pi}}{4t^{3/2}}\,\frac{(-1)^k(\beta^2/4t)^k}{k!},$$

from which one obtains

$$(13)\qquad p(z, w, t) = \frac{\sqrt{2}e^{-t/4}}{(4\pi t)^{3/2}}\int_{d(z,w)}^\infty \frac{\beta e^{-\beta^2/4t}}{\sqrt{\cosh \beta - \cosh d(z, w)}}\, d\beta.$$

Note that (13) displays the positivity of p for $z \neq w$, $t = 0$.

We leave it to the reader to verify that, in fact, $p(z, w, t)$ given by (13) is a fundamental solution to the heat equation on $\mathbb{H}^2$.

3. LÖBELL SURFACES AND THE ESTIMATES OF P. BUSER

In all that follows we use the term *surface* for any 2-dimensional orientable Riemannian manifold, and *Riemann surface* for a surface whose Gauss curvature is identically -1.

If M is a complete Riemann surface, then its Riemann universal covering is $\mathbb{H}^2$; and if M is also compact, it must have (by the Gauss–Bonnet formula) genus $g \geq 2$.

The reader will recall from Theorem II.5 that for any domain Ω in $\mathbb{H}^2$, the lowest Dirichlet eigenvalue of Ω is always greater than $\frac{1}{4}$. It is natural to ask whether the same estimate holds for the nonzero eigenvalues of any compact Riemann surface (McKean [2]). This question is important in, for example, hyperbolic lattice-point problems, and in the asymptotics of the length spectrum of the surface (cf. Chapter XI). Randol [1] showed that given any compact Riemann surface, it possesses a compact covering for which there are nonzero eigenvalues arbitrarily close to 0. (cf. Section XI.5). Here we shall prove a slightly different result using more elementary methods:

THEOREM 1 (Buser [1]). Given any $\varepsilon > 0$, and integer $g \geq 2$, there exists a compact Riemann surface of genus g such that

$$\lambda_{2g-3} < \varepsilon.$$

Before giving the proof of Theorem 1, it will be convenient to introduce *Fermi coordinates* on an arbitrary surface. A more general sketch can be found in Section XII.8. Let M be a surface, $\gamma\colon \mathbb{R} \to M$, $|\gamma'| = 1$ an embedded geodesic, and $\xi\colon \mathbb{R} \to TM$ a continuous unit vector field along γ, orthogonal to γ. We define the map

$$E(r, t) = \exp_{\gamma(r)} t\xi(r);$$

the general theory guarantees the existence of $\alpha < 0$, $\beta > 0$ so that E is a diffeomorphism of $\mathbb{R} \times (\alpha, \beta)$ onto its image in M. If we wish to consider Fermi coordinates based on a simple closed geodesic γ of length l, then we replace $\mathbb{R}$, above, by $\mathbb{S}^1(l/2\pi)$. In this case we are guaranteed the existence of $\alpha < 0$, $\beta > 0$ for which E is a diffeomorphism of $\mathbb{S}^1(l/2\pi) \times (\alpha, \beta)$ onto its image in M.

It is customary to define the *injectivity radius of E* as the largest $T > 0$ for which $E\,|\,\mathbb{R} \times (-T, T)$ (or $E\,|\,\mathbb{S}^1(l/2\pi) \times (-T, T)$ in the periodic case) is a diffeomorphism onto its image, *the collar in M about* γ.

Standard arguments verify that on the collar, we have

$$\text{(J)} \qquad \begin{gathered} \nabla_t \partial_t E = 0, \qquad |\partial_t E| = 1, \qquad \langle \partial_r E, \partial_t E \rangle = 0, \\ \nabla_t^2 \partial_r E + (K \circ E)\partial_r E = 0, \\ \partial_r E(r, 0) = \gamma'(r), \qquad \nabla_t \partial_r E(r, 0) = 0, \end{gathered}$$

where K denotes the Gauss curvature of M. Since the unit vector field $\partial_r E/|\partial_r E|$ must be parallel in the t direction, we may replace the vector Jacobi equation (J) by the scalar equation

$$\partial_t^2 \eta + K\eta = 0$$

for $\eta = |\partial_r E|$, with initial data

$$\eta(r, 0) = 1, \qquad \partial_t \eta(r, 0) = 0.$$

The Riemannian metric is written in these coordinates as

$$(14) \qquad ds^2 = dt^2 + \eta^2(r, t)\, dr^2,$$

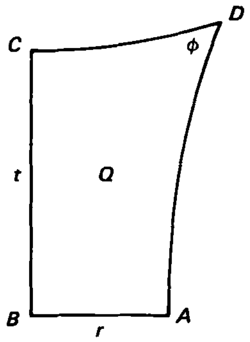


Fig. 4

and the density of the Riemannian measure is given by

$$dA = \eta(r, t)\, dt\, dr. \tag{15}$$

Note that when $K \equiv -1$ we have

$$\eta(r, t) = \cosh t. \tag{16}$$

PROOF OF THEOREM 1: Our surfaces consist of a collection first constructed by Löbell [1].

Consider the quadrilateral Q of Fig. 4, in $\mathbb{H}^2$, with $\varphi = \pi/3$, and place six copies of Q adjacent to one another as in Fig. 5. We call the resulting hexagon in $\mathbb{H}^2$, G, and label the geodesic segments bounding $G: \alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3$ as indicated, with

$$l(\alpha_j) = 2r, \qquad l(\beta_k) = 2t, \qquad j = 1, 2, 3.$$

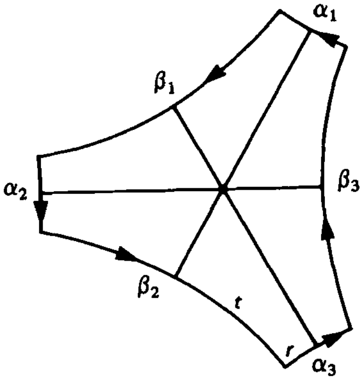


Fig. 5

Take another copy of G, $\tilde{G}$, with boundary $\tilde{\alpha}_1$, $\tilde{\beta}_1$, $\tilde{\alpha}_2$, $\tilde{\beta}_2$, $\tilde{\alpha}_3$, $\tilde{\beta}_3$, and identify β_j with $\tilde{\beta}_j$ for $j = 1, 2, 3$. The result is a Riemann surface Ω with compact closure and bounded by 3 simple closed geodesics, homeomorphic to $\mathbb{S}^2$ with 3 disks removed. Let γ_j be the bounding geodesic obtained from α_j and $\tilde{\alpha}_j$, $j = 1, 2, 3$. Clearly each of the geodesics γ_j has length $4r$. The Gauss–Bonnet theorem and formula (Section III.1) implies that

$$A(\Omega) = 2\pi.$$

Note that by formula (38) in Section 5, t and r are related by

$$\tfrac{1}{2} = (\sinh r)(\sinh t). \tag{17}$$

So we think of the family of surfaces Ω as parametrized by r. From (17), it is clear that we may guarantee $t > 1$ by picking

$$r < \operatorname{arcsinh}\{\tfrac{1}{2} \operatorname{csch} 1\}. \tag{18}$$

We refer to our surfaces Ω as *Löbell Y-pieces.*

LEMMA 2. For Löbell Y-pieces Ω, satisfying (18), we have

$$\lambda(\Omega) \le \frac{12r \sinh 1}{2\pi - 12r \sinh 1}, \tag{19}$$

where $\lambda(\Omega)$ denotes the lowest Dirichlet eigenvalue of Ω.

PROOF: First note that for each $j = 1, 2, 3$ the injectivity radius associated to each γ_j is greater than or equal to t, and that the collars

$$K_j = \{x \in \Omega : d(x, \gamma_j) < t\},$$

$j = 1, 2, 3$ (more precisely, half-collars) are pairwise disjoint. For $j = 1, 2, 3$ let D_j be the attenuated collars,

$$D_j = \{x \in \Omega : d(x, \gamma_j) < 1\}.$$

Then

$$A(D_j) = l(\gamma_j) \int_0^1 \cosh \tau \, d\tau = 4r \sinh 1. \tag{20}$$

Consider the function φ on Ω given by

$$\varphi(x) = \begin{cases} d(x, \partial\Omega), & d(x, \partial\Omega) \le 1, \\ 1 & \text{otherwise.} \end{cases}$$

Then

$$\int_\Omega |\text{grad } \varphi|^2 \, dA = 12r \sinh 1$$

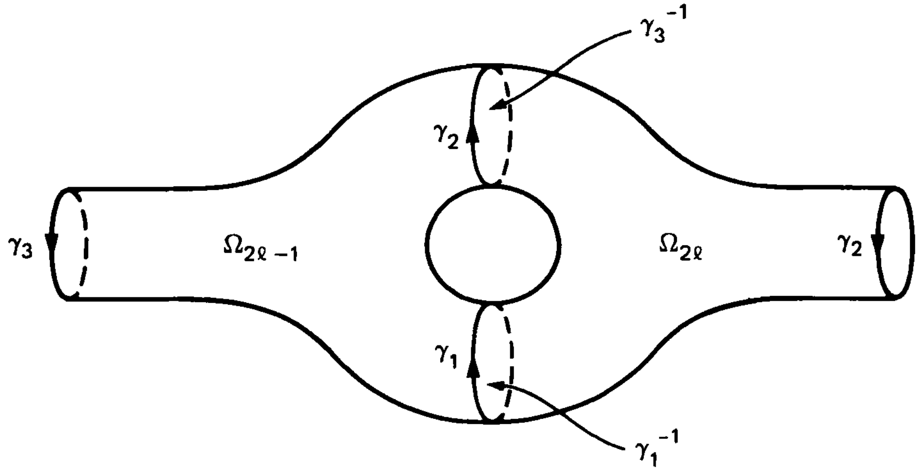


Fig. 6

and

$$\int_\Omega \varphi^2\,dA \geq A(\Omega) - \sum_{j=1}^{3} A(D_j) = 2\pi - 12r\sinh 1.$$

Estimate (19) now follows from Rayleigh's theorem (Section I.5).

To conclude the proof of Theorem 1, we take $2g-2$ copies of $\Omega: \Omega_1, \ldots, \Omega_{2g-2}$, with boundary geodesics $\gamma_{k,j}$, $k = 1, \ldots, 2g-2$, $j = 1, 2, 3$ (the k indicating that $\gamma_{k,j}$ is part of the boundary of Ω_k). We think of each geodesic $\gamma_{k,j}$ as oriented by Ω_k, and denote the same geodesic with reverse orientation by $(\gamma_{k,j})^{-1}$.

Now identify $\gamma_{k,2}$ with $(\gamma_{k+1,3})^{-1}$ for all $k = 1, 2, \ldots, 2g-3$ and $\gamma_{2g-2,2}$ with $(\gamma_{1,3})^{-1}$. Also identify $\gamma_{2l-1,1}$ with $(\gamma_{2l,1})^{-1}$ for all $l = 1, \ldots, g-1$. The result of these identifications is a compact Riemann surface M of genus g (Fig. 6). Since we used $2g-2$ copies of Ω to produce the *Löbell surface M*, Lemma 2, and the domain *monotonicity* of eigenvalues for vanishing Dirichlet data (Section I.5), combine to imply

$$\lambda_{2g-3}(M) \leq \lambda(\Omega) \leq \frac{12r\sinh 1}{2\pi - 12r\sinh 1}.$$

Given any $\varepsilon > 0$, we may pick r, at the outset, to be sufficiently small to guarantee, beside (18),

$$\frac{12r\sinh 1}{2\pi - 12r\sinh 1} < \varepsilon,$$

which implies the theorem.

The Löbell surfaces may also be used to prove

THEOREM 2 (Buser [1]). Given positive integers g and k and any $\varepsilon > 0$, there exists a compact Riemann surface M, of genus g, satisfying

$$\lambda_k(M) \leq \tfrac{1}{4} + \varepsilon.$$

PROOF: Recall from Theorem III.8 that for any geodesic disk $\mathsf{B}(\delta)$ in $\mathbb{H}^2$ we have

$$\tfrac{1}{4} < \lambda(\mathsf{B}(\delta)) \leq \tfrac{1}{4} + \pi^2/\delta^2,$$

where $\lambda(\mathsf{B}(\delta))$ is the lowest Dirichlet eigenvalue of $\mathsf{B}(\delta)$. By Corollary III.3 we therefore have, for any compact Riemann surface M, with diameter $d(M)$,

$$\lambda_k(M) \leq \tfrac{1}{4} + 4\pi^2 k^2/d^2(M). \tag{21}$$

If M has genus g, and l is the length of a simple closed geodesic in M, then by determining Fermi coordinates on M based on the geodesic, and using the Gauss–Bonnet theorem (cf. a similar argument in Corollary III.4), we have

$$d(M) \geq \operatorname{arcsinh}\{2\pi(g-1)/l\}. \tag{22}$$

But for the Löbell surfaces, it is always possible to let $l = 4r \downarrow 0$. In these examples, estimates (21), (22) combine to imply the theorem.

On the other hand, there are some restrictions on the number of small eigenvalues.

THEOREM 3 (Buser [1]). For any compact Riemann surface of genus g, we have

$$\lambda_{4g-2} > \tfrac{1}{4}.$$

PROOF: It is a result of Fricke–Klein [1, p. 318] that M can be triangulated with $4g - 2$ geodesic triangles. By the domain monotonicity of eigenvalues with vanishing Neumann data (Section I.5) the theorem will follow from

LEMMA 3. For any geodesic triangle T in $\mathbb{H}^2$, and $\mu(T)$ the first nonzero Neumann eigenvalue of T, we have

$$\mu(T) > \tfrac{1}{4}.$$

(i) (ii)

Fig. 7

PROOF: Before starting the proof, we emphasize that T denotes a domain in $\mathbb{H}^2$, homeomorphic to a 2-disk and bounded by 3 geodesic segments.

By Courant's nodal domain theorem (Section I.5) any eigenfunction φ of $\mu(T)$ has precisely 2 nodal domains.

If one of the nodal domains G of φ has compact closure in the interior of T, then $\mu(T) = \lambda(G) > \frac{1}{4}$ by Theorem II.5.

Now assume that both nodal domains of φ have closure intersecting ∂T. Then the nodal line $\varphi^{-1}[0]$ consists of a smooth curve in T (Cheng [4]), and must intersect ∂T precisely twice. We distinguish two possibilities: the two intersections of the nodal line with ∂T (i) are, or (ii) are not, on the closure of just 1 geodesic segment of ∂T (Fig. 7).

In the first case, let G denote the nodal domain bounded by the nodal line and the geodesic segment σ determined by the intersection of the nodal line with ∂T. Now pick Fermi coordinates $z = q(r, t)$ in $\mathbb{H}^2$ based on the full geodesic containing σ. Then the preimage of G in $\mathbb{R}^2$ is a domain of the form

$$\{(r, t) : t \in \bigcup_{j=1}^{N(r)} (\alpha_j(r), \beta_j(r)),\ r \in (a, b)\},$$

where for each r in (a, b), $N(r)$ may be finite or infinite, and $\alpha_j(r) \geq 0$ for all j. For the eigenfunction φ we then have, setting

$$\Phi = \varphi \circ q,$$

$$\iint_G \varphi^2\, dA = \int_a^b dr \sum_{j=1}^{N(r)} \int_{\alpha_j(r)}^{\beta_j(r)} \Phi^2(r, t) \cosh t\, dt. \tag{23}$$

To estimate this integral from above, we give a slight variation of the argument establishing (II.45) (that the lowest Dirichlet eigenvalue of any disk in $\mathbb{H}^2$ is greater than $\frac{1}{4}$). Fix j and r. Then (suppressing j and r from the notation) we have

$$\begin{aligned}\int_\alpha^\beta \Phi^2 \cosh &= \Phi^2 \sinh|_\alpha^\beta - 2\int_\alpha^\beta \Phi(\partial_t\Phi)\sinh\\ &= -2\int_\alpha^\beta \Phi(\partial_t\Phi)\sinh\\ &\le 2\left\{\int_\alpha^\beta \Phi^2\sinh\right\}^{1/2}\left\{\int_\alpha^\beta (\partial_t\Phi)^2\sinh\right\}^{1/2}\\ &\le 2\left\{\int_\alpha^\beta \Phi^2\cosh\right\}^{1/2}\left\{\int_\alpha^\beta (\partial_t\Phi)^2\cosh\right\}^{1/2}\\ &\le 2\left\{\int_\alpha^\beta \Phi^2\cosh\right\}^{1/2}\left\{\int_\alpha^\beta (|\mathrm{grad}\,\varphi|^2\circ q)\cosh\right\}^{1/2},\end{aligned}$$

that is, after canceling and squaring,

$$\int_\alpha^\beta \Phi^2\cosh \le 4\int_\alpha^\beta (|\mathrm{grad}\,\varphi|^2\circ q)\cosh. \tag{24}$$

The key to the string of inequalities is the vanishing of

$$\Phi^2 \sinh|_\alpha^\beta$$

—indeed, β will always correspond to a point of ∂G on the nodal line, in which case $\Phi = 0$; and α will correspond to either a point of ∂G on the nodal line, in which case $\Phi = 0$, or α will correspond to a point of ∂G on the geodesic segment of ∂T, in which case $\alpha = 0$ and $\sinh = 0$. We also note that one goes from the third to the fourth line via the Cauchy–Schwarz inequality.

One now substitutes (24) into (23) to obtain

$$\tfrac{1}{4}\iint_G \varphi^2\,dA \le \iint_G |\mathrm{grad}\,\varphi|^2\,dA = \mu(T)\iint_G \varphi^2\,dA,$$

which implies the theorem in this first case.

In the second case, where the nodal line intersects ∂T at 2 distinct geodesic segments, one determines geodesic polar coordinates centered at the intersection of the 2 segments and argues as in the first case (following the argument for (II.45) even more closely than in the first case).

4. LOW EIGENVALUES AND SHORT GEODESICS

We now examine the methods of Section 3 more closely. Estimate (22) shows that the existence of a short simple closed geodesic implies a large diameter for a compact Riemann surface. However, the estimate requires a global piece of information, namely, the genus of the compact Riemann surface. In what follows we start with an estimate of B. Randol for the injectivity radius associated to a simple closed geodesic. As the length of the simple closed geodesic becomes arbitrarily small, the estimate of the injectivity radius will be of the same order as the estimate of the diameter given by (22). This will allow for a more systematic investigation of the arguments of Section 3.

THEOREM 4 (Randol [5]). Let M be a compact Riemann surface with simple closed geodesic γ of length $l > 0$. Then the injectivity radius $\imath$, of the collar about γ in M, satisfies

$$\imath \geq \operatorname{arcsinh}\{\operatorname{csch} l/2\}. \tag{25}$$

COROLLARY 1. If C_γ denotes the collar about γ, then

$$A(C_\gamma) \geq 2l \operatorname{csch} l/2.$$

Note that (25) implies, using the concavity in C_γ of the distance function from γ, that intersecting simple closed geodesics cannot be simultaneously small.

Also note that

$$\lim_{l \downarrow 0} 2l \operatorname{csch} l/2 = 4.$$

We remark that estimates for bounding $A(C_\gamma)$ from below were first considered in Keen [1] and Halpern [1]. The generalization of Randol's argument to surfaces of variable nonpositive Gauss curvature can be found in Chavel–Feldman [3], and a still further generalization, with weaker definition of the collar, can be found in Buser [2].

PROOF OF THEOREM 4: Let $\gamma\colon \mathbb{S}^1(l/2\pi) \to M$, $|\gamma'| = 1$, be a simple closed geodesic of length l.

An argument of Klingenberg [1] can be easily adapted (cf. Grossman [1]) to show that there exist two distinct geodesic segments $\sigma_j\colon [0, \imath] \to M$, $|\sigma_j'| = 1$, with $\sigma_j(0)$ on γ, for $j = 1, 2$, and satisfying

$$\sigma_1(\imath) = \sigma_2(\imath), \qquad \sigma_1'(\imath) = -\sigma_2'(\imath).$$

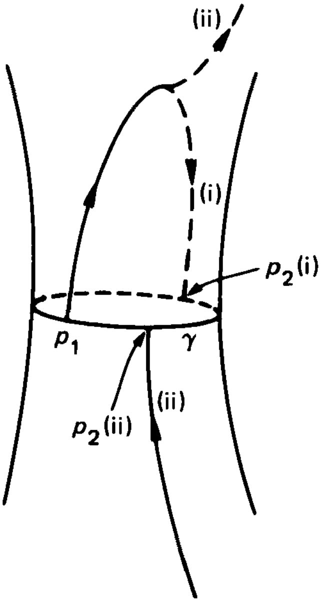


Fig. 8

So the union of the two geodesic segments is a smooth geodesic segment which can be written as $\sigma\colon [0, 2\iota] \to M$, $|\sigma'| = 1$, such that $\sigma(0) = \sigma_1(0) \equiv: p_1$, $\sigma(2\iota) = \sigma_2(0) \equiv: p_2$. We let q denote $\sigma(\iota)$.

Note that Fig. 8 contains the two possibilities of how σ intersect γ at p_2. Compare the discussion below.

Assume that γ is parametrized so that $\gamma(0) = p_1$, and that γ is oriented so that $\sigma'(0)$ is obtained by rotating $\gamma'(0)$ through $\pi/2$ radians in M_p, the tangent space to M at p.

Next, consider the geodesic τ in M for which $\tau(0) = q$, and $\tau'(0)$ is obtained by rotating $\sigma'(\iota)$ through $-\pi/2$ radians in M_q. The key to the proof is that *if* (25) *is false, then* τ *is a geodesic loop which is freely homotopic to* γ. Indeed, we lift the geodesics γ, σ to geodesics $\tilde{\gamma}$, $\tilde{\sigma}$ in $\mathbb{H}^2$ (the universal covering of M) starting at some lift $\tilde{p}$ of p_1. Note that $\tilde{\sigma}'(0)$ is obtained by rotating $\tilde{\gamma}'(0)$ through $\pi/2$ radians in $(\mathbb{H}^2)_{\tilde{p}}$. Next, we let Γ be the geodesic in M such that $\Gamma(0) = \gamma(-l/2) = \gamma(l/2)$, and $\Gamma'(0)$ is obtained by rotating $\gamma'(-l/2) = \gamma'(l/2)$ through $\pi/2$ radians in $M_{\Gamma(0)}$. When γ is lifted to $\mathbb{H}^2$ (starting at p_1) to $\tilde{\gamma}$, the lifts of $\gamma(-l/2) = \gamma(l/2)$ are now distinct points on $\tilde{\gamma}$, thereby determining 2 lifts of Γ—$\tilde{\Gamma}_1, \tilde{\Gamma}_2$—starting at $\tilde{\gamma}(-l/2)$, $\tilde{\gamma}(l/2)$, respectively. The lift, now, of the geodesic τ, $\tilde{\tau}$, starting at $\tilde{\sigma}(\iota)$ will be oriented so that $\tilde{\tau}'(0)$ is obtained by rotating $\tilde{\sigma}'(\iota)$ through $-\pi/2$ radians in $(\mathbb{H}^2)_{\tilde{\sigma}(\iota)}$.

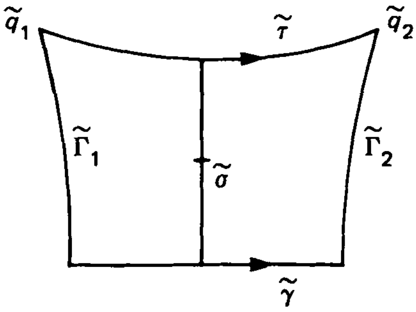


Fig. 9

By (3), if (25) is false, then $\tilde{\tau}$ will have to intersect $\tilde{\Gamma}_1$ and $\tilde{\Gamma}_2$. By symmetry, this intersection must be at the same arclength along $\tilde{\Gamma}_1$ and $\tilde{\Gamma}_2$ (cf. Fig. 9). The respective points of intersection, $\tilde{q}_1, \tilde{q}_2$, will then project, under the covering, to the same point in M. Therefore $\tilde{\tau}$ projects to the loop τ in M, freely homotopic to γ.

We now distinguish the two possibilities referred to in Fig. 8. For convenience, let ξ be the continuous unit vector field along γ, orthogonal to γ, for which $\sigma'(0) = \xi(p_1)$. The two possibilities are that (i) $\sigma'(2\iota) = -\xi(p_2)$, or (ii) $\sigma'(2\iota) = \xi(p_2)$.

We consider the first possibility, that $\sigma'(2\iota) = -\xi(p_2)$. Then we go through the same lifting construction in $\mathbb{H}^2$ as before, except that we now start at p_2. However, since the lift of the geodesic segment σ starting at p_2 is oriented in the direction opposite to the one obtained by starting at p_1, the lift of τ in $\mathbb{H}^2$ *now* appears with the orientation opposite to the left of τ obtained when we started the lifting procedure at p_1. We therefore have τ freely homotopic to γ^{-1}, which implies γ is homotopic to γ^{-1}, contradicting a theorem of Preissmann [1] (cf. Section XI.1).

For the second possibility, we have the lifting to $\mathbb{H}^2$ through $\tilde{p}_2$ appear as in Fig. 10. One then easily sees that if the surface M is cut along γ then the boundaries of the resulting surfaces are freely homotopic—an impossibility, since the genus of M is greater than or equal to 2.

We now apply Randol's estimate to give a more general form of the argument of Theorem 1 (Schoen–Wolpert–Yau [1]).

We are given a fixed number $\beta > 0$, and consider all noncompact Riemann surfaces Ω with compact closure, whose boundary consists of closed geodesic circles with total length $l(\partial\Omega)$, satisfying

$$l(\partial\Omega) \leq 2\beta.$$

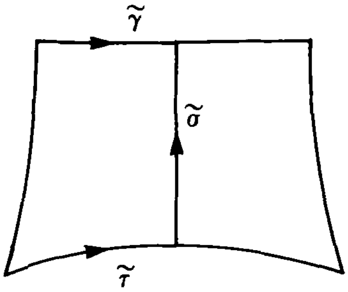


Fig. 10

We claim that

$$\lambda(\Omega) \le \frac{l(\partial M)\cosh\beta}{\alpha^2\{2\pi - l(\partial M)\operatorname{csch}\beta\}}, \tag{26}$$

where

$$\alpha \equiv\colon \operatorname{arcsinh}(\operatorname{csch}\beta). \tag{27}$$

To prove (26), let the geodesic circles $\gamma_1, \ldots, \gamma_k$, with respective injectivity radii $\iota_1, \ldots, \iota_k$ for the associated collars. Then since $l_j < 2\beta$ for all $j = 1, \ldots, k$, we have by (25),

$$\iota_j > \alpha$$

for all $j = 1, \ldots, k$.

For each $j = 1, \ldots, k$ consider the attenuated collar

$$D_j = \{x \in \Omega : d(x, \gamma_j) < \alpha\}.$$

LEMMA 4 (Matelski [1]). When $k > 1$, the collection of collars $D_1, \ldots, D_k$ is pairwise disjoint.

PROOF: Assume $D_r \cap D_s \neq \varnothing$ for a pair of distinct r and s. Then there exists α' in $(0, \alpha)$ such that the, still again, attenuated collars

$$D'_r = \{x : d(x, \gamma_r) < \alpha'\}, \qquad D'_s = \{x : d(x, \gamma_s) < \alpha'\}$$

still have nonempty intersection. Let x_0 be the closest point in $\overline{K'_r \cap K'_s}$ to γ_r. Then an easy argument shows that

$$d(x_0, \gamma_s) = \alpha',$$

and the distance circles

$$\Gamma_r = \{y : d(y, \gamma_r) = d(x_0, \gamma_r)\}, \qquad \Gamma_s = \{y : d(y, \gamma_s) = \alpha'\}$$

are tangent at x_0. The argument of Randol's theorem will then show that the geodesic in M through x_0, tangent to Γ_r and Γ_s at x_0, will be a geodesic loop simultaneously freely homotopic to γ_r, γ_s. So γ_r and γ_s are freely homotopic, which is impossible unless Ω is a cylinder bounded by γ_r and γ_s in which case the Gauss curvature could not always be negative (by the Gauss–Bonnet theorem). This implies the lemma.

We now estimate $\lambda(M)$ as in the case of the Löbell surfaces. Pick the function

$$\varphi(x) = \begin{cases} d(x, \partial\Omega), & d(x, \partial\Omega) \leq \alpha, \\ \alpha & \text{otherwise.} \end{cases}$$

Then one easily has

$$\int_\Omega |\mathrm{grad}\ \varphi|^2\, dA = l(\partial\Omega) \operatorname{csch} \beta,$$

and

$$\int_\Omega \varphi^2\, dA \geq \alpha^2\{A(\Omega) - \sum_{j=1}^{k} A(D_j)\} \geq \alpha^2\{2\pi - l(\partial\Omega) \operatorname{csch} \beta\}.$$

We have, above, $A(\Omega) \geq 2\pi$ since by the Gauss–Bonnet formula $A(\Omega)$ is an integral multiple of 2π. We therefore have (26), by Rayleigh's principle (Section I.5).

We now apply (26). Let M be a compact Riemann surface of genus g. For each $k = 1, \ldots, 2g - 3$, consider the family of curves $\mathscr{C}_k(M)$ consisting of the disjoint union of simple closed geodesics dividing M into $k + 1$ components. Then a combinatorial argument (cf. Keen [2]) implies that the maximal number of simple closed geodesics, dividing M into $\leq 2g - 2$ components, is $3g - 3$. If we let $l_k(M)$ denote the minimum length of all curves in $\mathscr{C}_k(M)$, then a result of Bers [1, Statement XV] implies that there exists a positive number $\beta = \beta(g)$, depending only on g, such that

$$l_k(M) < \beta(g), \qquad k = 1, \ldots, 2g - 3. \tag{28}$$

An imediate consequence of (26), (28) and max–min arguments (Section I.5) is

THEOREM 5 (Schoen–Wolpert–Yau [1]). If M is a compact Riemann surface of genus g then there exists a positive number β_g depending only on g, such that

$$\lambda_k(M) \leq l_k(M)\beta_g \tag{29}$$

for all $k = 1, \ldots, 2g - 3$.

We note that (21), (22) imply

(30) $$\lambda_{2g-2}(M) \leq \tfrac{1}{4} + 4\pi^2(2g-2)^2/\operatorname{arcsinh}^2\{2\pi(g-1)/\beta_g\},$$

the right-hand side depending only on g. Next, we give the argument in Schoen–Wolpert–Yau [1] for a lower bound of λ_n in terms of l_n, $n = 1, \ldots, 2g-3$.

THEOREM 6. For any integer $g \geq 2$ there exists a positive number $\alpha = \alpha_g$ such that for any compact Riemann surface of genus g we have

(31) $$\lambda_k(M) \geq l_k^2(M)\alpha_g,$$

$k = 1, \ldots, 2g-3$, and

(32) $$\lambda_{2g-2}(M) \geq \alpha_g.$$

We note that for small $l_n(M)$, (31) can be replaced by the stronger result,

(33) $$\lambda_k(M) \geq l_k(M)\alpha_g$$

(in fact, it requires that one first prove (31)) but we shall content ourselves, here, with the weaker result. See Schoen–Wolpert–Yau [1] for the argument for (33).

PROOF OF THEOREM 6: For any regular domain Ω in M, we define the (Neumann) Cheeger constant $\mathfrak{h}_N(\Omega)$ by

$$\mathfrak{h}_N(\Omega) = \inf \frac{L(\Gamma)}{\min\{A(\Omega_1), A(\Omega_2)\}},$$

where Γ ranges over embedded 1-manifolds in Ω, for which $\partial\Gamma = \bar{\Gamma} \cap \partial\Omega$, which divide Ω into two open sets Ω_1, Ω_2 with $\Gamma = \partial\Omega_1 \cap \Omega = \partial\Omega_2 \cap \Omega$. As in Theorem IV.5, one can show that one obtains the same infimum if one restricts Γ so that the associated Ω_1, Ω_2 are connected; and, in what follows, we always assume Γ has this property. Also, one can easily derive the Cheeger inequality, as in Theorems IV.3 and IV.11,

(34) $$\mu(\Omega) \geq \tfrac{1}{4}\mathfrak{h}_N^2(\Omega),$$

where $\mu(\Omega)$ is the lowest nonzero Neumann eigenvalue of Ω.

Now let M be divided, by a compact 1-manifold C consisting of simple closed geodesics $\gamma_1, \ldots, \gamma_s$, into $k+1$ domains $M_1, \ldots, M_{k+1}$, $k \leq 2g-3$, such that

$$l_k(M) = \sum_{r=1}^{s} L(\gamma_r).$$

Recall that

$$s \leq 3g - 3.$$

The domain monotonicity of eigenvalues with vanishing Neumann eigenvalues then implies that

$$\lambda_k(M) \geq \min\{\mu(M_1), \ldots, \mu(M_{k+1})\}. \tag{35}$$

From (34), it will suffice to bound $\mathfrak{h}_N(M_j)$, $j = 1, \ldots, k + 1$, from below. Fix one of the M_j and call it Ω.

When considering any Γ in the competition that defines $\mathfrak{h}_N(\Omega)$, we have, for each component of Γ_0, two possibilities:

(i) the component Γ_0 is not closed, and it therefore intersects $\partial\Omega$ at both endpoints; or

(ii) the component Γ_0 is closed.

In case (i), our first comment is that both endpoints of Γ_0 must lie on the same component γ of $\partial\Omega$, in which case there are, here, two possibilities: either Γ_0 is contained in the collar in Ω associated to γ, or it is not completely contained in the collar.

If Γ_0 is contained completely in the collar, then Γ_0 and a geodesic segment of γ bound a domain G contained in the collar, the geodesic segment having length less than or equal to that of Γ_0, and G homeomorphic to a 2-disk. Then $\Gamma = \Gamma_0$, and since G is homeomorphic to a 2-disk we have, by the argument of (IV.15),

$$A(G) \leq L(\partial G) \leq 2L(\Gamma);$$

so

$$L(\Gamma)/\min\{A(G), A(\Omega\backslash\bar{G})\} \geq L(\Gamma)/A(G) \geq \tfrac{1}{2} \geq l_k(M)/2\beta(g).$$

If Γ_0 is not completely contained in the collar, then

$$L(\Gamma_0) \geq 2\iota_\gamma \geq 2 \operatorname{arcsinh}\{\operatorname{csch} \beta(g)/2\},$$

by (25), (28). So here we have

$$\begin{aligned} L(\Gamma)/\min\{A(\Omega_1), A(\Omega_2)\} &\geq L(\Gamma_0)/A(M) \\ &\geq \frac{\operatorname{arcsinh}\{\operatorname{csch} \beta(g)/2\}}{2\pi(g-1)\beta(g)}\, l_k(M). \end{aligned}$$

We now consider case (ii), namely, all components of Γ are closed. If a component Γ_0 of Γ bounds a domain G in Ω homeomorphic to a 2-disk, then $\Gamma = \Gamma_0$, and by the argument of (IV.15) we have

$$A(G) \leq L(\partial G), \tag{36}$$

and one argues as above. Similarly, if Γ_0 is freely homotopic to a component γ of $\partial\Omega$, or to another component Γ_1 of Γ, then Γ_0 and the other circle bound a domain G homeomorphic to an annulus. But here, too, inequality (36) is valid (cf. Ionin [1]; Yau [1]). When Γ_0 is freely homotopic to γ, we have

$$L(\partial G) = L(\Gamma_0) + L(\gamma) \le 2L(\Gamma_0) \le 2L(\Gamma);$$

and when Γ_0 is freely homotopic to Γ_1, we have

$$L(\partial G) = L(\Gamma_0) + L(\Gamma_1) \le L(\Gamma).$$

So, again, we obtain

$$L(\Gamma)/\min\{A(G), A(\Omega\backslash G)\} \ge l_k(M)/2\beta(g).$$

The last possibility, therefore, is that all the components of Γ are closed; none of them bound a two-disk; none of them are freely homotopic to a component of $\partial\Omega$, nor are any pair of them freely homotopic to each other. We give a lower bound to $L(\Gamma)$ depending only on g and $l_k(M)$. First we replace each component of Γ by the shortest loop in Ω freely homotopic to it. This loop must be a nontrivial simple closed geodesic. Since this replacement does not increase length, we may assume from the outset that Γ consists of a finite collection of simple closed geodesics. If

$$L(\Gamma) < (1/s)l_k(M),$$

then there exists a component, say γ_1, of C with $L(\Gamma) < L(\gamma_1)$. This would imply that by reattaching along γ_1, and cutting along Γ we would have a subdivision of M into $k + 1$ domains by disjoint simple closed geodesics with total length $< l_k(M)$—an impossibility. We therefore have

$$L(\Gamma) \ge (1/s)l_k(M) \ge (3g - 3)^{-1}l_k(M).$$

So

$$L(\Gamma)/\min\{A(\Omega_1), A(\Omega_2)\} \ge L(\Gamma)/A(M) \ge l_k(M)/12\pi(g - 1)^2.$$

We therefore have, by (34) and (35), the estimate (32), where

$$\alpha_g = \frac{1}{4}\left[\min\left\{\frac{1}{2\beta(g)}, \frac{\operatorname{arcsinh}\{\operatorname{csch}\beta(g)/2\}}{2\pi(g - 1)\beta(g)}\right\}, \frac{1}{12\pi(g - 1)^2}\right],$$

and the theorem is proved.

5. THE UPPER HALF-SPACE MODEL OF HYPERBOLIC SPACE

We first discuss the 2-dimensional case. We let

$$\mathbb{C}^+ = \{z = x + iy \in \mathbb{C} : y > 0\},$$

and consider the map $\mathbb{C}^+ \to D$ given by

$$w = (z - i)/(z + i)$$

—a fractional linear transformation of $\mathbb{C}$, taking $\mathbb{C}^+$ to D. To pull back the $\mathbb{H}^2$ metric on D to $\mathbb{C}^+$, we have

$$dw = 2i\, dz/(z + i)^2, \qquad |dw| = 2|dz|/|z + i|^2, \qquad 1 - |w|^2 = 4y/|w|^2;$$

therefore

$$ds = |dz|/y.$$

Again, the geodesics of $\mathbb{H}^2$, in this model, are generalized circle segments in $\mathbb{C}^+$ intersecting the boundary, that is, the x axis, orthogonally.

One can check, now, that the orientation preserving isometries of $\mathbb{H}^2$, in the $\mathbb{C}^+$ model, are given by PSL(2; $\mathbb{R}$), the fractional linear transformations

$$w = (\alpha z + \beta)/(\gamma z + \delta),$$

where α, β, γ, δ are real constants satisfying $\alpha\delta - \beta\gamma = 1$. Thus the group is given by the real matrices

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

(with composition of mappings represented by matrix multiplication) with determinant equal to 1. To achieve an isomorphism between the isometry group and the matrix group one must factor the matrix group by $\pm$ (identity matrix). To include orientation reversing isometries, one has the group generated by PSL(2; $\mathbb{R}$) and the reflection in the y axis.

Now let $z = \rho e^{i\theta}$ determine polar coordinates in $\mathbb{C}$, and consider the domain Ω in $\mathbb{C}^+$ given by

$$0 < T \leq \rho \leq T + R, \qquad 0 < \Psi \leq \theta \leq \pi/2.$$

To calculate the hyperbolic distance r from iT to $i(T + R)$, we have

$$r = \int_T^{T+R} dy/y = \ln(1 + R/T);$$

so

$$r = \ln(1 + R/T), \qquad R = T(e^r - 1).$$

Next, we note that the geodesic segments given by

$$\{z = Te^{i\theta} : \Psi \leq \theta \leq \pi/2\}$$

and

$$\{z = (T + R)e^{i\theta} : \Psi \leq \theta \leq \pi/2\}$$

have the same hyperbolic length, since the metric ds is invariant with respect to homotheties of $\mathbb{C}$, centered on the x axis and restricted to $\mathbb{C}^+$. The specific hyperbolic length t is given by

$$t = \int_{\pi/2}^{\Psi} \csc\theta \, d\theta = \operatorname{arctanh} \cos \Psi.$$

These calculations suggest that we introduce in $\mathbb{C}^+$ the coordinates

$$r = \ln(1 + \rho), \qquad t = \operatorname{arctanh} \cos \theta, \tag{37}$$

for then we have

$$ds^2 = |dz|^2/y^2 = (dr^2 + r^2\, d\theta^2)/r^2 \sin^2\theta = dt^2 + \cosh^2 t\, dr^2$$

that is,

$$ds^2 = dt^2 + \cosh^2 t\, dr^2.$$

Thus the coordinates r, t are Fermi coordinates in $\mathbb{H}^2$ based on the geodesic represented by the y axis in $\mathbb{C}^+$, where r denotes directed hyperbolic distance from i, along the y axis, and t denotes the directed hyperbolic distance along the geodesic emanating, orthogonally, from the y axis. Of course, the injectivity radius, here, is $+\infty$.

We now consider the quadrilateral Q in Fig. 11, and establish the formula

$$\cos \varphi = (\sinh r)(\sinh t). \tag{38}$$

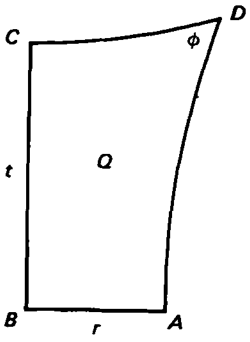


Fig. 11

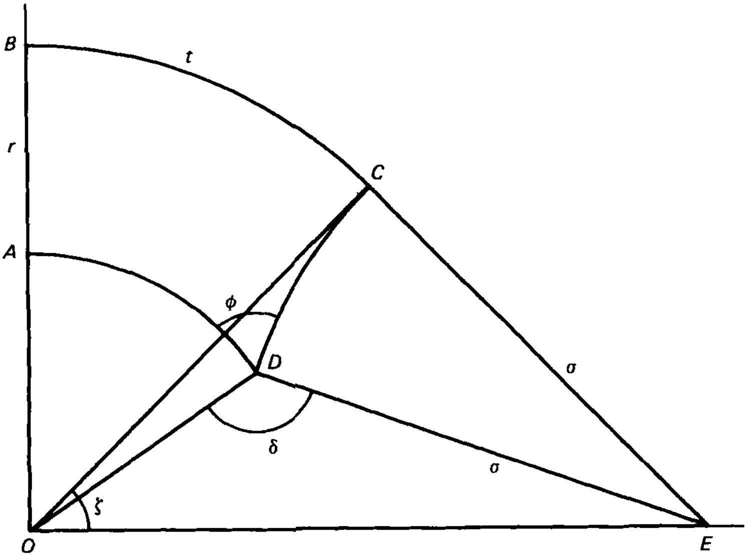


Fig. 12

To this end, we represent Q in $\mathbb{C}^+$ by Fig. 12, where E is the center of the circle corresponding to the geodesic CD; σ is the radius of CD;

$$\theta = \measuredangle EOC, \qquad \delta = \measuredangle ODE;$$

and R, T are the Euclidean lengths corresponding to the hyperbolic lengths r, t. We take $A = i$; so

$$\sigma = (1 + R)\tan\theta, \qquad E = (1 + R)\sec\theta.$$

By the law of cosines,

$$\cos\varphi = -\cos\delta = \frac{(1 + R)^2 - 1}{2(1 + R)}\operatorname{ctn}\theta,$$

and (38) follows from (37).

We now consider the upper half-space model of $\mathbb{H}^n$ for general $n \geq 2$. To map

$$\mathbb{R}^n_+ \equiv: \{x \in \mathbb{R}^n : x^n > 0\},$$

to $\mathbb{B}^n$, first let

(39) $$z = x + (\tfrac{1}{2} - 2x^n)e_n,$$

where e_n is the nth element of the standard basis of $\mathbb{R}^n$. Then $\{x : x^n > 0\}$ is mapped diffeomorphically onto $\{z : z^n < \frac{1}{2}\}$. Next, consider the inversion

$$y = e_n + (z - e_n)/|z - e_n|^2 \tag{40}$$

in the sphere of radius 1 centered at $z = e_n$. Then $\{z : z^n < \frac{1}{2}\}$ is mapped diffeomorphically onto $\{y : |y| < 1\}$. The composition of (39) followed by (40) therefore maps $\mathbb{R}^n_+$ diffeomorphically onto $\mathbb{B}^n$.

To determine the induced metric on $\mathbb{R}^n_+$ we have

$$dy = \frac{dz}{|z - e_n|^2} - \frac{2\langle dz, z - e_n\rangle}{|z - e_n|^4}(z - e_n),$$

$$|dy|^2 = \frac{|dz|^2}{|z - e_n|^4}, \qquad 1 - |y|^2 = \frac{(1 - 2z^n)}{|z - e_n|^2};$$

so

$$\frac{4|dy|^2}{(1 - |y|^2)^2} = \frac{4|dz|^2}{(1 - 2z^n)^2}.$$

From (39) we have on $\mathbb{R}^n_+$

$$ds^2 = |dx|^2/(x^n)^2. \tag{41}$$

We finally note that, in these coordinates, the Laplace operator is given (using (I.33)) by

$$\begin{aligned}\Delta &= (x^n)^n \sum_{j=1}^{n} \frac{\partial}{\partial x^j}\left((x^n)^{2-n}\frac{\partial}{\partial x^j}\right)\\ &= (x^n)^2 \sum_{j=1}^{n} \frac{\partial^2}{\partial x^j \partial x^j} + (2 - n)x^n \frac{\partial}{\partial x^n},\end{aligned}$$

that is,

$$\Delta = (x^n)^2 \sum_{j=1}^{n} \frac{\partial^2}{\partial x^j \partial x^j} + (2 - n)x^n \frac{\partial}{\partial x^n} \tag{42}$$

is our formula for Δ on $\mathbb{H}^n$ in the upper half-space model.

CHAPTER XI

The Selberg Trace Formula

Burton Randol

1. PRELIMINARIES

In this chapter we will discuss and apply some of the idea connected with the compact case of the Selberg trace formula (Selberg [1]), emphasizing the differential-geometric rather than the arithmetic side of the theory, and restricting our discussion to hyperbolic spaces. We begin with a very brief review of some hyperbolic geometry.

Recall that hyperbolic n-space $\mathbb{H}^n$ ($n \geq 2$) is a complete simply connected Riemannian manifold having constant sectional curvature equal to -1, and that for a given dimension, any two such spaces are isometric (Wolf [1]).

There are several models for $\mathbb{H}^n$, the most important being the half-space model, the ball model, and the hyperboloid or Lorentz model, with the ball model being especially useful for questions involving rotational symmetry. The half-space model consists of the open half-space of points $(x_1, \ldots, x_{n-1}, t)$ in $\mathbb{R}^n$ for which $t > 0$. The metric in this case is given by $(ds)^2/t^2$, where ds is the Euclidean distance element, and this metric is obviously conformal. The ball model consists of the open unit ball $|x| < 1$ ($x = (x_1, \ldots, x_n)$) in $\mathbb{R}^n$, and the metric for this model is given by $4(ds)^2/(1-|x|^2)^2$. We will not use the hyperboloid model in this chapter, but the algebraic structure of the isometry group is especially clear in this model, and it is useful, among other things, for constructing examples of compact hyperbolic space forms (cf. Borel [1]; Sullivan [1]).

We begin with a discussion of the half-space model. It is evident that if S is any $(n-1)$-dimensional hemisphere in $\mathbb{R}^n$ whose equatorial plane is the boundary $t = 0$, then inversion, or reflection, in S is an isometry of $\mathbb{H}^n$. This is obvious since the hyperbolic length element is given by ds/t, which is invariant under such reflections (recall that inversion in a sphere of radius r and center X_0 is given vectorially by $X \to r^2(X - X_0)/|X - X_0|^2 + X_0$). Clearly ordinary reflection in the intersection of the half-space with an

0-00-000000-0

$(n-1)$-dimensional plane perpendicular to the boundary is also an isometry. Note that by taking a product of two reflections in concentric hemispheres we can obtain any homothety about any point on the boundary, and by doing the same for two parallel half-planes, we can obtain any translation, from which it is easy to see that the group generated by such reflections, and hence the isometry group of $\mathbb{H}^n$, is transitive. We will henceforth call half-planes and hemispheres perpendicular to the boundary hyperplanes, or hyperbolic hyperplanes, if there is danger of confusion with Euclidean hyperplanes, and use the term *reflection* to mean reflection in such a hyperplane.

It is a simple exercise to see that the image of a hyperplane under reflection is again a hyperplane, and that any two hyperplanes can be transformed into each other by a finite sequence of such reflections. Additionally, any two generalized circles, that is, half-lines or semicircles perpendicular to the boundary can be transformed into each other by a finite sequence of reflections.

Suppose now that two points in the half-space model lie on a vertical line L perpendicular to the boundary. It is then evident, since the hyperbolic distance element is ds/t, that the unique shortest path between them is the segment of L connecting them, and it follows that the portion of any such line interior to the half-space is a geodesic. Since any two generalized circles can be brought into each other by an isometry, and since for any two points there is a generalized circle containing them, this shows that the geodesics in the half-space model are precisely the generalized circles, and that for any two points, there is a unique geodesic containing them.

Notice that a hyperplane has an intrinsic characterization as the locus of points equidistant from two points. This is clear if the hypersurface is a Euclidean half-plane perpendicular to the boundary, and since any two hyperplanes can be brought into one another by an isometry, it is true in general. Another intrinsic characterization which can be immediately verified is that a hyperplane consists of the union of geodesics perpendicular to and passing through a fixed point on a geodesic. Reflection in a hyperplane P can also be intrinsically described as follows: if a point $p \in \mathbb{H}^n$ lies on P it is not moved. Otherwise, there exists a unique geodesic g containing p and perpendicular to P. The image of p is then the point on the other side of P which lies on g and at a distance along g from P equal to that of p. Notice finally that a hyperplane in $\mathbb{H}^n$ with the induced metric is a totally geodesic subspace and a model for $\mathbb{H}^{n-1}$. This is clear from the metric if the hyperplane corresponds to a Euclidean hyperplane perpendicular to the boundary, and it then follows for all hyperplanes by the transitivity of isometries on the set of hyperplanes. Henceforth, we will think of hyperplanes as intrinsic objects in $\mathbb{H}^n$.

Passing now to the ball model, we note that the fractional linear transformation $z \to (z - i)/(z + i)$ of the complex plane maps the upper half-plane $\operatorname{Im} z > 0$ with the hyperbolic metric isometrically onto the unit disk $|z| < 1$ with the hyperbolic metric. Since fractional linear transformations are conformal on $\mathbb{C}^1$, it follows that the geodesics in the ball model for $\mathbb{H}^2$ are either open arcs of circles perpendicular to and having endpoints on the boundary $|z| = 1$, or open diameters. In dimension 2, the hyperplanes coincide with the geodesics. Additionally, in the ball model for $\mathbb{H}^n$ with n arbitrary, it is obvious from the metric that the intersection of the ball with a Euclidean $(n - 1)$-dimensional equatorial hyperplane is a locus of points equidistant from two points, and is therefore a hyperbolic hyperplane. Now it is clear that the intersection with the ball of any Euclidean k-plane through the origin is, with the induced metric, the ball model for $\mathbb{H}^k$. That such an intersection is totally geodesic can be seen by first noting that it is a hyperplane and hence totally geodesic in some ball model for $\mathbb{H}^{k+1}$, obtained by intersecting the $\mathbb{H}^n$-ball with an appropriate Euclidean $(k + 1)$-plane. Then working our way up until the process stops at $\mathbb{H}^n$, we eventually find that the intersection is the first stage in a tower of subspaces, each of which is totally geodesic in the next, and hence it is itself totally geodesic in the $\mathbb{H}^n$-ball. Since any two points in the $\mathbb{H}^n$-ball can be connected by a unique generalized circle lying in a Euclidean 2-plane through the origin, it follows that the geodesics in the ball model for $\mathbb{H}^n (n > 2)$ are given, as in the 2-dimensional case, by open diameters or by open arcs of circles perpendicular to and having endpoints on the boundary. Since any hyperplane is the union of the geodesics perpendicular to and intersecting some geodesic at a point, it follows that the hyperplanes in the ball model are either intersections with the ball of Euclidean equatorial planes, or intersections with Euclidean spheres perpendicular to the boundary.

Suppose T is an arbitrary isometry of the $\mathbb{H}^n$-ball. By the transitivity of the group of reflections we can find a reflection g such that $g \circ T(0) = 0$. Now it is clear, since the metric is radial, that the subgroup of the isometry group which stabilizes the origin is $O(n)$, and since $O(n)$ is generated by reflections in equatorial hyperplanes, it follows that $g \circ T$ is a product of reflections and so therefore is T; that is, the full isometry group of $\mathbb{H}^n$ is generated by reflections.

Suppose now that M is a compact n-dimensional hyperbolic manifold, that is, a compact Riemannian manifold having everywhere constant sectional curvature equal to -1 (for a discussion of such manifolds, which are numerous, see, for example, Borel [1], Ford [1], Sullivan [1], and Thurston [1]).

Now M can be regarded as the quotient of $\mathbb{H}^n$ by a discontinuous

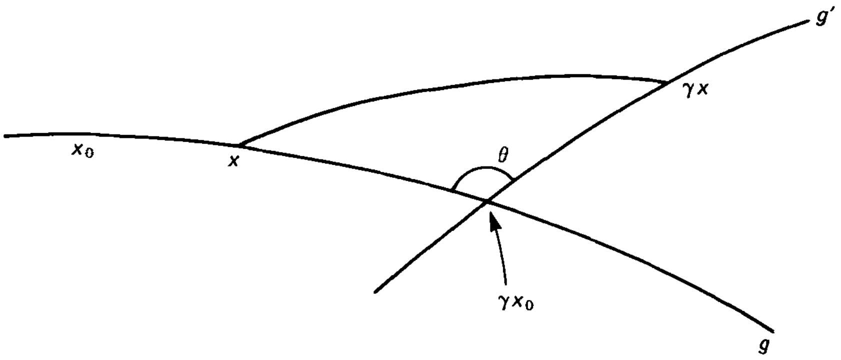


Fig. 13

compact group Γ of isometries, which are the covering transformations of M. Suppose $\gamma \neq e$ is a nontrivial element of Γ. Consider the function $d(x, \gamma x)$ on $\mathbb{H}^n$, where d denotes hyperbolic distance. Evidently it is bounded away from zero, since for $x \in \mathbb{H}^n$, any arc in $\mathbb{H}^n$ connecting x and γx projects to a closed homotopically nontrivial loop on M, and by compactness there is a smallest possible length associated with such a loop. Again using the compactness of the quotient manifold M, it is easy to see that there exists $x_0 \in \mathbb{H}^n$ such that $d(x_0, \gamma x_0)$ actually achieves the minimum. Let g be the geodesic passing through x_0 and γx_0. Then γ leaves g invariant, although it does not in general fix the points of g. In order to see that g is invariant, note that γ maps g onto some geodesic g', which clearly passes through the point γx_0. Suppose $g \neq g'$, that is, $\theta \neq \pi$ (cf. Fig. 13). Let x be a point on g between x_0 and γx_0. By the triangle inequality, it must be the case that the geodesic arc connecting x with γx is shorter than the sum of the lengths of the geodesic segments connecting x to γx_0, and γx_0 to γx. But this latter sum is clearly $d(x_0, \gamma x_0)$, which is the presumed minimum of the function $d(x, \gamma x)$. It follows that $\theta = \pi$, so g is invariant under γ. As we shall see, if $\gamma \neq e$ such a geodesic is unique, and is called the axis of γ. In order to obtain a canonical form for γ, which will in turn imply this last result, note to begin with that in the half-space model we can, by the transitivity of isometries on geodesics, conjugate γ into some γ' so that the geodesic I, which is the vertical half-line perpendicular to the boundary and emanating from the origin is invariant by γ'.

Now I, with the induced metric, is the half-line $0 < t < \infty$, with metric $(dt)^2/t^2$, and it follows from calculus that every isometry of I is either of the form S or TS, where S is a dilation $t \to \lambda t$ $(\lambda > 0)$, and T is inversion about the point $t = 1$; that is, T is the transformation $t \to 1/t$. For later convenience, we will denote by $\mathbb{H}^1$ the Riemannian 1-manifold of which I is a model. Now since every transformation of the form TS has a fixed point on I, namely, $t = \lambda^{-1/2}$, and since $d(x, \gamma' x)$ is bounded away from zero (since γ' is conjugate to γ), we conclude that the transformation γ' is a pure dilation on I, which can be regarded as the restriction of a dilation $S: (x_1, \ldots, x_{n-1}, t) \to (\lambda x_1, \ldots, \lambda x_{n-1}, \lambda t)$ on $\mathbb{H}^n$. It follows that $S^{-1}\gamma'$ is the identity on I. Passing to the ball model, and letting I correspond to a diameter, we see that $S^{-1}\gamma'$ in this model is an element of $O(n)$ fixing some 1-dimensional subspace; that is, $S^{-1}\gamma'$ has a 1-dimensional eigenspace with eigenvalue 1. It follows that $S^{-1}\gamma'$ can be identified in an obvious way with an element of $O(n-1)$. Returning to the half-space model, we summarize our conclusions.

If $M = \Gamma\backslash\mathbb{H}^n$ is a compact hyperbolic manifold, then every $\gamma \in \Gamma$ other than the identity is conjugate in the full isometry group of $\mathbb{H}^n$ to a transformation of the form KS, where S is a dilation $(x_1, \ldots, x_{n-1}, t) \to (\lambda x_1, \ldots, \lambda x_{n-1}, \lambda t)$, and K is an element of $O(n-1)$ acting on $(x_1, \ldots, x_{n-1})$. Moreover, it is clear from this canonical form that for every such γ, there is exactly one geodesic I_γ, such that $\gamma(I_\gamma) = I_\gamma$. This geodesic is called the *axis* of γ.

Suppose now that γ_1 and γ_2 are two nontrivial elements of Γ which commute. Then $\gamma_2(I_{\gamma_1}) = \gamma_2\gamma_1(I_{\gamma_1}) = \gamma_1\gamma_2(I_{\gamma_1})$; that is, the geodesic $\gamma_2(I_{\gamma_1})$ is invariant under γ_1. Since the axis of γ_1 is unique, this shows that $\gamma_2(I_{\gamma_1}) = I_{\gamma_1}$, so I_{γ_1} is the axis of both γ_1 and γ_2. In other words, any two nontrivial commuting transformations in Γ are coaxial.

Next suppose that Γ' is a nontrivial abelian subgroup of Γ. By the previous argument, all the nontrivial transformations in Γ' have the same axis, which we will call A. Since Γ is discontinuous, so is Γ', which can be regarded as acting discontinuously and in a fixed-point free manner on A, which is a model for $\mathbb{H}^1$. It follows that Γ' can be thought of as a group of dilations on $\mathbb{H}^1$, and it is well known that such a group, if it is discontinuous, is infinite cyclic. Let $\gamma_0 \in \Gamma'$ generate the transformations of Γ' on A. Then it is easy to see that γ_0 generates the transformations of Γ' on $\mathbb{H}^n$ as well, since otherwise there would be a $\gamma_1 \in \Gamma'$ such that $\gamma_0^k\gamma_1$ is the identity on A but not on $\mathbb{H}^n$, for some integer k. Since no element of Γ' can have a fixed point, this is impossible. Summing up, we have shown that any nontrivial abelian subgroup of Γ must be infinite cyclic. Note also, for later use, that the above arguments show that the centralizer of any nontrivial element of Γ must be infinite cyclic (cf. Preissmann [1]).

We next briefly describe the so-called Dirichlet fundamental domain D for Γ, taken about a point x_0 in $\mathbb{H}^n$.

Suppose x_0 is a point in $\mathbb{H}^n$. Define D to be the set of x's in $\mathbb{H}^n$ which satisfy $d(x, x_0) \le d(x, \gamma x_0)$ for any $\gamma \in \Gamma$. Equivalently, D can be constructed as follows. For a nontrivial $\gamma \in \Gamma$, consider the geodesic segment g_γ connecting x_0 to γx_0. Let P_γ be the hyperplane which is the perpendicular bisector of g_γ, and let H_γ be the closed half-space consisting of the points on the x_0 side of p_γ, together with P_γ itself. Then D is the intersection of all the H_γ's, and it is not difficult to see that it is in fact the intersection of finitely many of them and is compact; that is, D is a compact convex polyhedron in $\mathbb{H}^n$. Moreover, it is easy to see that the polyhedra of the form $\gamma(D)$ $(\gamma \in \Gamma)$ tesselate $\mathbb{H}^n$; that is, $\mathbb{H}^n$ is covered by the union of the $\gamma(D)$'s, and any two of them either have no points in common, or intersect only at their boundaries, and in such a way that vertices always intersect vertices, and the intersection is the convex span of the common vertices.

2. THE PRETRACE FORMULA

We begin by recalling a few facts from previous chapters.

(a) In the upper half-space model for $\mathbb{H}^n$, the Laplacian is given by

$$\Delta = t^2(\partial^2/\partial x_1^2 + \cdots + \partial^2/\partial x_{n-1}^2 + \partial^2/\partial t^2) - (n-2)t\,\partial/\partial t.$$

(b) The area $A(r)$ of a sphere in $\mathbb{H}^n$ of hyperbolic radius r is $\mathrm{c}_{n-1}\sinh^{n-1}r$, where c_{n-1} is the area of the unit sphere in $\mathbb{R}^n$, so the Laplacian in geodesic polar coordinates about a point p in $\mathbb{H}^n$ is given by

$$\Delta = \partial^2/\partial r^2 + (n-1)\coth r\,\partial/\partial r + \Delta_{S(p;r)},$$

where $\Delta_{S(p;r)}$ is the Laplacian on the geodesic sphere of radius r about p.

Suppose now that $k(p)$ $(-\infty < p < \infty)$ is an even C^∞ function having compact support. By setting $k(x,y) = k(d(x,y))$, where $(x,y) \in \mathbb{H}^n \times \mathbb{H}^n$, we obtain a smooth function of x and y which depends only on the distance between x and y. We will call such a function a point-pair invariant.

LEMMA 1. Suppose $k(x,y)$ is a point-pair invariant. Then $\Delta_x k(x,y) = \Delta_y k(x,y)$, where the subscript indicates the variable in which the Laplacian is applied.

PROOF: Let us look for a moment at $\Delta_x k(x,y)$. In geodesic polar coordinates about y, $k(x,y) = k(r)$, since $k(x,y)$ is radial about y, so $\Delta_x k(x,y) = k''(r) + (n-1)\coth r\,k'(r)$, since $\Delta_{S(y;r)}k(r) = 0$. That is, $\Delta_x k(x,y) = k''(d(x,y)) + (n-1)\coth(d(x,y))k'(d(x,y))$.To examine $\Delta_y k(x,y)$

we take geodesic polar coordinates about x, repeat the argument, and find that $\Delta_y k(x,y)$ is given by the same formula. q.e.d.

Suppose now that f is a smooth function on $\mathbb{H}^n$. Then the integral

$$\int_{\mathbb{H}^n} k(x,y) f(y)\, dy$$

makes sense, and we claim that the Laplacian commutes with this operation of f, that is,

$$\Delta_x \int_{\mathbb{H}^n} k(x,y) f(y)\, dy = \int_{\mathbb{H}^n} k(x,y)(\Delta_y f(y))\, dy,$$

or what is the same thing,

$$\int_{\mathbb{H}^n} (\Delta_x k(x,y)) f(y)\, dy = \int_{\mathbb{H}^n} k(x,y)(\Delta_y f(y))\, dy.$$

To see this, recall that $(\Delta F, G) = (F, \Delta G)$ if either F or G is compactly supported, so

$$\int_{\mathbb{H}^n} k(x,y)(\Delta_y f(y))\, dy = \int_{\mathbb{H}^n} (\Delta_y k(x,y)) f(y)\, dy,$$

and by the last lemma, the integral on the right equals

$$\int_{\mathbb{H}^n} (\Delta_x k(x,y)) f(y)\, dy,$$

which proves the assertion.

We next examine radial eigenfunctions of the Laplacian. Suppose φ satisfies $\Delta\varphi = \lambda\varphi$ for some λ, and that φ is radial about a point $x \in \mathbb{H}^n$. Introducing geodesic polar coordinates about x, we see that

$$\varphi''(r) + (n-1)\coth r\, \varphi'(r) - \lambda\varphi(r) = 0, \tag{1}$$

which is a second-order ordinary differential equation for φ. This equation has a singularity of the first kind at $r = 0$, and it follows from the standard theory of such equations that for any λ, real or complex, the general solution is of the form $c_1 f_1(r) + c_2 f_2(r)$, where $f_1(r)$ is an entire function of r with $f_1(0) = 1$, and $f_2(r)$ is defined for all $r > 0$, with a singularity at $r = 0$ of logarithmic type if $n = 2$, and of type $r^{-(n-2)}$ if $n > 2$ (Olver [1], pp. 148–153]). In particular, for any fixed λ, real or complex, and any point $x \in \mathbb{H}^n$, there is exactly one function $\omega_\lambda(x,y)$, which is radial about x, with $\omega_\lambda(x,x) = 1$, and which satisfies $\Delta_y \omega_\lambda(x,y) = \lambda\omega_\lambda(x,y)$. Since $\omega_\lambda(x,y)$ is a function only of $d(x,y)$, it follows that it also satisfies $\Delta_x \omega_\lambda(x,y) = \lambda\omega_\lambda(x,y)$.

Suppose now that f is a function on $\mathbb{H}^n$. We define its radialization about a point $x \in \mathbb{H}^n$, written $f_x^\#(y)$, by setting $f_x^\#(y) = \int_{S_x} f(Ty)\,dT$, where S_x is the subgroup of the isometry group of $\mathbb{H}^n$ which stabilizes x, and dT is normalized Haar measure on S_x. As we have already seen, the S_x's are all conjugate and can be identified with the orthogonal group $O(n)$.

It is clear that $f_x^\#(y)$ is a radial function of y, with $f_x^\#(x) = f(x)$.

Suppose now that $k(x, y)$ is a point-pair invariant. We claim that

$$\int_{\mathbb{H}^n} k(x,y) f(y)\,dy = \int_{\mathbb{H}^n} k(x,y) f_x^\#(y)\,dy.$$

To see this, note that

$$\begin{aligned}\int_{\mathbb{H}^n} k(x,y) f_x^\#(y)\,dy &= \int_{\mathbb{H}^n} k(x,y)\,dy \int_{S_x} f(Ty)\,dT\\ &= \int_{S_x} dT \int_{\mathbb{H}^n} k(x,y) f(Ty)\,dy\\ &= \int_{S_x} dT \int_{\mathbb{H}^n} k(x, T^{-1}y) f(y)\,dy\\ &= \int_{S_x} dT \int_{\mathbb{H}^n} k(x,y) f(y)\,dy\\ &= \int_{\mathbb{H}^n} k(x,y) f(y)\,dy,\end{aligned}$$

which proves the assertion.

THEOREM 1. Suppose φ is an eigenfunction of the Laplacian, with $\Delta\varphi = \lambda\varphi$. Then φ is also an eigenfunction for the integral operator corresponding to any point-pair invariant, and its eigenvalue in the latter context depends only on λ, and not on φ; that is, there exists a function h, defined on the set of all possible eigenvalues, which we have seen in this case is all of $\mathbb{C}^1$, such that

$$\int_{\mathbb{H}^n} k(x,y)\varphi(y)\,dy = h(\lambda)\varphi(x).$$

The fact that h does not depend on φ, while easy to establish, is of crucial importance for the development of the theory, because it shows that h can be computed by using a simple set of representative eigenfunctions, which makes unnecessary computations involving φ's about which we may a priori know very little.

PROOF OF THEOREM 1. We begin by noting that

$$\int_{\mathbb{H}^n} k(x,y)\varphi(y)\,dy = \int_{\mathbb{H}^n} k(x,y)\varphi_x^\#(y)\,dy = \varphi(x)\int_{\mathbb{H}^n} k(x,y)\omega_\lambda(x,y)\,dy.$$

Now it is easy to see that for a given λ, the quantity $\int_{\mathbb{H}^n} k(x,y)\omega_\lambda(x,y)\,dy$ does not depend on x. In more detail, bearing in mind that the isometry group of $\mathbb{H}^n$ is transitive, note that for any isometry T,

$$\begin{aligned}\int_{\mathbb{H}^n} k(Tx,y)\omega_\lambda(Tx,y)\,dy &= \int_{\mathbb{H}^n} k(x,T^{-1}y)\omega_\lambda(x,T^{-1}y)\,dy,\\ &= \int_{\mathbb{H}^n} k(x,y)\omega_\lambda(x,y)\,dy.\end{aligned}$$

It follows that if we set $h(\lambda) = \int_{\mathbb{H}^n} k(x,y)\omega_\lambda(x,y)\,dy$, the theorem is proved.

Let us now compute the function h.

For this purpose it is helpful to have a simple collection of representative eigenfunctions, and in the half-space model, which is convenient for this purpose, such a collection is provided by functions of the form $t^s = e^{s\log t}$, where s is complex and we take the real branch of the logarithm. It follows from the expression for the Laplacian in the half-space model that

$$\Delta t^s = s(s+1-n)t^s. \tag{2}$$

Since $s(s+1-n)$ takes on all complex values as s ranges over the complex plane, we can regard h as a function of a complex parameter r, related to s by the equation $s = ((n-1)/2) + ir$. This will be computationally very convenient. Note that in terms of r, (2) becomes

$$\Delta t^{((n-1)/2)+ir} = -(((n-1)/2)^2 + r^2)t^{((n-1)/2)+ir}.$$

It will also be computationally convenient to replace $k(x,y) = k(d(x,y))$ by an arbitrary function of the form $L(Q(x,y))$, where $L(\rho)$ $(0 \le \rho < \infty)$ is a C^∞ function with compact support and $Q(x,y) = |x-y|^2/t\sigma$, where $|x-y|$ is the Euclidean distance between x and y, with $x = (x_1,\ldots,x_{n-1},t)$ and $y = (y_1,\ldots,y_{n-1},\sigma)$. Note that $Q(x,y)$ is a C^∞ strictly increasing function of $d(x,y)$. Indeed, the relationship between d and Q is given by $Q = (2\sinh d/2)^2$, so any L is automatically an even C^∞ function of hyperbolic distance.

We will now determine $h(r)$ as a function of $L(x,y)$.

By Theorem 1,

$$\int_{\mathbb{H}^n} L(x,y)\sigma^{((n-1)/2)+ir}\,dy = h(r)t^{((n-1)/2)+ir}. \tag{3}$$

If we take $x = (0, \ldots, 0, 1)$, and note that by the conformality of the metric, $dy = \sigma^{-n}\, dy_1 \cdots dy_{n-1}\, d\sigma$, the left-hand side of (3) becomes

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dy_1 \cdots dy_{n-1} \int_0^{\infty} L((y_1^2 + \cdots + y_{n-1}^2 + (1-\sigma)^2)/\sigma)\sigma^{(1-n)/2}\sigma^{ir}\, d\sigma/\sigma$$

After the transformation $\sigma^{-1/2}y_j \to y_j$, this becomes

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dy_1 \cdots dy_{n-1} \int_0^{\infty} L(y_1^2 + \cdots + y_{n-1}^2 + (\sigma^{1/2} - \sigma^{-1/2})^2)\sigma^{ir}\, d\sigma/\sigma,$$

or

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dy_1 \cdots dy_{n-1} \int_{-\infty}^{\infty} L(y_1^2 + \cdots + y_{n-1}^2 + (2\sinh u/2)^2)e^{iru}\, du,$$

if we set $e^u = \sigma$.

Taking polar coordinates $\rho = (y_1^2 + \cdots + y_{n-1}^2)^{1/2}$, $\theta = \rho^{-1}(y_1, \ldots, y_{n-1})$, and changing the order of integration, this becomes

$$\int_{-\infty}^{\infty} e^{iru}\, du \int_{\mathbb{S}^{n-2}} d\theta \int_0^{\infty} L(\rho^2 + (2\sinh u/2)^2)\rho^{n-2}\, d\rho,$$

where $d\theta$ is the area element on $\mathbb{S}^{n-2}$. Since the integrand is independent of θ and even in u, we find, after the transformation $\rho^2 + (2\sinh u/2)^2 \to \rho$, that

$$h(r) = \mathbf{c}_{n-2} \int_0^{\infty} \cos ru\, du \int_{z(u)}^{\infty} L(\rho)(\rho - z(u))^{(n-3)/2}\, d\rho, \tag{4}$$

where $z(u) = (2\sinh u/2)^2$, and for $n = 2$ we define $\mathbf{c}_0 = 2$. This will be our expression for h in terms of L. Notice that if n is even, the inner integral is a fractional integral evaluated at $z(u)$, and if n is odd it is an ordinary iterated integral evaluated at $z(u)$. In particular, if $n = 3$, it is simply $-F(z(u))$, where F is the antiderivative of L which vanishes at infinity.

Let us next see how to evaluate h in terms of an arbitrary $k(x, y) = k(d(x, y))$.

If we think of k and L as functions of a single variable ρ $(0 \le \rho < \infty)$, then it is evident that the L which corresponds to a given k is obtained by setting $L(\rho) = k \circ \delta(\rho)$, where $\delta(\rho) = z^{-1}(\rho)$. Substituting $k \circ \delta$ for L in (4), and noting that $z'(\rho) = 2\sinh \rho$, we then find that $h(r)$ is given by

$$2\mathbf{c}_{n-2} \int_0^{\infty} \cos ru\, du \int_u^{\infty} k(\rho)(z(\rho) - z(u))^{(n-3)/2} \sinh \rho\, d\rho, \tag{5}$$

which is our expression for h in terms of k. Note that if $n = 3$, this becomes

$$h(r) = 4\pi \int_0^{\infty} \cos ru\, du \int_u^{\infty} k(\rho) \sinh \rho\, d\rho. \tag{6}$$

Suppose now that M is a compact hyperbolic manifold. Then M can be regarded as a quotient of $\mathbb{H}^n$ by a fixed-point free discontinuous group Γ of isometries, and we have seen in the last section that Γ is torsion-free as well, since any nontrivial cyclic subgroup of Γ must be infinite. Suppose f is an L^2 function on M. Then f can be regarded as a function on $\mathbb{H}^n$ which is L^2 on compact sets, and automorphic with respect to Γ, that is, $f(\gamma x) = f(x)$. Conversely any such function can be regarded as an L^2 function on M. Now it is very natural to ask how the operator corresponding to a point-pair invariant acts on such an f. Accordingly, suppose k is a point-pair invariant corresponding to an even C^∞ compactly supported function on $(-\infty, \infty)$. For convenience, we will suppose that k is real, although this requirement can be easily dropped. Consider the operator

$$f \to \int_{\mathbb{H}^n} k(x,y) f(y)\, dy \tag{7}$$

on the Hilbert space of functions on $\mathbb{H}^n$ which are L^2 on compact sets and automorphic with respect to Γ, where the Hilbert space inner product is

$$(f,g) = \int_M f\bar{g}$$

(remember that Γ-automorphic function which are L^2 on compact subsets of $\mathbb{H}^n$ can be regarded as L^2 functions on M).

Suppose now that F is a fundamental domain for Γ. Then the images of F under Γ tesselate $\mathbb{H}^n$, so we can write

$$\begin{aligned}\int_{\mathbb{H}^n} k(x,y) f(y)\, dy = \sum_{\gamma\in\Gamma} \int_{\gamma F} k(x,y) f(y)\, dy &= \sum_{\gamma\in\Gamma} \int_F k(x,\gamma y) f(\gamma y)\, dy \\ &= \int_F \left(\sum_{\gamma\in\Gamma} k(x,\gamma y) \right) f(y)\, dy.\end{aligned}$$

Consider the function $K(x,y) = \sum_\gamma k(x, \gamma y)$ (since k is compactly supported there are only finitely many nonzero terms in the sum for any given x and y).

It is easily checked that $K(x,y)$ is biautomorphic, that is, $K(\gamma_1 x, \gamma_2 y) = K(x,y)$ for any $\gamma_1, \gamma_2 \in \Gamma$. In addition, $K(x,y)$ is C^∞ and $K(x,y) = \bar{K}(y,x)$ (since $k(x,y)$ is a real point-pair invariant). It follows from this that $K(x,y)$ can be regarded as the kernel of a Hilbert–Schmidt operator on $L^2(M)$, given by

$$f \to \int_M K(x,y) f(y)\, dy. \tag{8}$$

On the other hand, the action of K on a function $f \in L^2(M)$ can be computed by (7), if we lift f to $\mathbb{H}^n$.

Suppose now that $\varphi_0, \varphi_1, \varphi_2, \ldots$ is a complete orthonormal sequence of eigenfunctions, real or complex, of the Laplacian on M, corresponding to eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$. (Following previous convention, we write $\Delta\varphi_n + \lambda_n\varphi_n = 0$.)

If we regard the φ_j's as automorphic eigenfunctions of the Laplacian on $\mathbb{H}^n$, we have already seen that they satisfy

$$\int_{\mathbb{H}^n} k(x, y)\varphi_j(y)\, dy = h(r_j)\varphi_j(x),$$

where, following our convention about the sign of λ_j, r_j is either of the two roots of $((n-1)/2)^2 + r_j^2 = \lambda_j$.

In terms of $K(x, y)$, this says that the φ_n's satisfy

$$\int_M K(x, y)\,\varphi_n(y)\, dy = h(r_n)\,\varphi_n(x).$$

That is, the φ_n's are eigenfunctions of the Hilbert–Schmidt operator corresponding to the kernel $K(x, y)$, and since they are complete, they constitute a complete orthonormal set of eigenfunctions for this operator. It therefore follows from standard Hilbert–Schmidt theory (Riesz–Nagy [1]) that

$$2K(x, y) = \sum_n h(r_n)\,\varphi_n(x)\overline{\varphi}_n(y),$$

where the convergence is in $L^2(M \times M)$ and we count both r_n's corresponding to a given λ_n. (If $\lambda_n = ((n-1)/2)^2$, we count $r_n = 0$ twice.)

In other words,

$$\sum_\gamma k(x, \gamma y) = \tfrac{1}{2}\sum_n h(r_n)\,\varphi_n(x)\overline{\varphi}_n(y). \tag{9}$$

What can we say about the validity of (9) in the sense of pointwise convergence? From Theorem IV.14 we have that in two dimensions, $\|\varphi_n\|_\infty = O(\lambda_n) = O(r_n^2)$, and in dimension $m \geq 3$, $\|\varphi_n\|_\infty = O(\lambda_n^{m/4}) = O(r_n^{m/2})$ (cf. also deRham [1]). Since k comes from an even C^∞ function with compact support, it is easily checked that h is entire and that $h(r_n) = O(r_n^{-j})$ for any $j \geq 0$, so it follows that the right-hand side of (9) must converge uniformly and absolutely to the left-hand side, since the equality is already true in $L^2(M \times M)$. This fact can also be derived from Mercer's theorem (Riesz–Nagy [1, p. 245]).

In the event k is complex, the truth of (9) follows, in both the L^2 and pointwise sense, by splitting k into its real and imaginary parts. We will call the identity (9), in either its L^2 or pointwise form, the *pretrace formula*. The function h is computed from (5). Equivalently, if we start with a function L of $Q(x, y)$, the pretrace formula is

$$\sum_{\gamma} L(x, \gamma y) = \tfrac{1}{2} \sum_{n} h(r_n)\varphi_n(x)\overline{\varphi}_n(y), \tag{10}$$

where h is given by (4).

A natural question to ask in connection with the pretrace formula is: What are conditions on k, less restrictive than compact support, which will guarantee the L^2 and/or uniform pointwise truth of (9)? We will not enter into a detailed discussion of this question, but remark that the hypothesis of compact support can be considerably weakened. In general, if we have a $k(x, y)$ for which we want to establish, say, the pointwise version of (9), we define h by (5), verify that the right-hand side of (9) is absolutely and uniformly convergent with this h, and then approximate k by a sequence of smooth compactly supported point-pair invariants, for which the sum on the left-hand side in (9) tends to the corresponding sum for k. We next try to prove, for example, by dominated convergence, that the associated sequence of the right-hand sides of (9) converges to the right-hand side corresponding to k. It should be pointed out in this context that $k(\rho)$ should be $O(e^{-c\rho})$, for some $c > n - 1$, in order to guarantee the absolute convergence of the left-hand side of (9), since the volume of the ball of radius r in $\mathbb{H}^n$ grows like $e^{(n-1)r}$, which implies that for a given y, the number of lattice points γy which are within distance r of a given x grows like $e^{(n-1)r}$. We leave the following as exercises for the reader.

(a) If k is given by $k(\rho) = e^{-c\rho^2}$, for $c > 0$, then (9) is valid, with absolute and uniform convergence.

(b) If

$$\begin{aligned} k(\rho) &= 1 \quad \text{for} \quad 0 \le \rho \le r, \\ &= 0 \quad \text{for} \quad \rho > r, \end{aligned}$$

then (9) is valid in the L^2 sense.

3. APPLICATIONS OF THE PRETRACE FORMULA

We will sketch three applications of the pretrace formula. The first, due to Wolfe [1], is a hyperbolic version of Kendall's celebrated result on the number of Euclidean lattice points in a random oval (Kendall [1]). The second, due to the author (Randol [9]), deals with ergodic properties of the projection to a quotient manifold of dilating spheres in $\mathbb{H}^n$. The third, due to the author, describes analogs of classical electrostatics in hyperbolic manifolds. Although the discussion of each of these examples can be carried out in arbitrary dimensions without any essential difficulty, we will in all cases work in $\mathbb{H}^3$, because of the exceptional simplicity of formula (6), which links $k(x, y)$ to $h(r)$ in that case.

A. Lattice-Point Estimates[1]

Suppose M is a compact 3-dimensional hyperbolic manifold. As we have seen, we can regard M as a quotient of $\mathbb{H}^3$ by a discontinuous group Γ of isometries. For a given pair of points (x, y) in $\mathbb{H}^3 \times \mathbb{H}^3$, let $N_T(x, y)$ be the number of lattice-points of the form γy ($\gamma \in \Gamma$) which lie on the closed ball of radius T about x. Then if, for arbitrary x and y we define $k_T(x, y)$ to be 1 if $d(x, y) \leq T$ and 0 otherwise, it is clear that $N_T(x, y) = \sum_\gamma k_T(x, \gamma y)$, and that $N_T(x, y)$ is biautomorphic; that is, for any $\gamma, \gamma_2 \in \Gamma$, $N_T(\gamma_1 x, \gamma_2 y) = N_T(x, y)$, so $N_T(x, y)$ can be regarded as a function on $M \times M$. By the exercise at the end of the last section, the L^2 pretrace formula is valid for $k_T(x, y)$, so we conclude that in the sense of L^2 equivalence

$$N_T(x, y) = \tfrac{1}{2} \sum_n h_T(r_n) \varphi_n(x) \overline{\varphi}_n(y), \tag{11}$$

where the subscript on h indicates its dependence on the parameter T. Suppose now we regard $N_T(x, y)$ as a random variable on $M \times M$, where $M \times M$ is equipped with the product of normalized hyperbolic measure on each of the factors. Then it follows from (11) that the mean of $N_T(x, y)$ is $A^{-1} h_T(i)$, where A is the volume of M. (Recall that in three dimensions the two r_n's which correspond to $\lambda_0 = 0$ are i and $-i$, and that $\varphi_0 \equiv A^{-1/2}$.) Again by (11), bearing in mind that the functions $\varphi_n(x)\overline{\varphi}_n(y)$ are orthonormal on $M \times M$ with respect to the product of ordinary hyperbolic measure on each factor, we find that the variance of $N_T(x, y)$ is given by

$$(2A)^{-2} \sum_n{}' |h_T(r_n)|^2, \tag{12}$$

where the prime means that the values $r_n = \pm i$ are omitted from the sum.

---

[1] Compare Günther [1], Patterson [1], and Wolfe [1].

We wish to estimate the variance as a function of T, and for this we need to compute $h_T(r)$. By (6), $h_T(r)$ will be given in this case by

$$\begin{aligned} &4\pi \int_0^\infty \cos ru\,du \int_u^\infty k(\rho)\sinh\rho\,d\rho \\ &\quad = 4\pi \int_0^T (\cosh T - \cosh u)\cos ru\,du \\ (13)\quad &\quad = (4\pi/r(1+r^2))(\cosh T \sin rT - r\sinh T\cos rT) \qquad (r \neq 0, \pm i) \\ &\quad = (4\pi)(T\cosh T - \sinh T) \qquad (r = 0) \\ &\quad = \pi\sinh 2T - 2\pi T \qquad (r = \pm i). \end{aligned}$$

Note that the value of $h_T(\pm i)$ coincides with

$$V(T) = 4\pi \int_0^T \sinh^2 u\,du,$$

which is the volume of the ball of radius T in hyperbolic 3-space, and that

$$h_T(r) = O(e^{(1+|r|)T})$$

if r is imaginary and $|r| < 1$, and

$$h_T(r) = O(r^{-3}e(T))$$

if r is real, where $e(T) = Te^T$ if $r_n = 0$ is present, and equals e^T otherwise. Note also that by Weyl's asymptotic formula for the eigenvalues, $\sum_{n\neq 0} |r_n|^{-(3+\varepsilon)}$ is convergent, for any $\varepsilon > 0$.

Combining these facts, we are led to the following conclusions:

(a) The expected value of $N_T(x, y)$ is $A^{-1}V(T)$.

(b) The variance of $N_T(x, y)$ is given by

$$(2A)^{-2}\sum_n{}' |h_T(r_n)|^2, \tag{14}$$

where $h_T(r)$ is given by (13). In more detail, each imaginary r_n which is present in (14) contributes a term which is asymptotic to

$$4A^{-2}\pi^2|r_n(1 + r_n^2)|^{-2}e^{2(1+|r_n|)T},$$

while the infinitely many terms corresponding to real r_n's in (14) are in their aggregate $O((e(T))^2)$. In particular, the standard deviation is asymptotic to

$$2A^{-1}\pi(\alpha(1 + \alpha^2))^{-1}e^{(1+\alpha)T},$$

where α is the largest absolute value in the interval $(0, 1)$ of an imaginary r_n. If no such r_n is present, the standard deviation is $O(e(T))$.

Remark 1: Notice the special role played by the imaginary r_n's with $|r_n| < 1$ in this result, or what amounts to the same thing, the λ_n's satisfying $0 < \lambda_n < 1$. In general, the eigenvalues (if any) of a compact n-dimensional hyperbolic manifold which satisfy $0 < \lambda_n < ((n-1)/2)^2$ are called small eigenvalues of the manifold (cf. Randol [1]), and as we shall see, they play a prominent and striking role in many aspects of the theory.

Remark 2: The pretrace formula can also be used to derive lattice-point estimates for fixed values of x and y (cf. Günther [1]; Patterson [1]).

B. Ergodic Properties of Projections of Dilating Spheres

As before, we assume for simplicity that $n = 3$. Keeping the previous notation, define, for fixed $x \in \mathbb{H}^3$, $d\mu_T(x, y)$ to be normalized hyperbolic area measure on the sphere of radius T about x, and for fixed $x \in M$, denote by $dm_T(x, y)$ $(y \in M)$ the projection of $d\mu_T(x, y)$ onto M, that is, the dm_T measure of a set in M is the measure with respect to $d\mu_T$ of its total preimage under the projection from $\mathbb{H}^3$ to M. Then

$$dm_T(x, y) = \sum_{\gamma} d\mu_T(x, \gamma y), \tag{15}$$

and the right-hand side of (15) can be analyzed using the techniques we have developed. By approximating the measure $d\mu_T(x, y)$ by a family of smooth compactly supported functions, we find that the effect of $dm_T(x, y)$ on a C^∞ function f on M is given by the formula

$$\int_M f(y)\, dm_T(x, y) = \tfrac{1}{2} \sum_n c_n h_T(r_n) \varphi_n(x), \tag{16}$$

where $c_n = \int_M f(x)\overline{\varphi}_n(x)\, dx$ is the nth Fourier coefficient of f, and $h_T(r)$ is computed in accordance with (6), with the function $k(\rho)$ replaced by $(4\pi \sinh^2 T)^{-1}$ times a δ-function concentrated at $\rho = T$. Carrying out this computation, we find that

$$h_T(r) = (\sin rT)(r \sinh T)^{-1}.$$

To estimate the c_n's, note that

$$\begin{aligned} |c_n| = |(f, \varphi_n)| &= \lambda_n^{-i} |(f, \Delta^i \varphi_n)| \qquad \text{for any} \quad i \geq 1 \\ &= \lambda_n^{-i} |(\Delta^i f, \varphi_n)| \\ &\leq \lambda_n^{-i} \|\Delta^i f\| = O(r_n^{-2i}), \end{aligned}$$

by the Cauchy–Schwarz inequality, where the inner products and L^2 norm are taken over M.

On the other hand, we have already remarked (Theorem IV.14) that $\|\varphi_n\|_\infty \leq r_n^{3/2}$, so for large i, the series on the right in (16) converges absolutely and uniformly. Moreover, since it is easily checked that

$$c_0 h_T(\pm i)\varphi_0(x) = A^{-1}\int_M f,$$

which is the integral of f over M with respect to normalized hyperbolic measure, we conclude that the difference $D_T(f)$ between the integral of f with respect to normalized hyperbolic measure and the integral of f with respect to $dm_T(x, y)$ is given by the formula

$$D_T(f) = \sum_n{}' c_n h_T(r_n)\varphi_n(x),$$

where as before, the prime means that the values $r_n = \pm i$ are omitted from the sum. Since as we have seen, $|c_n \varphi_n(x)| = O(\lambda_n^{-j})$ for any fixed $j \geq 1$, this implies that the asymptotic behavior of $D_T(f)$, which is a standard measure of equidistribution of $dm_T(x, y)$, can be studied in terms of the $h_T(r_n)$'s. In particular, remembering that $h_T(r) = (\sin rT)(r \sinh T)^{-1}$, our analysis leads to the following conclusions:

(a) If M has small eigenvalues, then $D_T(f) = O(e^{(\alpha-1)T})$, where α is defined as in the previous example.

(b) If M has no small eigenvalues, then $D_T(f) = O(e^{-T})$.

Remark 3: This account is adapted from Randol [9], in which projections of dilations of more general sets in $\mathbb{R}^n$ onto the n-torus are also discussed.

Remark 4: Peter Sarnak has also studied projections of dilating spheres in $\mathbb{H}^3$, using the wave equation.

C. Electrostatics

We retain the notation of the previous examples, and again for simplicity assume that M is 3-dimensional. In what follows, we will think of M as a conductor of unit conductivity. Additionally, we postulate a force law for M, that is, a smooth real-valued function $H(\rho)$, defined on $[0, \infty)$, which we will assume arises from a potential $k(\rho)$ by the relation $k'(\rho) = -H(\rho)$, where

$k(\rho)$ is a smooth function on $[0, \infty)$ which vanishes at infinity and for which the pretrace formula (9) is valid, with uniform and absolute convergence. We will assume that the effects created by a charge concentrated at a point propagate along geodesics, with an attenuation given by the force law for M, and obey the law of superposition. More precisely, if a unit of charge is located at $y \in M$, then its effect at $x \in M$ $(x \neq y)$ is given by $\sum_n H(L_n)V_n$, where L_n ranges over the lengths of the geodesic segments g_n connecting y to x, and V_n is the unit tangent vector to g_n at x. If a unit of charge with negative sign is located at y, we define its effect at x to be $-\sum_n H(L_n)V_n$. Additionally, we think of vector fields on M as force fields, and assume that such a field F induces a current on M. The direction of the current at a point $x \in M$ is that of F at x. Under the influence of F, the instantaneous rate of flow of charge out of a small piece Q of M having boundary ∂Q is assumed to be given by integrating the dot product of F with the outer normal to ∂Q over ∂Q, and therefore by the divergence theorem is equal to $\int_Q \operatorname{div} F$.

Ohm's law suggests that we define the instantaneous power generated within S by the current induced by F to be $\int_Q (F, F)$, that is, the square of the L^2 norm of F over Q.

Let us examine some of the consequences of these assumptions.

It is evident from the definitions that if we lift everything to $\mathbb{H}^3$, then the effect at a point $x \in \mathbb{H}^3$ of a unit charge concentrated at $y \in \mathbb{H}^3$ is given by $\sum_\gamma H(d(x, \gamma y))V_\gamma$, where V_γ is the unit tangent vector at x to the geodesic segment connecting γy to x. Now it is easily checked that $H(d(x, \gamma y))V_\gamma = -\nabla_x k(x, \gamma y)$, where "$\nabla_x$" denotes the gradient taken with respect to the x variable, and $k(x, y) = k(d(x, y))$, so it follows that

$$\sum_\gamma H(d(x, \gamma y))V_\gamma = -\nabla_x \sum_\gamma k(x, \gamma y),$$

and by the pretrace formula,

$$-\nabla_x k(x, \gamma y) = -\nabla_x \sum_n h(r_n)\varphi_n(x)\overline{\varphi}_n(y).$$

Passing back to M, we conclude that the vector field corresponding to a unit charge located at $y \in M$ is given by

$$F(x, y) = -\sum_n{}' h(r_n)(\nabla\varphi_n(x))\overline{\varphi}_n(y) \tag{17}$$

(the gradient removes the constant function).

Suppose now that we have a distribution $m(y)$ of charge on M. If we define the vector field it produces to be

$$F_m(x) = \int_M F(x, y)m(y)\,dy,$$

it follows from (17) that

$$(18) \qquad F_m(x) = -\sum_n{}' a_n h(r_n) \nabla\varphi_n(x),$$

where $a_0, a_1, a_2, \ldots$ is the sequence of Fourier coefficients of $m(x)$ with respect to $\varphi_0, \varphi_1, \varphi_2, \ldots$.

Suppose now that we have force fields $F_m(x)$ and $F_{m'}(x)$, corresponding to two charge distributions m and m' with Fourier coefficients $a_0, a_1, a_2, \ldots$ and $b_0, b_1, b_2, \ldots$. What is the inner product of F_m and $F_{m'}$, that is, $\int_M (F_m, F_{m'})$?

It follows from (18) that the inner product will be given by

$$(19) \qquad \sum_i{}' \sum_j{}' a_i \bar{b}_j h(r_i) \bar{h}(r_j) \int_M (\nabla\varphi_i, \nabla\varphi_j) = \sum_n{}' a_n \bar{b}_n |h(r_n)|^2 \lambda_n,$$

where the last conclusion follows from Green's theorem for M.

In more detail, because of Green's theorem, the nondiagonal terms in (19) are zero, and

$$\int_M (\nabla\varphi_n, \nabla\varphi_n) = -\int_M \varphi_n \overline{\Delta\varphi_n} = \lambda_n.$$

In particular, we conclude that the square of the L^2 norm of the force field F_m, corresponding to a distribution with Fourier coefficients $a_0, a_1, a_2, \ldots$, is given by the formula

$$(20) \qquad \|F\|^2 = \sum_n{}' |a_n|^2 |h(r_n)|^2 \lambda_n.$$

This leads at once to an interesting conclusion, which we express as a theorem.

THEOREM 2. For any force law, a constant charge distribution on M always gives rise to the zero force field on M. If the function h does not vanish at any $r_n \neq \pm i$, then the constant distributions are the only distributions having this property, which we will call neutrality. If, on the other hand, h vanishes at one or more of the r_n's $\neq \pm i$, then there are nontrivial neutral distributions. These distributions are precisely those whose nonzero Fourier coefficients occur at indices corresponding to r_n's at which h vanishes.

Remark 5: It is possible to exhibit force laws for which there are nontrivial neutral distributions. Interestingly, there are an infinite number of Gaussian potentials, that is, k's of the form $e^{-B\rho^2}$ $(B > 0)$ for which there exist such distributions. In order to see this, note that if we integrate (6) by parts, we find that

$$h(r) = (4\pi/r) \int_0^\infty k(\rho) \sinh \rho \sin r\rho \, d\rho.$$

It follows that if we define $k(\rho) = e^{-\rho^2/4A}$ $(A > 0)$, then up to a multiplicative constant depending on A, $h(r) = r^{-1}e^{-Ar^2} \sin 2Ar$ (Bateman [1, p. 92, formula 37]). From this it is clear that by appropriately selecting A, we can produce a zero of h at any nonzero real r_n. On the other hand, it is easy to see that the conditions $k(\rho) > 0$ and $(k(\rho) \sinh \rho)' < 0$ will ensure that $h(r)$ is positive on the real axis, and on the segment of the imaginary axis between $-i$ and i. To avoid later anomalies involving the potential energy of distributions, we will in what follows assume that $h(r_n) \geq 0$ for all $r_n \neq \pm i$.

Suppose now we begin with a distribution $m(x)$ which is not neutral, and wish to describe its evolution toward a neutral distribution. What is the differential equation of this process? If we denote the distribution at time t by $m(x, t)$ $(m(x, 0) = m(x))$, the corresponding force field by $F_m(x, t)$, and recall that the instantaneous rate of flow of charge out of a small ball Q in M is given by $\int_Q \operatorname{div}_x F_m(x, t) \, dx$, we easily conclude, by considering this quantity divided by the volume of Q as Q shrinks to a point, that

$$\partial/\partial t \, m(x, t) = -\operatorname{div}_x F_m(x, t).$$

Denoting the Fourier coefficients of $m(x, t)$ by $a_0(t), a_1(t), a_2(t), \ldots$, and recalling that by (18),

$$F_m(x, t) = -\sum_n{}' a_n(t) h(r_n) \nabla \varphi_n(x),$$

we conclude that

$$\begin{aligned} \partial/\partial t \, m(x, t) &= \sum_n{}' a_n'(t) \varphi_n(x) = \operatorname{div}_x \sum_n{}' a_n(t) h(r_n) \nabla \varphi_n(x) \\ &= \sum_n{}' a_n(t) h(r_n) \operatorname{div} \nabla \varphi_n(x) = \sum_n{}' a_n(t) h(r_n) \Delta \varphi_n(x) \\ &= -\sum_n{}' a_n(t) h(r_n) \lambda_n \varphi_n(x). \end{aligned}$$

This implies that

$$a_n'(t) = -a_n(t) h(r_n) \lambda_n,$$

or

$$a_n(t) = a_n(0) e^{-h(r_n)\lambda_n t}.$$

We collect these conclusions into a theorem.

THEOREM 3.

$$a_n(t) = a_n(0)e^{-h(r_n)\lambda_n t},$$

$$m(x, t) = -\sum_n{}' a_n(0)e^{-h(r_n)\lambda_n t}\varphi_n(x),$$

and so, by (18),

$$F_m(x, t) = -\sum_n{}' a_n(0)e^{-h(r_n)\lambda_n t}h(r_n)\nabla\varphi_n(x).$$

From this we see that $m(x)$ decays exponentially into the neutral distribution whose Fourier coefficients are zero except at those n for which $h(r_n) = 0$, where the Fourier coefficient is simply $a_n(0)$. We are, of course, using the assumption that $h(r_n) \geq 0$ for all $r_n \neq \pm i$.

What is the energy expended in this process, that is, what is the potential energy of $m(x)$?

Recall that the instantaneous power produced over all of M is given by $\int_M (F_m, F_m)$, and it follows from Theorem 3 and (20) that this equals

$$\sum_n{}' |a_n(0)|^2 e^{-2h(r_n)\lambda_n t}(h(r_n))^2\lambda_n.$$

The total energy expended in moving to a neutral distribution is accordingly given by

$$\int_0^\infty \sum_n{}' |a_n(0)|^2 e^{-2h(r_n)\lambda_n t}(h(r_n))^2\lambda_n \, dt = \tfrac{1}{2}\sum_n{}' |a_n(0)|^2 h(r_n).$$

THEOREM 4. If the force law is such that $h(r_n) \geq 0$ for $r_n \neq \pm i$, then the potential energy of a distribution $m(x)$ having Fourier coefficients $a_0, a_1, a_2, \ldots$ is finite, and is given by

$$\tfrac{1}{2}\sum_n{}' |a_n|^2 h(r_n).$$

Remark 6: This result can also be derived from the counterpart of the standard formula (Wermer [1, p. 33]) for the energy of a distribution in terms of the potential function.

COROLLARY. Under the circumstances of Theorem 4, the maximum potential energy which an L^2 distribution of norm 1 can have is $\frac{1}{2}\max_n h(r_n)$.

This maximum is achieved for any normalized distribution whose nonzero Fourier coefficients occur at indices for which $h(r_n)$ is a maximum.

Remark 7: This is only one of several possible corollaries of this type, each of which corresponds to a different admissible class of Fourier coefficients. The corollary suggests the possibility of an intrinsic notion of capacity for all of M, if one defines the capacity of M to be the maximum potential energy that M can hold, as a result of the presence of a suitably normalized charge distribution. Alternatively, one can minimize rather than maximize the potential energy. This will not give anything interesting for M itself, but for subsets it can. For a subset S one can, for example, take the infimum of the potential energy over the set of nonnegative measures supported on S and having suitably normalized L^∞ Fourier coefficients.

There are other avenues of investigation suggested by Theorem 4. For example, suppose we take a finite number of δ-functions on M, supported at points $x_1, \ldots, x_N$. It would be interesting to know something about the stable configurations for such a collection, that is, the configurations for which the associated force field vanishes at $x_1, \ldots, x_N$. The previous analysis predicts that with the passage of time, the charges will diffuse out and cease to be pointlike, but it is intuitively plausible, and can be proved using (17), that a configuration which minimizes potential energy is stable, and the potential energy in the case at hand can be described. In particular, the nth Fourier coefficient of such a distribution is $\overline{\varphi}_n(x_1) + \cdots + \overline{\varphi}_n(x_N)$, and so by Theorem 4, the question becomes one of finding local minima in the N-fold Cartesian product of M with itself for the function

$$\tfrac{1}{2}\sum{}'\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\varphi_n(x_i)\overline{\varphi}_n(x_j)\right)h(r_n). \tag{21}$$

This is probably in general difficult, although there are some easy consequences of (21) that one can derive immediately. For example, if the x_j's are identically and uniformly distributed throughout M, and if we normalize the volume of M, then the expected value of the energy is

$$(N/2)\sum_n{}' h(r_n).$$

Remark 8: Although we have carried out the analysis in the context of $\mathbb{H}^3$ because of the simplicity of the transforms which arise in that case, there is no intrinsic dimensional restriction involved, and similar conclusions are true for any $\mathbb{H}^n$. Additionally, an analysis of this kind can be carried out, using the Poisson summation formula, for quotients of Euclidean spaces by lattices of translations.

Remark 9: There are, of course, many applications of the pretrace formula besides those which we have sketched in this section. For example, it can be used to study eigenvalue and eigenfunction asymptotics of M (cf. Randol [8]).

4. THE TRACE FORMULA

Suppose we set $y = x$ in either form of the pretrace formula and integrate over M, thereby eliminating the eigenfunctions. Is the resulting trace formula of interest?

In order to study this question, we will work in the covering space and take as our starting point the form of the pretrace formula given by (10). Setting $y = x$ in (10) and integrating over a fundamental domain F of Γ, we obtain

$$\sum_n h(r_n) = 2AL(0) + 2\sum_\gamma{}' \int_F L(x, \gamma x)\, dx. \tag{22}$$

Let us group the terms on the right of (22) by conjugacy classes of Γ. Let $\{\gamma\}$ denote the conjugacy class of Γ determined by the element γ. We are then led to consider sums of the form

$$\sum_{\gamma\in\{\gamma_0\}} \int_F L(x, \gamma x)\, dx.$$

A typical term of the last sum is of the form

$$\int_F L(x, \gamma_1^{-1}\gamma_0\gamma_1 x)\, dx = \int_F L(\gamma_1 x, \gamma_0\gamma_1 x)\, dx = \int_{\gamma F} L(x, \gamma_0 x)\, dx.$$

On the other hand, if $\gamma_1^{-1}\gamma_0\gamma_1 = \gamma_2^{-1}\gamma_0\gamma_2$, then $\gamma_0(\gamma_1\gamma_2^{-1}) = (\gamma_1\gamma_2^{-1})\gamma_0$, so that $\gamma_1\gamma_2^{-1} \in \Gamma_{\gamma_0}$, where Γ_{γ_0} is the centralizer of γ_0 in Γ; that is, $\gamma_2 = \gamma'\gamma_1$, with $\gamma' \in \Gamma_{\gamma_0}$.

It follows from this that the sum on the right-hand side of (22) can be replaced by

$$2\sum{}^* \int_{D_\gamma} L(x, \gamma x)\, dx, \tag{23}$$

where the sum is now over one representative from each nontrivial conjugacy class of Γ, and $D_\gamma = \bigcup_{\gamma_1} \gamma_1 F$, where γ_1 ranges over a set of representatives from the orbits of the left action of Γ_γ on Γ. Now it is easy to check that D_γ is a fundamental domain for $\Gamma_\gamma/\mathbb{H}^n$, and that the integral in (23) is the same for any two fundamental domains of Γ_γ which are not too pathological. In order to select a particularly advantageous fundamental domain, let us look at Γ_γ. As we have seen in the first section of this chapter, Γ_γ is infinite cyclic. Suppose δ is a generator for Γ_γ such that $\gamma = \delta^m$ for some positive integer m. By conjugation within the full isometry group, if necessary, we can assume that δ is of the form K_0S_0, where S_0 is a dilation $(y_1, \ldots, y_{n-1}, t) \rightarrow (N_\delta y_1, \ldots, N_\delta y_{n-1}, N_\delta t)$, with $N_\delta > 1$, and $K_0 \in O(n-1)$ operates on the $y_1, \ldots, y_{n-1}$ variables. Such a change of coordinates on $\mathbb{H}^n$ will of course have no effect on the computation of a typical integral in (23). Note that in such a coordinate system, γ itself is of the form KS, where $K = K_0^m$ and $S = S_0^m$ (since K_0 and S_0 commute), and by analogy with N_δ, we define $N_\gamma = N_\delta^m$, which is the dilation factor for γ in this coordinate system. This is clearly a well-defined function on conjugacy classes.

It is evident from this that the subset of $\mathbb{H}^n$ defined by $1 \leq t \leq N_\delta$ is a fundamental domain for Γ_γ, and we will now define D_γ to be this set.

In view of this, we need to compute

$$\int_{D_\gamma} L(x, \gamma x)\,dx = \int_{t=1}^{N_\delta} t^{-n}\,dt \int_{\mathbb{R}^{n-1}} L((|w - KSw|^2 + (1 - N_\gamma)^2 t^2)/N_\gamma t^2)\,dw,$$

where we have set $(y_1, \ldots, y_{n-1}, y) = (w, t)$, recalled that $L(x, y) = L(|x - y|^2/\sigma t)$, and restricted KS to R^{n-1} in the obvious way.

If we make the transformation $(N_\gamma^{1/2} t)^{-1} w \rightarrow w$, this becomes

$$N_\gamma^{(n-1)/2} \int_{t=1}^{N_\delta} t^{-1}\,dt \int_{\mathbb{R}^{n-1}} L(|(I - KS)w|^2 + (N_\gamma^{1/2} - N_\gamma^{-1/2})^2)\,dw,$$

where I is the identity transformation on $\mathbb{R}^{n-1}$. Setting $l_\delta = \log N_\delta$, $l_\gamma = \log N_\gamma$, this becomes

$$N_\gamma^{(n-1)/2} l_\delta \int_{\mathbb{R}^{n-1}} L(|(I - KS)w|^2 + z(l_\gamma))\,dw.$$

After the transformation $(I - KS)w \rightarrow w$, this becomes

$$N_\gamma^{(n-1)/2} l_\delta |\det(I - KS)|^{-1} \int_{\mathbb{R}^{n-1}} L(|w|^2 + z(l_\gamma))\,dw.$$

If, now, we introduce polar coordinates (ρ, θ) in $\mathbb{R}^{n-1}$, and then make the transformation $\rho^2 + z(l_\gamma) \rightarrow \rho$, exactly as in the steps leading up to (4), we find that the last quantity becomes

$$\tfrac{1}{2} N_\gamma^{(n-1)/2} l_\delta |\det(I - KS)|^{-1} \mathbf{c}_{n-2} \int_{z(l_\gamma)}^{\infty} L(\rho)(\rho - z(l_\gamma))^{(n-3)/2}\,d\rho.$$

By (4), the integral in this expression is the cosine transform of $2(\pi c_{n-2})^{-1}h(r)$, evaluated at l_γ, so if we define

$$g(u) = \pi^{-1}\int_0^\infty h(r)\cos ru\,dr,$$

we have shown that

$$\int_{D_\gamma} L(x, \gamma x)\,dx = N_\gamma^{(n-1)/2} l_\delta |\det(I - KS)|^{-1} g(l_\gamma).$$

Thus, if we define $\Lambda(\gamma) = l_\delta$, and note that the quantity $|\det(I - KS)|^{-1}$ is a well-defined function on conjugacy classes, we obtain

$$\sum_n h(r_n) = 2AL(0) + 2\sum_{\{\gamma\}} N_\gamma^{(n-1)/2}\Lambda(\gamma)|\det(I - KS)|^{-1} g(l_\gamma), \tag{24}$$

where the sum on the right is over the nontrivial conjugacy classes of Γ. Notice that in (24) the only remaining vestige of the function L is the first term on the right, which suggests the possibility of regarding h or its transform g, rather than L, as the primary function and thereby directly studying the effects on (24) of various transform pairs.

Suppose, for example, we regard g as the primary function. Then as we have remarked, (24) implies the relation

$$c_{n-2}^{-1} g(\delta(u)) = \int_u^\infty L(\rho)(\rho - u)^{(n-3)/2}\,d\rho, \tag{25}$$

where as before, $\delta(u) = z^{-1}(u)$.

If we make the assumption that g is an even C^∞ function of compact support, then $G(u) = c_{n-2}^{-1} g(\delta(u))$ is C^∞ with compact support on $[0, \infty)$, and if $n \geq 3$ is odd, we conclude from (25) that

$$L(u) = b_n G^{((n-1)/2)}(u), \tag{26}$$

while if $n \geq 2$ is even, we find that

$$\int_u^\infty L(\rho)(\rho - u)^{-1/2}\,d\rho = b_n G^{((n-2)/2)}(u), \tag{27}$$

where we have set $b_n = (-1)^{(n-1)/2}(((n-3)/2)!)^{-1}$ if n is odd, and $b_n = (-1)^{(n-2)/2}(((n-3)/2)((n-5)/2)\cdots\frac{1}{2})^{-1}$ if n is even ($b_2 = 1$).

Now (27) is an instance of Abel's integral equation, and it is well known (cf. Courant [1]) that (27) implies that

$$L(u) = -\pi^{-1} b_n \int_u^\infty G^{(n/2)}(\rho)(\rho - u)^{-1/2}\,d\rho. \tag{28}$$

It follows directly from (26) if n is odd, and easily from (28) if n is even, that under the hypotheses on g, $L(u)$ is C^∞ with compact support on $[0, \infty)$. For the case n even, one integrates the right-hand side of (28) repeatedly by parts, from which it is easily seen that L has arbitrarily many derivatives.

That is, if g is an even C^∞ function of compact support, then g corresponds to a valid L in the pretrace formula. Moreover, by the Paley–Wiener theorem, if the function h is an even entire function of exponential type, and of rapid decrease on the real line, then the corresponding g will be an even C^∞ function having compact support, so there are realistic conditions on either h or g which will guarantee the validity of (24), with absolute convergence on both sides.

Remark 10: It can be shown more generally that if h is even and holomorphic in a strip of width greater than $n - 1$ about the real axis, and if $h(r) = O(r^{-(n+\varepsilon)})$ uniformly in this strip as $r \to \infty$, then (24) is valid, with absolute convergence on both sides.

Let us now look more closely at $L(0)$.

Suppose n odd. Then since

$$g(u) = \pi^{-1} \int_0^\infty h(r) \cos ru \, dr,$$

it follows from (26) that

$$L(0) = \int_0^\infty h(r)\Phi_n(r) \, dr,$$

where

$$\Phi_n(r) = (\pi c_{n-2})^{-1} b_n (\partial^{(n-1)/2}/\partial u^{(n-1)/2} \cos r\delta(u))|_{u=0}. \tag{29}$$

If n is even, similar reasoning based on (28) leads to the conclusion that

$$L(0) = \int_0^\infty h(r)\Phi_n(r) \, dr,$$

where

$$\Phi_n(r) = -(\pi^2 c_{n-2})^{-1} b_n \int_0^\infty \rho^{-1/2} (\partial^{n/2}/\partial \rho^{n/2} \cos r\delta(\rho)) \, d\rho. \tag{30}$$

Let us calculate $\Phi_2(r)$ and $\Phi_3(r)$. We begin with $\Phi_3(r)$. By (29),

$$\begin{aligned}\Phi_3(r) = -(2\pi^2)^{-1}(\partial/\partial u \cos r\delta(u))|_{u=0} &= (2\pi^2)^{-1} \lim_{u\to 0} r(\sin r\delta(u))\delta'(u) \\ &= (2\pi^2)^{-1} \lim_{u\to 0} r(\sin r\delta(z(u)))\delta'(z(u)) \\ &= (2\pi^2)^{-1} \lim_{u\to 0} r(\sin ru)(\operatorname{csch} u) \\ &= (2\pi^2)^{-1} r^2.\end{aligned}$$

We next calculate $\Phi_2(r)$. By (30),

$$\Phi_2(r) = -(2\pi^2)^{-1} \int_0^\infty \rho^{-1/2}(\partial/\partial\rho \cos r\delta(\rho))\, d\rho.$$

Now for $\rho > 0$,

$$\partial/\partial\rho \cos r\delta(\rho) = -(r \sin \delta(\rho))\delta'(\rho),$$

so if we set $u = \delta(\rho)$, $\rho = z(u)$, the last integral becomes

$$(2\pi^2)^{-1} \int_0^\infty r(\sin ru)(\operatorname{csch} u/2)\, du,$$

where we have used the fact that $z'(u) = \sinh u$.

The last integral is a sine transform, and it follows from Bateman [1, p. 88], that it equals $(2\pi)^{-1} r \tanh \pi r$.

Recapitulating, we have shown that

$$\Phi_2(r) = (2\pi)^{-1} r \tanh \pi r,$$

and

$$\Phi_3(r) = (2\pi^2)^{-1} r^2.$$

In general, $\Phi_n(r)$ will be a polynomial of degree $(n - 1)/2$ in r^2 if n is odd, and will be a polynomial of degree $(n - 2)/2$ in r^2 times $r \tanh \pi r$, if n is even. We mention without proof that it can be shown using Harish-Chandra's Plancherel formula for spherical functions (Harish-Chandra [1]) that in general

$$\Phi_n(r) = n\omega_n(2\pi)^{-2}|\Gamma(ir + (n-1)/2)|^2 |\Gamma(ir)|^{-2},$$

where ω_n is the volume of the unit ball in $\mathbb{R}^n$. It would be an interesting exercise to derive this from (29) and (30).

Summarizing, for suitable transform pairs h and g,

$$\sum_n h(r_n) = 2A \int_0^\infty h(r)\Phi_n(r)\, dr + 2 \sum_{\{\gamma\}} N_\gamma^{-(n-1)/2} \Lambda(\gamma) |\det(I - S^{-1}K^{-1})|^{-1} g(l_\gamma).$$

or what is the same thing,

$$\sum_n h(r_n) = 2A\int_0^\infty h(r)\Phi_n(r)\,dr + 2\sum_{\{\gamma\}} N_\gamma^{-(n-1)/2}\Lambda(\gamma)|\det(I - S^{-1}K^{-1})|^{-1}g(l_\gamma), \tag{31}$$

with absolute convergence of both sides. We will call the identity (31) the *Selberg trace formula*.

Note that in 2 dimensions, if Γ is a group of directly conformal (i.e., holomorphic) transformations, which we can always assume to be the case by the uniformization theorem, then K is the identity for all $\{\gamma\}$, and

$$N_\gamma^{-1/2}|\det(I - S^{-1}K^{-1})|^{-1} = N_\gamma^{-1/2}|1 - N_\gamma^{-1}|^{-1} = \tfrac{1}{2}\operatorname{csch} l_\gamma/2,$$

so in two dimensions, the trace formula becomes

$$\sum_n h(r_n) = (A/\pi)\int_0^\infty h(r)r\tanh \pi r\,dr + \sum_{\{\gamma\}} \Lambda(\gamma)(\operatorname{csch} l_\gamma/2)g(l_\gamma). \tag{32}$$

For reasons of notational economy, we will henceforth designate the quantity $|\det(I - S^{-1}K^{-1})|^{-1}$, which occurs on the right-hand side of (31), by $D(\gamma)$ ($D(\gamma)$ of course depends only on the conjugacy class of γ).

5. APPLICATIONS OF THE TRACE FORMULA

A. The Geometric Information Contained in the Laplace Spectrum

Suppose M is a compact hyperbolic n-manifold. We will call the sequence of eigenvalues $0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \cdots$, with repetitions for multiplicities, the *Laplace spectrum* of M. Knowing the Laplace spectrum of M is clearly equivalent to knowing the numbers $\{r_n\}$ which occur on the left side of the pretrace and trace formulas.

Now it follows immediately from the Weyl asymptotic law for the eigenvalues that two manifolds having the same Laplace spectrum have the same dimension and volume. We remark that the Weyl law in the case at hand can be easily obtained from the trace formula (cf. Hejhal [1]).

Let us examine the geometric meaning of the numbers $\{l_\gamma\}$ which occur on the right-hand side of the trace formula. Suppose $\gamma \in \Gamma$ is an element of the conjugacy class associated to l_γ. Then the axis I_γ of γ projects to a geodesic on M which must be closed, since it follows from the definition of l_γ that any two points on I_γ which are distance l_γ apart are identified by γ. In this way, we can associate to γ a closed geodesic on M of length l_γ, and it is evident that the closed geodesic resulting from this process depends only on the conjugacy class of γ.

Conversely, every closed geodesic g on M corresponds to a unique conjugacy class of Γ in the above way, for if we lift g to a geodesic I in $\mathbb{H}^n$, it follows from the fact that g is closed, that there must be some $\gamma \in \Gamma$ which takes a given point p_1 on I to a point p_2 on I, with $d(p_1, p_2) = \text{length}(g)$. It is then clear that if S denotes the segment of I from p_1 to p_2, that S and $\gamma(S)$ must join smoothly at p_2, since otherwise the projection of S to M would have a corner; that is, I must be invariant under γ, so $I = I_\gamma$ and $l_\gamma = \text{length}(g)$. Clearly such a γ is unique for a given lift I of g, since if γ_1 and γ_2 both have these properties, then $\gamma_2^{-1}\gamma_1(p_1) = p_1$, which is impossible unless $\gamma_1 = \gamma_2$. It follows easily that if we consider all possible lifts of g, it is precisely the elements of $\{\gamma\}$ which correspond to g in the above way.

That is, we can identify the nontrivial conjugacy classes of Γ with the closed geodesics of M, and the numbers $\{l_\gamma\}$ are precisely the lengths of the closed geodesics. It is customary to call the sequences of the l_γ's, arranged in nondecreasing order and with repetitions for multiplicities, the *length spectrum* of M.

Suppose now that two compact hyperbolic manifolds M and M' have the same Laplace spectrum. If we apply the trace formula (31) to both of them and subtract the resulting identities from each other, we conclude, since the volumes of M and M' must be identical, that the sum which is the second term on the right-hand side of the trace formula must be identical for M and M'. If now for a given $l > 0$, we let the function g in the trace formula range through a family of even functions whose restriction to $[0, \infty)$ is supported in progressively smaller intervals about l, and such that $g(l) = e^{((n-1)/2)l}$ for any g in the family, we find that the following quantities are identical for M and M': The numbers l, which can arise as the length of some closed geodesic, and associated with each such l, the sum of the corresponding numbers $\Lambda(\gamma)D(\gamma)$.

Let us examine the implications of this result for the 2-dimensional case.

In this case, as we have seen, we may assume that $D(\gamma) = |1 - N_\gamma^{-1}|^{-1}$, which is computable from l_γ alone.

Now among the nontrivial conjugacy classes of Γ there are the so-called primitive ones, which are the conjugacy classes not of the form $\{\gamma^k\}$, for any $k \geq 2$. These correspond to the primitive closed geodesics on M, that is, the closed geodesics which do not wrap around themselves.

Suppose now that two compact hyperbolic 2-manifolds have the same Laplace spectrum. By our previous remarks, the corresponding length spectra must both have the same first element l_1, and any conjugacy class corresponding to l_1 is clearly primitive. Now in 2 dimensions, we can compute the multiplicity m_1 with which l_1 occurs, since as we have remarked, in this case $D(\gamma)$ is computable from l_γ, and for a primitive conjugacy class, $\Lambda(\gamma) = l_\gamma$. For each $j \geq 1$, we can then remove from each length spectrum m_1 elements of the form jl_1. The first element of each of the two remaining sequences clearly corresponds to a primitive conjugacy class. It follows from repetition of this process that the length spectra of M and M' must be identical, since any nontrivial conjugacy class is a positive integral power of a primitive conjugacy class; that is, in two dimensions, the Laplace spectrum determines the length spectrum. We therefore conclude that two compact hyperbolic 2-manifolds having the same Laplace spectrum have the same volume and the same length spectrum. It follows from the Gauss–Bonnet theorem that they must also have the same genus, and therefore isomorphic fundamental groups, since they are of constant curvature -1 and have the same volume. Conversely, it is not difficult to show that two such manifolds having the same length spectrum and volume must have the same Laplace spectrum (cf. McKean [2]). We remark that there is considerable redundancy of information in these data. For example, it is not difficult to show that in the case at hand, the length spectrum determines the volume, and there are other redundancies as well (McKean [2]).

Is it perhaps the case that two compact hyperbolic 2-manifolds with the same Laplace spectrum are isometric?

Wolpert [1] has shown that this is close to being correct, in the sense that the Riemann surfaces not characterized by their Laplace spectrum are, in a precise sense, very infrequent in an appropriate space of such surfaces. On the other hand, Vignéras [1] has shown that in two and three dimensions, there exist nonisometric pairs of compact hyperbolic manifolds having the same Laplace spectrum (cf. also Sunada [1]). In three dimensions this has the striking corollary that the Laplace spectrum does not even determine the fundamental group, since by the Mostow rigidity theorem (Mostow [1]), any two compact hyperbolic 3-manifolds having the same fundamental group must be isometric.

B. The Asymptotic Behavior of the l_γ's[2]

For $T > 0$, define $G(T) = \sum_{l_\gamma \leq T} 1$.
What is the behavior of $G(T)$ as $T \to \infty$?

[2] DeGeorge [1], Hejhal [1], Huber [1], Langlands [1], and Randol [3].

We will study this question by studying the asymptotic behavior of a sequence of auxiliary functions culminating in $G(T)$. The first auxiliary function is $H(T)$, defined by

$$H(T) = \sum_{l_\gamma \leq T} \Lambda(\gamma)(1 + N_\gamma^{-(n-1)})D(\gamma).$$

In order to describe the asymptotic behavior of $H(T)$, set $\beta = (n-1)(1 - \frac{1}{2}n)$, and introduce numbers $\alpha_1, \ldots, \alpha_m$, by defining α_j to be $|r_j| + (n-1)/2$, where $r_1, \ldots, r_m$ is the sequence, in order of decreasing absolute value, of those r_n's in the trace formula which are located on the open segment of the imaginary axis between $((n-1)/2)i$ and $((n-1)/2)(1 - 1/n)i$. If there are no such r_n's present, we simply omit the terms arising from the α_j's in all subsequent estimates. Finally, for typographic convenience, introduce a function $E_T(x)$ $(T, x > 0)$, by setting $E_T(x) = x^{-1}e^{Tx}$.

LEMMA 2. As $T \to \infty$,

$$H(T) = E_T(n-1) + E_T(\alpha_1) + \cdots + E_T(\alpha_m) + O(e^{\beta T}).$$

We will defer the proof of this lemma for the present and now examine some of its consequences.

If we recall that $D(\gamma) = |\det(I - S^{-1}K^{-1})|^{-1}$, that K^{-1} and S^{-1} commute, and that the powers of K^{-1} are all orthogonal matrices and thus have columns which are unit vectors in $\mathbb{R}^{n-1}$, it follows easily from the Neumann series for $(I - S^{-1}K^{-1})^{-1}$, that $D(\gamma) = 1 + O(N_\gamma^{-1})$; that is

$$\begin{aligned} H(T) &= \sum_{l_\gamma \leq T} \Lambda(\gamma)(1 + N_\gamma^{-(n-1)})(1 + O(N_\gamma^{-1})) \\ &= \sum_{l_\gamma \leq T} \Lambda(\gamma) + O\left(\sum_{l_\gamma \leq T} \Lambda(\gamma) N_\gamma^{-1}\right). \\ &\sim E_T(n-1), \qquad \text{by Lemma 2.} \end{aligned}$$

Thus, if we define

$$\psi(T) = \sum_{l_\gamma \leq T} \Lambda(\gamma),$$

it follows from the last estimate and the fact that only a finite number of the N_γ's are less than any fixed bound, that $\psi(T) \sim E_T(n-1)$.

We therefore find that

$$\sum_{l_\gamma \leq T} \Lambda(\gamma) N_\gamma^{-1} = \int_0^T e^{-x}\, d\psi(x) = O(e^{(n-2)T})$$

if $n > 2$, and $=O(T)$ if $n = 2$.

Thus, by Lemma 2 and the fact that the last two estimates are in all dimensions less than $O(e^{\beta T})$, we find that in fact

$$\psi(T) = E_T(n-1) + E_T(\alpha_1) + \cdots + E_T(\alpha_m) + O(e^{\beta T}). \tag{33}$$

Next consider the function

$$\vartheta(T) = \sum_{l_\gamma \le T}^{\#} \Lambda(\gamma),$$

where the sum is taken over the l_γ's corresponding to primitive geodesics.

Evidently

$$\psi(T) = \vartheta(T) + \vartheta(T/2) + \cdots + \vartheta(T/k), \tag{34}$$

where, for a fixed manifold M, k is of the order of T, since once k is greater than or equal to T/l_1, (34) will certainly be valid. (As before, we are denoting the length of the smallest closed geodesic of M by l_1.)

It now follows easily from (33) and (34) that

$$\vartheta(T) = E_T(n-1) + E_T(\alpha_1) + \cdots + E_T(\alpha_m) + O(e^{\beta T}). \tag{35}$$

Next define $\pi(T) = \sum_{l_\gamma \le T}^{\#} 1$, where as in the case of $\vartheta(T)$, the sum is taken over the l_γ's corresponding to primitive geodesics.

Now $\pi(T) = \int_\delta^T x^{-1}\,d\vartheta(x)$ if $\delta < l_1$, so it follows from (35) that

$$\pi(T) = \operatorname{li}(e^{(n-1)T}) + \operatorname{li}(e^{\alpha_1 T}) + \cdots + \operatorname{li}(e^{\alpha_m T}) + O(T^{-1}e^{\beta T}), \tag{36}$$

where $\operatorname{li}(x) = \int_2^x dt/\log t$.

Finally, note that exactly as before, our original function $G(T)$ satisfies

$$G(T) = \pi(T) + \pi(T/2) + \cdots + \pi(T/k),$$

so (36) implies the following theorem:

THEOREM 5.

$$G(T) = \operatorname{li}(e^{(n-1)T}) + \operatorname{li}(e^{\alpha_1 T}) + \cdots + \operatorname{li}(e^{\alpha_m T}) + O(T^{-1}e^{\beta T}).$$

Note that since $\operatorname{li}(x) \sim x/\log x$, this implies that

$$G(T) \sim (n-1)^{-1}T^{-1}e^{(n-1)T}.$$

Remark 11: The principal term in the asymptotic estimate of Theorem 5 does not depend on the manifold M, but only on the dimension.

Remark 12: The "O" error term in all the estimates up to and including Theorem 5 can be very slightly improved by using, in the proof of Lemma 2, known estimates for the remainder in Weyl's asymptotic law for the Laplace spectrum, but we will not carry this out.

We must now prove Lemma 2.

For this purpose, define a family $g_T^\varepsilon(x)$ of functions for use as g's in the trace formula as follows:

(a) Let $I_T(x)$ denote the indicator function of the interval $[-T, T]$; that is, $I_T(x) = 1$ if $|x| \leq T$ and $I_T(x) = 0$ if $|x| > T$.

(b) Suppose $\varphi(x)$ is a nonnegative C^∞ function supported on $[-1, 1]$, and such that $\int_{-1}^{1} \varphi(x)\,dx = 1$. For $\varepsilon > 0$, define $\varphi_\varepsilon(x)$ to be $\varepsilon^{-1}\varphi(x/\varepsilon)$. Then $\varphi_\varepsilon(x)$ is supported on $[-\varepsilon, \varepsilon]$, and $\int_{-\varepsilon}^{\varepsilon} \varphi_\varepsilon(x)\,dx = 1$; that is, the φ_ε's constitute a family of approximate δ-functions.

(c) Finally, set $g_T^\varepsilon(x) = 2(\cosh((n-1)/2)x)(I_T * \varphi_\varepsilon)(x)$. Then for any $\varepsilon, T > 0$, $g_T^\varepsilon(x)$ is a valid g in the trace formula, and the corresponding h in the trace formula, which we will denote by $h_T^\varepsilon(r)$, is given by $S(r + ((n-1)/2)i) + S(r - ((n-1)/2)i)$, where $S(w) = (2w^{-1} \sin Tw)(\hat{\varphi}_\varepsilon(w))$, with the understanding that $S(0) = 2T$.

Now define $H_\varepsilon(T)$, which will be an approximation to $H(T)$, by setting

$$H_\varepsilon(T) = \sum_{\{\gamma\}} N_\gamma^{-(n-1)/2} \Lambda(\gamma) D(\gamma) g_T^\varepsilon(l_\gamma).$$

Note that for any $\varepsilon > 0$, $H_\varepsilon(T - \varepsilon) \leq H(T) \leq H_\varepsilon(T + \varepsilon)$.

On the other hand, it follows from the trace formula that

(37)

$$H_\varepsilon(T) = \tfrac{1}{2} \sum_j h_T^\varepsilon(r_j) - A \int_0^\infty h_T^\varepsilon(r) \Phi_n(r)\,dr = \sum_j{}^{\#} h_T^\varepsilon(r_j) + \int_0^\infty h_T^\varepsilon(r)\,d\mu_n(r),$$

where $\sum^{\#}$ denotes the finite sum over those r_j's on the nonnegative imaginary axis, and $d\mu_n(r)$ is the measure on $[0, \infty)$ given by $dN(r) - A\Phi_n(r)\,dr$, where $N(r) = \sum_{0 < r_j \leq r} 1$.

We now set $\varepsilon = e^{-((n-1)/2n)T}$.

Then if we note that by Taylor series considerations, $\hat{\varphi}_\varepsilon(x) = \hat{\varphi}(\varepsilon x) = 1 + O(\varepsilon x)$ for fixed x as $\varepsilon \to 0$, it follows that the term in $\sum_j^{\#} h_T^\varepsilon(r_j)$ corresponding to r_j $(1 \leq j \leq m)$ is of the form

$$(2\alpha_j^{-1} \sinh \alpha_j T)\,\hat{\varphi}_\varepsilon(i\alpha_j) = E_T(\alpha_j) + O(\varepsilon e^{\alpha_j T}),$$

so we easily conclude that

$$(38)\qquad \sum_j{}^{\#} h_T^\varepsilon(r_j) = E_T(n-1) + E_T(\alpha_1) + \cdots + E_T(\alpha_m) + O(\varepsilon e^{(n-1)T})$$
$$= E_T(n-1) + E_T(\alpha_1) + \cdots + E_T(\alpha_m) + O(e^{\beta T}).$$

We still need to estimate the term $\int_0^\infty h_T^\varepsilon(r)\,d\mu_n(r)$ in (37).

By Weyl's law, the total variation of $d\mu_n(r)$ over an interval $[0,x]$ is $O(x^n)$.

Now it follows from the form of $h_T^\varepsilon(r)$, that for real r, and any fixed $k > 0$,

$$|h_T^\varepsilon(r)| \le M(k)e^{((n-1)/2)T}(1+r)^{-1}(1+\varepsilon r)^{-k};$$

that is,

$$\left|\int_0^\infty h_T^\varepsilon(r)\,d\mu_n(r)\right| \le M(k)e^{((n-1)/2)T}\int_0^\infty (1+r)^{-1}(1+\varepsilon r)^{-k}\,|d\mu_n(r)|.$$

To estimate the last integral, we break it up into two ranges, by setting $\int_0^\infty = \int_0^{1/\varepsilon} + \int_{1/\varepsilon}^\infty$.

In the range $[0, 1/\varepsilon]$,

$$h_T^\varepsilon(r) = O(e^{((n-1)/2)T}|1+r|^{-1}),$$

while in the range $[1/\varepsilon, \infty)$,

$$h_T^\varepsilon(r) = O(e^{((n-1)/2)T}\varepsilon^{-k}r^{-(k+1)}).$$

If we now integrate by parts, we easily obtain the estimate $O(e^{\beta T})$ for both integrals; that is,

$$(39)\qquad \int_0^\infty h_T^\varepsilon(r)\,d\mu_n(r) = O(e^{\beta T}).$$

Combining (38) and (39), we conclude that

$$(40)\qquad H_\varepsilon(T) = E_T(n-1) + E_T(\alpha_1) + \cdots + E_T(\alpha_m) + O(e^{\beta T}).$$

But

$$H_\varepsilon(T-\varepsilon) \le H(T) \le H_\varepsilon(T+\varepsilon),$$

and

$$E_{T\pm\varepsilon}(n-1) = (n-1)^{-1}e^{(n-1)(T\pm\varepsilon)} = (n-1)^{-1}e^{(n-1)T}(1+O(\varepsilon))$$
$$= E_T(n-1) + O(e^{\beta T}),$$

so it follows from (40) that

$$H(T) = E_T(n-1) + E_T(\alpha_1) + \cdots + E_T(\alpha_m) + O(e^{\beta T}),$$

which concludes the proof of Lemma 2.

C. Small Eigenvalues

In the discussion of the last topic the small eigenvalues of M, if present, have once again played a major role. Do there exist manifolds having small eigenvalues? Can one characterize such manifolds? We will very briefly outline a method for obtaining an answer to the first question. It turns out that such manifolds exist in all dimensions, with arbitrarily numerous small eigenvalues as close to zero as one wishes (Randol [1]).

Satisfactory answers to the second question are not known. There are interesting classes of hyperbolic manifolds which may not possess small eigenvalues. For example, it is known (Jacquet–Langlands [1]) that the Laplace spectrum of a compact hyperbolic 2-manifold corresponding to a quaternion lattice embeds in the discrete Laplace spectrum of the (noncompact) manifold associated with some congruence subgroup of $SL(2, Z)$, and it is known (Selberg [2]) that for congruence subgroups, $\lambda_1 \geq \frac{3}{16}$. It is quite possible, though unproved, that for such groups $\lambda_1 \geq \frac{1}{4}$.

In order to discuss the first question we need a form of the trace formula somewhat more general than the version we have developed. In more detail, suppose χ is a character of the discontinuous group Γ corresponding to a compact n-dimensional hyperbolic manifold M.

Consider the problem $\Delta f + \lambda f = 0$ on $\mathbb{H}^n$, where the function f is required to transform under Γ by $f(\gamma x) = \chi(\gamma) f(x)$. Then it can be shown (cf. Hejhal [1]; Selberg [1]) that there is associated with this problem a sequence $0 \leq \lambda_0(\chi) \leq \lambda_1(\chi) \leq \cdots$ of nonnegative eigenvalues tending to infinity, and that the set of associated eigenfunctions is complete in the Hilbert space of measurable functions on $\mathbb{H}^n$ which transform in the above manner and are L^2 over a fundamental domain F of Γ, with the inner product given by $(f, g) = \int_F f\bar{g}$.

It is possible to develop the pretrace and trace formulas within this more general context. In particular, the trace formula takes the form

$$(41) \quad \sum_n h(r_n(\chi)) = 2A \int_0^\infty h(r)\Phi_n(r)\,dr + 2 \sum_{\{\gamma\}} \chi(\gamma) N_\gamma^{-(n-1)/2} \Lambda(\gamma) D(\gamma) g(l_\gamma),$$

where the $r_n(\chi)$'s correspond to the $\lambda_n(\chi)$'s in the previously discussed way.

Now by using (41) with appropriate transform pairs, it can be shown that the $\lambda_n(\chi)$'s vary continuously with χ. In particular, and this is the only fact that is needed, if Γ admits a sequence of nontrivial characters converging to the trivial character χ_0, then $\lambda_0(\chi) \to 0$ as $\chi \to \chi_0$ (Randol [1]).

Suppose now that $G = \Gamma/[\Gamma, \Gamma]$ has elements of infinite order; that is, suppose the first Betti number of M is positive. Let g be a nontorsion element of a basis B of G, and define a character χ of G by setting $\chi(g) = \exp(2\pi i/2^N)$, for some very large positive integer N, and $\chi(g') = 1$ for $g' \in B$ and $g' \neq g$. Let K be a large positive integer which is nevertheless much smaller than N, and define a finite sequence $\chi_1, \ldots, \chi_K$ of characters of G by setting $\chi_1 = \chi, \chi_2 = \chi_1^2, \ldots, \chi_K = \chi_{K-1}^2$.

Then the characters $\chi_1, \ldots, \chi_K$ are also characters of Γ in an obvious way, and if $1 \leq m \leq K - 1$, there exists $\gamma \in \Gamma$ such that $\chi_m(\gamma) \neq 1$ and $\chi_j(\gamma) = 1$ for $j > m$.

Now given a manifold M and any small $\varepsilon > 0$, we can, by the continuity of the eigenvalues of χ, find N such that $0 \leq \lambda_0(\chi_j) \leq \varepsilon$ for $j = 1, \ldots, K$.

Suppose accordingly that $F_1(x), \ldots, F_K(x)$ are eigenfunctions corresponding to such $\lambda_0(\chi_1), \ldots, \lambda_0(\chi_K)$. Then the functions $F_1(x), \ldots, F_K$ are linearly independent, for if not, there exists an integer m such that $1 \leq m \leq K - 1$, a constant $c_m \neq 0$, and constants $c_{m+1}, \ldots, c_K$, such that

$$c_m F_m(x) + \cdots + c_K F_K(x) \equiv 0. \tag{42}$$

Now suppose $\gamma \in \Gamma$ is chosen so that $\chi_m(\gamma) \neq 1$, but $\chi_j(\gamma) = 1$ for $j > m$, and suppose x_0 is such that $F_m(x_0) \neq 0$. Then

$$c_m F_m(\gamma x_0) + c_{m+1} F_{m+1}(\gamma x_0) + \cdots + c_K F_K(\gamma x_0) = 0,$$

or

$$c_m \chi_m(\gamma) F_m(x_0) + c_{m+1} F_{m+1}(x_0) + \cdots + c_K F_K(x_0) = 0. \tag{43}$$

Subtracting (43) from (42), with the latter specialized to $x = x_0$, we obtain a contradiction.

Now let Γ' be the kernel in Γ of the homomorphism χ_1. Then Γ' is of finite index in Γ, since χ_1 is a homomorphism into a finite group. Furthermore, for any $F_j(x)$, $\gamma \in \Gamma'$ implies that $F_j(\gamma x) = F_j(x)$. Thus the $F_j(x)$'s are eigenfunctions of the Laplacian on the manifold M' corresponding to Γ', and since the $F_j(x)$'s are linearly independent, we conclude that the number of λ_n's which lie in $[0, \varepsilon]$ is at least K. It follows easily from the maximum principle for harmonic functions that none of these eigenvalues is zero, so M' has at least K small eigenvalues, if $\varepsilon < (n - 1)/2$.

Now it is proved in Millson [1] that in every dimension there are compact hyperbolic manifolds with arbitrarily large first Betti number, so it follows that the phenomenon of small eigenvalues occurs in all dimensions.

Remark 13: In two dimensions, every compact hyperbolic manifold has positive first Betti number, so any compact hyperbolic 2-manifold has finite covers with arbitrarily numerous small eigenvalues.

6. CONCLUDING REMARKS

The Selberg formulas, in the spirit in which we have presented them, were outlined in Selberg [1]. There is also a little-known and prescient note of Delsarte [Delsarte] which strikingly anticipates some of the later developments. A forthcoming book by Cohen and Sarnak [1] will present a detailed treatment of various aspects of the general hyperbolic case, and the books of Hejhal [1, 2], which are encyclopedic in character, contain an enormous amount of detailed information about the 2-dimensional case.

In order to maintain a reasonable size for this chapter, we have not discussed several important aspects of the pretrace and trace formulas; for example, the noncompact case, the Selberg zeta function, applications to more general groups, and connections with automorphic forms and arithmetic, as found, for example, in Jacquet–Langlands [1]. For a comprehensive bibliography of the subject we refer the interested reader to the reference sections of Hejhal [1, 2].

CHAPTER XII

Miscellanea

In this chapter we collect various facts, problems, references, and proofs, which are either used in the previous chapters, or supplement and amplify earlier material.

1. VOLUMES OF DISKS AND SPHERES

We recall, here, how to explicitly calculate the volume $\boldsymbol{\omega}_n$ of the unit disk in $\mathbb{R}^n$. First note that if $\mathbf{c}_{n-1}$ denotes the $(n-1)$-volume of $\mathbb{S}^{n-1}$, then one has, using spherical coordinates on $\mathbb{R}^n$,

$$\boldsymbol{\omega}_n = \mathbf{c}_{n-1}/n.$$

So it suffices to calculate $\mathbf{c}_{n-1}$.

Define the *gamma function* $\Gamma(x)$, $x > 0$, by

$$\Gamma(x) = \int_0^\infty e^{-t}t^{x-1}\,dt,$$

and verify that

$$\Gamma(x+1) = x\Gamma(x)$$

for all $x > 0$, and

$$\Gamma(n+1) = n!,$$

$$\int_0^\infty t^n e^{-t^2}\,dt = \tfrac{1}{2}\Gamma((n+1)/2),$$

for all nonnegative integers n.

Next, use spherical coordinates in $\mathbb{R}^n$ to obtain

$$\left\{\int_{\mathbb{R}} e^{-t^2}\,dt\right\}^n = \tfrac{1}{2}c_{n-1}\Gamma(n/2)$$

for all $n \geq 1$. Apply to $n = 2$, and conclude that

$$\int_{\mathbb{R}} e^{-t^2}\,dt = \sqrt{\pi} = \Gamma(\tfrac{1}{2}),$$

and

$$\mathbf{c}_{n-1} = 2\pi^{n/2}/\Gamma(n/2).$$

Another method of calculating $\mathbf{c}_{n-1}$ is to use geodesic spherical coordinates in $\mathbb{S}^{n-1}$ to prove

$$\mathbf{c}_{n-1} = \mathbf{c}_{n-2}\int_0^{\pi} \sin^{n-2}t\,dt;$$

one can now iterate the formula.

2. THE FOURIER TRANSFORM

For f, g in $L^1(\mathbb{R}^n)$, the *convolution of f and g, $f * g$*, defined by

$$(f * g)(x) = (2\pi)^{-n/2}\int_{\mathbb{R}^n} f(x - y)g(y)\,dV(y),$$

converges absolutely for almost all x, and satisfies

$$\|f * g\|_{L^1} \leq (2\pi)^{-n/2}\|f\|_{L^1}\|g\|_{L^1}.$$

Given $f\colon \mathbb{R}^n \to \mathbb{C}$, $y \in \mathbb{R}^n$, one defines the function $\tau_y f$ by

$$(\tau_y f)(x) = f(x - y),$$

and verifies that if $f \in L^p$, $p \geq 1$, then the map $\mathbb{R}^n \to L^p$ given by

$$y \mapsto \tau_y f$$

is uniformly continuous.

To every $f \in L^1$ is associated the *Fourier transform of f, $\hat{f}$* by

$$\hat{f}(\xi) = (2\pi)^{-n/2}\int_{\mathbb{R}^n} f(x)e^{-i\langle x,\xi\rangle}\,dV(x).$$

One establishes directly that

$$|\hat{f}(\xi)| \leq (2\pi)^{-n/2}\|f\|_{L^1}$$

for all $\xi \in \mathbb{R}^n$, and

$$\widehat{\tau_y f}(\xi) = e^{-i\langle y,\xi\rangle}\hat{f}(\xi),$$
$$\widehat{e^{i\langle y,\ \rangle}f}(\xi) = (\tau_y \hat{f})(\xi),$$
$$\widehat{f(\lambda x)}(\xi) = \lambda^{-n}\hat{f}(\xi/\lambda).$$

One also has that if $g \in L^1$, then

$$\widehat{f * g}(\xi) = \hat{f}(\xi)\hat{g}(\xi).$$

Finally, the Riemann–Lebesgue lemma states that

$$\lim_{\xi \to \infty} f(\xi) = 0$$

for all $\xi \in \mathbb{R}^n$.

We now let $\mathscr{S}(\mathbb{R}^n)$ be the *Schwartz space of functions on* $\mathbb{R}^n$, that is, $\mathscr{S}$ consists of those $f \in C^\infty$ such that for any integer $N \geq 0$, and multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$, we have

$$(1 + |x|^2)^N (D^\alpha f)(x)$$

bounded on $\mathbb{R}^n$, where

$$D^\alpha \equiv: (-i)^{\alpha_1 + \cdots + \alpha_n}(\partial/\partial x_1)^{\alpha_1} \cdots (\partial/\partial x_n)^{\alpha_n}.$$

Certainly, we have

$$C_c^\infty \subseteq \mathscr{S} \subseteq L^p$$

for all $p \geq 1$; so $\mathscr{S}$ is dense in L^p. One checks that if $f, g \in \mathscr{S}$, then $f * g \in \mathscr{S}$, and

$$D^\alpha(f * g) = (D^\alpha f) * g, \qquad \widehat{D^\alpha f}(\xi) = \xi^\alpha \hat{f}(\xi), \qquad (-D)^\alpha f = \widehat{x^\alpha f(x)}.$$

Thus the Fourier transform $f \mapsto f$ maps $\mathscr{S}$ to $\mathscr{S}$.

One also checks that for

$$\phi(x) = e^{-|x|^2/2}$$

we have

$$\hat{\phi}(\xi) = e^{-|\xi|^2/2}.$$

Finally, one has for $f \in \mathscr{S}$, the *Fourier inversion formula*

$$f(x) = (2\pi)^{-n/2}\int_{\mathbb{R}^n} \hat{f}(\xi)e^{i\langle x,\xi\rangle}\, dV(\xi),$$

and the *Plancherel formula*

$$\|f\|_{L^2} = \|\hat{f}\|_{L^2}.$$

3. THE POISSON SUMMATION FORMULA

Given the lattice Γ, of rank n, in $\mathbb{R}^n$, recall from Section II.2 that the eigenfunctions of the associated torus T are given by

$$\phi_y(x) = e^{-2\pi i\langle y,x\rangle},$$

where $y \in \Gamma^*$ and x ranges over T. Show that

$$(\phi_y, \phi_{y'}) = \begin{cases} \operatorname{vol} T & \text{if} \quad y = y' \\ 0 & \text{if} \quad y \neq y', \end{cases}$$

and conclude that for any $f \in L^2(T)$ we have

$$f = \sum_{y\in\Gamma^*} c_y\phi_y, \tag{1}$$

in $L^2(T)$, where

$$c_y = (\operatorname{vol} T)^{-1}\int_T f\phi_y \, dV.$$

In a fashion similar to that of Fourier series, one finds that the convergence in (1) is uniform for any $f \in C^\infty(T)$.

Given $f \in \mathscr{S}$, one derives the *Poisson summation formula*

$$\sum_{\gamma\in\Gamma} f(\gamma) = \frac{(2\pi)^{n/2}}{\operatorname{vol} T}\sum_{y\in\Gamma^*} f(2\pi y) \tag{2}$$

as follows: Set

$$g(x) = \sum_{\gamma\in\Gamma} f(x+\gamma)$$

and show that $g \in C^\infty(T)$. Then expand g as a function on T by

$$g = \sum_{y\in\Gamma^*} c_y\phi_y,$$

from which one obtains

$$\sum_{\gamma\in\Gamma} f(\gamma) = \sum_{y\in\Gamma^*} c_y,$$

and, finally, verify that

$$c_y = \{(2\pi)^{n/2}/\operatorname{vol} T\}f(2\pi y).$$

Next, we remark that Weyl's asymptotic formula for eigenvalues of tori can be seen as a consequence of the Poisson summation formula. The argument is to pick, for $t > 0$,

$$f(x) = e^{-|x|^2/4t},$$

show that (2) implies

$$\sum_{\gamma\in\Gamma^*} e^{-4\pi^2|\gamma|^2 t} = \frac{\operatorname{vol} T}{(4\pi t)^{n/2}} \sum_{\gamma\in\Gamma} e^{-|\gamma|^2/4t}, \tag{3}$$

and then show that

$$\lim_{t\to 0^+} \sum_{\gamma\in\Gamma,\gamma\neq 0} e^{-|\gamma|^2/4t} = 0.$$

Thus, by (II.3), we have

$$\sum_{k=0} e^{-\lambda_k t} \sim (\operatorname{vol} T)/(4\pi t)^{-n/2} \tag{4}$$

as $t \to 0^+$, where $\{0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots\}$ are the eigenvalues of T. Weyl's formula now follows from (4) and Karamata's Tauberian theorem (Feller [1, p. 442 ff.]).

It is not surprising that the Poisson formula implies the Weyl formula, since (3) is nothing but a rewrite of

$$\sum_{j=0} e^{-\lambda_j t} = \int_M p(x, x, t)\, dV(x), \tag{VI.14}$$

where p is the heat kernel of T. Indeed, if $\mathfrak{q}: \tilde{M} \to M$ is a Riemannian covering, with deck transformation group Γ, and $\tilde{p}, p$ are the respective heat kernels of $\tilde{M}, M$ (we are in a situation where the heat kernels are unique), then for all $(\tilde{x}, \tilde{y}, t) \in \tilde{M} \times \tilde{M} \times (0, +\infty)$ we have

$$p(\mathfrak{q}(\tilde{x}), \mathfrak{q}(\tilde{y}), t) = \sum_{\gamma\in\Gamma} \tilde{p}(\tilde{x}, \gamma\cdot\tilde{y}, t).$$

We therefore have, for the torus T,

$$p(x, y, t) = \sum_{\gamma\in\Gamma} \mathfrak{e}(x, y + \gamma, t),$$

where $\mathfrak{e}$ is the euclidean heat kernel, which implies, easily,

$$\int_T p(x, x, t)\, dV(x) = \frac{\operatorname{vol} T}{(4\pi t)^{n/2}} \sum_{\gamma\in\Gamma} e^{-|\gamma|^2/4t}.$$

4. THE FOURIER TRANSFORM AND THE HEAT EQUATION

We sketch the formal calculations which motivate guessing

$$\mathfrak{e}(x, y, t) = (4\pi t)^{-n/2} e^{-|x-y|^2/4t}$$

as the fundamental solution of the heat equation on $\mathbb{R}^n$.

For a function v: $\mathbb{R}^n \times [0, +\infty) \to \mathbb{R}$ we set

$$v(\hat{\xi}, t) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} v(x, t) e^{-i\langle x, \xi \rangle} \, dV(x),$$

that is, $v(\ , t)$ is the Fourier transform, with respect to the space variable x, only, of v.

Given u: $\mathbb{R}^n \times [0, +\infty) \to \mathbb{R}$ with

$$u(\ , 0) = f,$$

and satisfying the heat equation on $\mathbb{R}^n \times (0, +\infty)$, we have

$$-|\xi|^2 u(\hat{\xi}, t) = (\Delta u)(\hat{\xi}, t) = (\partial u/\partial t)(\hat{\xi}, t) = (\partial/\partial t)(u(\hat{\xi}, t)),$$

from which one concludes the existence of $c(\xi)$ such that

$$u(\hat{\xi}, t) = c(\xi) e^{-|\xi|^2 t}.$$

But

$$c(\xi) = u(\hat{\xi}, 0) = \hat{f}(\xi);$$

so

$$\begin{aligned} u(\xi, t) = \hat{f}(\xi) e^{-|\xi|^2 t} &= \hat{f}(\xi) e^{-|2\xi\sqrt{t}|^2/2} \\ &= \hat{f}(\xi) \{\widehat{(2t)^{-n/2} e^{-|\ |^2/4t}}\}(\xi) \\ &= \{\widehat{f * (2t)^{-n/2} e^{-|\ |^2/4t}}\}(\xi). \end{aligned}$$

Thus

$$\begin{aligned} u(x, t) &= \{f * (2t)^{-n/2} e^{-|\ |^2/4t}\}(x) \\ &= (4\pi t)^{-n/2} \int_{\mathbb{R}^n} e^{-|x-y|^2/4t} f(y) \, dV(y), \end{aligned}$$

which is (VI.16).

5. EIGENFUNCTIONS ON SPHERES AND HYPERBOLIC SPACE

We give here a more explicit representation of solutions to

$$T'' + (n - 1)(\mathrm{C}_\kappa/\mathrm{S}_\kappa)T' + \{\lambda - l(l + n - 2)\mathrm{S}_\kappa^{-2}\}T = 0 \tag{5}$$

(cf. (II.33), (II.34)) for the radial factor of eigenfunctions, on spheres and hyperbolic space, obtained via separation of variables. We are, therefore, considering the case $\kappa \neq 0$. The case $\kappa = 0$ corresponds to the euclidean case, and is discussed in Theorem II.4.

First set

$$q(t) = \mathbf{S}_\kappa^{n/2-1}(t)T(t).$$

Then $q(t)$ satisfies the differential equation

$$q'' + (\mathbf{C}_\kappa/\mathbf{S}_\kappa)q' + \{\lambda + \kappa(n/2 - 1) - [(n/2 - 1)^2 C_\kappa^2 + l(l + n - 2)]\mathbf{S}_\kappa^{-2}\}q = 0.$$

Upon changing the independent variable by

$$x = \mathbf{C}_\kappa(t), \qquad p(x) = q(t),$$

we obtain, for $p(x)$, the *associated Legendre equation*

$$(1 - x^2)\frac{d^2p}{dx^2} - 2x\frac{dp}{dx} + \left\{\nu(\nu + 1) - \frac{\mu^2}{1 - x^2}\right\}p = 0,$$

of degree ν given by

(6) $$\nu = -\tfrac{1}{2} \pm \sqrt{\lambda/\kappa + (n - 1)^2/4},$$

and order μ given by

(7) $$\mu = \sqrt{(n - 2)^2/4 + l(l + n - 2)}.$$

The solution of (5), bounded in the neighborhood of $t = 0$, is therefore given by

$$T(t) = \mathbf{S}_\kappa^{1-n/2}(t)P_\nu^\mu(\mathbf{C}_\kappa(t)),$$

where P_ν^μ is Legendre's associated function of the first kind with ν, μ given by (6), (7), respectively.

For more details, we refer the reader to the treatises by Hobson [1], Lebedev [1], and Olver [1].

6. MINIMAL SUBMANIFOLDS OF EUCLIDEAN SPACES AND SPHERES

Let $\hat{M}^k$, M^n be fixed Riemannian manifolds, $k < n$, and $\Phi: \hat{M} \to M$ a Riemannian immersion of $\hat{M}$ into M, or, in the language of Section II.1, Φ is a local isometry. Because all our calculations are local, we shall simply assume that Φ is an imbedding—more particularly, we shall think of Φ as the inclusion map of a closed manifold $\hat{M}$ in M.

To each $p \in \hat{M}$, $\hat{M}_p$ denotes the tangent space to $\hat{M}$ at p, and $\hat{M}_p^\perp$ the orthogonal complement of $\hat{M}_p$ in M_p. Given any $\xi \in M_p$, we let ξ^T, ξ^N denote the projections of ξ to M_p, $\hat{M}_p^\perp$, respectively.

It is standard that if $\hat{\nabla}, \nabla$ denote the Levi–Civita connections of the respective Riemannian metrics on $\hat{M}$, M, then for any $p \in \hat{M}$, $\xi \in \hat{M}_p$, and Y a tangent vector field of M defined on a neighborhood, in M, of p, we have

$$\hat{\nabla}_\xi Y = (\nabla_\xi Y)^T.$$

Also, it is standard that in this situation, $(\nabla_\xi Y)^N$ only depends on the value of Y at p, not on the behavior of Y on a neighborhood of p. So, given ξ, η in M_p, one extends η to a tangent vector field Y on a neighborhood of p in $\hat{M}$, and defines

$$B(\xi, \eta) = (\nabla_\xi Y)^N.$$

Then $B\colon \hat{M}_p \times \hat{M}_p \to \hat{M}_p^\perp$ is easily seen to be symmetric bilinear, and is referred to as the *second fundamental form of M in M*, the *first fundamental form* consisting of the restriction of the Riemannian metric of M to $\hat{M}$. The *mean curvature vector H of $\hat{M}$ in M* is defined to be the trace of the second fundamental form with respect to the first fundamental form.

If $x\colon U \to \mathbb{R}^k$ is a chart on $\hat{M}$, then

$$H = \sum_{r,s=1}^{k} g^{rs}(\nabla_r \partial_s)^N,$$

where $\partial_1, \ldots, \partial_k$ are the coordinate vector fields on U associated to the chart, and g^{rs} is given by (I.21).

Now let $M = \mathbb{R}^n$. Then the map $(X^1, \ldots, X^n) \mapsto X^A$, for each $A = 1, \ldots, n$, is a C^∞ function on $\mathbb{R}^n$, and the function

$$\phi^A = X^A \circ \Phi$$

is a C^∞ function on $\hat{M}$. We denote by $\Delta\Phi$ the $\mathbb{R}^n$-valued mapping

$$\Delta\Phi = (\Delta\phi^1, \ldots, \Delta\phi^n)$$

with the Laplacian calculated on $\hat{M}$. (We shall take for granted the identification of tangent vectors to $\mathbb{R}^n$ with vectors in $\mathbb{R}^n$.) The key formula is

$$\Delta\Phi = H \tag{8}$$

for any Riemannian immersion $\Phi\colon \hat{M}^k \to \mathbb{R}^n$.

To prove (8) we first show that $\Delta\Phi$ is always normal to $\hat{M}$. We do the calculation with respect to a coordinate chart $x\colon U \to \mathbb{R}^k$ on $\hat{M}$, and take advantage of the identification of the abstract coordinate vector field ∂_j with $\partial\Phi/\partial x^j \equiv: \partial_j\Phi$. We have (using the notation of Section I.1)

$$\begin{aligned}\langle \Delta\Phi, \partial_l\Phi\rangle &= (1/\sqrt{g})\sum_{r,s}\partial_r(\sqrt{g}g^{rs})g_{sl} + \sum_{r,s} g^{rs}\langle\partial_r\partial_s\Phi, \partial_l\Phi\rangle \\ &= (1/\sqrt{g})\partial_l\sqrt{g} + \sum_{r,s}(\partial_r g^{rs})g_{sl} + \sum_{r,s,t} g^{rs}\Gamma^t_{st}g_{tl} \\ &= \sum_{r,s}[\tfrac{1}{2}g^{rs}\partial_l g_{sr} - g^{rs}\partial_r g_{sl} + \tfrac{1}{2}\{g^{rs}\partial_s g_{lr} + g^{rs}\partial_r g_{ls} - g^{rs}\partial_l g_{sr}\}] \\ &= 0;\end{aligned}$$

in passing from the second term of the right-hand side of the first line to the third term in the second line we used the fact that

$$\langle\partial_r\partial_s\Phi, \partial_l\Phi\rangle = \langle\nabla_r\partial_s, \partial_l\rangle,$$

and Eq. (I.23); from the second to the third line we used the formula for differentiating determinants, and the explicit calculation of the Christoffel symbols in (I.31).

Therefore,

$$\begin{aligned}\Delta\Phi = (\Delta\Phi)^N &= \left[\sum_{r,s}\{(1/\sqrt{g})\partial_r(\sqrt{g}g^{rs})\partial_s\Phi + g^{rs}\partial_r\partial_s\Phi\}\right]^N \\ &= \left[\sum_{r,s}(1/\sqrt{g})\partial_r(\sqrt{g}g^{rs})\partial_s\Phi\right]^N + \left[\sum_{r,s}g^{rs}(\nabla_r\partial_s)\right]^N \\ &= H,\end{aligned}$$

which concludes the derivation of (8).

What if $M = \mathbb{S}^n(\alpha)$, the n-sphere of radius α in $\mathbb{R}^{n+1}$? We claim that (8) becomes

$$\Delta\Phi = H - (k/\alpha^2), \tag{9}$$

where Φ denotes $n + 1$ coordinates of position in $\mathbb{R}^{n+1}$ and H denotes the mean curvature vector of $\hat{M}$ in $\mathbb{S}^n(\alpha)$. Indeed, we now have

$$\partial_r\partial_s\Phi = \hat{\nabla}_r\partial_s + (\nabla_r\partial_s)^N + \langle\partial_r\partial_s\Phi, \Phi/\alpha\rangle\Phi/\alpha. \tag{10}$$

Using (8) as though $\hat{M}$ were in $\mathbb{R}^{n+1}$, and applying it to (10), we obtain

$$\Delta\Phi = H + \sum_{r,s} g^{rs}\langle\partial_r\partial_s\Phi, \Phi\rangle\Phi/\alpha^2.$$

But, since $|\Phi| \equiv \alpha$,

$$\langle\partial_r\partial_s\Phi, \Phi\rangle = -\langle\partial_s\Phi, \partial_r\Phi\rangle = -g_{rs},$$

which implies (9).

A Riemannian immersion $\Phi\colon \hat{M}^k \to M^n$ is said to be *minimal* if the mean curvature vector vanishes identically on M. (The name derives from the fact that the Euler–Lagrange equation associated to the variational problem of finding immersions of least k-volume is $H = 0$.) From (8) we immediately have that a Riemannian immersion Φ of $\hat{M}^k$ into $\mathbb{R}^n$ is minimal if and only if the coordinate functions of the immersion are harmonic. From (9) we immediately conclude that if a Riemannian immersion Φ of $\hat{M}^k$ in $\mathbb{S}^n(\alpha)$ is minimal, then the coordinate functions of the immersion are eigenfunctions of the Laplacian on $\hat{M}$ with eigenvalue k/α^2.

A theorem of Takahashi [1] gives the converse of this last result, namely, if $\Phi\colon \hat{M}^k \to \mathbb{R}^n$ is a Riemannian immersion such that

$$\Delta\Phi + \lambda\Phi = 0 \tag{11}$$

on all of $\hat{M}$, where λ is a nowhere vanishing function on $\hat{M}$, then Φ is a minimal Riemannian immersion into $\mathbb{S}^{n-1}(\alpha)$ for some α, and $\lambda \equiv k/\alpha^2$. Indeed, because $\Delta\Phi$ is always normal to $\hat{M}$, we have from (11) that

$$0 = \langle \xi, \Delta\Phi \rangle = -\lambda\langle \Phi, \xi \rangle$$

for all $\xi \in T\hat{M}$. Since λ never vanishes, we have $\langle \Phi, \xi \rangle = 0$ for all $\xi \in T\hat{M}$, from which one concludes immediately that $|\Phi|$ is constant on $\hat{M}$. So, Φ maps into $\mathbb{S}^{n-1}(\alpha)$ for some $\alpha > 0$. Therefore,

$$\begin{aligned} -\lambda\alpha^2 = \langle \Phi, \Delta\Phi \rangle &= \sum_{A=1}^{n} \Phi^A \Delta\Phi^A \\ &= \sum_{A=1}^{n} \{\operatorname{div}(\Phi^A \operatorname{grad} \Phi^A) - |\operatorname{grad} \Phi^A|^2\} \\ &= \tfrac{1}{2}\Delta|\Phi|^2 - \sum_A |\operatorname{grad} \Phi^A|^2 \\ &= -\sum_A |\operatorname{grad} \Phi^A|^2. \end{aligned}$$

But for any Riemannian immersion, one has

$$\sum_A |\operatorname{grad} \Phi^A|^2 = k. \tag{12}$$

The minimality of Φ in $\mathbb{S}^{n-1}(\alpha)$ follows easily.

A closer look at this last argument indicates (using (12)) that

$$\tfrac{1}{2}\Delta|\Phi|^2 = \langle \Phi, \Delta\Phi \rangle + k$$

for any Riemannian immersion $\Phi\colon \hat{M}^k \to \mathbb{R}^n$, the formula being employed in the above situation. For H the mean curvature vector of the immersion, we have

$$\tfrac{1}{2}\Delta|\Phi|^2 = \langle \Phi, H \rangle + k. \tag{13}$$

One easily concludes from (13) that if Ω is a regular domain in $\hat{M}$, with boundary Γ carrying outward (with respect to Ω and still tangent to $\hat{M}$) unit vector field ν, then

$$kV(\Omega) + \iint_{\Omega} \langle \Phi, H \rangle \, dV = \int_{\Gamma} \langle \Phi, \nu \rangle \, dA, \tag{14}$$

a formula of J. H. Jellet [1]. We may now leave, as an exercise for the reader, the following generalization of A. Hurwitz' proof of the classical isoperimetric inequality (Chavel [2]): Assume Γ is connected and $\hat{M}$ is minimal in $\mathbb{R}^n$. Then

$$\sqrt{\lambda(\Gamma)}\, V(\Omega)/A(\Gamma) \leq \sqrt{k-1}/k \tag{15}$$

with equality if and only if Ω is the intersection of a disk in $\mathbb{R}^n$ with a k-dimensional affine space. (Formulas (14), (15) and the result also apply when $k = n$, with H defined to be identically 0.) Note that when $k = 2$, formula (15) is, in fact,

$$L^2(\Gamma) - 4\pi A(\Omega) \geq 0, \tag{16}$$

a result originally due to Carleman [1]. We refer to Chavel [2] for other references, and for the extension of (15) to simply connected ambient spaces of nonpositive sectional curvature.

Takahashi's theorem may be used to construct minimal immersions of homogeneous spaces into spheres, as follows: M is a Riemannian manifold such that its group of isometries G acts transitively on M, that is, given distinct $x, y \in M$, there exists $g \in G$ such that $g \cdot x = y$. For any $g \in G$, we have the usual action g^* acting on functions on M. Since G is a group of isometries, it preserves the volume element of M, and, therefore, its action on functions preserves the L^2-inner product of functions on M.

Let W be a finite-dimensional subspace of $C^\infty(M)$, invariant under the action of G. Then, with the inner product on W inherited from $L^2(M)$, W is viewed as a finite-dimensional Euclidean space with a group G^* of orthogonal transformations. The space M may be mapped into W, by assigning to every $x \in M$, the uniquely determined element of W, $\Lambda(x)$, for which

$$(\Lambda(x), \phi) = \phi(x)$$

for all $\phi \in W$. If $\{\phi_1, \ldots, \phi_l\}$ is an orthonormal basis of W, then

$$\Lambda(x) = \sum_{j=1}^{l} \phi_j(x)\phi_j. \tag{17}$$

One checks that for any $g \in G$, we have

$$g^*(\Lambda(x)) = \Lambda(g^{-1} \cdot x), \tag{18}$$

from which one proves that $\|\Lambda\|$ is constant on M—so Λ maps M into a sphere in W. In fact, from (17), one has that the radius of the sphere is $\sqrt{\dim W/V(M)}$.

From (18) one has that Λ is C^∞. One checks that the pullback of the metric on W to M via Λ is an invariant Riemannian metric with respect to G. Therefore, if for any $x \in M$ the fixed-point group of x acts irreducibly on M_x, then the pullback of the metric on W to M via is conformal to the original metric on M. The homogeneity of M then implies that Λ is a homothety. The homothety is nontrivial as long as Λ is not the constant map, that is, as long as W does not consist exclusively of constant functions. When Λ is a nontrivial homothety, we may rescale the metric on W to make Λ into a local isometry. One finally shows that Λ is a Riemannian covering of M onto $\Lambda(M)$.

For W we pick finite-dimensional eigenspaces corresponding to nonzero eigenvalues of the Laplacian all of whose elements are in L^2. Then Takahashi's theorem will guarantee that Λ is a minimal immersion. The example of the first two eigenspaces on the standard sphere (corresponding to nonzero eigenvalues) are discussed in Berger–Gauduchon–Mazet [1, pp. 175–178].

A closer analysis of the function $\Lambda: M \to W$ yields the result (Li [2]) that if M is a compact homogeneous Riemannian manifold, with diameter d, then

$$\lambda(M) \geq \pi^2/4d^2;$$

and if the fixed-point action on tangent spaces is irreducible, then the estimate can be improved to

$$\lambda(M) \geq \pi^2 n/4d^2.$$

General introductions to minimal submanifolds are to be found in Lawson [1] and Osserman [1]. Papers in which the above discussion is developed are, for example, Calabi [2], doCarmo–Wallach [1], Li [3], and Wallach [1].

7. NORMALIZATION OF GEOMETRIC DATA

We collect here various formulas describing the effect of changing a given Riemannian metric $\langle\ ,\ \rangle$, on an n-dimensional manifold M, to the Riemannian metric

$$\langle\langle\ ,\ \rangle\rangle = \sigma^2\langle\ ,\ \rangle. \tag{18}$$

We denote the new data with the same notation as the original data except for the ^ over the appropriate symbol.

We start by assuming that σ is a positive function on M, thereby describing a *conformal change of metric*, that is, the measurement of angles in the new metric is the same as in the original one.

For the expressions of the Riemannian metric, given in local coordinates by (I.21), we have

$$\hat{g}_{jk} = \sigma^2 g_{jk}, \qquad \hat{g}^{jk} = \sigma^{-2} g^{jk}, \qquad \hat{g} = \sigma^{2n} g. \tag{20}$$

Therefore, we have for n-dimensional volumes

$$\widehat{dV} = \sigma^n \, dV, \tag{21}$$

and for $(n - 1)$-dimensional volumes

$$\widehat{dA} = \sigma^{n-1} \, dA. \tag{22}$$

We also have, for a function f on M,

$$\widehat{\operatorname{grad}} \, f = \sigma^{-2} \operatorname{grad} f. \tag{23}$$

When $n = 2$, the Dirichlet integral (I.77) satisfies

$$\hat{D}[f, h] = D[f, h]. \tag{24}$$

One also has for the Laplacian

$$\hat{\Delta} = \sigma^{-2}\Delta; \tag{25}$$

so the metrics have the same collection of harmonic functions. From among the many, and far-reaching, consequences of this fact we cite one striking example: The Poisson integral formula for the unit disk in the Euclidean plane, solving the Dirichlet problem for the unit 2-disk, becomes, when the unit 2-disk is endowed with the hyperbolic metric, the solution to the Dirichlet problem on the complete hyperbolic plane for given "boundary values at infinity." In particular, Liouville's theorem, that there are no bounded nonconstant entire harmonic functions, is false for the hyperbolic plane. (Actually, Poisson integral formulas also exist for higher-dimensional hyperbolic spaces, but they are not nearly as effortless as the 2-dimensional case.)

We also note that when $n = 2$, the Gauss curvatures are related by

$$\hat{K} = \sigma^{-2}(K - 2\Delta \ln \sigma).$$

We now assume that σ is a positive constant; so the new Riemannian metric is *homothetic* to the original one, that is, in addition to the same measurement of angles, one has that the distortion of length of tangent vectors is constant on all of TM. Of course we have as mentioned (20)–(23). Equation (24) becomes

$$\hat{D}[f, h] = \sigma^{n-2} D[f, h]. \tag{26}$$

Equation (25) still remains valid. So any eigenfunction of the Laplacian Δ remains an eigenfunction of $\hat{\Delta}$, except that if λ was the original eigenvalue, the new eigenvalue $\hat{\lambda}$ will be

$$\hat{\lambda} = \sigma^{-2}\lambda. \tag{27}$$

The Levi–Civita connection $\hat{\nabla}$ is the same as the original one ∇, and the Riemann curvature tensor $\hat{R}$ is equal to R. For sectional curvatures one has

$$\hat{K} = \sigma^{-2}K, \tag{28}$$

for Ricci curvatures

$$\widehat{\mathrm{Ric}} = \sigma^{-2}\,\mathrm{Ric}, \tag{29}$$

and for scalar curvatures

$$\hat{S} = \sigma^{-2}S. \tag{30}$$

For Cheeger constants (Definition IV.1) one has

$$\hat{\mathfrak{h}} = \sigma^{-1}\mathfrak{h}, \tag{31}$$

and for the Sobolev and isoperimetric constants (Definitions IV.2, 3) one has

$$\hat{\mathfrak{s}} = \mathfrak{s}, \qquad \hat{\mathfrak{I}} = \mathfrak{I}. \tag{32}$$

8. GEODESIC COORDINATES

We are given the Riemannian manifold M and a point $p \in M$. Recall from Section III.1 that $\mathfrak{D}_p$ denotes the largest open set in M_p such that for any $\xi \in \mathfrak{D}_p$, the geodesic

$$\gamma_\xi(t) = \exp t\xi$$

minimizes the distance from p to $\gamma_\xi(t)$, for all $t \in [0, 1]$. The image of $\mathfrak{D}_p$ under the exponential map is denoted by D_p, and exp maps $\mathfrak{D}_p$ diffeomorphically onto D_p.

For fixed $\xi \in \mathfrak{S}_p$, let τ_t denote parallel translation by t units along γ_ξ,

$$\mathcal{R}(t)\eta = \tau_{-t}\{R(\gamma_\xi'(t), \tau_t\eta)\gamma_\xi'(t)\}$$

for any $\eta \in M_p$ (where R is the Riemann curvature tensor), the orthogonal complement of $\mathbb{R}\xi$ in M_p, $\mathscr{A}(t; \xi)$ the path of linear transformations of $\xi^{\perp}$ satisfying

$$\mathscr{A}'' + \mathscr{R}\mathscr{A} = 0 \tag{33}$$

with initial conditions

$$\mathscr{A}(0; \xi) = 0, \qquad \mathscr{A}'(0; \xi) = I, \tag{34}$$

and

$$\sqrt{\mathfrak{g}}(t; \xi) = \det \mathscr{A}(t; \xi).$$

Then the Riemannian metric on D_p can be expressed by

$$ds^2(\exp t\xi) = dt^2 + |\mathscr{A}(t; \xi)\, d\xi|^2,$$

with volume element

$$dV(\exp t\xi) = \sqrt{\mathfrak{g}}(t; \xi)\, dt\, d\mu_p(\xi),$$

where $d\mu_p$ is the $(n-1)$-dimensional volume element of $\mathfrak{S}_p$. The $(n-1)$-dimensional volume element of $\mathsf{S}(p; t) \cap \mathsf{D}_p$ is given by

$$dA(\exp t\xi) = \sqrt{\mathfrak{g}}(t; \xi)\, d\mu_p(\xi).$$

Note that for small t we have the Taylor expansions

$$\mathscr{A}(t; \xi) = tI - t^3\mathscr{R}(0)/6 + O(t^4),$$

$$\sqrt{\mathfrak{g}}(t; \xi) = t^{n-1}(1 - t^2 \operatorname{Ric}(\xi, \xi)/6 + O(t^3)),$$

from which one obtains the classical formula

$$S(p) = \lim_{t \to 0} \frac{\mathfrak{c}_{n-1}t^{n-1} - A(\mathsf{S}(p; t))}{\mathfrak{c}_{n-1}t^{n+1}/6n} = \lim_{t \to 0} \frac{\boldsymbol{\omega}_n t^n - V(\mathsf{B}(p; t))}{\mathfrak{c}_{n-1}t^{n+2}/6n(n+2)}.$$

We now consider *Riemann normal coordinates on* D_p. Fix an orthonormal basis $\{e_1, \ldots, e_n\}$ of M_p and map $\mathbf{n}\colon \mathfrak{D}_p \to \mathbb{R}^n$ by

$$\mathbf{n}^j(q) = \langle (\exp|\mathsf{D}_p)^{-1}(q), e_j \rangle,$$

that is,

$$\zeta = \sum_j \zeta^j e_j \quad \Rightarrow \quad \mathbf{n}^j(\exp \zeta) = \zeta^j.$$

For fixed $\xi \in \mathfrak{S}_p$ we have

$$\mathbf{n}^j(\gamma_\xi(t)) = t\xi^j, \qquad \gamma'_\xi(t) = \sum_j \xi^j \partial_j(\gamma_\xi(t));$$

and the vector field $\partial_j(\gamma_\xi(t))$ along $\gamma_\xi(t)$ is given by

$$\partial_j(\gamma_\xi(t)) = (\exp_p)_{*|t\xi} = t^{-1} Y_j(t),$$

where $Y_j(t)$ is the Jacobi field along γ_ξ satisfying

$$Y_j(0) = 0, \qquad (\nabla_t Y)(0) = e_j.$$

One now easily shows that for all $\zeta \in \mathfrak{D}_p$,

$$g_{jk}(\exp \zeta) = \delta_{jk} - \tfrac{1}{3}\langle R(\zeta, e_j)\zeta, e_k\rangle + O(|\zeta|^3),$$

$$g^{jk}(\exp \zeta) = \delta_{jk} + \tfrac{1}{3}\langle R(\zeta, e_j)\zeta, e_k\rangle + O(|\zeta|^3),$$

$$\sqrt{\mathfrak{g}}(\exp \zeta) = 1 - \tfrac{1}{6}\operatorname{Ric}(\zeta, \zeta) + O(|\zeta|^3).$$

Note that

$$\partial_l g_{jk}(p) = 0$$

for all j, k, l and that for any C^2 function f, we have

$$(\Delta f)(p) = \sum_{j=1}^{n} (\partial^2 f/\partial \mathbf{n}^{j2})(p).$$

As an exercise, we consider the eigenvalues of small geodesic disks, namely, we consider $\delta > 0$ for which $\overline{\mathsf{B}(p; \delta)} \subseteq \mathsf{D}_p$, and fix Riemann normal coordinates on $\mathsf{B}(p; \delta)$. An easy argument based on the above formulas, the Rayleigh theorem, and the max–min methods, show that if Λ_k is the kth Dirichlet eigenvalue of the unit disk in $\mathbb{R}^n$, then

$$\lambda_k(\mathsf{B}(p; \delta)) \sim \Lambda_k/\delta^2 \tag{35}$$

as $\delta \downarrow 0$.

A slightly trickier result is that

$$\lambda_k(\mathsf{B}(p; \delta)) \sim \Lambda_k/\delta^2 - S(p)/6 \tag{36}$$

as $\delta \downarrow 0$. To prove (36) we note that for $\varphi = \sqrt{\mathfrak{g}}$ on D_p we have

$$\Delta(\varphi^{-1/2})(\exp \zeta) = \tfrac{1}{6}S(p) + O(|\zeta|)$$

by the above formulas. Next, associate to any $f \in C_c^\infty(\mathsf{B}(p; \delta))$ the function $F \in C_c^\infty(\mathsf{B}(p; \delta))$ given by

$$F = \varphi^{1/2} f.$$

Then

$$\Delta f = F\Delta(\varphi^{-1/2}) + \varphi^{-1/2} \sum_{j,k} \partial_j(g^{jk}\partial_k F).$$

Since φ never vanishes on $\mathsf{B}(p;\delta)$, we conclude that the eigenvalue problem

$$\Delta f + \lambda f = 0$$

is equivalent to the eigenvalue problem

$$LF + \lambda F = 0, \qquad L = \sum_{j,k} \partial_j(g^{jk}\partial_k) + \varphi^{1/2}\Delta(\varphi^{-1/2}).$$

Working with L, and the above arguments, one can obtain (36) as above.

We note that (36) might be viewed as a variant infinitesimal (with respect to δ) Cheng theorem (cf. Theorems III.5, 7) in that raising the "curvature" lowers the eigenvalue.

We now consider *Fermi coordinates based on a submanifold.* Let $\hat{M}^k$ be a closed submanifold of M^n $k < n$, and $\mathfrak{N}\hat{M}$ the *normal bundle of* $\hat{M}$ *in* M, that is, $\mathfrak{N}M = \bigcup_{p\in\hat{M}} \hat{M}_p^{\perp}$ endowed with a natural differentiable structure. Let $\pi\colon \mathfrak{N}\hat{M} \to \hat{M}$ be the projection map and $\mathrm{Exp}\colon \mathfrak{N}\hat{M} \to M$ the exponential map, that is,

$$\mathrm{Exp} = \exp|\mathfrak{N}\hat{M}.$$

We also let $\mathfrak{S}\mathfrak{N}\hat{M}$ denote the unit normal bundle, that is,

$$\mathfrak{S}\mathfrak{N}\hat{M} = \mathfrak{N}\hat{M} \cap \mathfrak{S}M.$$

Now fix $\xi \in \mathfrak{S}\mathfrak{N}\hat{M}$, $p = \pi(\xi)$. Then there exists a symmetric $L_\xi\colon \hat{M}_p \to \hat{M}_p$ such that

$$\langle B(u, v), \xi\rangle = \langle L_\xi u, v\rangle,$$

for all $u, v \in \hat{M}_p$, where B is the second fundamental form of M in M (cf. Section 6). L_ξ is often referred to as the *Weingarten map of* ξ. Note that L_ξ is given explicitly by

$$L_\xi u = -(\nabla_u Y)^T,$$

where Y is *any* extension of ξ to a normal vector field along M.

A vector field Y along γ_ξ is said to be *transverse* if

(i) $Y(0) \in M_{\pi(\xi)}$,
(ii) $\langle (\nabla_t Y)(0) + L_\xi(Y(0)), M_p\rangle = 0$.

A transverse vector field arises from a *transverse variation of* γ_ξ as follows: Let $p = \pi(\xi)$, $\sigma(\varepsilon)$ a path in $\hat{M}$ with $\sigma(0) = p$, $\xi(\varepsilon)$ a unit vector field along σ, orthogonal to $\hat{M}$, with $\xi(0) = \xi$, and let

$$v(t, \varepsilon) = \mathrm{Exp}\; t\xi(\varepsilon).$$

Then $\eta \equiv: \partial_\varepsilon v|_{\varepsilon=0}$ is transversal. Indeed, $\eta(0) = \sigma'(0) \in \hat{M}_p$. Furthermore, since $L_\xi(\eta(0)) \in \hat{M}_p$, we have at $t = \varepsilon = 0$,

$$\begin{aligned}\langle (\nabla_t \eta)(0) + L_\xi(\eta(0)), \hat{M}_p \rangle &= \langle \nabla_t \partial_\varepsilon v - (\nabla_\varepsilon \partial_t v)^T, \hat{M}_p \rangle \\ &= \langle \nabla_t \partial_\varepsilon v - \nabla_\varepsilon \partial_t v, \hat{M}_p \rangle = 0.\end{aligned}$$

It is standard that the collection of transverse Jacobi fields along γ_ξ is an n-dimensional vector space; that $t\gamma'_\xi(t)$ is always a transverse Jacobi field; that the Wronskian

$$\langle \nabla_t X, Y \rangle - \langle \nabla_t Y, X \rangle$$

of transverse Jacobi fields X, Y along γ_ξ is constant; and that the transverse Jacobi fields along γ_ξ, always orthogonal to γ_ξ, are an $(n-1)$-dimensional subspace of all the transverse Jacobi fields along γ_ξ.

The point $\gamma_\xi(\beta)$, $\beta > 0$, is said to be a *focal point of* $\hat{M}$ *along* γ_ξ if there exists a nonzero transverse Jacobi field Y, along γ_ξ, vanishing for $t = \beta$. The classical argument of Jacobi shows that γ_ξ cannot minimize distance to $\hat{M}$ past a focal point of $\hat{M}$ along γ_ξ. It is also standard that to each $\xi \in \mathfrak{S}\mathfrak{N}\hat{M}$ the distance

$$c(\xi) \equiv: \sup\{t > 0 : d(\hat{M}, \gamma_\xi(t)) = t\}$$

from $\pi(\xi)$ to its *cut point* $\gamma_\xi(c(\xi))$ *along* γ_ξ, is positive and that γ_ξ is the only minimizing geodesic from $\hat{M}$ to $\gamma_\xi(t)$ for all $t \in (0, c(\xi))$.

We set

$$\begin{aligned}\mathfrak{D}\hat{M} &= \{t\xi \in \mathfrak{N}\hat{M} : 0 \le t < c(\xi), \xi \in \mathfrak{S}\mathfrak{N}\hat{M}\}, \qquad \mathsf{D}\hat{M} = \operatorname{Exp} \mathfrak{D}M, \\ \mathfrak{B}\hat{M}(\delta) &= \{\zeta \in \mathfrak{N}\hat{M} : |\zeta| < \delta\}, \\ \mathfrak{S}\hat{M}(\delta) &= \{\zeta \in \mathfrak{N}\hat{M} : |\zeta| = \delta\}, \\ \mathsf{S}\hat{M}(\delta) &= \{q \in M : d(q, \hat{M}) = \delta\}.\end{aligned}$$

Then we always have

$$\mathsf{B}\hat{M}(\delta) = \operatorname{Exp} \mathfrak{B}\hat{M}(\delta), \qquad \mathsf{S}\hat{M}(\delta) \cap \mathsf{D}\hat{M} = \operatorname{Exp} \mathfrak{S}\hat{M}(\delta) \cap \mathfrak{D}\hat{M}.$$

To describe the Riemannian metric on $\mathsf{D}\hat{M}$ in terms of distance from $\hat{M}$ we let $\xi(\varepsilon)$ be a path in $\mathfrak{N}\hat{M}$, $\sigma(\varepsilon) = (\pi \circ \xi)(\varepsilon)$, and

$$v(t, \varepsilon) = \operatorname{Exp} t\xi(\varepsilon).$$

Then

$$\partial_t v = \gamma'_{\xi(\varepsilon)}, \qquad \nabla_t \partial_t v = 0,$$

and $\partial_t v$ has constant length equal to 1. For $t = 0$ we naturally have

$$(\partial_t v)(0, \varepsilon) = \xi(\varepsilon), \qquad (\partial_\varepsilon v)(0, \varepsilon) = \sigma'(\varepsilon);$$

so

$$\langle \partial_t v, \partial_\varepsilon v\rangle(0, \varepsilon) = 0.$$

On the other hand,

$$\begin{aligned}\partial_t\langle \partial_t v, \partial_\varepsilon v\rangle &= \langle \partial_t v, \nabla_t \partial_\varepsilon v\rangle \\ &= \langle \partial_t v, \nabla_\varepsilon \partial_t v\rangle \\ &= \tfrac{1}{2}\partial_\varepsilon|\partial_t v|^2 \\ &= 0.\end{aligned}$$

So

$$\langle \partial_\varepsilon v, \partial_t v\rangle = 0$$

for all t, ε.

The study of $\partial_\varepsilon v$ is based on the fact that $\partial_\varepsilon v$ is a solution to Jacobi's equation along $\gamma_{\xi(\varepsilon)}$. The initial data are given by

$$(\partial_\varepsilon v)(0, \varepsilon) = \sigma'(\varepsilon)$$

(as mentioned above), and

$$(\nabla_t \partial_\varepsilon v)(0, \varepsilon) = (\nabla_\varepsilon \partial_t v)(0, \varepsilon) = \nabla_\varepsilon \xi(\varepsilon) = -L_{\xi(\varepsilon)}\sigma'(\varepsilon) + (\nabla_\varepsilon \xi(\varepsilon))^N.$$

For $\varepsilon = 0$ set $\xi(0) = \xi$, $\sigma'(0) = \zeta$, and $[(\nabla_\varepsilon \xi)(0)]^N = \eta$. Then $(\partial_\varepsilon v)(t, 0)$ is the Jacobi field along γ_ξ which is the sum of the transverse Jacobi fields Z, Y along γ_ξ, where

(37) $$Z(0) = \zeta, \qquad (\nabla_t Z)(0) = -L_\xi \zeta,$$

(38) $$Y(0) = 0, \qquad (\nabla_t Y)(0) = \eta.$$

When $k = n - 1$, that is, $\dim \hat{M} = \dim M - 1$, then $\hat{M}_p^\perp$ is 1-dimensional, and the Jacobi field Y of (38) vanishes identically. Then for $\xi \in \mathfrak{S}\mathfrak{N}M$, $p = \pi \circ \xi$, we may express the Riemannian metric on $\mathsf{D}\hat{M}$ by

$$ds^2(\operatorname{Exp} t\xi) = dt^2 + |\mathscr{A}(t; \xi)\, dp|^2,$$

where dp denotes the generic element of $\hat{M}_p = \xi^\perp$, and $\mathscr{A}(t; \xi)$ is the solution of (33) satisfying

(39) $$\mathscr{A}(0; \xi) = I, \qquad \mathscr{A}'(0; \xi) = -L_\xi.$$

The volume element on $\mathsf{D}\hat{M}$ is given by

$$dV(\operatorname{Exp} t\xi) = \sqrt{\mathfrak{g}}(t; \xi)\, dt\, \widehat{dA}(\pi \circ \xi),$$

where $\widehat{dA}$ is the (n − 1)-dimensional volume element of $\hat{M}$. If we write dA for the $(n - 1)$-dimensional volume element on $\mathsf{S}\hat{M}(\delta) \cap \mathsf{D}\hat{M}$, then we have

$$dA(\operatorname{Exp} \delta\xi) = \sqrt{\mathfrak{g}}(\delta; \xi)\, \widehat{dA}(\pi \circ \xi).$$

9. THE LÉVY–GROMOV ISOPERIMETRIC INEQUALITY

We first require a Bishop comparison theorem for Fermi coordinates based on a codimension 1 closed submanifold $\hat{M}$ of a given Riemannian manifold M. Such estimates go back to Berger [1], Grossman [1], and, more recently, to Heintze–Karcher [1]. We keep the notation from the end of the previous section.

We assume, for convenience, that M is complete, that κ is a real constant satisfying

$$\operatorname{Ric}(\zeta, \zeta) \geq (n-1)\kappa|\zeta|^2$$

for all $\zeta \in TM$, and that δ is a real constant, and ξ an element of $\mathfrak{S}\mathfrak{N}M$ for which

$$\operatorname{tr} L_\xi \geq (n-1)\delta.$$

Let $\beta_{\kappa,\delta}$ be the first zero of

$$\mathbf{J}_{\kappa,\delta}(t) \equiv: \{\mathbf{C}_\kappa(t) - \delta \mathbf{S}_\kappa(t)\}^{n-1}$$

in $(0, +\infty]$. Then for $\mathscr{A}(t; \xi)$ satisfying (33) with initial conditions (39), and β the first zero of $\sqrt{\mathfrak{g}}(t; \xi)$ in $(0, +\infty]$, we have

$$\beta \leq \beta_{\kappa,\delta} \tag{40}$$

and

$$\sqrt{\mathfrak{g}}(t; \xi) \leq \mathbf{J}_{\kappa,\delta}(t) \tag{41}$$

on $[0, \beta]$, with equality in (41) at $t_0 \in (0, \beta]$ if and only if

$$L_\xi = \delta I, \tag{42}$$

and

$$\mathscr{R} = \kappa I, \qquad \mathscr{A}(\ ; \xi) = (\mathbf{C}_\kappa - \delta \mathbf{S}_\kappa) I \tag{43}$$

on all of $[0, t_0]$.

Note that in the spaceform M_κ with the hypersurface $\hat{M}$ consisting of a geodesic sphere of radius $\beta_{\kappa,\delta}$, we have

$$\sqrt{\mathfrak{g}}(t; \xi) = \mathbf{J}_{\kappa,\delta}(t)$$

for ξ pointing into the sphere and

$$\sqrt{\mathfrak{g}}(t; \xi) = \mathbf{J}_{\kappa,-\delta}(t)$$

for ξ pointing outward from the sphere.

To prove (40), (41), set

$$\tau = \min\{\beta, \beta_{\kappa,\delta}\}, \qquad U = \mathscr{A}'\mathscr{A}^{-1}, \qquad \varphi = \operatorname{tr} U = \sqrt{\mathbf{g}}'/\sqrt{\mathbf{g}};$$

then, as in the proof of the Bishop theorem (Section III.3), φ satisfies the differential inequality

$$\varphi' + \varphi^2/(n-1) + \kappa(n-1) \leq 0. \tag{44}$$

For

$$\psi = (n-1)\{\mathbf{C}_\kappa - \delta\mathbf{S}_\kappa\}'/\{\mathbf{C}_\kappa - \delta\mathbf{S}_\kappa\},$$

we have

$$\psi' + \psi^2/(n-1) + \kappa(n-1) = 0,$$
$$\psi' = -(n-1)(\kappa + \delta^2)/\{\mathbf{C}_\kappa - \delta\mathbf{S}_\kappa\}^2,$$
$$\varphi(0) \leq -(n-1)\delta = \psi(0).$$

If $\kappa + \delta^2 \neq 0$, then ψ' never vanishes on $[0, \tau)$, so there exists a C^∞ function $\theta(t)$ defined on $[0, \tau)$ by

$$\varphi(t) = \psi(\theta(t)),$$

which implies

$$(\psi' \circ \theta)(1 - \theta') \geq 0.$$

Should $\kappa + \delta^2 > 0$, then ψ' is always negative, and $\theta(0) \geq 0$, $\theta' \geq 1$ on $[0, \tau)$, from which one has $\theta(t) \geq t$ and

$$\varphi(t) \leq \psi(t),$$

which implies (41) on $[0, \tau)$. In particular, $\beta = \tau$, which implies (40), (41). If we have $\varphi(t_0) = \psi(t_0)$ for any $t_0 \in (0, \beta)$, then $\theta' = 1$ and $\varphi = \psi$ on all of $[0, t_0]$. The conclusions (42), (43) then follow as in the Bishop theorem.

If $\kappa + \delta^2 < 0$ then one argues in a similar fashion.

If $\kappa + \delta^2 = 0$, then for sufficiently small $\varepsilon > 0$, $\kappa + (\delta - \varepsilon)^2 \neq 0$, and if we let ψ_ε denote the function ψ but with δ replaced by $\delta - \varepsilon$, we then have

$$\varphi \leq \psi_\varepsilon$$

on $[0, \tau)$ for all $\varepsilon > 0$, which implies

$$\varphi(t) \leq \psi(t) = -(n-1)\delta$$

on $[0, \tau)$ which yields inequality (41) with

$$\mathbf{J}_{\kappa,\delta}(t) = e^{-(n-1)\delta t}.$$

If there exists $t_0 \in [0, \beta]$ for which $\sqrt{\mathfrak{g}}(t_0) = e^{-(n-1)\delta t_0}$, then $\varphi(t) = -(n-1)\delta$ and $\varphi'(t) = 0$ on all of $[0, t_0)$. Thus we have equality in (43) on all of $[0, t_0)$, which implies (42), and (43) on all of $[0, t_0]$.

Now also assume that M is compact, and $\hat{M}$ a compact, codimension 1, submanifold of M dividing M into domains M_1, M_2. Then there exist $\delta \in (-\infty, +\infty)$ and

$$\rho_1 \leq \beta_{\kappa,\delta}, \qquad \rho_2 \leq \beta_{\kappa,-\delta}$$

such that

$$\rho_1 + \rho_2 \leq d(M), \tag{45}$$

$$V(M_1) \leq A(\hat{M}) \int_0^{\rho_1} \mathbf{J}_{\kappa,\delta}, \tag{46}$$

$$V(M_2) \leq A(\hat{M}) \int_0^{\rho_2} \mathbf{J}_{\kappa,-\delta}. \tag{47}$$

Indeed, let $\hat{M}'$ vary over all compact, codimension 1, submanifolds of M which divide M into domains M'_1, M'_2 satisfying

$$V(M_1) = V(M'_1).$$

Let $\hat{M}_0$ be the submanifold in this collection for which $A(\hat{M}')$ achieves a minimum. Then the regular points of $\hat{M}_0$ have constant mean curvature, say δ. (We leave the discussion of the existence of $\hat{M}_0$, and the lack of effect of possible singularities on the ensuing argument, to Gromov [1] and Schoen [1].) Furthermore, the domains $M_{1,0}$, $M_{2,0}$ are covered by Exp $\mathfrak{N}\hat{M}$. Then the claim follows easily from (40).

An easy application of the above result is that the Cheeger constant $\mathfrak{h}(M)$ satisfies

$$\mathfrak{h}(M) \geq \left\{ \int_0^{d(M)} \mathbf{C}_\kappa^{n-1} \right\}^{-1},$$

which by Cheeger's inequality (Theorem IV.11) provides a lower bound for $\lambda(M)$ in terms of the diameter of M, and a lower bound of the Ricci curvature of M—a result originally due to Li–Yau [1] (their estimate being sharper).

A second application is the isoperimetric inequality discussed in Remarks 1, 2 of Section IV.2. Again M is compact with Ricci curvature bounded below by $(n-1)\kappa$, and $\kappa > 0$. Set

$$\beta = V(M)/V(\mathbb{M}_\kappa).$$

Then $\beta < 1$ by the Bonnet–Myers theorem (Section III.3). Let Ω be a domain in M with smooth boundary, and D the geodesic disk in $\mathbb{M}_\kappa$ for which

$$\beta = V(\Omega)/V(D).$$

Then

$$A(\partial\Omega) \geq \beta A(\partial D), \tag{48}$$

with equality in (48) if and only if M is isometric to $\mathbb{M}_\kappa$, and Ω is isometric to D. Indeed, there exist ρ_1, ρ_2 and δ so that

$$A(\partial\Omega) \geq V(\Omega)\left\{\int_0^{\rho_1} \mathbf{J}_{\kappa,\delta}\right\}^{-1} \geq V(\Omega)\,\frac{A(\mathsf{S}_\kappa(\beta_{\kappa,\delta}))}{V(\mathsf{B}_\kappa(\beta_{\kappa,\delta}))},$$

that is,

$$A(\partial\Omega) \geq V(\Omega)A(\mathsf{S}_\kappa(\beta_{\kappa,\delta}))/V(\mathsf{B}_\kappa(\beta_{\kappa,\delta})), \tag{49}$$

and, at the same time,

$$A(\partial\Omega) \geq V(M\backslash\Omega)A(\mathsf{S}_\kappa(\beta_{\kappa,-\delta}))/V(\mathsf{B}_\kappa(\beta_{\kappa,-\delta})). \tag{50}$$

Furthermore, the function $A(\mathsf{S}_\kappa(\tau))/V(\mathsf{B}_\kappa(\tau))$ is strictly decreasing with respect to τ. So, if $\beta_{\kappa,\delta}$ is less than the radius of D, then (49) implies

$$A(\partial\Omega) \geq V(\Omega)A(\partial D)/V(D) = \beta A(\partial D).$$

If, on the other hand, $\beta_{\kappa,\delta}$ is greater than the radius of D, then $\pi/\sqrt{\kappa} - \beta_{\kappa,\delta} = \beta_{\kappa,-\delta}$ is less than the radius of $\mathbb{M}_\kappa\backslash D$, and (50) implies

$$A(\partial\Omega) \geq V(M\backslash\Omega)A(\partial(\mathbb{M}_\kappa\backslash D)/V(\mathbb{M}_\kappa\backslash D) = \beta A(\partial D).$$

So (48) is proven. The case of equality is easy.

10. HEAT CONDUCTION ON THE EUCLIDEAN UPPER HALF-SPACE

In this section we give an explicit solution to the initial-boundary value problem for the heat equation on the Euclidean upper half-space:

$$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x^n > 0\}.$$

Although this example does not fit into the theory of Chapter VII, it has the advantage of possessing easily obtainable formulas.

The idea is that to each $x = (x^1, \ldots, x^n)$ in $\mathbb{R}^n_+$ we associate its reflection in the plane $x^n = 0$, $x^* = (x^1, \ldots, -x^n)$; and, for

$$\mathrm{e}(x, y, t) = (4\pi t)^{-n/2}e^{-|x-y|^2/4t}$$

the heat kernel of $\mathbb{R}^n$, and

$$\psi(x, y) = \begin{cases} 1/(n-2)\mathbf{c}_{n-1}|x-y|^{n-2}, & n > 2, \\ -(1/2\pi)\ln|x-y|, & n = 2, \end{cases}$$

the Newtonian potential of $\mathbb{R}^n$, we define

$$q(x, y, t) = \mathfrak{e}(x, y, t) - \mathfrak{e}(x, y^*, t),$$

$$G(x, y) = \psi(x, y) - \psi(x, y^*),$$

to be our candidates for the Dirichlet heat kernel and Green's function, respectively, on $\mathbb{R}^n_+$. Both functions satisfy (i) their appropriate differential equation in the x variable, (ii) are perturbations of the free space kernels by global solutions of the respective differential equations, (iii) are symmetric in the space variables, (iv) are positive on $\mathbb{R}^n_+$, and vanish on $\partial\mathbb{R}^n_+ = \mathbb{R}^{n-1}$.

For the moment, assume that $n > 2$. Then from the substitution

$$\tau = |x - y|^2/4t$$

we have

$$\begin{aligned} \int_0^\infty \mathfrak{e}(x, y, t)\,dt &= \frac{|x-y|^{-n+2}}{4\pi^{n/2}} \int_0^\infty e^{-\tau}\tau^{n/2-2}\,d\tau \\ &= |x-y|^{-n+2}\Gamma(n/2-1)/4\pi^{n/2} = \psi(x, y) \end{aligned}$$

(cf. Section XII.1), which is the formula (VII.49) for $\Omega = \mathbb{R}^n$, $n > 2$, which implies, in turn, (VII.49) for $\Omega = \mathbb{R}^n_+$ with q, G above.

Now suppose that f: $\mathbb{R}^n_+ \to \mathbb{R}$, φ: $\mathbb{R}^{n-1} \to \mathbb{R}$ are bounded and continuous, and seek a solution $u(x, t)$ to the heat equation on $\mathbb{R}^n_+ \times (0, +\infty)$ with

$$\lim_{t \downarrow 0} u(\ , t) = f, \qquad \lim_{x \to w \in \partial\Omega} u(x,\ ) = \varphi(w).$$

Then (VII.27) suggests that we consider

$$\begin{aligned} u(x, t) = &\iint_{\mathbb{R}^n_+} q(x, y, t) f(y)\,dV(y) \\ &- \int_0^t d\tau \int_{\mathbb{R}^{n-1}} \frac{\partial q}{\partial \nu_w}(x, w, t)\varphi(w)\,dA(w). \end{aligned}$$

To treat the two integrals, we set

$$u_1(x, t) = \iint_{\mathbb{R}^n_+} q(x, y, t) f(y)\,dV(y),$$

$$u_2(x, t) = -\int_0^t d\tau \int_{\mathbb{R}^{n-1}} \frac{\partial q}{\partial \nu_w}(x, w, t)\varphi(w)\,dA(w).$$

First note that

$$|x - y^*| \geq |x - y|,$$

so that

$$0 < q < \mathbf{e}$$

on all of $\mathbb{R}^n_+$. Also, whenever x is restricted to a compact subset K in $\mathbb{R}^n_+$, then

$$|x - y^*| \geq d(K, \mathbb{R}^{n-1}) > 0$$

for all y in $\mathbb{R}^n_+$. Using these comments and the arguments of Section VI.2, one can easily verify the convergence of the integral u_1; that differentiation of u_1 with respect to x and t may be carried out under the integral sign; and, therefore, u_1 is a solution to the heat equation; and that

$$\lim_{t \downarrow 0} u_1(\ , t) = f, \qquad \lim_{x \to \partial\Omega} u_1 = 0.$$

To estimate u_1 for large values of t, one has

$$\begin{aligned} u_1(x, t) &\leq \{\sup|f|\} \iint_{\mathbb{R}^n_+} \{\mathbf{e}(x, y, t) - \mathbf{e}(x, y^*, t)\}\, dV(y) \\ &= \frac{\{\sup|f|\}}{\sqrt{4\pi t}} \int_0^\infty \{e^{-(x^n - y^n)^2/4t} - e^{-(x^n + y^n)/4t}\}\, dy^n \\ &= \frac{\{\sup|f|\}}{\sqrt{4\pi t}} \left[\int_0^A + \int_A^\infty\right] \end{aligned}$$

for any $A > 0$. Given any $\varepsilon > 0$, there exists $A > 0$ such that

$$(1 - \varepsilon)(y^n + x^n) \leq y^n - x^n \leq y^n + x^n$$

for all $y^n \geq A$. Fix this value of A; then

$$\begin{aligned} \int_A^\infty \{e^{-(y^n - x^n)^2/4t} - e^{-(y^n + x^n)^2/4t}\}\, dt \leq \{\sqrt{4t}/(1 - \varepsilon)\} \int_{(1-\varepsilon)(A + x^n)/\sqrt{4t}}^\infty e^{-z^2}\, dz \\ - \sqrt{4t} \int_{(A + x^n)/\sqrt{4t}}^\infty e^{-z^2}\, dz, \end{aligned}$$

from which one concludes

$$\limsup_{t \uparrow +\infty} u_1 \leq \{\sup|f|\}\{(1 - \varepsilon)^{-1} - 1\}/2$$

for all $\varepsilon > 0$. Therefore

$$\lim_{t\uparrow+\infty} u_1 = 0.$$

To study u_2 note that

$$(\partial q/\partial v_w)(x, w, t) = -(x^n/t)\mathrm{e}(x, w, t);$$

so

$$u_2(x,t) = \int_0^t d\tau \int_{\mathbb{R}^{n-1}} (x^n/t)\mathrm{e}(x, w, \tau)\varphi(w)\, dA(w). \tag{51}$$

We leave it to the reader to establish that u_2 is a solution to the heat equation, identically equal to 0 at time $t = 0$.

If we consider the substitutions

(i) $w = \hat{x} + r\xi, \quad r > 0, \quad \xi \in \mathbb{S}^{n-2},$

(ii) $r = 2\sqrt{\tau}s,$

(iii) $\rho = (x^n)^2/4t,$

then

$$u_2(x,t) = \pi^{-n/2}\int_{(x^n)^2/4t}^{\infty} \rho^{-1/2}e^{-\rho}\, d\rho \int_0^\infty e^{-s^2}s^{n-2}\, ds \cdot \int_{\mathbb{S}^{n-2}} \varphi(\hat{x} + x^n s\xi/\sqrt{\rho})\, d\mathbb{S}^{n-2}(\xi),$$

where $d\mathbb{S}^{n-2}$ is the $(n-2)$-dimensional volume element on $\mathbb{S}^{n-2}$. We leave it to the reader to legitimize the formal limit:

$$\lim_{x^n\downarrow 0} u_2(x,t) = \varphi(\hat{x})\mathrm{c}_{n-2}\pi^{-n/2}\int_0^\infty \rho^{-1/2}e^{-\rho}\, d\rho \int_0^\infty e^{-s^2}s^{n-2}\, ds = \varphi(\hat{x})$$

by Section 1.

To investigate what happens to $u_2(x, t)$ as $t \uparrow +\infty$, we apply the substitution

$$\rho = |x - w|^2/4t$$

directly to (50), and obtain

$$u_2(x,t) = \frac{x^n}{\pi^{n/2}}\int_{\mathbb{R}^{n-1}} \frac{\varphi(w)}{|x - w|^n}\, dA(w)\int_{|x-w|^2/4t}^{\infty} e^{-\rho}\rho^{n/2-1}\, d\rho.$$

It is not too hard to justify the formal limit

$$\begin{aligned}\lim_{t\uparrow+\infty} u_2(x,t) &= \frac{x^n}{\pi^{n/2}}\int_{\mathbb{R}^{n-1}} \frac{\varphi(w)}{|x-w|^n}\,dA(w)\int_0^\infty e^{-\rho}\rho^{n/2-1}\,d\rho \\ &= \frac{x^n\Gamma(n/2)}{\pi^{n/2}}\int_{\mathbb{R}^{n-1}} \frac{\varphi(w)}{|x-w|^n}\,dA(w) \\ &= \frac{2x^n}{\mathbf{c}_{n-1}}\int_{\mathbb{R}^{n-1}} \frac{\varphi(w)}{|x-w|^n}\,dA(w) \\ &= -\int_{\mathbb{R}^{n-1}} (\partial G/\partial \nu_w)(x,w)\varphi(w)\,dA(w),\end{aligned}$$

which is the classical solution to the Dirichlet problem for Laplace's equation $\Delta u = 0$ on the Euclidean upper half-space.

Our calculation, therefore, yields Theorem VII.6 for $\mathbb{R}^n_+$.

11. THE MAXIMUM PRINCIPLE FOR THE LAPLACIAN

The strong maximum principle. Let M be a Riemannian manifold with Laplace–Beltrami operator Δ, and let u be a bounded C^2 function on M which satisfies

$$\Delta u \geq 0 \tag{52}$$

on M. If there exists x_0 in M for which

$$u(x_0) = \sup_M u \tag{53}$$

then

$$u(x) = u(x_0)$$

on all of M.

Furthermore, if $u \in C^2(M) \cap \bar{C}^1(M)$ (cf. Definition VII.2) satisfies (52) on all of M, and M has a nonempty boundary with $x_0 \in M$ satisfying (53), and the interior sphere condition at x_0 (i.e., there exist $y \in M$, $r > 0$, such that $\mathsf{B}(y;r) \subseteq \mathsf{D}_y \subseteq M$, $x_0 \in \overline{\mathsf{B}(y;r)}$), then

$$\partial u/\partial \nu > 0$$

at x_0.

We refer the reader to Protter–Weinberger [1, Section II.3] for a proof of the theorem when M is diffeomorphic to a domain in Euclidean space. A standard continuation argument then extends the result to all of M.

Of course, if one is given

$$\Delta u \leq 0$$

instead of (52), on all of M, then one has a corresponding minimum principle. For solutions of Laplace's equation, both principles are valid.

12. RECENT PROGRESS ON EIGENVALUE AND HEAT KERNEL ESTIMATES

We start by noting that for Neumann eigenvalues $\{0 = \mu_0 < \mu_1 \leq \mu_2 \leq \cdots\}$ of a regular domain M in $\mathbb{R}^n$, the Polya conjecture (cf. Remark II.1) reads as

$$(\mu_k)^{n/2} \leq \{(2\pi)^n/\omega_n\} k/\mathrm{vol}\, M$$

for all $k = 0, 1, 2, \ldots$. The conjecture has yet to be proved or disproved. Partial results can be found in Li-Yau [4]. Compare also the discussion below.

Next, recall that Theorem IV.9 gives estimates for lower bounds of Dirichlet eigenvalues in terms of the Sobolev(-isoperimetric) constant and the volume. Furthermore, once one knows the Sobolev constant, the estimates of the eigenvalues have the form

$$\lambda_k \geq \mathrm{const} \cdot k^{1/(n-1)} V^{-2/n}$$

(for example, when $n > 2$). Of course, it is desirable to improve the above estimates to the form

$$\lambda_k \geq \mathrm{const} \cdot (k/V)^{2/n},$$

which will reflect the growth indicated by the Weyl formula (naturally, for regular domains in $\mathbb{R}^n$, the Polya conjecture is that the constant *is* that of the Weyl formula.) As remarked in Remark IV.3, any estimate of the form

$$\sum_{k=1}^{\infty} e^{-\lambda_k t} \leq A t^{-n/2} \tag{54}$$

implies

$$\lambda_k \geq \{Ae\}^{-2/n} k^{2/n}$$

for all $k = 1, 2, \ldots$. We summarize the argument of Cheng-Li [1], in the spirit of Section IV.4, to obtain an estimate of the form (54).

The idea is that for the Dirichlet heat kernel q of M, we have

$$q(x, x, t) = \int_M q^2(x, y, t/2)\, dV(y),$$

which implies

$$\begin{aligned}(\partial_t q)(x, x, t) &= \int_M (\partial_t q)(x, y, t/2) q(x, y, t/2)\, dV(y)\\ &= \int_M (\Delta_y q)(x, y, t/2) q(x, y, t/2)\, dV(y)\\ &= -\int_M \operatorname{grad}_y q^2(x, y, t/2)\, dV(y).\end{aligned}$$

But from the inequality (IV.26) (we are assuming, for convenience, that $n > 2$) we have, for

$$f(y) = q(x, y, t/2),$$

the inequality

$$\int_M |\operatorname{grad} f|^2\, dV \geq \mathrm{c}(M) \| f^2 \|_{n/(n-2)},$$

where c is given by (IV.27). Hölder's inequality implies

$$\| f^2 \|_{n/(n-2)} \geq \left\{ \int_M f^2 \right\}^{(n+2)/n} \left\{ \int_M |f| \right\}^{-4/n},$$

and naturally, we have

$$\int_M |f|\, dV \leq 1.$$

One concludes

$$(\partial_t q)(x, x, t) \leq -\mathrm{c}(M) \{ q(x, x, t) \}^{(n+2)/n}.$$

From

$$\lim_{t \downarrow 0} q(x, x, t) = +\infty,$$

one now has

$$q(x, x, t) \leq \{ \mathrm{c}(M) t \}^{-n/2} = \{ c(n)/\mathrm{s}(M) \} t^{-n/2}, \tag{55}$$

where $c(n)$ is a constant depending only on n, and $\mathrm{s}(M)$ is the Sobolev(-isoperimetric) constant of M. Thus (55) gives an upper bound on the trace of the heat kernel depending only on the Sobolev(-isoperimetric) constant of M. Therefore,

$$\sum_{k=1}^{\infty} e^{-\lambda_k t} = \int_M q(x, x, t)\, dV(x) \leq \{c(n)V(M)/\mathrm{s}(M)\}t^{-n/2},$$

which implies

$$\lambda_k(M) \geq c(n)\{\mathrm{s}(M)k/V(M)\}^{2/n} \tag{56}$$

for all $k = 1, 2, \ldots$.

Next, let M be a compact Riemannian manifold of dimension $n \geq 2$, diameter d, and volume V having Ricci curvature bounded below by $(n-1)\kappa$, $\kappa \leq 0$. In Section III.3, we derived Cheng's comparison theorem (Theorem III.7) and gave (Remark III.4) a crude upper bound

$$\lambda_j(M) \leq c(n, \kappa, d)j^2$$

(where $c(n, \kappa, d)$ denotes a constant depending on n, κ, and d) for large values of j. For $\kappa = 0$ we had, explicitly,

$$\lambda_j(M) \leq 4\mathrm{c}_D j^2/d^2$$

for all $j = 1, 2, \ldots$. A better estimate was obtained in Li-Yau [1]; namely, for all $j = 1, 2, \ldots$, we have

$$\lambda_j \leq c_1(n, \kappa, d) + c_2(n, \kappa, d)\{(j+1)/V\}^{2/n}, \tag{57}$$

which is consistent with the growth with respect to j, given by the Weyl formula. For $\kappa = 0$, the estimate becomes

$$\lambda_j \leq c(n)\{(j+1)/V\}^{2/n} \tag{58}$$

for all $j = 1, 2, \ldots$.

For lower bounds on eigenvalues, we have already mentioned, immediately following Theorem V.4, that λ_1 may be bounded below in terms of n, κ, d alone (Li-Yau [1]). A weaker estimate in terms of the same data (due to Gromov [1]) was derived, here in Section 9. For $\kappa = 0$, H. Z. Yang and J. Q. Zhong [1] have recently proved the sharp estimate

$$\lambda_1 \geq \pi^2/d^2.$$

For lower bounds on the higher eigenvalues, preliminary estimates were obtained in Li-Yau [1]. These were improved, using upper bounds for the heat kernel, in Li-Yau [4], to yield

$$\lambda_j \geq c(n, j, \kappa, d)$$

for all $j = 0, 1, 2, \ldots$, which, when $\kappa = 0$, becomes

$$\lambda_j \geq c(n)j^{2/n}/d^2$$

for all $j = 0, 1, 2, \ldots$.

Although we have only discussed the eigenvalues of a compact manifold, the above papers of Li and Yau also contain results (and references) for the Dirichlet and Neumann eigenvalues.

In Section VIII.3 we derived lower bounds for the Dirichlet heat kernel on geodesic disks when the Ricci curvature is bounded from below. These were then extended to lower bounds for the heat kernel of a complete noncompact Riemannian manifold with Ricci curvature bounded from below. When M is compact with nonnegative Ricci curvature and heat kernel p, Li and Yau [4] have shown that

$$p(x, x, t) \geq (4\pi t)^{-n/2} \tag{59}$$

for all (x, t) in $M \times (0, +\infty)$. Inequality (59) is also valid when M is a regular domain, with convex boundary, and nonnegative Ricci curvature and p is the Neumann heat kernel. One has, then, for the eigenvalues,

$$\sum_{k=0} e^{-\lambda_k t} \geq (4\pi t)^{-n/2} V(M)$$

when M is compact, and

$$\sum_{k=0} e^{-\mu_k t} \geq (4\pi t)^{-n/2} V(M)$$

for the Neumann eigenvalues of M a regular domain with convex boundary and nonnegative Ricci curvature. (This is the "Polya conjecture" for the Laplace transform of Neumann eigenvalues.)

New upper bounds on the heat kernel have been developed in Li-Yau [4]; namely, if M is a complete Riemannian manifold with Ricci curvature bounded below by $(n - 1)\kappa$, $\kappa \leq 0$, and p is the heat kernel, then

$$p(x, y, t) \leq \frac{c(n, \delta)}{V^{1/2}(B(x; \sqrt{t}))V^{1/2}(B(y; \sqrt{t}))} \exp\left\{\frac{-d^2(x, y)}{(4 + \delta)t} - c(n)\delta\kappa t\right\} \tag{60}$$

for all $\delta > 0$, where $c(n, \delta) \to \infty$ as $\delta \to 0$. Note that the estimate (60) is sharp (except for the $c(n, \delta)$ in the front) when $M = \mathbb{S}^k \times \mathbb{R}^l$, $k + l = n$ (in which case, $\kappa = 0$). If the Ricci curvature is not bounded from below, then one also has a corresponding estimate for x, y restricted to geodesic disks centered at some fixed point.

A corresponding lower bound of the form

$$p(x, y, t) \geq \frac{c(n, \varepsilon)}{V^{1/2}(B(x; \sqrt{t}))V^{1/2}(B(y; \sqrt{t}))} \exp\left\{\frac{-d^2(x, y)}{(4 - \varepsilon)t} + c(n)\varepsilon\kappa t\right\} \tag{61}$$

is also derived in Li–Yau [4].

APPENDIX

Laplacian on Forms

Jozef Dodziuk

The Laplace operator acting on functions has a natural generalization to differential forms. In the theory of this new operator, also called the Laplacian, the interplay among analysis, topology, and geometry is even more striking. One of the most exciting aspects of this interplay is the heat equation approach to index theorems, which will be outlined in Section B for the special case of generalized Gauss–Bonnet formula (cf. Atiyah–Bott–Patodi [1] and Gilkey [2]). References for the classical theory of the Laplacian on forms include deRahm [1], Yano–Bochner [1], and Warner [1]. The aim of this section is to describe the flavor of this beautiful subject and to entice the reader to further exploration. No attempt is made to give proofs or to make an exhaustive survey of the area.

A. DERAHM–HODGE THEORY, VANISHING THEOREMS

For a smooth manifold M^n, let $A^p(M)$ denote the space of C^∞ differential forms of degree $p = 0, 1, 2, \ldots, n$ with real coefficients. The exterior derivative d maps $A^p(M)$ into $A^{p+1}(M)$ and the resulting complex (the deRham complex of M)

$$\text{(A1)} \qquad 0 \to A^0(M) \xrightarrow{d} A^1(M) \xrightarrow{d} \cdots \xrightarrow{d} A^{n-1}(M) \xrightarrow{d} A^n(M) \to 0$$

contains a great deal of topological information about M. Namely, we have

THEOREM A1 (deRham). The cohomology of the deRham complex of M is isomorphic to the singular cohomology of M with real coefficients (more precisely, the isomorphism is given by integration of forms over smooth singular chains).

From now until further notice we require M to be compact, oriented, Riemannian, and without boundary. The Riemannian metric $\langle\ ,\ \rangle$ induces inner products in fibers of various tensor bundles over M, also denoted by $\langle\ ,\ \rangle$. It also induces the $*$ operator mapping forms of degree p to forms of degree $n-p$. For a form $\alpha \in \Lambda^p M_x$, $*\alpha$ is uniquely determined by requiring that

(A2) $$\langle \alpha, \beta \rangle_x \, dV = \beta \wedge * a,$$

for all $\beta \in \Lambda^{n-p} M_x$. Identifying decomposable covectors $\eta_1 \wedge \eta_2 \wedge \cdots \wedge \eta_p$ with $\operatorname{span}\{\eta_1, \eta_2, \ldots, \eta_p\}$ one can interpret $*$ as taking the orthogonal complement. For two smooth forms $\alpha, \beta \in A^p(M)$, we define their L^2 inner product as

(A3) $$(\alpha, \beta) = \int_M \langle \alpha, \beta \rangle \, dV = \int_M \alpha \wedge * \beta.$$

The operator $\delta = (-1)^{np+n+1} * d*$ is the formal adjoint of d in the sense that

(A4) $$(d\alpha, \beta) = (\alpha, \delta\beta)$$

whenever $\alpha \in A^{p-1}(M)$, $\beta \in A^{n-p}(M)$. This is an easy consequence of Stokes's theorem. It is worth mentioning that, under the duality $M_x \cong M_x^*$ induced by the Riemannian metric, $-\delta$ on forms of degree one corresponds to the divergence of vector fields.

The Laplace operator $\Delta_p \colon A^p(M) \to A^p(M)$ is now defined as

(A5) $$\Delta_p = -(d\delta + \delta d), \qquad p = 0, 1, 2, \ldots, n.$$

Note that $\Delta_0 = -\delta d$ is just the Laplacian on functions, since $\delta \,|\, A^0(M) \equiv 0$. One often writes $\Delta_p = \Delta$ if keeping track of degree is not crucial. A form ω is called harmonic if $\Delta\omega = 0$. The space of all harmonic p-forms will be denoted by $\mathscr{H}^p(M)$. By (A4) ω is harmonic if and only if $d\omega = \delta\omega = 0$. Another consequence of (A4) is that Δ is symmetric, that is, $(\Delta\alpha, \beta) = (\alpha, \Delta\beta)$ for $\alpha, \beta \in A^*(M)$. We mention that in case of the flat torus $\mathbb{R}^n/\mathbb{Z}^n$, in the natural coordinate system induced from $\mathbb{R}^n$, the Laplacian is computed by applying the ordinary Euclidean Laplacian to the coefficients of a differential form.

A fundamental result relating the Laplacian to the topology of M is the theorem of Hodge (cf. deRham [1] or Warner [1]).

THEOREM A2. For every $p = 0, 1, 2, \ldots, n$, $A^p(M)$ admits the orthogonal direct sum decomposition

$$A^p(M) = dA^{p-1}(M) \oplus \mathscr{H}^p(M) \oplus \delta A^{p+1}(M),$$

that is, every form $\omega \in A^p(M)$ can be written as $\omega = d\alpha + h + \delta\beta$, with $\alpha \in A^{p-1}(M)$, $\beta \in A^{p+1}(M)$, and $\Delta h = 0$. $d\alpha$, h, and $\delta\beta$ are determined uniquely. As a consequence $\mathscr{H}^p(M)$ is isomorphic to the deRham cohomology of M in dimension p.

One can see easily that Theorem A2 is implied by existence of the orthogonal sum decomposition

$$A^p(M) = \mathscr{H}^p(M) \oplus \Delta A^p(M).$$

In case of functions this reduces to the familiar fact that the equation $\Delta u = f$ can be solved for u if and only if f is orthogonal to constants, that is, if $\int_M f\,dV = 0$. One can see the way the Hodge theorem comes about by pretending that the spaces of differential forms are finite dimensional. If this were so, $(\Delta_p A^p(M))^\perp = \ker \Delta_p^* = \ker \Delta_p = \mathscr{H}^p(M)$ and $A^p(M) = \mathscr{H}^p(M) \oplus \Delta_p A^p(M)$. In general such heuristic, "finite-dimensional" arguments produce correct results for the Laplacian on a compact manifold. However, to justify them one has to work considerably harder. A proof of the Hodge theorem based on the heat equation for Δ_p will be sketched below.

The Hodge theory gives a relation between the topology and analysis on M. One way in which geometry enters is via the following formula which goes back to Bochner (cf. Yano–Bochner [1], or deRham [1]).

PROPOSITION A3. For $0 < p \leq n$ and $\omega \in A^p$

(a) $$\Delta\omega = -(\nabla^*\nabla\omega + F_p\omega),$$

(b) $$\tfrac{1}{2}\Delta|\omega|^2 = \langle\Delta\omega, \omega\rangle + \langle F_p\omega, \omega\rangle + |\nabla\omega|^2.$$

Here ∇ denotes the covariant derivative induced by the Levi–Civita connection, ∇^* is its formal adjoint, and F_p is an algebraic operator involving the Riemann curvature tensor. For $p = 1$, F_p is given by the Ricci tensor. Thus applying the second formula above to a differential df one obtains formula (III.56).

This proposition leads to a method (due to Bochner) of proving "vanishing theorems," that is, theorems which assert that under certain assumptions on curvature some cohomology groups are trivial. A scheme for proving such theorems is given by

THEOREM A4. Suppose $F_p \geq 0$, that is, $\langle F_p\omega, \omega\rangle_x \geq 0$ for all $x \in M$ and all $\omega \in \Lambda^p M_x$. Suppose further that there exists $x_0 \in M$ for which

$\langle F_p\omega, \omega\rangle_{x_0} > 0$ for all nonzero p-forms $\omega \in \Lambda^p\mathrm{M}_x$. Then every harmonic form of degree p is identically zero.

This theorem gives a method for proving vanishing theorems. The real work is always in proving that $F_p > 0$. We indicate how the method works. One can use either (a) or (b) of Proposition A3. Using (a) and $\Delta\omega = 0$, we obtain

$$0 = (\nabla\omega, \nabla\omega) + (F_p\omega, \omega).$$

Both terms are nonnegative so that $\nabla\omega \equiv 0$. It follows that $\langle\omega, \omega\rangle$ is a constant, which must be zero in view of strict positivity of F_p at one point of M. Alternatively, if ω is not identically zero, $|\omega|$ attains a positive maximum at a point $x_0 \in M$. Using (b) we see that

$$\tfrac{1}{2}\Delta|\omega|^2 = \langle F_p\omega, \omega\rangle + \langle\nabla\omega, \nabla\omega\rangle \geq 0$$

at x_0. By the maximum principle $|\omega| = \text{const}$ so that

$$\langle F_p\omega, \omega\rangle + \langle\nabla\omega, \nabla\omega\rangle \equiv 0.$$

As above ω must be zero identically.

It is rather surprising that such a simple result has so many deep consequences (cf. Yano–Bochner [1]). It applies not only to the Laplace operator, but also to the $\bar{\partial}$-Laplacian in Kählerian geometry (cf. Morrow–Kodaira [1]) and to the Dirac operator on spinors (cf. Hitchin [1]). It exhibits very nicely the interaction between spectral properties (the kernel of Δ), topology (cohomology), and geometry (curvature entering via F_p). We quote just one result proved fairly recently by Gallot and Meyer [1].

THEOREM A5. Suppose the curvature operator $\rho\colon \Lambda^2M_x \to \Lambda^2M_x$ has a positive lower bound at every point $x \in M$. Then M is a real homology sphere, that is, $H^*(M, \mathbb{R}) \cong H^*(S^n, \mathbb{R})$.

B. THE HEAT EQUATION ON FORMS

Consider the initial value problem

$$\text{(A6)} \qquad \Delta_p\omega = \partial\omega/\partial t, \qquad \omega(x, 0) = \omega_0(x),$$

where ω is a C^∞ p-form on M depending smoothly on a parameter $t \in [0, \infty)$. This problem admits a fundamental solution, often called the heat kernel, $e_p(x, y, t)$ whose value at a point $(x, y, t) \in M \times M \times (0, \infty)$ is a linear map from Λ^pM_y to Λ^pM_x. Using the duality $M_y \cong M_y^*$ given by the

Riemannian metric we can also think of $e_p(x, y, t)$ as an element of $\Lambda^p M_x \otimes \Lambda^p M_y$, that is, a double form. The $e_p(x, y, t)$ depends smoothly on $(x, y, t) \in M \times M \times (0, \infty)$, and its defining property is the fact that

$$\omega(x, t) = \int_M e_p(x, y, t) \wedge *\omega_0(y)\, dV(y)$$

is a solution of (A5). To be precise we require that $\omega(x, t)$ satisfies the heat equation for $t > 0$ and extends to a C^∞ function on $M \times [0, \infty)$ if ω_0 is smooth. Solutions of (A5) are unique (e.g., by the energy method). Therefore $e_p(x, y, t)$ is unique if it exists.

A modification of the method of Minakshisundaram employed in Chapter VI to construct $e_0(x, y, t) = p(x, y, t)$ yields existence of the fundamental solution of the heat equation on forms for $0 < p \leq n$ (cf. Patodi [1]).

THEOREM B1. Given a C^∞ form $\omega_0 \in A^p(M)$, the unique solution of (A5) is given by

$$\omega(x, t) = \int_M e_p(x, y, t) \wedge *\omega_0(y)\, dV(y), \tag{A7}$$

where the kernel e_p is a symmetric double form on $M \times M$ depending smoothly on a parameter $t > 0$. The kernels $e_p(x, y, t)$ have the properties

$$d_x e_p(x, y, t) = \delta_y e_{p+1}(x, y, t) \tag{A8}$$

for

$$(x, y, t) \in M \times M \times (0, \infty),$$

$$e_p(x, x, t) \sim \sum_{j=0}^{\infty} t^{-(n/2)+j} C_{p,j}(x), \tag{A9}$$

that is, for every integer $k \geq 0$

$$e_p(x, x, t) = \sum_{j=0}^{k} t^{-(n/2)+j} C_{p,j}(x) + O(t^{k-(n/2)})$$

as t approaches 0. Each $C_{p,j}$ is a smooth double form, the estimates are uniform on M, and the coefficients of $C_{p,j}$ are given by universal (but not unique) polynomials in Riemann curvature tensor and its covariant derivatives.

Milgram and Rosenbloom [1] showed that the Laplacian itself can be studied via the heat equation. In particular the Hodge theorem follows from basic properties of the heat kernels $e_p(x, y, t)$. We outline the argument. Let $P(t)\omega(x) = \int_M e_p(x, y, t) \wedge {*\omega(y)}\, dV_y$, for a form $\omega \in A^p(M)$. One first shows that the limit

$$H\omega = \lim_{t \to \infty} P(t)\omega$$

exists for every $\omega \in A^p(M)$. Clearly $P(t)H\omega = H\omega$ and therefore

$$0 = \frac{\partial}{\partial t} H\omega = \frac{\partial}{\partial t} P(t)H\omega = \Delta H\omega,$$

so that $H\omega$ is harmonic. Next consider $G\omega = \int_0^\infty (P(t)\omega - H\omega)\, dt$, and compute

$$\Delta G\omega = \int_0^\infty \Delta P(t)\omega = \int_0^\infty \frac{\partial}{\partial t} P(t)\omega = \lim_{t \to \infty} P(t)\omega - \lim_{t \to 0} P(t)\omega = H\omega - \omega.$$

Of course, in this formal argument convergence of the integral defining $G\omega$ and differentiation under the integral sign have to be justified. This is not difficult since, in fact, $P(t)\omega - H\omega$ decays exponentially as t goes to infinity. The equalities prove that $\omega = H\omega - \Delta G\omega = H\omega + d\delta\omega + \delta d\omega$, which is the Hodge decomposition of ω. Compare Sections VI.1, VII.3.

As in the case of functions one can also derive the spectral decomposition of the Laplacian and H. Weyl's asymptotic formula from the properties of the heat kernel. The argument differs only in notational detail from the proofs of the Sturm–Liouville decomposition and the Weyl's formula in Chapter VI.

THEOREM B2. The completion $L^2A^p(M)$ of $A^p(M)$ with respect to the L^2 norm has an orthonormal basis $\varphi_{1,p}, \varphi_{2,p}, \varphi_{3,p}, \ldots$ consisting of eigenforms of Δ_p. One can order the eigenforms so that the corresponding eigenvalues $\lambda_{k,p}$ satisfy

(A10) (i) $0 \le \lambda_{1,p} \le \lambda_{2,p} \le \lambda_{3,p} \le \cdots \to \infty$,

in particular the multiplicities are finite;

(ii) $$N_p(\lambda) = \#\{k \mid \lambda_{k,p} \le \lambda\} \sim \frac{\binom{n}{p} \operatorname{vol}(M)}{(4\pi)^{n/2}\Gamma((n/2) + 1)} \lambda^{n/2};$$

(iii) $$e_p(x, y, t) = \sum_{k=1}^{\infty} e^{-\lambda_{k,p} t} \varphi_{k,p}(x) \otimes \varphi_{k,p}(y).$$

We next describe how the heat kernels can be used to obtain a link between local and global invariants of the manifold M. Taking the trace and combining the asymptotic expansion (A9) with (A10)(iii) we obtain

$$\text{(A11)} \qquad \int_M \operatorname{tr} e_p(x, x, t)\, dV(x)$$

$$= \sum_{k=1}^{\infty} e^{-\lambda_{k,p} t} \underset{t \to 0^+}{\sim} t^{-n/2}(a_{0,p} + a_{1,p} + t^2 a_{2,p} + \cdots),$$

where $a_{j,p} = \int_M \operatorname{tr} C_{p,j}(x)\, dV(x)$. On one side of this formula we have eigenvalues which are global invariants of M. The other side contains integrals of locally defined functions.

To sketch the heat equation approach to generalized Gauss–Bonnet theorem we need some notation. For $\lambda \geq 0$, and $0 \leq p \leq n$ let $A^p(\lambda)$ be the eigenspace of Δ_p belonging to the eigenvalue λ. Of course $A^p(\lambda)$ is finite dimensional and it is nonzero only for countably many values of λ. Clearly $A^p(\lambda)$ is perpendicular to $A^p(\mu)$ for $\lambda \neq \mu$, and, since $d\Delta = \Delta d$, $\delta\Delta = \Delta\delta$, $dA^p(\lambda) \subset A^{p+1}(\lambda)$ and $\delta A^{p+1}(\lambda) \subset A^p(\lambda)$. It is an easy exercise in linear algebra to see that, for $\lambda > 0$, the sequence

$$0 \to A^0(\lambda) \xrightarrow{d} A^1(\lambda) \xrightarrow{d} \cdots \xrightarrow{d} A^n(\lambda) \to 0$$

is exact. Therefore, for $\lambda > 0$,

$$\sum_{p=1}^{n} (-1)^p \dim A^p(\lambda) = 0.$$

It follows that

$$\begin{aligned} \sum_{p=1}^{n} (-1)^p \sum_{\lambda \geq 0} e^{-\lambda t} \dim A^p(\lambda) &= \sum_{\lambda \geq 0} e^{-\lambda t} \sum_{p=0}^{n} (-1)^p \dim A^p(\lambda) \\ &= \sum_{p=0}^{n} (-1)^p \dim A^p(0) \\ &= \sum_{p=0}^{n} (-1)^p \dim \mathscr{H}^p(M) = \chi(M), \end{aligned}$$

where $\chi(M)$ is the Euler characteristic of M. This expression for $\chi(M)$ combined with (A11) gives

$$\begin{aligned} \chi(M) = \sum_{p=0}^{n} (-1)^p \int_M \operatorname{tr} e_p(x, x, t)\, dV(x) &\underset{t \to 0^+}{\sim} \sum_{p=0}^{n} (-1)^p t^{-n/2} \sum_{j=0}^{\infty} a_{j,p} t^j \\ &= t^{-n/2} \sum_{j=0}^{\infty} t^j \left(\sum_{p=0}^{n} (-1)^p a_{j,p} \right). \end{aligned}$$

It follows immediately that for $j < n/2$, $\sum_{p=0}^{n}(-1)^p a_{j,p} = 0$, that $\chi(M) = 0$ for n odd and that for even n

$$\chi(M) = \sum_{p=0}^{n} (-1)^p a_{n/2,p}.$$

We recall that

$$a_{j,p} = \int_M \operatorname{tr} C_{p,j}(x)\, dV(x).$$

The following theorem was conjectured by McKean and Singer [1] as a "fantastic cancellation." It was proved independently by Gilkey [1] and Patodi [1].

THEOREM B3.

$$\left(\sum_{p=0}^{n} (-1)^p \operatorname{tr} C_{p,j}(x)\right) dV(x) = \begin{cases} 0 & \text{if } j < n/2, \\ \Omega & \text{if } j = n/2, \end{cases}$$

where Ω is the integrand in generalized Gauss–Bonnet formula (cf. Chern [1]).

Gilkey's proof is particularly elegant. He proves that the construction of the heat kernels forces the forms $(\sum_{p=0}^{n} (-1)^p \operatorname{tr} C_{p,j}(x))\, dV$ to have certain functorial properties. He then shows that the dimension of the space of forms with these properties is zero when $j < n/2$, and is one for $j = n/2$. The proof is completed by verifying that Ω has the required properties and that $\int_{S^{2k}} \Omega = 2$ for all k.

We remark that other classical index theorems (Hirzebruch signature formula, Riemann–Roch formula for compact Kähler manifolds) can be given similar proofs (cf. Gilkey [1]). One can also prove the Lefschetz fixed-point formula (as well as Atiyah–Bott [1] generalization of it) using the heat equation (cf. Kotake [1]).

C. EIGENVALUE ESTIMATES

The situation here is much worse than in the case of functions. Such tools as the maximum principle or Harnack inequality which hold for elliptic (or parabolic) equations are not available for systems (e.g., for the Laplacian on forms). Similarly, the beautiful relationship between geometric invariants (e.g., isoperimetric constant or Cheeger's constant) and analytic ones

(Sobolev constant, the smallest eigenvalue) have not been generalized to the Laplacian on forms. The case of Cheeger's inequality deserves some discussion. The following example is due to E. Calabi and its understanding leads to Cheeger's inequality. Let M be the surface of the dumbbell pictured in Fig. 1 on page 79. If the radius of the tube tends to zero, then λ_1, the first positive eigenvalue of Δ_0, approaches zero. One can think of the space M splitting into union M' of two disjoint manifolds. The multiplicity of zero as an eigenvalue of Δ_0 is one for M and two for M'. Assuming that the spectrum varies continuously under such drastic deformation of the manifold, $\lambda_1(M)$ has to approach zero. Cheeger's insight is that this is the only way to make λ_1 approach zero. Similarly, in the case of forms, the convergence of the smallest positive eigenvalue of Δ_p to zero ought to correspond to creation of a new cycle of dimension p. It appears that some kind of isoperimetric constant involving cycles of codimension other than one might give a lower bound for the lowest eigenvalue of Δ_p. No such result has been proved.

Finally, as another indication that much less is known about the Laplacian on forms, no comparison theorems for heat kernels exist (cf. Sections VIII.3 and 4). It is not even clear that any comparisons are possible since one deals with vector-valued functions.

Having pointed out the difficulties one encounters we review some of the known eigenvalue estimates for Δ_p.

The simplest lower bound for $\lambda_{0,p}$ follows from Weitzenböck identity, Proposition A3 (Lichnerowicz's theorem, (III.55), is a special case).

THEOREM C1. Suppose $F_p \geq \alpha > 0$ at every point of M. Then $\lambda_{1,p} \geq \alpha$.

The proof is straightforward. The theorem may be regarded as a quantitative version of the vanishing Theorem A4. Concrete results are proved by examining which geometric conditions guarantee that F_p is bounded below by a positive constant. The following is an example.

THEOREM C2 (Gallot and Meyer [1]). Suppose that the curvature operator $\rho_x\colon \Lambda^2 M_x^* \to \Lambda^2 M_x^*$ satisfies $\rho_x \geq \kappa > 0$ for all $x \in M$. Then for every p, $1 \leq p \leq n/2$,

$$\lambda_{1,p} \geq \kappa p(n - p + 1).$$

The equality is attained for the sphere of constant curvature k.

We remark that this gives a lower bound of $\lambda_{1,p}$ for all p, $1 \leq p \leq n - 1$, since $*$ maps $A^p(M)$ isomorphically onto $A^{n-p}(M)$ and $*\Delta = \Delta*$.

Lower bounds for higher eigenvalues of Δ_p were obtained by Li [1]. His departure point is also the Weitzenböck identity A3, which easily leads to the inequality

$$|\omega|\,|\Delta\omega| \leq (\lambda - p(n - p)\kappa)|\omega|^2$$

for every eigenform $\omega \in A^p(M)$ belonging to the eigenvalue λ. As above κ denotes the lower bound of the curvature operator. Setting $f = |\omega|$ and integrating, one obtains

$$\int_M f^{2l-1}\Delta f \leq (\lambda - p(n - p)\kappa)\|f\|_{2l}^{l}.$$

On the other hand,

$$\int_M f^{2l-1}\Delta f = (2l - 1)\int \langle f^{2l-2}\nabla f, \nabla f\rangle = \frac{2l - 1}{l^2}\|\nabla f^l\|_2^2.$$

The scheme described in Section IV.4 applies to the function f and yields (among other things) a lower bound for $\lambda_{j,p}$ in terms of the volume, the lower bound of the curvature operator, the Sobolev constant, p, j and the dimension of M.

D. BOUNDARY-VALUE PROBLEMS

There are boundary conditions for Δ_p, which are natural generalizations of Dirichlet and Neumann boundary conditions for Δ_0. As in the case of functions their understanding helps even in the study of Δ_p for manifolds without boundary. To describe these boundary conditions we need some notation. Let M be a smooth, compact, oriented, Riemannian manifold with smooth boundary ∂M. Suppose η is the unit inward pointing normal in the cotangent space at a boundary point $x \in \partial M$. A differential form $\omega \in \Lambda^p M_x^*$ can be decomposed uniquely into its normal and tangential components $\omega = \omega_{\text{tan}} + \omega_{\text{norm}}$, where $\omega_{\text{norm}} = \mu \wedge \eta$, $\omega_{\text{tan}} \in \Lambda^p(\partial M)_x^*$, $\mu \in \Lambda^{p-1}(\partial M)_x^*$.

Analogous to Dirichlet boundary conditions are absolute boundary conditions

$$\omega_{\text{tan}} = 0, \qquad (\delta\omega)_{\text{tan}} = 0.$$

Relative boundary conditions

$$\omega_{\text{norm}} = 0, \qquad (d\omega)_{\text{norm}} = 0$$

generalize the Neumann conditions.

Either of these boundary conditions gives rise to an elliptic, self-adjoint boundary value problem. Formally, the theory is very similar to what we outlined for the boundaryless case. For example, the deRham–Hodge theory holds (cf. Conner [1]; Ray and Singer [1]). In particular, harmonic forms which satisfy absolute (respectively, relative) boundary conditions represent the absolute cohomology of M (respectively, the relative cohomology of the pair $(M, \partial M)$). Indeed, this is the source of the terminology.

Heat equation methods have been used successfully to study these boundary value problems (e.g., by Ray and Singer [1] and by Cheeger [2]) and to some other nonlocal boundary value problems in connection with index theorem for manifolds with boundary (cf. Atiyah–Patodi–Singer [1]). H. Weyl's asymptotic formula holds in this context and gives the asymptotic rate of growth of the eigenvalues but almost no eigenvalue estimates are known for Δ_p with either absolute or relative boundary conditions.

Bibliography

R. A. ADAMS

1. "Sobolev Spaces." Academic Press, New York, 1955.

AMERICAN MATHEMATICAL SOCIETY PROCEEDINGS OF SYMPOSIA IN PURE MATHEMATICS

1. "Global Analysis." **16** (1970).
2. "Partial Differential Equations." **23** (1973).
3. "Differential Geometry." **27** (1975).
4. "Geometry of the Laplace Operator." **36** (1980).

N. ARONSZAJN

1. A unique continuation theorem for solutions of elliptic partial differential equations or inequalities of second order. *J. Math. Pures Appl.* **36**, 237–249 (1957).

M. F. ATIYAH and R. BOTT

1. A Lefschetz fixed point formula for elliptic complexes. *Ann. Math.* **86**, 374–407 (1967).

M. F. ATIYAH, R. BOTT, and V. K. PATODI

1. On the heat equation and the index theorem. *Invent. Math.* **19**, 279–330 (1973).

M. F. ATIYAH, V. K. PATODI, and I. M. SINGER

1. Spectral asymmetry and Riemannian geometry. *Math. Proc. Cambridge Philos. Soc.* I: **77**, 43–69 (1975); II: **78**, 405–432 (1975); III: **79**, 71–99 (1976).

T. AUBIN

1. Problèmes isopérimétriques et espaces de Sobolev. *J. Differ. Geom.* **11**, 573–598 (1976).
2. "Nonlinear Analysis on Manifolds. Monge-Ampere Equations." Springer Verlag, Berlin and New York, 1982.

R. AZENCOTT

1. Behaviour of diffusion semi-groups at infinity. *Bull. Soc. Math. Fr.* **102**, 193–240 (1974).

A. BAIDER and E. A. FELDMAN

1. On oscillatory elliptic equations on manifolds. *Trans. Am. Math. Soc.* **258**, 495–504 (1980).

C. BANDLE

1. "Isoperimetric Inequalities and Applications." π-Pitman, London, 1980.

J. L. BARBOSA and M. doCARMO

1. On the size of a stable minimal surface in $\mathbb{R}^3$. *Am. J. Math.* **98**, 515–528 (1976).
2. Stability of minimal surfaces and eigenvalues of the Laplacian. *Math. Z.* **173**, 13–28 (1980).

J. BARTA

1. Sur la vibration fundamentale d'une membrane. *C.R. Acad. Sci.* **204**, 472–473 (1937).

BATEMAN MANUSCRIPT PROJECT

1. "Tables of Integral Transforms," Vol. I. McGraw-Hill, New York, 1954.

P. BÉRARD
1. Spectre et groupes cristallographique. I. Domaines euclidiens. *Invent. Math.* **58**, 179–199 (1980).
2. Remarques sur la conjecture de Weyl. *Compositio Math.* **48**, 35–53 (1983).
P. BÉRARD and M. BERGER
1. Le spectre d'une variété riemanniene en 1981 (preprint).
P. BÉRARD and G. BESSON
1. Spectres et groupes cristallographiques. II. Domaines spherique. *Ann. Inst. Fourier, Grenoble* **30**, 237–248 (1980).
2. Remarque sue un article de Marcel Berger: Sur une inégalité pour la première valeur propre du laplacien. *Bull. Soc. Math. Fr.* **108**, 333–336 (1980).
P. BÉRARD and D. MEYER
1. Inégalités isopérimétrique et applications. *Ann. Sci. Éc. Norm. Super., Paris* **15**, 513–542 (1982).
L. BÉRARD BERGERY
1. Laplacien et géodésiques fermées sur les formes d'espace hyperbolique compactes. *Semin. Bourbaki* **24**, exposé 406 (1971–1972).
L. BÉRARD BERGERY and J. P. BOURGUIGNON
1. Laplacians and Riemannian submersions with totally geodesic fibres. *Ill. J. Math.* **26**, 181–200 (1982).
M. BERGER
1. An extension of Rauch's metric comparison theorem and some applications. *Ill. J. Math.* **6**, 700–711 (1962).
2. Le spectres des variétés riemannienes. *Rev. Roum. Math. Pures Appl.* **13**, 915–931 (1968).
3. Eigenvalues of the Laplacian. *Am. Math. Soc. Proc. Symp. Pure Math.* **16**, 121–125 (1970).
4. Geometry of the spectrum. I. *Am. Math. Soc. Proc. Symp. Pure Math.* **27**, Part II, 129–152 (1975).
5. Une inégalité universelle pour la première valeur propre du laplacien. *Bull. Soc. Math. Fr.* **109**, 3–9 (1979).
6. Une borne inférieure pour le volume d'une variété riemannienes en fonction du rayon d'injectivité. *Ann. Inst. Fourier, Grenoble* **30**, 259–265 (1980).
7. H. E. Rauch, géomètre differentiel. "Differential Geometry and Complex Analysis," H. E. Rauch Memorial Volume (to appear).
M. BERGER and J. L. KAZDAN
1. A Sturm-Liouville inequality with applications to an isoperimetric inequality for volume, injectivity radius, and to wiedersehn manifolds. *In* "General Inequalities" (E. F. Beckenbach, ed.), Vol. 2, pp. 367–377. Birkhaeuser, Basel, 1980.
M. BERGER, P. GAUDUCHON, and E. MAZET
1. "Le Spectre d'une Variété Riemannienes," Lect. Notes Math. **194**. Springer-Verlag, Berlin and New York, 1974.
L. BERS
1. Spaces of degenerating Riemann surfaces, "Discontinuous Groups and Riemann Surfaces." *Ann. Math. Stud.* **79**, 43–55 (1974).
A. BESSE
1. "Manifolds all of whose Geodesics are Closed." Springer-Verlag, Berlin and New York, 1978.
G. BESSON
1. Sur la multiplicité de la première valeur propre des surfaces riemannienes. *Ann. Inst. Fourier, Grenoble* **30**, 109–128 (1980).

R. BISHOP and R. CRITTENDEN

1. "Geometry of Manifolds." Academic Press, New York, 1964.

G. BOL

1. Isoperimetrische Ungleichung für Bereiche auf Flächen. *Jahresber. Dtsch. Math. Ver.* **51**, 219–257 (1941).

A. BOREL

1. Compact Clifford-Klein forms of symmetric spaces. *Topology* **2**, 111–122 (1963).

H. J. BRASCAMP and E. LIEB

1. On extensions of the Brunn-Minkowski and Prekopa-Leindler theorems, including inequalities for log concave functions, and with an application to a diffusion equation. *J. Funct. Anal.* **22**, 366–389 (1976).

R. BROOKS

1. The fundamental group and the spectrum of the Laplacian. *Comment. Math. Helv.* **56**, 581–598 (1981).
2. A relation between growth and the spectrum of the Laplacian. *Math. Z.* **178**, 501–508 (1981).
3. On the spectrum of noncompact manifolds with finite volume (preprint).
4. The spectral geometry of the Appollonian packing. *Commun. Pure Appl. Math.* (to appear).

J. BRÜNING

1. Über Knoten von Eigenfunktionen des Laplace-Beltrami Operators. *Math. Z.* **158**, 15–21 (1978).

P. BUSER

1. Riemannsche Flächen mit Eigenwerte in $(0, \frac{1}{4})$. *Comment. Math. Helv.* **52**, 25–34 (1977).
2. The collar theorem and examples. *Manuscr. Math.* **25**, 349–357 (1978).
3. On Cheeger's inequality: $\lambda_1 \geq \frac{1}{4}h^2$. *Am. Math. Soc. Proc. Symp. Pure Math.* **36**, 29–77 (1980).
4. A note on the isoperimetric constant. *Ann. Sci. Ec. Norm. Super., Paris* **15**, 213–230 (1982).

L. A. CAFARELLI and J. SPRUCK

1. Convexity properties of solutions to some classical variational problems. *Commun. Partial Differ. Equations.* **7**, 1337–1379 (1982).

E. CALABI

1. An extension of E. Hopf's maximum principle with an application to geometry. *Duke Math. J.* **25**, 285–303 (1965).
2. Minimal immersions of surfaces in Euclidean spheres. *J. Differ. Geom.* **1**, 111–126 (1967).

T. CARLEMAN

1. Zur theorie der Minimalflächen. *Math. Z.* **9**, 154–160 (1921).

I. CHAVEL

1. "Riemannian Symmetric Spaces of Rank One." Marcel Dekker, New York, 1972.
2. On A. Hurwitz' method in isoperimetric inequalities. *Proc. Am. Math. Soc.* **71**, 275–279 (1978).
3. Lowest eigenvalue inequalities. *Am. Math. Soc. Proc. Symp. Pure Math.* **36**, 79–89 (1980).

I. CHAVEL and E. A. FELDMAN

1. The first eigenvalue of the Laplacian on manifolds of nonnegative curvature. *Compositio Math.* **29**, 43–53 (1974).
2. An optimal Poincaré inequality for convex domains of nonnegative curvature. *Arch. Ration. Mech. Anal.* **65**, 263–273 (1977).
3. Cylinders on surfaces. *Comment. Math. Helv.* **53**, 439–447 (1978).
4. Spectra of domains in compact manifolds. *J. Funct. Anal.* **30**, 198–222 (1978).

5. Isoperimetric inequalities on curved surfaces. *Adv. Math.* **37**, 83–98 (1980).
6. Spectra of manifolds with small handles. *Comment. Math. Helv.* **56**, 83–102 (1981).
7. Isoperimetric constants of manifolds with small handles. *Math. Z.* **184**, 435–448 (1983).
8. Diffusion on manifolds with small handles (preprint).
9. The Riemannian Wiener sausage (preprint).

J. CHEEGER

1. A lower bound for the smallest eigenvalue of the Laplacian. "Problems in Analysis," pp. 195–199. Princeton Univ. Press, Princeton, New Jersey, 1970.
2. Analytic torsion and the heat equation. *Ann. Math.* **109**, 259–322 (1979).

J. CHEEGER and D. EBIN

1. "Comparison Theorems in Riemannian Geometry." North-Holland Publ., Amsterdam, 1975.

J. CHEEGER and S. T. YAU

1. A lower bound for the heat kernel. *Commun. Pure Appl. Math.* **34**, 465–480 (1981).

J. CHEEGER, M. GROMOV, and M. TAYLOR

1. Finite propagation speed, kernel estimates for functions of the Laplace operator, and the geometry of complete Riemannian manifolds. *J. Differ. Geom.* **17**, 15–54 (1982).

S. Y. CHENG

1. Eigenfunctions and eigenvalues of the Laplacian. *Am. Math. Soc. Proc. Symp. Pure Math.* **27**, Part II, 185–193 (1975).
2. Eigenvalue comparison theorems and its geometric applications. *Math. Z.* **143**, 289–297 (1975).
3. A characterization of the 2-sphere by eigenfunctions. *Proc. Am. Math. Soc.* **55**, 379–381 (1976).
4. Eigenfunctions and nodal sets. *Comment. Math. Helv.* **51**, 43–55 (1976).
5. A note on the first eigenvalue of Riemann surfaces with genus greater than one (preprint).
6. On the Hayman-Osserman-Taylor inequality (preprint).

S. Y. CHENG and P. LI

1. Heat kernel estimates and lower bounds of eigenvalues. *Comment. Math. Helv.* **56**, 327–338 (1981).

S. Y. CHENG and S. T. YAU

1. Differential equations on Riemannian manifolds and their geometric applications. *Commun. Pure Appl. Math.* **28**, 333–354 (1975).

S. Y. CHENG, P. LI, and S. T. YAU

1. On the upper estimate of the heat kernel of a complete Riemannian manifold. *Am. J. Math.* **103**, 1021–1063 (1981).
2. Heat equations on minimal submanifolds and their applications. *Am. J. Math.* (to appear).

S. S. CHERN

1. A simple intrinsic proof of the Gauss-Bonnet formula for closed Riemannian manifolds. *Ann. Math.* **45**, 747–752 (1944).

E. A. CODDINGTON and N. LEVINSON

1. "Theory of Ordinary Differential Equations." McGraw-Hill, New York, 1955.

P. COHEN and P. SARNAK

1. "Discontinuous Groups and Harmonic Analysis" (in preparation).

P. E. CONNER

1. The Neumann problem for differential forms on Riemannian manifolds. *Mem. Am. Math. Soc.* **20** (1956).

R. COURANT

1. "Differential and Integral Calculus," Vol. II. Wiley (Interscience), New York, 1947.

R. COURANT and D. HILBERT

1. "Methods of Mathematical Physics," Vol. I. Wiley (Interscience), New York, 1953; Vol. II, 1967.

C. B. CROKE

1. Some isoperimetric inequalities and eigenvalue estimates. *Ann. Sci. Ec. Norm. Super., Paris* **13**, 419–435 (1980).
2. An eigenvalue pinching theorem. *Invent. Math.* **68**, 253–256 (1982).
3. The first eigenvalue of the Laplacian for plane domains (preprint).
4. A sharp 4-dimensional isoperimetric inequality (preprint).
5. Curvature-free volume estimates (preprint).

E. B. DAVIES

1. Some norm bounds and quadratic form inequalities for Schrodinger operators (preprint).

A. DEBIARD, B. GAVEAU, and E. MAZET

1. Théorèmes de comparisons en géométrie riemanniene. *Publ. R.I.M.S., Kyoto Univ.* **12**, 391–425 (1976).

D. L. DeGEORGE

1. Length spectrum for compact locally symmetric spaces of strictly negative curvature. *Ann. Sci. Ec. Norm. Super., Paris* **10**, 133–152 (1977).

G. Del GROSSO and F. MARCHETTI

1. Asymptotic estimates for the principal eigenvalue of the Laplacian in a geodesic ball. *Appl. Math. Optim.* **10**, 37–50 (1983)

J. DELSARTE

1. Sur le gitter fuchsien. *C. R. Acad. Sci., Paris* **241**, 147–149 (1942).

G. deRHAM

1. "Variétés Différentiables." Hermann, Paris, 1960.

M. doCARMO and N. R. WALLACH

1. Minimal immersions of spheres into spheres. *Ann. Math.* **95**, 43–62 (1971).

J. DODZIUK

1. Eigenvalues of the Laplacian and the heat equation. *Am. Math. Mon.* **9**, 686–695 (1981).
2. Maximum principle for parabolic inequalities and the heat flow on open manifolds. *Indiana Univ. Math. J.* **32**, 703–716 (1983).
3. Every covering of a compact Riemann surface of genus greater than one carries a nontrivial L^2 harmonic differential. *Acta Math.* (to appear).
4. Difference equations, isoperimetric inequality, and transcience of certain random walks. *Trans. Am. Math. Soc.* (to appear).

H. DONNELLY

1. Asymptotic expansions for the compact quotients of proper discontinuous group actions. *Ill. J. Math.* **23**, 485–496 (1979).
2. On the essential spectrum of a complete Riemannian manifold. *Topology* **20**, 1–14 (1981).

H. DONNELLY and P. LI

1. Lower bounds for eigenvalues of negatively curved manifolds. *Math. Z.* **172**, 29–40 (1980).
2. Lower bounds for eigenvalues of Riemannian manifolds. *Mich. Math. J.* **29**, 149–161 (1982).
3. Heat equation and compactifications of complete Riemannian manifolds (preprint).

G. F. D. DUFF

1. "Partial Differential Equations." Univ. of Toronto Press, Toronto, 1965.

J. J. DUISTERMAAT and V. W. GUILLEMIN

1. The spectrum of positive elliptic operators and periodic bicharacteristics. *Invent. Math.* **29**, 39–79 (1975).

H. DYM and H. P. McKEAN

1. "Fourier Series and Integrals." Academic Press, New York, 1972.

C. FABER

1. Beweiss, dass unter allen homogenen Membrane von gleicher Fläche und gleicher Spannung die kreisförmige die tiefsten Grundton gibt. *Sitzungsber.—Bayer. Akad. Wiss., Math.-Phys. Munich.* pp. 169–172 (1923).

H. FEDERER and W. H. FLEMING

1. Normal and integral currents. *Ann. Math.* **72**, 458–520 (1960).

W. FELLER

1. "An Introduction to the Theory of Probability," 2nd ed., Vol. II. Wiley, New York, 1971.

F. FIALA

1. Le problème des isopérimètres sur les surfaces ouvertes à courbure positive. *Comment. Math. Helv.* **13**, 293–346 (1940–1941).

V. A. FOCK

1. On the representation of an arbitrary function by an integral involving Legendre's function with a complex index. *Dokl. Akad. Nauk SSSR* **39**, 253–256 (1943).

L. FORD

1. "Automorphic Functions." McGraw-Hill, New York, 1929.

R. FRICKE and F. KLEIN

1. "Vorlesung über die Theorie der Automorphen Functionen." G. Teubner, Leipzig, 1897.

S. FRIEDLAND and W. K. HAYMAN

1. Eigenvalue inequalities for the Dirichlet problem on spheres, and the growth of subharmonic functions. *Comment. Math. Helv.* **51**, 133–161 (1976).

M. GAFFNEY

1. The conservation property of the heat equation on Riemannian manifolds. *Commun. Pure Appl. Math.* **12**, 1–11 (1959).

M. GAGE

1. Upper bounds for the first eigenvalue of the Laplace-Beltrami operator. *Indiana Univ. Math. J.* **29**, 897–912 (1980).

S. GALLOT and D. MEYER

1. Opérateur de courbure et laplacien des formes différentielles d'une variété riemanniene. *J. Math. Pures Appl.* **54**, 259–284 (1975).

D. GILBARG and N. S. TRUDINGER

1. "Elliptic Partial Differential Equations of Second Order." Springer-Verlag, Berlin and New York, 1977.

P. B. GILKEY

1. Curvature and the eigenvalues of the Laplacian for elliptic complexes. *Adv. Math.* **10**, 311–325 (1973).
2. "The Index Theorem and the Heat Equation." Publish or Perish, Boston, Massachusetts, 1974.

C. S. GORDON and E. N. WILSON

1. Isospectral deformations of compact solvmanifolds (preprint).

L. W. GREEN

1. Aufwiedersehnsflächen. *Ann. Math.* **78**, 289–299 (1963).

D. GROMOLL, W. KLINGENBERG, and W. MEYER

1. "Riemannsche Geometrie im Grossen," Lect. Notes Math., **55**. Springer-Verlag, Berlin and New York, 1968.

M. GROMOV

1. Paul Levy's isoperimetric inequality (preprint).

N. GROSSMAN

1. The volume of a totally geodesic hypersurface in a pinched manifold. *Pac. J. Math.* **23**, 257–262 (1967).

K. GROVE and K. SHIOHAMA

1. A generalized sphere theorem. *Ann. Math.* **106**, 201–211 (1977).

P. GÜNTHER

1. Gitterpunktprobleme in Symmetrischen Riemannsche Räumen von Rang 1. *Math. Nachr.* **94**, 5–27 (1980).

N. HALPERN

1. Collars on Riemann surfaces (preprint).

HARISH-CHANDRA

1. Spherical functions on semi-simple Lie groups. I. *Am. J. Math.* **80**, 241–310 (1958).

W. HAYMAN

1. Some bounds for principal frequency. *Applicable Anal.* **7**, 247–254 (1978).

J. J. HEBDA

1. Universal upper bounds for the first eigenvalue of the Laplacian (preprint).

E. HEINTZE and H. KARCHER

1. A general comparison theorem with applications to volume estimates for submanifolds. *Ann. Sci. Ec. Norm. Super., Paris* **11**, 451–470 (1978).

D. A. HEJHAL

1. "The Selberg Trace Formula for PSL(2, ℝ). Vol. I," Lect. Notes Math. **548**. Springer-Verlag, Berlin and New York, 1976.
2. "The Selberg Trace Formula for PSL(2, ℝ). Vol. II," Lect. Notes Math. **1001**. Springer-Verlag, Berlin and New York, 1983.

S. HELGASON

1. "Topics in Harmonic Analysis on Homogeneous Spaces." *Prog. Math.* **13** (1981).

J. HERSCH

1. Quatre propriétés isopérimétrique de membranes sphérique. *C.R. Acad. Sci., Paris* **270**, 1645–1648 (1970).

R. HERSH

1. The method of transmutation. "Partial Differential Equations and Related Topics," Lect. Notes Math. **446**, 264–282. Springer-Verlag, Berlin and New York, 1975.

N. HITCHIN

1. Harmonic spinors. *Adv. Math.* **14**, 1–55 (1974).

E. W. HOBSON

1. "The Theory of Spherical and Ellipsoidal Harmonics." Cambridge Univ. Press, London and New York, 1931.

H. HUBER

1. Zur analytichen Theorie hyperbolisher Raumformen und Bewgungsgruppen. II. *Math. Ann.* **142**, 385–398 (1961).

A. HURWITZ

1. Sur le probleme des isopérimètres. *C.R. Acad. Sci., Paris* **132**, 401–403 (1901).

S. ILIAS

1. Constantes explicites pour les inégalités de Sobolev sur les variétés riemannienes compactes. *Ann. Inst. Fourier, Grenoble* **33**, 151–165 (1983).

V. K. IONIN

1. Isoperimetric and various other inequalities on a manifold of bounded curvature. *Sib. Math. J.* **10**, 329–342 (1969).

H. JACQUET and R. P. LANGLANDS

1. "Automorphic Forms on GL(2)," Lect. Notes Math. **114**. Springer-Verlag, Berlin and New York, 1970.

J. H. JELLET

1. Sur la surface dont la courbure moyenne est constante. *J. Math. Pures Appl.* **18**, 163–167 (1853).

M. KAC

1. Can one hear the shape of a drum? *Am. Math. Mon.* **73**(4), Part II, 1–23 (1966).
2. Probabilistic methods in some problems of scattering theory. *Rocky Mt. Math. J.* **4**, 511–537 (1974).

Y. KANNAI

1. Off diagonal short time asymptotics for fundamental solutions of heat equations. *Commun. Partial Differ. Equations* **2**, 781–830 (1977).

L. KARP

1. Subharmonic functions on real and complex manifolds. *Math. Z.* **179**, 535–554 (1982).
2. Noncompact manifolds with purely continuous spectrum (preprint).
3. On the spectra of noncompact manifolds (preprint).
4. Differential inequalities on Riemannian manifolds (preprint).
5. Heat diffusion, capacity, and Brownian motion on noncompact manifolds (to appear).

L. KARP and P. LI

1. The heat equation on complete Riemannian manifolds (preprint).

L. KEEN

1. Collars on Riemann surfaces. "Discontinuous Groups and Riemann Surfaces." *Ann. Math. Stud.* **79**, 263–268 (1974).
2. A rough fundamental domain for Teichmuller spaces. *Bull. Am. Math. Soc.* **83**, 1199–1226 (1977).

D. KENDALL

1. On the number of lattice-points inside a random oval. *Q. J. Math.* **19**, 1–26 (1948).

W. KLINGENBERG

1. Contributions to Riemannian geometry in the large. *Ann. Math.* **69**, 654–666 (1959).

E. T. KORNHAUSER and I. STAKGOLD

1. A variational theorem for $\Delta u + \lambda u = 0$ and its applications. *J. Math. Phys.* **31**, 45–54 (1952).

T. KOTAKE

1. The fixed-point theorem of Atiyah-Bott via parabolic operators. *Commun. Pure Appl. Math.* **22**, 789–806 (1969).

E. KRAHN

1. Über eine von Rayleigh formulierte Minmaleigenschaft des Kreises. *Math. Ann.* **94**, 97–100 (1925).

R. P. LANGLANDS

1. Unpublished lecture notes on the trace formula.

H. B. LAWSON

1. "Lecture Notes on Minimal Submanifolds." I.M.P.A., Rio de Janeiro, 1970 (reprinted: Publish or Perish, Berkeley, California, 1980).

N. N. LEBEDEV

1. "Special Functions and their Applications" (R. A. Silverman, ed. and transl.). Dover, New York, 1972.

S. S. LEE and S. H. CRANDALL

1. "The Eigenmodes of an Equilateral Triangle, Mélanges Theodore Vogel," pp. 255–268. Univ. Libre de Bruxelles, 1978.

P. LI

1. On the Sobolev constant and the p-spectrum of a compact Riemannian manifold. *Ann. Sci. Ec. Norm. Super., Paris* **13**, 419–435 (1980).

2. Eigenvalue estimates on homogeneous spaces. *Comment. Math. Helv.* **55**, 347–363 (1980).
3. Minimal immersions of compact irreducible homogeneous Riemannian manifolds. *J. Differ. Geom.* **16**, 105–116 (1981).
4. Poincaré inequalities on Riemannian manifolds. "Seminar on Differential Geometry," *Ann. Math. Stud.* **102**, 73–84 (1982).

P. LI and S. T. YAU

1. Estimates of eigenvalues of a compact Riemannian manifold. *Am. Math. Soc. Proc. Symp. Pure Math.* **36**, 205–240 (1980).
2. A new conformal invariant and its applications to the Willmore conjecture and the first eigenvalue for compact surfaces. *Invent. Math.* **69**, 269–291 (1982).
3. On the Schrödinger equation and the eigenvalue problem. *Commun. Math. Phys.* **88**, 309–318 (1983).
4. On the parabolic kernel of the Schrödinger operator (preprint).

A. LICHNEROWICZ

1. "Geometrie des Groupes des Transformationes." Dunod, Paris, 1958.

E. H. LIEB

1. The number of bound states of one-body Schrödinger operators and the Weyl problem. *Am. Math. Soc. Proc. Symp. Pure Math.* **36**, 241–252 (1980).
2. On the lowest eigenvalue of the Laplacian for the intersection of two domains. *Invent. Math.* **74**, 441–448 (1983).

F. LÖBELL

1. Die überall regulären begrenzten Flächen fester Krümmung. Dissertation, Tübingen Universität, 1927.

J. M. LUTTINGER

1. Generalized isoperimetric inequalities. *J. Math. Phys.* **14**, 586–593 (1973).

H. M. MACDONALD

1. Zeroes of the spherical harmonic P_n^m considered as a function of n. *Proc. London Math. Soc.* **31**, 264–278 (1899).

H. P. McKEAN

1. An upper bound for the spectrum of Δ on a manifold of negative curvature. *J. Differ. Geom.* **4**, 359–366 (1970).
2. Selberg's trace formula applied to a compact Riemann surface. *Commun. Pure Appl. Math.* **25**, 225–246 (1972).

H. P. McKEAN and I. M. SINGER

1. Curvature and eigenvalues of the Laplacian. *J. Differ. Geom.* **1**, 43–70 (1967).

P. J. McKENNA and M. RAO

1. Lower bounds for the first eigenvalue of the Laplacian, with Dirichlet boundary conditions, and a theorem of Hayman (preprint).

J. P. MATELSKI

1. A compactness theorem for Fuchsian groups of the second kind. *Duke Math. J.* **43**, 829–840 (1976).

T. MATSUZAWA and S. TANNO

1. Estimates of the first eigenvalue of a big cup domain of a 2-sphere. *Comp. Math.* **47**, 95–100 (1982).

F. G. MEHLER

1. Über eine mit den Kugel und Cylinderfunctionen und ihre Anwendung in der Theorie der Elektricitätsvertheilung. *Math. Ann.* **18**, 161–194 (1881).

A. N. MILGRAM and P. C. ROSENBLOOM

1. Harmonic forms and heat conduction. I. *Proc. Natl. Acad. Sci. U.S.A.* **37**, 180–184 (1951).

J. MILLSON

1. On the first Betti number of a constant negatively curved manifold. *Ann. Math.* **104**, 235–247 (1976).

S. MINAKSHISUNDARAM

1. A generalization of Epstein zeta functions. *Can. J. Math.* **1**, 320–329 (1949).
2. Eigenfunctions on Riemannian manifolds. *J. Indian Math. Soc.* **17**, 158–165 (1953).

S. MINAKSHISUNDARAM and A. PLEIJEL

1. Some properties of the eigenfunctions of the Laplace operator on Riemannian manifolds. *Can. J. Math.* **1**, 242–256 (1949).

J. MORROW and K. KODAIRA

1. "Complex Manifolds." Holt, New York, 1971.

J. MOSER

1. A Harnack inequality for parabolic differential equations. *Commun. Pure Appl. Math.* **17**, 101–134 (1964).

G. D. MOSTOW

1. "Strong Rigidity of Locally Symmetric Spaces." *Ann. Math. Stud.* **78** (1973).

R. NARASIMHAN

1. "Analysis on Real and Complex Manifolds." North-Holland Publ., Amsterdam, 1968.

M. OBATA

1. Certain conditions for a Riemannian manifold to be isometric with a sphere. *J. Math. Soc. Jpn.* **14**, 333–340 (1962).

F. J. W. OLVER

1. "Asymptotics and Special Functions." Academic Press, New York, 1974.

R. OSSERMAN

1. "A Survey of Minimal Surfaces." Van Nostrand-Reinhold, Princeton, New Jersey, 1969.
2. A note on Hayman's theorem on the bass note of a drum. *Comment. Math. Helv.* **52**, 545–555 (1977).
3. The isoperimetric inequality. *Bull. Am. Math. Soc.* **84**, 1182–1238 (1978).
4. Bonnesen-style isoperimetric inequalities. *Am. Math. Mon.* **86**, 1–29 (1979).
5. Minimal surfaces, Gauss maps, total curvature, eigenvalue estimates, and stability. "The Chern Symposium," pp. 199–228. Springer-Verlag, Berlin and New York, 1979.

S. OZAWA

1. Singular Hadamard's variation of domains and eigenvalues of the Laplacian. *Proc. Jpn. Acad.* **56**, 306–310 (1980).
2. Singular variation of domains and eigenvalues of the Laplacian. *Duke Math. J.* **48**, 767–778 (1981).
3. Electrostatic capacity and eigenvalues of the Laplacian. *J. Fac. Sci., Univ. Tokyo* **30**, 53–62 (1983).

G. C. PAPANICOLAOU and S. R. S. VARADHAN

1. "Diffusion in Regions with Many Small Holes," Lect. Notes in Control and Information, **75**, 190–206. Springer-Verlag, Berlin and New York, 1980.

V. K. PATODI

1. Curvature and the eigenforms of the Laplace operator. *J. Differ. Geom.* **5**, 233–249 (1971).

S. J. PATTERSON

1. A lattice-point problem in hyperbolic space. *Mathematika* **22**, 81–88 (1975).

L. E. PAYNE

1. Isoperimetric inequalities and their applications. *SIAM Rev.* **9**, 453–488 (1967).

M. A. PINSKY

1. The spectrum of the Laplacian on a manifold of negative curvature. I. *J. Differ. Geom.* **13**, 87–91 (1978).

2. On the spectrum of Cartan-Hadamard manifolds. *Pac. J. Math.* **94**, 223–230 (1981).
3. Eigenvalues of an equilateral triangle. *SIAM J. Math. Anal.* **11**, 819–827 (1980).
4. Completeness of the eigenfunctions of the equialteral triangle (preprint).
5. The first eigenvalue of a spherical cap. *Appl. Math. Opt.* **7**, 137–139 (1981).

A. PLEIJEL

1. Remarks on Courant's nodal line theorem. *Commun. Pure Appl. Math.* **9**, 543–550 (1956).

G. POLYA

1. On the eigenvalues of vibrating membranes. *Proc. London Math. Soc.* **11**, 419–433 (1961).

G. POLYA and G. SZEGÖ

1. "Isoperimetric Inequalities in Mathematical Physics." *Ann. Math. Stud.* **27** (1951).

S. C. PORT and C. J. STONE

1. Brownian Motion and Classical Potential Theory." Academic Press, New York, 1978.

A. PREISSMANN

1. Quelques propriétés globales des espaces de Riemann. *Comment. Math. Helv.* **15**, 175–216 (1943).

M. H. PROTTER and H. F. WEINBERGER

1. "Maximum Principles in Differential Equations." Prentice-Hall, Englewood Cliffs, New Jersey, 1967.

B. RANDOL

1. Small eigenvalues of the Laplace operator on compact Riemann surfaces. *Bull. Am. Math. Soc.* **80**, 996–1000 (1974).
2. On the analytic continuation of the Minakshisundaram-Pleijel zeta-function for compact Riemann surfaces. *Trans. Am. Math. Soc.* **201**, 241–246 (1975).
3. On the asymptotic distribution of closed geodesics on compact Riemann surfaces. *Trans. Am. Math. Soc.* **233**, 241–247 (1977).
4. The Riemann hypothesis for Selberg's zeta-function and the asymptotic behavior of eigenvalues of the Laplace operator. *Trans. Am. Math. Soc.* **236**, 209–233 (1978).
5. Cylinders in Riemann surfaces. *Comment. Math. Helv.* **54**, 1–5 (1979).
6. A remark on the multiplicity of the discrete spectrum of congruence subgroups. *Proc. Am. Math. Soc.* **81**, 339–340 (1981).
7. The length spectrum of a Riemann surface is always of unbounded multiplicity. *Proc. Am. Math. Soc.* **81**, 455–456 (1981).
8. A Dirichlet series of eigenvalue type with applications to asymptotic estimates. *Bull. London Math. Soc.* **13**, 309–315 (1981).
9. The behavior under projection of dilating sets in a covering space, *Trans. Am. Math. Soc.* (to appear).

H. E. RAUCH

1. A contribution to differential geometry in the large. *Ann. Math.* **54**, 38–55 (1951).

J. RAUCH

1. Mathematical theory of crushed ice. "Partial Differential Equations and Related Topics," Lect. Notes Math. **446**, pp. 370–379. Springer-Verlag, Berlin and New York, 1975.

J. RAUCH and M. TAYLOR

1. Penetrations into shadow regions and unique continuation properties in hyperbolic mixed problems. *Indiana Univ. Math. J.* **22**, 277–285 (1972).
2. Potential and scattering on wildly perturbed domains. *J. Funct. Anal.* **18**, 27–59 (1975).

D. B. RAY and I. M. SINGER

1. R-torsion and the Laplacian on Riemannian manifolds. *Adv. Math.* **7**, 145–210 (1971).

(LORD) J. W. S. RAYLEIGH

1. "The Theory of Sound." Macmillan, New York, 1877 (reprinted: Dover, New York, 1945).

R. C. REILLY
1. Applications of the Hessian operator in a Riemannian manifold. *Indiana Univ. Math. J.* **26**, 459–472 (1977).
F. RIESZ and B. SZ. NAGY
1. "Functional Analysis" (transl. by L. F. Boron). Frederick Ungar Publ. Co., New York, 1955.
T. SAKAI
1. On eigenvalues of the Laplacian, and the curvature of a Riemannian manifold. *Tohuku Math. J.* **23**, 589–603 (1971).
L. A. SANTALO
1. "Integral Geometry and Geometric Probability." Addison-Wesley, Reading, Massachusetts, 1976.
S. SATO
1. Barta's inequalities and the first eigenvalue of a cap domain of a 2-sphere. *Math. Z.* **181**, 313–318 (1982).
M. SCHIFFER and D. C. SPENCER
1. "Functionals of Finite Riemann Surfaces." Princeton Univ. Press, Princeton, New Jersey, 1954.
E. SCHMIDT
1. Die Brunn-Minkowski Ungleichung und ihr Spiegelbild sowie die isoperimetrische Eigenschaft der Kugel in der euklidischen und nichteuklidischen Geometrie. *Math. Nachr.* **1**, 81–157 (1948); **2**, 171–244 (1949).
R. SCHOEN
1. A lower bound for the first eigenvalue of a negatively curved manifold. *J. Differ. Geom.* **17**, 233–238 (1982).
R. SCHOEN, S. WOLPERT, and S. T. YAU
1. Geometric bounds on the low eigenvalues of a compact Riemann surface. *Am. Math. Soc. Proc. Symp. Pure Math.* **36**, 279–285 (1980).
R. SEELEY
1. Some remarks on diffusion with drift (preprint).
A. SELBERG
1. Harmonic analysis and discontinuous groups in weakly symmetric spaces with applications to Dirichlet series. *J. Indian Math. Soc.* **20**, 47–87 (1956).
2. On the estimation of Fourier coefficients of modular forms. *Am. Math. Soc. Proc. Symp. Pure Math.* **8**, 1–15 (1965).
B. SIMON
1. "Functional Integration and Quantum Physics." Academic Press, New York, 1979.
U. SIMON and H. WISSNER
1. Geometry of the Laplace operator (preprint).
E. M. STEIN and G. WEISS
1. "Introduction to Fourier Analysis on Euclidean Spaces." Princeton Univ. Press, Princeton, New Jersey, 1971.
N. SUBIA
1. Formule de Selberg et formes d'espaces hyperbolique compacte. "Analyse Harmonique sur les Groupes de Lie," Lecture Notes Math. **497**, 674–700. Springer-Verlag, Berlin and New York, 1975.
D. SULLIVAN
1. "Seminar on Conformal and Hyperbolic Geometries." I.H.E.S., 1982.
2. λ-potential theory on manifolds (preprint).
T. SUNADA
1. Riemannian coverings and isospectral manifolds (preprint).

C. A. SWANSON

1. Asymptotic variational formulae for eigenvalues. *Can. Math. Bull.* **6**, 15–25 (1963).
2. A domain perturbation problem for elliptic operators. *Ann. Mat. Pura Appl.* **116**, 229–240 (1977).

G. SZEGÖ

1. Inequalities for certain eigenvalues of a membrane of given area. *J. Ration. Mech. Anal.* **3**, 343–356 (1954).

T. TAKAHASHI

1. Minimal immersions of Riemannian manifolds. *J. Math. Soc. Jpn.* **18**, 380–385 (1966).

M. TAYLOR

1. Estimate on the fundamental frequency of a drum. *Duke Math. J.* **46**, 447–453 (1979).

A. TERRAS

1. Noneuclidean harmonic analysis. *SIAM Rev.* **24**, 159–193 (1982).
2. Noneuclidean harmonic analysis, the central limit theorem, and long transmission lines with random inhomogeneities (preprint).

W. THURSTON

1. "The Geometry and Topology of 3-manifolds," Lect. Notes. Dept. of Math., Princeton University, Princeton, New Jersey, 1980.

K. UHLENBECK

1. Eigenfunctions of Laplace operators. *Bull. Am. Math. Soc.* **78**, 1073–1076 (1972).
2. Generic properties of eigenfunctions. *Am. J. Math.* **98**, 1059–1078 (1976).

H. URAKAWA

1. On the least eigenvalue of the Laplacian for compact group manifolds. *J. Math. Soc. Jpn.* **31**, 209–226 (1979).
2. Bounded domains which are isospectral but not isometric. *Ann. Sci. Ec. Norm. Super., Paris* **15**, 441–456 (1982).

S. R. S. VARADHAN

1. Diffusion processes in small time intervals. *Commun. Pure Appl. Math.* **20**, 659–685 (1967).

J. VAUTHIER

1. Diffusion sur une variété riemanniene complète à courbure négative. *C.R. Acad. Sci., Paris* **275**, 925–926 (1972).

M.-F. VIGNÉRAS

1. Variétés riemannienes isospectrales et non isométriques. *Ann. Math.* **112**, 21–32 (1980).

N. R. WALLACH

1. Minimal immersions of symmetric spaces into spheres. *In* "Symmetric Spaces" (W. M. Boothby and G. L. Weiss, eds.), pp. 1–40. Marcel Dekker, New York, 1972.

F. WARNER

1. "Foundations of Differential Manifolds and Lie Groups." Scott, Foresman & Co., Glenview, Illinois, 1971.

H. F. WEINBERGER

1. On the isoperimetric inequality for the *n*-dimensional free membrane problem. *J. Ration. Mech. Anal.* **5**, 633–636 (1956).

A. WEINSTEIN

1. On the volume of manifolds all of whose geodesics are closed. *J. Differ. Geom.* **9**, 513–517 (1974).

J. WERMER

1. "Potential Theory," Lect. Notes Math. **408**. Springer-Verlag, Berlin and New York, 1974.

H. WEYL

1. Der Asymptotische Verteilungsgesetz der Eigenwerte Linearer partieller Differentialgleichungen. *Math. Ann.* **71**, 441–469 (1912).

H. WHITNEY

1. "Geometric Integration Theory." Princeton Univ. Press, Princeton, New Jersey, 1957.

J. WOLF

1. "Spaces of Constant Curvature." McGraw-Hill, New York, 1967.

W. WOLFE

1. The asymptotic distribution of lattice points in hyperbolic space. *J. Funct. Anal.* **31**, 333–340 (1979).

S. WOLPERT

1. The length spectra as moduli for compact Riemann surfaces. *Ann. Math.* **109**, 323–351 (1979).

B. WONG, S. T. YAU, and S. S. T. YAU

1. An estimate of the gap of the first two eigenvalues in the Schrödinger operator (preprint).

C. T. YANG

1. Odd dimensional Wiedersehn manifolds are spheres. *J. Differ. Geom.* **15**, 91–96 (1980).

H. Z. YANG and J. Q. ZHONG

1. On the estimates for the first eigenvalue of a compact Riemannian manifold of non-negative Ricci curvature (in Chinese). *Acad. Sin.* October (1983).

P. C. YANG and S. T. YAU

1. Eigenvalues of the Laplacian of compact Riemann surfaces and minimal submanifolds. *Ann. Sci. Ec. Norm. Super., Pisa* **7**, 55–63 (1980).

K. YANO and S. BOCHNER

1. "Curvature and Betti Numbers." Princeton Univ. Press, Princeton, New Jersey, 1953.

S. T. YAU

1. Isoperimetric constants and the first eigenvalue of a compact manifold. *Ann. Sci. Ec. Norm. Super., Paris* **8**, 487–507 (1975).
2. Harmonic functions on complete Riemannian manifolds. *Commun. Pure Appl. Math.* **28**, 201–228 (1975).
3. Some function-theoretic properties of complete Riemannian manifolds, and their applications to geometry. *Indiana Univ. Math. J.* **25**, 659–670 (1976); also cf. *ibid.* **31**, 607 (1982).
4. On the heat kernel of a complete Riemannian manifold. *J. Math. Pures Appl.* **57**, 191–201 (1978).
5. Survey on partial differential equations in Riemannian geometry. "Seminar on Differential Geometry," *Ann. Math. Stud.* **102**, 3–72 (1982).

Index

A

Associated Legendre equation and function, 309

B

Berger isoembolic inequality, 120
Berger–Kazdan inequality, 114
Bochner–Lichnerowicz formula, 83

C

Cheeger inequality
  lowest Dirichlet eigenvalue, 95
  lowest nonzero eigenvalue, closed problem, 109
  lowest nonzero Neumann eigenvalue, 259
Co-area formula, 86
Comparison theorem
  Bishop, 68, 71
  Bonnet–Myers, 73
  Cheeger–Yau, 194
  Cheng, 70, 74
  Debiard–Gaveau–Mazet, 194
  Hadamard–Cartan, 68
  McKean, 70
  Morse–Schonberg, 68
  Obata, 82
  H. E. Rauch, 67, 74
  Toponogov, 83
Conjugate point, 62
Constant
  Cheeger, 95, 109
  isoperimetric, 96, 110
  Sobolev, 96, 110
Convergence of heat flow to steady state
  compact manifold, 141
  Dirichlet boundary data, 170–172
  Euclidean upper half-space, 325–329
  forms, 339
Coordinates
  Fermi
    based on submanifold, 319
    surface, 247
  geodesic spherical, 65ff.
  Riemann normal, 317
  spherical
    Euclidean space, 34
    hyperbolic space, 37
    sphere, 37
Cut locus, 64
Cut point, 64

D

Differential forms
  Cheeger's inequality and, 342
  deRham complex, 334
  eigenforms, 339
  eigenvalue estimates, 342
  Gauss–Bonnet theorem and heat equation, 340–341
  heat kernel, 338
  Hodge decomposition, 335
  Laplacian on, 335
Dirichlet integral, 14
Divergence theorem, 6, 7
Duhamel's principle, 137, 145, 165

E

Eigenvalues
  closed problem, 7
    bounds
      lower, 112, 133, 251, 332
      upper, 77, 332
        Löbell surfaces, 251

Eigenvalues, con't.
    calculation
      projective space, 36
      sphere, 33
      torus, 28
    lower, relation to short geodesics, Riemann surfaces, 254–262
    lowest nonzero
      bounds
        lower, 82, 109, 127, 314, 324, 332
        upper, surface, 94
      pinching theorem, 92
    small, Riemann surface, 247, 300–302
  Dirichlet problem, 8
    bounds, lower, 95, 107, 330
    calculation
      disk, in space form, 40–42
      domain, crystallographic group, 33
      equilateral triangle, 33
    lowest
      bounds
        lower, 70, 79, 82, 87, 95, 129
          hyperbolic space, 46
        upper, 74, 79, 129
          disk, hyperbolic space, 80–82
          Löbell *Y*-piece, 249
      disk, space form, 42–45
        Euclidean space, 45
        hyperbolic space, 46
        sphere, 50–54
  domain monotonicity
    vanishing Dirichlet data, 17
    vanishing Neumann data, 18
  lower bounds from upper bounds on heat kernel, 157
  max–min characterization, 17
  minimal immersions and, 312–314
  mixed eigenvalue problem, 8
  multiplicity of, 24, 35, 104
  Neumann problem, 8
    calculation, disks in space form, 40–42
    lowest nonzero
      bounds
        lower, 251
        upper, 94
      disks in space form, 42–45
        Euclidean space, 45
        hyperbolic space, 46
        spheres, 50
  1-dimensional problems, 10
  Rayleigh's characterization, 16
  Sturm–Liouville decomposition
    closed eigenvalue problem, 139
    Dirichlet eigenvalue problem, 169
  topological perturbations with negligible effect on, 207–235
Eigenfunction, L^2–L^x estimates, 102, 112

F

Fourier transform, 304 ff.
  Fourier inversion formula, 305
  heat equation, 307
  Plancherel formula, 305
  Poisson summation formula, 306
  Weyl asymptotic formula, torus, 307
Function
  admissible for eigenvalue problems, 15
  directional derivative of, 1
  gradient of, 2
  Hessian of, 82
  Laplacian of, 3
  Sobolev space of, 14
  weak derivative of, 14
Fundamental solution, heat equation, *see* Heat kernel

G

Gamma function, 303
Green's formulas, 6, 7, 165
Green's function, 173
  Dirichlet problem, 175
  forms, 339
  relation to heat kernel, 177

H

Heat equation
  homogeneous, 134
  inhomogeneous, 135
  separation of variables, 12
Heat kernel
  almost Euclidean, 159
  Brownian motion associated to, 210ff.
  comparison theorems, 192–196
  compact manifold
    bounds
      lower, 333

      upper, 112, 133
    existence 151–154
    uniqueness, 138
  comparison theorems, 192–196
  Dirichlet, 158
    bounds
      lower, 195
      upper, 108, 195, 331–332
    existence, 164
    uniqueness, 167
  Euclidean space, 142–148, 307–308
  Fourier transform and, 306–308
  noncompact manifolds, *see also* Heat kernel, Euclidean space
    bounds
      lower, 196
      upper, 180, 196–206
    Euclidean upper half-space, 326
    existence, 188
    space form, 150, 151
      hyperbolic plane, 242–246
  topological perturbations with negligible effect on, 207–235
Hyperbolic plane
  geometry, 239–241
  heat kernel, 242–246
  upper half-plane model, 262–264
Hyperbolic space
  disk model, 37
  geometry, 266–271
  upper half-space model, 264–265

I

Injectivity radius, 118, 120, 247
Isoperimetric inequality
  geometric, 85, 89, 90, 92, 313
    Lévy–Gromov, 91, 322-325
  physical
    closed eigenvalue problem, surfaces, 94
    Faber–Krahn, 87
      minimal surfaces and, 93
      variation of, 91, 93
    Neumann eigenvalue problem, 94

J

Jump relation, 159, 161

K

Kinematic measure, 117
  Liouville's theorem, 117
  Santalo's formula, 125, 130

M

Manifold
  connection on, 2
    Christoffel symbols of, 4
    curvature
      operator along geodesic, 63
      tensor, 59
    exponential map, 57
    geodesic, 57
    geodesic completeness, 57
  Riemannian metric, 2
  tangent bundle of, 1
  tangent space to, 1
Mathematics of crushed ice, 233–238
Maximum principle
  heat operator, 139, 166, 169, 180, 182
  Laplace operator, 20, 127, 329–330
Metaphysical principle, 77
Minimal immersions and eigenvalues, 312–314

N

Nodal domains, 19
  Courant's theorem, 19, 92
  Pleijel's theorem, 24, 92
Nodal sets, 19
  convexity of, 24
  regularity of, 23

P

Polya conjecture, 33, 330, 333

R

Ricatti equation, 72
Riemannian covering, deck transformation group, 27
Riemannian manifold
  curvature
    Gauss, 60

Riemannian manifold, con't.
        Ricci, 60
        scalar, 60
        sectional, 59
    distance between points of, 39
    Gauss's lemma, 59
    length of paths in, 39
    Levi–Civita connection on, 2
    open disk, 40
    sphere in, 40
    submanifold
        cut point, 320
        focal point, 320
        mean curvature vector, 310
        normal bundle, 224
        second fundamental form, 310
        totally geodesic, 131
        Weingarten map, 319
    surfaces
        Gauss–Bonnet formula and theorem, 60–61
        Gauss curvature, 60
        theorema egregium, 60

S

Selberg trace formula, 266–302
    pretrace formula, 277
        applications
            electrostatics, 282–288
            lattice-point estimates, 279–281
            projections of dilating spheres, 281–282
    trace formula, 293
        applications
            Laplace and geodesic spectra, 293–300
            small eigenvalues, 300–302
Separation of variables
    eigenfunctions and eigenvalues of disk in space form, 40
    heat equation, 12
    wave equation, 11
Surface, 246
    collar about geodesic, 247
    Riemann, 246
        Löbell, 248–250

V

Vector field
    along path
        derivative of, 55
        parallel, 56
        parallel translation of, 57
    covariant derivative of, 2
    divergence of, 3
    transverse, 319
Visibility angle, 125
Volumes, disks and spheres, 303–304

W

Wave equation
    finite propagation speed, 198
    separation of variables, 11
Weyl's asymptotic formula, 9
    bounded domain in $\mathbb{R}^n$, 31
    closed eigenvalue problem, 155
    Dirichlet eigenvalue problem, 172
    forms, 339
    1-dimensional problems, 10
    Poisson summation formula and, 306–307
    torus, 30
Wiedersehnsraum, 123
    Blaschke conjecture, 123
Wirtinger inequality, 130

Pure and Applied Mathematics

A Series of Monographs and Textbooks

Editors **Samuel Eilenberg and Hyman Bass**
Columbia University, New York

RECENT TITLES

Carl L. DeVito. Functional Analysis
Michiel Hazewinkel. Formal Groups and Applications
Sigurdur Helgason. Differential Geometry, Lie Groups, and Symmetric Spaces
Robert B. Burckel. An Introduction to Classical Complex Analysis: Volume 1
Joseph J. Rotman. An Introduction to Homological Algebra
C. Truesdell and R. G. Muncaster. Fundamentals of Maxwell's Kinetic Theory of a Simple Monatomic Gas: Treated as a Branch of Rational Mechanics
Barry Simon. Functional Integration and Quantum Physics
Grzegorz Rozenberg and Arto Salomaa. The Mathematical Theory of L Systems
David Kinderlehrer and Guido Stampacchia. An Introduction to Variational Inequalities and Their Applications
H. Seifert and W. Threlfall. A Textbook of Topology; H. Seifert. Topology of 3-Dimensional Fibered Spaces
Louis Halle Rowen. Polynominal Identities in Ring Theory
Donald W. Kahn. Introduction to Global Analysis
Dragos M. Cvetkovic, Michael Doob, and Horst Sachs. Spectra of Graphs
Robert M. Young. An Introduction to Nonharmonic Fourier Series
Michael C. Irwin. Smooth Dynamical Systems
John B. Garnett. Bounded Analytic Functions
Eduard Prugovečki. Quantum Mechanics in Hilbert Space, Second Edition
M. Scott Osborne and Garth Warner. The Theory of Eisenstein Systems
K. A. Zhevlakov, A. M. Slin'ko, I. P. Shestakov, and A. I. Shirshov. Translated by Harry Smith. Rings That Are Nearly Associative
Jean Dieudonné. A Panorama of Pure Mathematics; Translated by I. Macdonald
Joseph G. Rosenstein. Linear Orderings
Avraham Feintuch and Richard Saeks. System Theory: A Hilbert Space Approach
Ulf Grenander. Mathematical Experiments on the Computer
Howard Osborn. Vector Bundles: Volume 1, Foundations and Stiefel-Whitney Classes
K. P. S. Bhaskara Rao and M. Bhaskara Rao. Theory of Charges
Richard V. Kadison and John R. Ringrose. Fundamentals of the Theory of Operator Algebras, Volume I
Edward B. Manoukian. Renormalization
Barrett O'Neill. Semi-Riemannian Geometry: With Applications to Relativity
Larry C. Grove. Algebra
E. J. McShane. Unified Integration
Steven Roman. The Umbral Calculus
John W. Morgan and Hyman Bass (Eds.). The Smith Conjecture

Sigurdur Helgason. Groups and Geometric Analysis: Integral Geometry, Invariant Differential Operators, and Spherical Functions

Isaac Chavel. Eigenvalues in Riemannian Geometry

IN PREPARATION

Robert B. Burckel. An Introduction to Classical Complex Analysis: Volume 2

Richard V. Kadison and John R. Ringrose. Fundamentals of the Theory of Operator Algebras, Volume II

A. P. Morse. A Theory of Sets, Second Edition

E. R. Kolchin. Differential Algebraic Groups

W. D. Curtis and F. R. Miller. Differential Manifolds and Theoretical Physics

Jean Berstel and Dominique Perrin. Theory of Codes