9 Riemannian Manifolds as Quantum Mechanical Worlds: The Spectrum and Eigenfunctions of the Laplacian

Contents

M. Berger, *A Panoramic View of Riemannian Geometry*

9.1 History

While harmonic analysis on domains in Euclidean space is a long established field, as seen in §1.8, the study of the Laplace operator on Riemannian manifolds (together with the heat and wave equations, and the spectrum and eigenfunctions) seems to have begun only quite recently. Some of the earliest accomplishments were the computation of the spectrum of $\mathbb{CP}^n$ (see §§9.5.4) and Lichnerowicz's inequality for the first eigenvalue (see §§9.10.1). The first paper to address the Laplacian on general Riemannian manifolds was Minakshisundaram & Pleijel 1949 [930]. More narrowly, Maaß 1949 [889] investigated the Laplacian on Riemann surfaces. Also, one can turn to Avakumović 1956 [90]. But a spark was lit in the 1960's when Leon Green asked if a Riemannian manifold was determined by its spectrum (the complete set of eigenvalues of the Laplacian).

In the special case of Riemann surfaces, a deep study of the spectrum can be found as early as 1954 in Selberg 1954 [1120, 1121] and 1955 in Huber 1956,1959,1961 [744, 745, 747].

Green's isospectral question was answered in the negative in Milnor 1964 [922]. Kac 1966 [775] in 1966 was also very influential. But the two major events were the papers of McKean & Singer 1967 [910] and Hörmander 1968 [734]. We will meet them below; let us just say that the first paper pioneered the study of the heat kernel expansion in Riemannian geometry, and its consequences. The second was more general, treating the case of a general elliptic operator, without reference to any Riemannian structure on the manifold under consideration. But it introduced the wave equation technique, microlocal analysis and symplectic geometry. This technique is indispensable when studying the relations between the spectrum and the geodesic flow; see §9.9. Thereafter the subject became a vast field of inquiry.

Note 9.1.0.3 (On the bibliography) As we go on in this book, we will have to give less and less detail, in order to keep the book of reasonable size. Then the reader will want to ask for more and more references, especially those of general character, as opposed to research articles. There are now quite a few books which addressing the topic of the present chapter. Some people still like Berger, Gauduchon, & Mazet 1971 [174] for an introduction and basic facts. But on most of the more advanced topics that book is completely outdated. New texts are: Chavel 1984,1993 [325, 326], Buser 1992,1997 [292, 293] which discusses only the spectral geometry of surfaces, Sakai 1996 [1085] chapter VI, Gilkey 1995 [564] (try to get this second edition), Bérard 1986 [135] which is very expository but outdated on some advanced topics, Guillemin & Sternberg 1977 [670]. In particular, Gilkey 1995 [564] is important and contains an amazing collection of mathematics in a single book, e.g. the η invariant which is hard to find in books. The heat equation, in a very general context, is also analyzed in Berline, Getzler & Vergne 1992 [179] and in Gilkey, Leahy & Park 2000 [566].

There are few completely expository books on the wave equation technique and microlocal analysis. The bible of Hörmander 1983 [737, 738] is hard to read, but Trèves 1980 [1198, 1199] is very informative. With a view toward physics, Guillemin & Sternberg 1977 [670] is fascinating. Note also that Bérard 1986 [135] contains a very extensive bibliography, but up to date only to 1986. ♦

9.2 Motivation

Why should a geometer, whose principal concern is in measurements of distance, desire to engage in analysis on a Riemannian manifold? For example, pondering the Laplacian, its eigenvalues and eigenfunctions? Here are some reasons, chosen from among many others. We note also here that the existence of a canonical elliptic differential operator on any Riemannian manifold, one which is moreover easy to define and manipulate, is one of the motivations to consider Riemannian geometry as a basic field of investigation. For Laplacians on more general spaces, see §14.5 and §14.6.

Riemannian geometry is by its very essence differential, working on manifolds with a differentiable structure. This automatically leads to analysis. It is interesting to note here that, historically, many great contributions to the field of Riemannian geometry came from analysts. Let us present a few names (we do not pretend to be exhaustive). Hadamard's contribution in quotation 10.1 on page 434 goes back to 1901 and Poincaré's in §§10.3.1 to 1905. Élie Cartan was an analyst; see Chern & Chevalley 1952 [368]. More recently, let us mention Bochner, (see theorem 345 on page 594), Nirenberg (see §§4.6.1), Chern, Calabi, Aubin, Yau and Gromov.

We will see deep links between the spectrum (especially the first eigenvalue λ_1) and periodic geodesics in §9.9 as well as in theorem 205 on page 447. The proof of Colding's L^1 theorem 76 on page 262 rests essentially on analysis, as we have briefly seen there. Harmonic coordinates turn out to be a godsend when studying convergence of Riemannian manifolds: see §§6.4.3 and §§12.4.2.

The deformation of a Riemannian metric via a parabolic evolution equation, which is based on hard techniques from the theory of partial differential equations, is extremely useful. We will see this in more than one instance: see §§11.4.3, the smoothing techniques in §§12.4.2 and the new proof of the conformal representation theorem 70 on page 254. This is one of many evolution equations arising in Riemannian geometry. Another type of evolution equation is the heat equation which will turn out not only to be useful in establishing the existence and some of the first properties of the spectrum and of the eigenfunctions (see §1.8), but has become a basic tool in a large number of contexts; see §9.7.

Finally let us mention harmonic maps (see §14.3), minimal submanifolds (e.g. for the theorem on manifolds with positive curvature operator in §§§12.3.1.4), and the use of geometric measure theory. And do not forget harmonic coordinates.

From the point of view of theoretical physics, it is very natural to consider the *semiclassical limit*, which is the limiting behaviour of the solutions of the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \hbar^2 \Delta f$$

as $\hbar \to 0$. In Euclidean space, this is equivalent to rescaling the spatial coordinates outward, looking at the large scale physics. The hope is that classical mechanics will emerge from this limit in some sense. This suggests looking at the asymptotic expansion of the eigenvalues λ_i as $i \to \infty$. This explains why we so often mention results such as theorems 164 on page 386, 172 on page 401, 174 on page 403, and 175 on page 404.

Every manifold is COMPACT and connected unless otherwise stated.

9.3 Setting Up

9.3.1 Xdefinition

Recalling §1.8 and §1.9, even before tackling the heat equation, the first thing to do is to define the Laplacian Δ on a Riemannian manifold M with metric g. It is a second order elliptic differential operator, attached intrinsically to M. It is not surprising that one can give many equivalent definitions of it. We start with the most natural, as soon as one knows that the Riemannian metric enables us to define an intrinsic second derivative (which is not the case for a manifold with "only" a smooth structure). To every smooth numerical function

$$f : M \to \mathbb{R}$$

we attach its Hessian

$$\operatorname{Hess} f$$

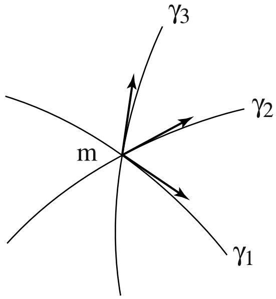

Fig. 9.1. Calculating the Laplacian by differentiating along an orthogonal system of geodesics and taking the sum

which is the bilinear symmetric quadratic differential form made up by the second derivatives of f. Namely, using the covariant derivative D, we set

$$\operatorname{Hess} f = Ddf$$

(see §15.5 if needed). To get a numerical function from this Hessian, we need only take its trace with respect to the metric g. For technical reasons, we add a minus sign. Beware that this is a matter of convention, and the convention depends on the author. The negative sign insures us that the eigenvalues will be nonnegative (in fact, positive except the 0^{th} whose eigenvalues are the constant functions). The *Laplacian* of f is then defined as

$$\Delta f = -trace_g \operatorname{Hess} f \ .$$

Since along geodesics, the (covariant) second derivative coincides with the ordinary numerical second derivative, by the definition of the trace with respect to g, the geometer will define the Laplacian of f at a point $m \in M$ as

$$\Delta f(m) = -\sum_{i=1}^{d} \frac{d^2}{dt^2} f\left(\gamma_i(t)\right)\Big|_{t=0}$$

where the γ_i are geodesics through m whose velocities at m form an orthonormal basis of $T_m M$. In particular, at the center m of a system of normal coordinates, this is written

$$\Delta f(m) = -\sum_{i=1}^{d} \frac{\partial^2 f}{\partial x_i^2}(m) \ .$$

This cannot be used as a definition directly, since one needs to show that such a description yields a well defined differential operator. Two other definitions can be given. The first uses the Hodge $*$ operation applied to differential forms, which will be defined in §§9.3.2. Then

$$\Delta f = - * d * d \,.$$

Using any definition, in general coordinate systems we find

$$\Delta f = -\frac{1}{\sqrt{\det g}} \sum_{j,k} \frac{\partial}{\partial x_j} \left(g^{jk} \sqrt{\det g} \frac{\partial f}{\partial x_k} \right) \tag{9.1}$$

where

$$\det g = \det \left(g_{ij} \right)$$

and the g^{jk} are the matrix elements of the inverse matrix to g_{jk}. We won't need to use this complicated formula. From all of these definitions, one sees that this Δ extends to any Riemannian manifold the Δ of Euclidean space as defined in equation 1.27 on page 94 and the Δ of the sphere defined in equation 1.28 on page 95. In this formula, one sees that the Laplacian involves the metric g and its first derivatives; this makes it an invariant which is not C^0 robust, only C^1 robust. However, §9.4 will show that the spectrum is C^0 robust. This is the beginning of spectral analysis for more general geometries; see §14.5 and §14.6.

If you are familiar with the notion of *symbol* of a differential operator, then the best way to define and to see the uniqueness of the Laplacian is to say that Δ is the second order differential operator whose principal symbol is $-g$ (the quadratic form giving the metric g) and which has no term of order zero.

If we construct a function measuring distance from some point then, when written in polar geodesic coordinates centered at that point, the Ricci curvature comes into the formula giving the Laplacian of this distance function. We employed this fact when proving Colding's L^2 theorem 77 on page 264:

$$\frac{d}{ds} \Delta f \circ \gamma + \frac{1}{d-1} \left(\Delta f \circ \gamma \right)^2 \leq - \operatorname{Ricci} \left(\gamma', \gamma' \right) \,.$$

9.3.2 The Hodge Star

To present many of the foundational facts in spectral geometry[1] we need the definition of the Laplace operator Δ on differential forms and the concept of adjoint operator. We first denote by $\Omega^p (M)$ the space of differential forms of degree p on the differentiable manifold M, which is defined on any differentiable manifold, without need for a Riemannian metric; see §§4.2.2. But if M is moreover equipped with a Riemannian metric and oriented, then there is an linear operator

$$* : \Omega^p (M) \to \Omega^{\dim(M)-p} (M)$$

called the Hodge star operator. Choosing a positive orthonormal basis

$$\left\{ e_i \right\}_{i=1,\dots,d}$$

[1] For example, theorems 338 on page 588 and 405 on page 665; also see §9.14.

for the tangent space $T_m M$ at a point $m \in M$, define

$$*\alpha\left(e_{p+1},\ldots,e_d\right)=\alpha\left(e_1,\ldots,e_p\right) .$$

This turns out to be independent of the choice of oriented orthonormal basis. The square of $*$ is plus or minus the identity on $\Omega^p(M)$:

$$*^2=(-1)^{p(\dim(M)-p)} .$$

The differential operator d is transformed by $*$ into another first order operator, denoted by d^* (sometimes also by δ)

$$d^*=(-1)^{1+d(p+1)} * d*$$

which is not dependent on the choice of orientation, hence is intrinsic. The reason for the notation

$$\delta=d^*$$

is that it is the adjoint of d:

$$\int_M d\alpha\wedge\beta=\int_M \alpha\wedge d^*\beta \tag{9.2}$$

for any

$$\alpha,\beta\in\Omega^p(M)$$

and any $p=0,\ldots,\dim(M)$. We can define a Laplacian for exterior forms of any degree by

$$\Delta=-\left(dd^*+d^*d\right)=-\left(d+d^*\right)^2 . \tag{9.3}$$

For the moment, we will only use the Laplacian on functions, i.e. $p=0$. This Δ is the same as the one previously defined in this chapter. A useful formula, valid for any pair of functions, is

$$\begin{aligned}\int_M g\Delta f&=\int_M\langle df,dg\rangle\\&=\int_M f\Delta g\end{aligned} \tag{9.4}$$

in particular

$$\int_M \Delta f=0$$

for any function f.

When using integrals like the above on compact Riemannian manifolds, we will often omit the Riemannian canonical measure:

$$\int_M f=\int_M f\,dV_M .$$

9.3.3 Facts

The Laplacian on any compact Riemannian manifold provides us with all the tools of Fourier analysis on our Riemannian manifold. Let us call a function ϕ an *eigenfunction* with *eigenvalue* the number λ if

$$\Delta f = \lambda f \ .$$

The set of all eigenvalues of Δ is an infinite discrete subset of $\mathbb{R}^+$ called the *spectrum* of Δ

$$\operatorname{Spec}(M) = \{\lambda_k\} = \{0 < \lambda_1 < \lambda_2 < \dots\} \tag{9.5}$$

with λ_k tending to infinity with k.

For each eigenvalue λ_i, the vector space of eigenfunctions ϕ satisfying

$$\Delta f = \lambda_i f$$

is always finite dimensional and its dimension is called the *multiplicity* of λ_i. Once we have a basis of the eigenfunctions with this eigenvalue written out, it is trivial to find an orthonormal basis

$$\{\phi_k\}$$

(where k runs from 1 to the multiplicity) of eigenfunctions. Here the orthonormalcy is to be understood for the global scalar product

$$\langle f, g \rangle_{L^2(M)} = \int_M fg \ .$$

Note that equation 9.4 on the preceding page shows (a classical fact) that eigenfunctions with different eigenvalues are automatically orthogonal. Unlike the domains in Euclidean space which we treated in chapter 1, our compact Riemannian manifolds have no boundary. This explains why we get the "extra" eigenvalue

$$\lambda_0 = 0$$

whose eigenfunctions are the constant functions.[2]

Note 9.3.3.1 Beware now that there are two different ways of writing the eigenvalues and the eigenfunctions when making sums. In the first one, we understand that a sum over the spectrum sums each eigenvalue a number of times given by its multiplicity. In the other notation, the indices are not those used in equation 9.5, but instead the index moves up at each eigenvalue through the entire multiplicity. Which sort of summation is required will always be clear from the context, as in what follows for example. ♦

[2] Since the manifold M is assumed to be connected, the multiplicity of λ_0 is exactly one.

As for classical Fourier series, any reasonable function

$$f : M \to \mathbb{R}$$

has Fourier coefficients

$$a_i = \int_M f\phi_i$$

and f is recovered from these coefficients by the converging series

$$f = \sum_i a_i\phi_i \ .$$

In the same spirit, the scalar product of two functions is the sum of products of their coefficients:

$$\int_M fg = \sum_i a_i b_i$$

where

$$f = \sum_i a_i\phi_i$$

$$g = \sum_i b_i\phi_i \ .$$

9.3.4 Heat, Wave and Schrödinger Equations

We will follow the same steps that we did in §1.8: defining heat, wave and Schrödinger equations on Riemannian manifolds. The heat equation for the heat $f(m,t)$ at time t at a point m of the Riemannian manifold M is

$$\Delta f = -\frac{\partial f}{\partial t} \ . \tag{9.6}$$

The wave equation for the height $f(m,t)$ of the "water" after time t at a point m is

$$\Delta f = -\frac{\partial^2 f}{\partial t^2} \ . \tag{9.7}$$

where if M were a surface, you would consider M covered in a thin sheet of water, or for M of three dimensions, M is a place through which sound is propagating. The wave equation can also be considered as describing the manifold M as a vibrating membrane object. Finally the Schrödinger equation uses complex valued functions and is written

$$\hbar^2 \Delta f = i\hbar\frac{\partial f}{\partial t} \tag{9.8}$$

where $i = \sqrt{-1}$ and $\hbar$ is Planck's constant.

To solve these equations, at least formally, one uses the same trick as in §§1.8.1. To solve such an equation depending both on time t and a point $m \in M$, the initial idea is to use the fact that, roughly by the Stone–Weierstraß approximation theorem, we need only to consider product functions

$$f(m,t) = g(m)h(t) \; .$$

One will subsequently consider series of them (as in the theory of Fourier series). Look for example at the heat equation. The function $f = gh$ satisfies the heat equation precisely when the functions g and h satisfy

$$\frac{\Delta g}{g} = -\frac{h'}{h} \tag{9.9}$$

where

$$h'(t) = \frac{dh}{dt}$$

is the usual derivative.

Since the first fraction depends only on the point $m \in M$ and the second only on the time t their common value has to be a constant, call it λ. Then the function

$$g : M \to \mathbb{R}$$

is an eigenfunction of Δ with eigenvalue λ, while h is an exponential decay at rate λ. If all eigenfunctions and eigenvalues of Δ are known, we can then solve the heat equation explicitly. Note that the time dependence $h(t)$ is

$$h(t) = \begin{cases} e^{-\lambda t} & \text{for the heat equation} \\ e^{i\lambda t} & \text{for the Schrödinger equation} \\ e^{i\sqrt{\lambda} t} & \text{for the wave equation.} \end{cases}$$

Physically, the product motions $g(m)h(t)$ are the stationary ones—they are the ones we can observe through some kind of "Riemannian stroboscopy."

As we did in Euclidean space, we will begin our analysis with the *fundamental solution of the heat equation*, denoted $K(m,n,t)$. One also calls it the *heat kernel*. It is a function

$$K : M \times M \times \mathbb{R}^+ \to \mathbb{R} \; .$$

It has the property that the solution $f(m,t)$ of the heat equation with initial temperature $f(m,0)$ at time zero is

$$f(m,t) = \int_M K(m,n,t) f(n,0)\, dn$$

and one can prove that the heat kernel is the sum of the convergent series

$$K(m,n,t) = \sum_{k=1}^{\infty} \phi_k(m)\phi_k(n)e^{-\lambda_k t}. \tag{9.10}$$

The reader can check this formally, ignoring convergence, by just plugging the series into the integral. The hard part, which required analysts' efforts, is to prove the convergence.

Another way to write the solution $f(m,t)$ with initial temperature $f(m,0)$ is to compute the Riemannian Fourier series

$$f(m,0) = \sum_{k=1}^{\infty} a_k \phi_k$$

and then

$$f(m,t) = \sum_{k=1}^{\infty} a_k \phi_k(m) e^{-\lambda_k t} .$$

For the wave equation, the fundamental solution similar to equation 9.10 on the facing page requires imaginary terms, i.e.

$$e^{i\sqrt{\lambda_k}t}$$

which are linear combinations of

$$\cos\left(\sqrt{\lambda_k}t\right) \text{ and } \sin\left(\sqrt{\lambda_k}t\right) .$$

But the dramatic difference between the heat equation and the wave equation is that waves demand not converging series, but distributions. Heat spreads out uniformly with time, while waves bounce up and down forever. This major difference explains why working with the wave equation (in Riemannian manifolds, but also in Euclidean spaces) is much more expensive mathematically. We refer to our bibliographical introduction for references. Note that the conservative nature of waves will provide an amazing source of information in §9.8. Another major difference between the heat equation and the wave equation is that for the waves one does really need to work in the tangent bundle and use the tools of microlocal analysis; a most informative book on the subject is Trèves 1980 [1198, 1199].

9.4 The Cheapest (But Most Robust) Method to Obtain Eigenfunctions: The Minimax Principle

9.4.1 The Principle

Analysis and convergence problems (which we will not attempt to explain) are very well exposed in Bérard 1986 [135]. We will begin as we did in §§1.8.3. One way to identify and then study the eigenfunctions is as follows. One pulls out the first one by the so-called *Dirichlet principle.* Among all functions, one looks for one minimizing the ratio

$$\text{Dirichlet}(f) = \frac{\int_M \|df\|^2}{\int_M f^2} \tag{9.11}$$

called the *Dirichlet quotient.*

The infimum value of zero is trivially attained for constant functions. So we look next to minimize this quotient among functions which are "not constant," more precisely among those functions orthogonal to constants, i.e. functions f with

$$\int_M f = 0 \; .$$

Let us compute the derivative of this ratio with respect to a variation $f + \varepsilon g$ of the function f (assuming it exists) achieving such a minimum, and use formula 9.3 on page 379 together with the Lagrange multiplier technique. We find that f necessarily satisfies

$$\Delta f = \lambda_1 f$$

for the constant

$$\lambda_1 = \inf \left\{ \frac{\int_M \|df\|^2}{\int_M f^2} \; : \; \int_M f = 0 \right\} \; . \tag{9.12}$$

Rescale f to have unit norm

$$\int_M f^2 = 1 \; .$$

This yields the first (nontrivial) eigenfunctions together with the first eigenvalue. Unlike a Euclidean domain, where there was only one first eigenfunction, here there may be a finite dimensional vector space of them; for example the sphere of dimension d has a $d+1$ dimensional space of eigenfunctions with the same eigenvalue λ_1.

To get the next eigenfunctions and values, one just applies the same trick, but restricting the set of functions f into consideration to the set of functions which are orthogonal to the first eigenfunctions

$$\int_M f \phi_i = 0$$

for $\phi_1, \ldots, \phi_h$ a basis for the eigenfunctions with eigenvalue λ_1. And keep going on in this way.

But this procedure necessitates calculating all of the eigenfunctions preceding the one that we might be looking for. To get around this obstacle, a wonderful trick was invented, the *minimax principle.* We first state the result, and then explain it geometrically on ordinary ellipsoids in $\mathbb{E}^3$. The eigenvalue λ_{k+1} is exactly

$$\lambda_k = \inf_V \sup_{f \in V} \text{Dirichlet}(f) \tag{9.13}$$

where V runs through all $k+1$ dimensional vector subspaces of the vector space of real valued functions on M.

The proof is detailed in the beginning of Bérard 1989 [135]. It involves a little linear algebra (geometrically pictured in figure 9.2) and of course some analysis, since we are working in an infinite dimensional space.

In Dirichlet(f) there are two positive definite quadratic forms. In $\mathbb{R}^3$, say that the first has as its unit level set an ellipsoid, and the second is the Euclidean structure (i.e. its unit level set is the unit sphere). Then the eigenfunctions correspond to the three principal axes of the ellipsoid, and the eigenvalues are their lengths. To find the length of the second principal axis, consider all of the ellipses obtained by cutting the ellipsoid by planes through the origin. The largest principal axis that occurs among all of the ellipses is the largest axis of the ellipsoid. The second largest axis of the ellipsoid is the largest number that occurs among all ellipses as the smaller of the two axes.[3]

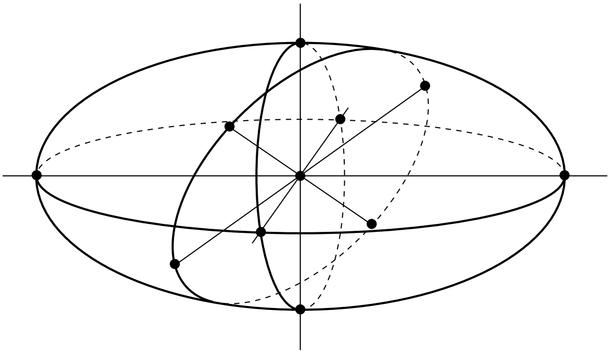

Fig. 9.2. The Dirichlet quotient is a quadratic function on the unit sphere in the infinite dimensional space of functions

The above method heuristically explains why every function is equal to a series of eigenfunctions and, since the space of functions is infinite dimensional, why the spectrum goes to infinity.

Note 9.4.1.1 A theoretical, but important, consequence of the minimax principle is that the spectrum is a robust invariant of the Riemannian metric; it depends only on the metric g, not on its derivatives (unlike the Laplacian itself); see equation 9.1 on page 378. Therefore the spectrum can be defined in a more general context; see §14.6. ♦

9.4.2 An Application

One of the main tasks when studying the spectrum of Riemannian manifolds is to relate the spectrum to the Riemannian invariants, for example the curva-

[3] It is harder to say than to see.

tures, the volume, the diameter, the injectivity radius, etc. This is the central objective of this chapter. So we start right away with an application of the minimax principle, given in Gromov 1999 [633]; for details, improvement and explicit constants we refer the reader to Bérard, Besson, & Gallot 1985 [139].

Theorem 164 *There is a universal constant*

$$\mathrm{univ}(d, r)$$

depending only on the dimension d of a compact Riemannian manifold M and the lower bound r of the Ricci curvature, such that for every k the eigenvalue λ_k of M obeys the upper bound

$$\lambda_k \leq \frac{\mathrm{univ}(d, r)}{\mathrm{Vol}(M)^{2/d}} k^{2/d} \ .$$

The asymptotic behavior in $k^{2/d}$ agrees with that which we will see in theorem 172 on page 401. Upper bounds are in general easier to get than lower ones. The reason is that the minimax principle, as we are going to see, shows that one can use upper bounds on the Dirichlet quotient for suitable functions to control the asymptotics of eigenvalues. For the proof, let us think of large indices k. The idea is to pack in M, as densely as possible, a set of metric balls

$$B_i = B\left(p_i, R\right) \ .$$

The number N of balls is controlled first by the usual metric trick of lemma 125 on page 333: if it is as dense as possible, then the balls

$$B\left(p_i, 2R\right)$$

will completely cover M. This enables us to estimate N with Ricci curvature thanks to Bishop's theorem 107 on page 310.

Now on every ball $B\left(p_i, R\right)$ we define a function f_i vanishing at the boundary of $B\left(p_i, R\right)$ and with a low Dirichlet quotient. This can be done by transferring (in polar coordinates on $B\left(p_i, R\right)$) the first eigenfunction g for the Dirichlet problem in the manifold with boundary

$$B\left(\mathbb{S}^d\left(\frac{r}{d-1}\right), R\right)$$

which is the metric ball of radius R in the comparison space

$$\mathbb{S}^d\left(\frac{r}{d-1}\right)$$

of constant curvature and whose Ricci curvature is our lower bound r. Knowledge of Ricci curvature permits us to control the Dirichlet quotient during the transfer (compare this with the geodesic transfer for Rauch–Toponogov theorems of chapter 6). This was done in Cheng 1975 [362]. There is a very nice proof today of Cheng's result, which is put in a very general context with a beautiful formula in Savo 1996 [1098]. On these balls in $\mathbb{S}^d\left(\frac{r}{d-1}\right)$ the first eigenvalue is known. This transplantation is similar, but not quite the same, as that of the Faber–Krahn inequality 1.22 on page 81. One finishes the estimate by applying the minimax principle to the N dimensional vector space of functions which is spanned by the f_i.

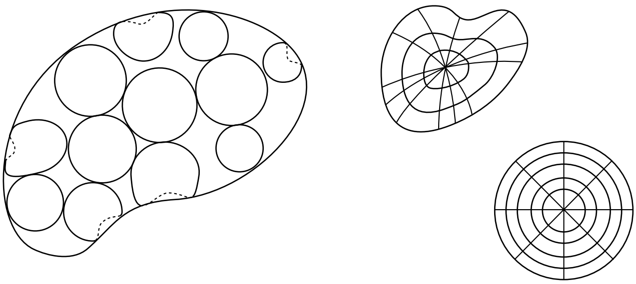

Fig. 9.3. Pack balls into your manifold, and transfer eigenfunctions into them from space forms $\mathbb{S}^d\left(\frac{r}{d-1}\right)$

9.5 Some Extreme Examples

Let us describe the spectral geometry of the most tractable Riemannian manifolds. We will follow more or less the geometric hierarchy of §6.6.

9.5.1 Square Tori, Alias Several Variable Fourier Series

The theory of eigenfunctions on tori, square or rectangular, is very much like that which we met in equation 1.21 on page 76 for the plane rectangle, except that now we use a periodic boundary condition, and of course we work in any dimension d. The variables $x_1, \dots, x_d$ separate for the Laplacian and we still have the Stone–Weierstraß theorem enabling us to look only at product functions

$$f(x_1, \dots, x_d) = f_1(x_1) \cdots f_d(x_d) \ .$$

Our torus is the quotient of $\mathbb{R}^d$ by the group $\mathbb{Z}^d$ of integral translations (this means that all sides of the box have unit length). The Riemannian structure we consider on it is of course the locally Euclidean one just obtained by the quotient operation. The functions $f_j(x_j)$ are linear combinations of

$$\cos(2\pi m_j x_j) \ \text{and} \ \sin(2\pi m_j x_i)$$

with the m_j any integers; the resulting product function is an eigenfunction of Δ with eigenvalue

$$4\pi^2\left(m_1^2+\cdots+m_d^2\right) .$$

As in the classical theory of Fourier series, these functions are rich enough so that there are no other eigenfunctions except appropriate linear combinations of these ones. The set of eigenvalues is thus known. But their multiplicity is another story; it leads to many problems in number theory, far from being finished today. Indeed the question of multiplicity is the question as to how many ways an integer can be written as a sum of d squares: see the Gauß circle problem on page 76 which is an unsolved problem in number theory. Some references on the circle problem: Erdös, Gruber & Hammer 1989 [491], Gruber & Lekkerkerker 1987 [660] page 135, Gruber & Wills 1993 [661], Walfisz 1957 [1227] and Krätzel 1988 [833]. However the first order asymptotic estimate of

$$N(\lambda) = \text{number of eigenvalues (with multiplicity) smaller than } \lambda$$

is very easy geometrically. We look for the number of points with integral coordinates which are located inside the ball $B(0,r)$ (centered at the origin) of radius

$$r=\frac{\sqrt{\lambda}}{2\pi} ,$$

see figure 1.84 on page 76. This figure shows that, up to an error term which becomes negligible because it is "only" of order R^{d-1}, we find that $N(\lambda)$ is asymptotic to the volume of the ball of radius $2\pi R$, namely

$$\frac{\beta(d)}{(2\pi)^d}\lambda^{2/d} .$$

Hence the second term in the expansion is again connected to the circle problem, and so is unknown.

9.5.2 Other Flat Tori

This time we quotient our vector space $\mathbb{R}^d$ by any lattice Λ. A lattice is the set of all integral linear combinations of a basis of $\mathbb{R}^d$. Motivated by the preceding "cube" case, we look for functions which are eigenvalues of Δ and Λ periodic. We search for them among the imaginary exponentials of linear functions, which can be always written in the form

$$f(x)=e^{2\pi i\langle\xi,x\rangle}$$

where $i=\sqrt{-1}$ and ξ is a vector which we will try to find. We will have Λ periodicity exactly when the scalar product

$$\langle\xi,x\rangle$$

is an integer for each $x \in \Lambda$. Those ξ form a lattice, called the dual lattice of Λ and denoted by Λ^*. It is trivial to see that

$$\Lambda^{**} = \Lambda$$

and that

$$\mathrm{Vol}\left(\mathbb{R}^d/\Lambda^*\right) = \frac{1}{\mathrm{Vol}\left(\mathbb{R}^d/\Lambda\right)}$$

The eigenfunctions of Δ are

$$e^{2\pi i\langle \xi, x\rangle}$$

for $\xi \in \Lambda^*$ and the eigenvalue of this eigenfunction is

$$4\pi^2|\xi|^2 .$$

But the precise description of the dual lattice is not so easy. It is only in dimension 2 that the dual lattice is always deduced from the original lattice by a similarity. Analysts know how to relate Λ and Λ^*, at least theoretically, with the *Poisson formula*:

$$\frac{1}{(4\pi t)^{d/2}} \mathrm{Vol}(\Lambda) \sum_{\lambda \in \Lambda} e^{-\|\lambda\|^2/4t} = \sum_{\xi \in \Lambda^*} e^{-4\pi^2\|\xi\|^2 t}. \tag{9.14}$$

Stated another way, the set of eigenvalues of our torus is the set of $4\pi^2$ multiples of square norms (distance to the origin) of the points in the dual lattice Λ^*. It is important for future developments in this book that the distance to the origin from a point of Λ is precisely the length of a periodic geodesic of our torus. So the Poisson formula yields a relation between the spectrum and the length spectrum.

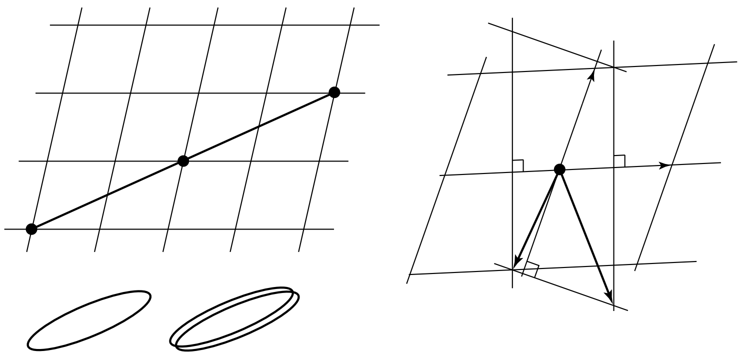

Fig. 9.4. **(a)** The same periodic geometric geodesic **(b)** Λ and Λ^* are similar in dimension 2 (only)

The proof of the Poisson formula is not very difficult. We can explicitly write down the heat kernel K^* of $\mathbb{R}^d$ (see §9.7). One then puts together the heat kernel $K(x,y)$ of our flat torus as a summation

$$\sum_i K^*(x, y + \lambda)$$

where λ runs through the lattice defining the torus. Using roughly the same idea, but with considerably more difficulty one can obtain Selberg's trace formula for space forms of negative curvature; see §§9.5.5.

9.5.3 Spheres

Harmonic analysis on spheres is a small miracle: we explained it in §§1.9.2 but the reader might like to see it again here. A polynomial p of degree k on $\mathbb{R}^{d+1}$ is said to be harmonic if

$$\Delta p = 0$$

for the Laplacian Δ on $\mathbb{R}^{d+1}$. The restriction

$$f = p|_{S^d}$$

to the sphere turns out to be an eigenfunction of the Laplacian on S^d with eigenvalue $k(k + d - 1)$. Its multiplicity is just the dimension of the vector space of harmonic polynomials of degree d, namely

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

Again as above, the Stone–Weierstraß theorem tells us that we have no other eigenfunctions and a complete orthonormal basis of eigenfunctions. This does not say that we know everything today about spherical harmonics, even if many people think we do. We turn now to the next objects in the hierarchy of §6.6.

9.5.4 $\mathbb{KP}^n$

Fourier analysis on $\mathbb{CP}^n$ goes back to Élie Cartan in his 1931 monograph Cartan 1992 [322]. The trick is the same as for the sphere, but here one starts with $\mathbb{C}^{n+1}$ and uses harmonic polynomials in the variables

$$z_j, \bar{z}_j .$$

Details can also be found in Berger, Gauduchon & Mazet 1971 [174].

Unhappily this trick does not work with the quaternions. This is linked with the following fact which we mention here because it is rarely known. It is impossible to define on $\mathbb{H}^n$ useful quaternionic derivatives analogous to the complex derivatives

$$\frac{\partial}{\partial z} \text{ and } \frac{\partial}{\partial \bar{z}} .$$

A related phenomenon: quaternionic structures on manifolds can be integrable only in the flat case. For all this, and a good notion of quaternionic functions, see Joyce 1997 [773] and the references there or the note 13.5.3.1 on page 653.

There are at least two ways to compute the spectra of the remaining $\mathbb{KP}^n$. One is to use a very general formula due to Hermann Weyl, and valid for all symmetric spaces. But the formula is explicit only in the sense that it is a summation over the roots of a certain Lie algebra. To get explicit expressions is hard. The other way is to use the general link between periodic geodesics and the spectrum, a quite deep result (unavoidably using the wave equation) which we will meet in §9.9.

The explicit result for all $\mathbb{KP}^n$ can be found on page 202 of Besse 1978 [182]. It is important to note the spectrum. Its square roots are in all cases included in intervals whose centers make up an arithmetic progression:

$$\operatorname{Spec}(\mathbb{KP}^n) \subset \bigcup_{k=0}^{\infty} \left[\frac{2\pi}{L}\left(k+\frac{\alpha}{4}\right)^2 - M, \frac{2\pi}{L}\left(k+\frac{\alpha}{4}\right)^2 + M \right] \tag{9.15}$$

where L is the common length of all of the geodesics (which are all periodic), M is some fixed constant and the "index" α is

$$\alpha(\mathbb{KP}^n) = \begin{cases} 0 & \mathbb{K}=\mathbb{R} \\ 1 & \mathbb{K}=\mathbb{C} \\ 3 & \mathbb{K}=\mathbb{H} \\ 7 & \mathbb{K}=\mathbb{C}\mathrm{a} \end{cases} .$$

As for the sphere, the multiplicities are very high but this is necessary to match the asymptotic behaviour of equation 9.20 on page 397. We will meet this special form of spectrum, as in equation 9.15, again in theorem 177 on page 406.

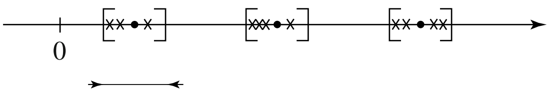

Fig. 9.5. The spectrum $\operatorname{Spec}(\mathbb{KP}^n)$

9.5.5 Other Space Forms

The spaces whose spectra we will look for include not only symmetric spaces of higher rank, but also space forms of negative curvature (any rank). For space forms of positive curvature and more generally for homogeneous spaces, the spectrum can be more or less handled in various cases, or only controlled in some instances. We do not give any details; they can be found in the various references which we will give later on.

The very hard but fascinating case is that of space forms of negative curvature. Then one needs to understand not only the Lie group and Lie algebra but also the discrete subgroup of isometries of the simply connected forms (of negative curvature) which yield compact quotients under study. The basic tool was discovered in 1956: it is the *Selberg trace formula*. This an entire subject in itself, intimately connected with number theory. We can only afford to give references on the subject. We choose to offer more or less expository references as opposed to partial results. We suggest for the Selberg trace formula on surfaces, which is quite special and exceptionally powerful: Buser 1992 [292] chapter 9, but the formula permeates a great deal of the book. Add of course the references given there. For higher dimensions, see Bunke & Olbrich 1995 [279]. For more about hyperbolic surfaces see §§9.13.2.

9.6 Current Questions

We can either concentrate on the eigenvalues or on the eigenfunctions. In each case, we can then ask how to derive information about the eigentheory from geometric information, and vice versa.

9.6.1 Direct Questions About the Spectrum

A typical result about eigenvalues is theorem 164 on page 386. It provides practically perfect upper control on the eigenvalues. It is optimal in the sense that none of the ingredients can be removed. Simple examples show that one needs a lower bound on the Ricci curvature and on the volume to obtain upper bounds on eigenvalues.

So the next natural question is to look for lower bounds. We will see below that lower bounds involve the diameter instead of the volume, and beyond that no more than a Ricci curvature lower bound; see §§9.7.3.

As explained in §9.2, the main question, vital for many physicists, is the asymptotic behavior of the spectrum. We will see that the first order term in the asymptotic expansion is easy to get. The next order term is another story, as we already saw for the flat torus case. The repartition of the spectrum about the asymptotic formula, the way the eigenvalues arrange themselves, is of equal significance in physics. Whatever a precise definition might be, one feels that the $\mathbb{KP}^n$ spectra given in equation 9.15 on the preceding page is an atypical distribution, with very high multiplicities, and poorly behaved if we want to tell different vibrations apart by hearing how they differ in frequencies. Looking at that equation, one might be led to wonder about *gaps* in the spectrum. We will meet some answers to this question, but some elementary questions of this sort are still completely open. Another important problem, also interesting for applications, is to have a lower bound for the first eigenvalue λ_1. It controls "resonances" and can prevent them. Control of all of the spectral data we have just discussed cannot be obtained only with a lower bound on Ricci curvature, volume and diameter. One will need to know more on the curvature, the injectivity radius, etc. For the behaviour of the spectrum when the metric varies, see Lott 2000 [881].

9.6.2 Direct Problems About the Eigenfunctions

There are very few results about eigenfunctions. It is natural to ask for control on the sup norm of the eigenfunctions, which amounts among other things to studying the asymptotic behavior of

$$\int_M \phi_i^2 \phi_j$$

for a fixed i with j going to infinity. The *nodal sets*, defined to be the zero sets of eigenfunctions, are of clear physical significance. Outside singularities, the nodal sets are hypersurfaces in the manifold. Do they have large measure (say $d-1$ dimensional Hausdorff measure)? How are they located? Think of the spreading out of nodal sets as a kind of even repartition in space. Today's harvest is quite meager: see §9.11.

9.6.3 Inverse Problems on the Spectrum

The literature on recovering Riemannian geometry from the spectrum is immense, this subject having excited people tremendously when it was triggered by Milnor 1964 [922]. There it was proven that two Riemannian manifolds which are not isometric can have the same spectrum. We will give below a brief account of the state of affairs today.

A completely different (still inverse) topic is to try to recover the Riemannian manifold from its geodesic flow. This can be asked in different ways. Suppose you know the lengths of all of the periodic geodesics (this is the so-called *length spectrum*); can you find the metric? But you might know even more, namely the complete structure of the flow on the unit tangent bundle (the phase space). See §9.12 and chapter 10 for the state of current knowledge.

9.7 First Tools: The Heat Kernel and Heat Equation

9.7.1 The Main Result

Theorem 165 (Minakshisundaram 1953 [929], McKean & Singer 1967 [910]) Let M be a compact Riemannian manifold. There is a function

$$K : M \times M \times \mathbb{R}_+^* \to \mathbb{R}$$

which is C^∞ and

1. Given any initial data $f : M \to \mathbb{R}$ the solution of the heat equation

$$-\frac{\partial F}{\partial t} = \Delta F$$

with

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

is given by

$$F(x) = \int_M K(x,y,t)\, f(y)\, dy$$

2. K is given by the convergent series

$$K(x,y,t) = \sum_i \exp\left(-\lambda_i t\right) \phi_i(x)\phi_i(y)$$

(where the eigenfunctions ϕ_i of the Laplace operator Δ are chosen so that they form an orthonormal basis of the square integrable functions on M)

3. For every $x \in M$ there is an asymptotic expansion as $t \to 0$ of the form

$$K(x,x,t) \sim \frac{1}{(4\pi t)^{d/2}} \sum_{k=0}^{\infty} u_k(x) t^k$$

where the $u_k : M \to \mathbb{R}$ are functions given by universal formulae expressing $u_k(x)$ in terms of the curvature tensor of M and its covariant derivatives at the point x.

The three argument function K is called the *fundamental solution of the heat equation* on M, or the *heat kernel* of M.

If one assumes existence of the heat kernel, it is easy to check the properties 1 and 2. Note the surprising symmetry, which has no reason a priori to hold:

$$K(x,y,t) = K(y,x,t)\ .$$

We recall that the physical interpretation of the heat kernel is the following: $K(x,y,t)$ is the temperature at time t and at the point y when a unit of heat (a *Dirac δ function*) is placed at the point x.

To find the proof and to get a feeling for why property 3 is reasonable, we recall what we saw in equation 1.26 on page 88, namely that the fundamental solution of the heat equation for the Euclidean plane was explicitly determined as

$$K^*(m,n,t) = \frac{1}{4\pi t} e^{-\|m-n\|^2/4t}$$

For a Euclidean space of general dimension d it is also explicit and easy to find by formal computation, namely:

$$K^*(x,y,t) = \frac{1}{(4\pi t)^{d/2}} e^{-d(x,y)^2/4t} \tag{9.16}$$

where we have replaced the square norm by the distance. To study heat on more general Riemannian manifolds, the idea is to get some function analogous to the above on a compact Riemannian manifold. It makes sense to consider equation 9.16 on the facing page in any Riemannian manifold, provided we cut it with a step function η. So we will set

$$H_0 = hS_0$$

for

$$S_0 = K^*$$

above and measure distance according to our Riemannian metric. This is a sort of first order approximation of the K that we are looking for. We have reason to hope that we can carry on in this direction, because the exponential decays very quickly with time t. In analysis, a function like S_0 (which approximates a kernel) is called a *parametrix*.

The sketch of the complete proof is as follows. We build up an exact solution in two steps. In the first step, one defines local parametrices with higher and higher orders of approximation by an induction formula and a sum as follows:

$$S_k = \frac{1}{(4\pi t)^{d/2}} e^{-d(x,y)^2/4t} \sum_{i=0}^{k} u_i(x,y)t^i$$

so that

$$\left(\Delta_x + \frac{\partial}{\partial t}\right) S_k = \frac{1}{(4\pi t)^{d/2}} e^{-d(x,y)^2/4t} \Delta_x u_k \tag{9.17}$$

But these functions are only define locally. We now define global functions

$$H_k$$

on our manifold with the above and a step function η by setting

$$H_k = \eta S_k \ .$$

These functions are certainly not what we are looking, since for example they depend on the choice of η. The trick is to define K again as a series by a double convolution process which will "forget" the η function. The two variables in the convolution are the space and the time. We define the *convolution* $A * B$ of two functions of (x,y,t) by

$$(A * B)(x,y,t) = \int_0^t d\tau \int_M A(x,z,\tau)B(z,y,t-\tau)dV_M\,(z)$$

and the desired fundamental solution is

$$K = \sum_i \left(\Delta_x - \frac{\partial}{\partial t}\right) (Hk^*)^i \tag{9.18}$$

which works as soon as k is large enough, namely

$$k > \frac{d}{2} \,.$$

For details of the proof we refer to III.E of Berger, Gauduchon & Mazet 1971 [174], Chavel 1984 [325] chapter VI, Gilkey 1995 [564] or chapter 2 of Berline, Getzler & Vergne 1992 [179]. Formal verification is trivial; the problems are principally in the convergence of the series and in the smoothness of the objects; smoothness is where we use the condition

$$k > \frac{d}{2}$$

which will not surprise readers used to Sobolev inequalities; see theorem 118 on page 325.

The universality of property 3 on page 394 is simply due to Élie Cartan's philosophy of normal coordinates. We saw one aspect of this philosophy when commenting on Jacobi's field equation 6.11 on page 248 in §§6.3.1. The second aspect is that Jacobi's equation can be differentiated as many times as we wish. In the result only the curvature tensor and its covariant derivatives of various orders will appear, and each of these in some universal polynomial expression. The Laplacian is also universal, involving only various derivatives of the Riemannian metric. It remains only to remark that, by construction of the kernel K in equation 9.18 on the preceding page, the u_k are the same as in property 3.

We mention here that the heat kernel is explicitly known for some special manifolds, as is the fundamental solution of the wave equation. Among these special manifolds are of course Euclidean spaces, spheres, and hyperbolic spaces. For example, in the case of the hyperbolic plane, these kernels are employed in Huber's result (theorem 192 on page 421). For the spheres, we can find the kernel in Cheeger & Taylor 1982 [355, 356]; see this text for previous results. For space forms, see the recent Bunke & Olbrich 1995 [279]. For symmetric spaces see Helgason 1992 [704].

9.7.2 Great Hopes

If in theorem 165 on page 393 we integrate over the manifold M and use property 2 we get the basic formula

$$\sum_k e^{-\lambda_k t} \sim \frac{1}{(4\pi t)^{d/2}} \operatorname{Vol}(M) \qquad \text{as } t \to \infty \tag{9.19}$$

which gives us the first order term of the asymptotic behavior of the eigenvalues (counted with multiplicity). This is called the *Hermann Weyl estimate*, although Weyl was only interested in domains with boundary in Euclidean spaces as seen in §§1.8.5. If one is only interested in obtaining this estimate, it can be obtained less expensively with the minimax principle.

Regarding inverse problems, the knowledge of the spectrum gives you the dimension of the manifold and its volume.

As in §§1.8.5, the Hardy–Littlewood–Karamata theorem applies to yield what we are really interested in, namely

$$\begin{aligned} N(\lambda) &= \# \{\lambda_i < \lambda\} \\ &= \frac{\beta(d)}{(2\pi)^d} \operatorname{Vol}(M,g) \lambda^{d/2} + o\left(\lambda^{d/2}\right) \end{aligned} \tag{9.20}$$

as $\lambda \to \infty$. From here, completely elementary calculus yields

$$\lambda_k \sim \left(\frac{(2\pi)^d}{\beta(d) \operatorname{Vol}(M,g)} \right)^{2/d} k^{2/d} \tag{9.21}$$

Note the perfect compatibility of this formula with theorem 164 on page 386.

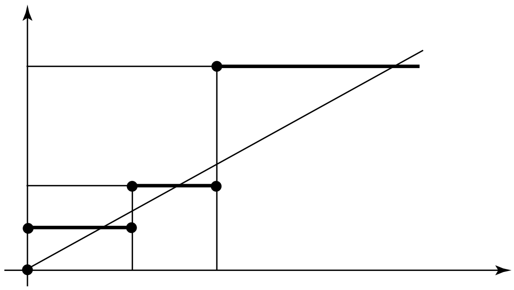

Fig. 9.6. The Weyl asymptotic for surfaces

The function $N(\lambda)$ is a step function. The next natural question on the spectrum is

Question 166 *How does the function $N(\lambda)$ distribute itself around the continuous function giving the asymptotic behaviour?*

Today we know very little about this question. But we will see in §9.9 that with the wave equation technique one can replace the little o by a capital

$$O\left(\lambda^{(d-1)/2}\right) .$$

This will permit some rudimentary control, on the gaps for example.

There is also a heuristic principle to the effect that there is a deep relation between the jumps in the spectrum and the lengths of the periodic geodesics.

See §9.9, theorem 176 on page 405, note 9.9.0.1 on page 408, the proof of theorem 189 on page 420, and §§§ 9.13.2.1 on page 426.

The reader might wonder why we did not use the full asymptotic expansion obtained from parts 2 and 3 of theorem 165 on page 393. Let us look at it:

$$\sum_k e^{-\lambda_k t} \sim \frac{1}{(4\pi t)^{d/2}} \left(\operatorname{Vol}(M,g) + U_1 t + U_2 t^2 + \cdots\right) \qquad \text{as } t \to \infty \quad (9.22)$$

where

$$U_k = \int_M u_k \ .$$

We did not do so first because the Hardy–Littlewood–Karamata theorem does not provide any information beyond the first order term. That is to say, the knowledge of the U_k is strictly useless for finding the higher order terms in $N(\lambda)$. We will need more than the above expansion—either a much more subtle analysis of the heat kernel or, better, the wave equation.

Still one can try to use theorem 165 on page 393 and see what one can extract from it. As expected the first u_k expressions should be simple. In fact various authors have computed the two first; if we write scalar for the scalar curvature of our manifold, and R for its Riemann curvature tensor, then

$$\begin{aligned} u_1(x) &= \frac{1}{6}\operatorname{scalar}(x) \\ u_2(x) &= \frac{1}{360}\left(2\|R\|^2 - 2\|\operatorname{Ricci}\|^2 + 5\operatorname{scalar}^2\right) \end{aligned} \qquad (9.23)$$

Beginning with the third term, the expressions become more and more complicated. For example, the third term involves the covariant derivative of the curvature tensor. We refer for those and their applications to: the end of this section for the uniqueness of the spectrum of low dimensional spheres, to theorem 188 on page 418 for the compactness of the sets of isospectral metrics on compact surfaces, to Gilkey 1995 [564] and Berline 1992 [179] for very general references.

If you integrate u_1 you get

$$\frac{1}{6}\int_M \operatorname{scalar} \ .$$

If M is a surface, the Blaschke–Gauß–Bonnet formula 28 on page 138 yields

$$\frac{1}{6}\int_M \operatorname{scalar} = \frac{\pi}{3}\chi(M) \ .$$

Although this is of no use for calculating $N(\lambda)$, it is very helpful for the inverse problem—it implies that the knowledge of the spectrum (1) tells you that you are on a surface (see above) but moreover (2) we now know its genus.

An important (but which will turn out eventually to be a “useless”) remark: the fact we get a topological invariant, in particular an invariant of rescaling the metric, is not surprising because $U_1 t$ has to be divided by $t^{2/d} = t$ here, so it is dimensionless. So the natural question is: is U_2 a topological invariant of four dimensional manifolds? This question brought great excitement to spectral geometry in 1966, and was one of the reasons for serious study of the U_k. The answer is *no*. A simple reason is that the generalization of the Gauß–Bonnet theorem in four dimensions (see §§11.3.6 or §15.7) is

$$\chi(M) = \frac{1}{32\pi^2} \int_M \|R\|^2 - 4\| \operatorname{Ricci} \|^2 + \operatorname{scalar}^2 \tag{9.24}$$

Asking U_2 to be invariant as well (even if not linked to the characteristic) is too much, as trivial examples show.

From the opposite point of view, it is easy to apply equations 9.23 on the facing page to a surface to prove that the round (constant curvature) sphere S^2 is characterized by its spectrum, as are flat tori. On the other hand, we will see in §§9.13.2 that there are isospectral nonisometric Riemannian surfaces of constant negative curvature. Still, using the higher U_k, it is proven in Tanno 1980 [1181] that round spheres of up to six dimensions are characterized by their spectra. The same question for higher dimensional spheres is still open today. This shows how far we are today toward understanding the spectra of Riemannian manifolds. Another nice application of the U_k is to be found in §§9.12.3.

Note 9.7.2.1 (Spectra of space forms) Let us reconsider that the knowledge of the spectrum yields the knowledge of all the U_k integrals. Look at the case of space forms (of constant sectional curvature). Then all the u_k are known and in particular the U_k are all known as soon as one knows the volume of the manifold. This does not yield the space form (up to isometry) except in one dimension. ♦

The spectral determination of the Euler characteristic χ above for surfaces is exceptional: today there is no known topological information in the spectrum in dimensions three and higher. Of course, so far we are discussing the spectra of the Laplace operator on functions. For the spectra of the Laplace operator on more general tensors, e.g. differential forms, see §9.14.

Note 9.7.2.2 (Futility of the U_k) Besides the theoretical interest of establishing a solid foundation for Fourier analysis on a Riemannian manifold, at this moment the heat equation technique seems to be of little use. It might seem that this is because the curvature appears in the asymptotic expansion in a too algebraically complicated manner. Except for the second term, the expansion involves not only the curvature but also its covariant derivatives; in particular geometric invariants (the volume excepted) like the diameter, the injectivity radius, the geodesic flow, do not enter into it. But one "explanation" for the impotence of the U_k is given by the following, which is a strong generalization of theorem 186 on page 415:

Theorem 167 (Lohkamp 1996 [874]) *Consider any compact manifold M of dimension larger than two, and any infinite sequence of positive numbers*

$$0 < \lambda_1 < \lambda_2 < \cdots .$$

Then there is a sequence of metrics g_m on M of fixed volume and fixed integral of scalar curvature such that not only does the spectrum of g_m coincide with the given sequence from 0 to λ_m, but all of the U_{2k} go to $+\infty$ and all of the U_{2k+1} go to $-\infty$. Under the same conditions, there is also another sequence of metrics with the same spectral condition but this time the volume is fixed and the Ricci curvature satisfies

$$\mathrm{Ricci}\,(g_m) < -m^2 .$$

This explains the near inefficacy of the U_k and the poverty of the hypothesis of negativity of Ricci curvature (see §§12.3.5). For the nature of the proof, see §§9.12.1. ♦

There is also an important geometric formula which deserves to be mentioned, even if at the moment it has no geometric application:

Theorem 168 (Varadhan 1967 [1206])

$$\lim_{t\to 0} t \log K(x,y,t) = -\frac{d\,(x,y)^2}{2}$$

for any x, y close enough.

Varadhan's formula works within the injectivity radius. What happens when y moves to the cut locus of x is the subject of Malliavin & Stroock 1996 [890]; dramatic changes take place, for example on the standard sphere strange events occur at antipodal points. But theorem 168 is fundamental to modern probability theory, and in particular to the Malliavin stochastic calculus on infinite dimensional Riemannian manifolds (e.g. path spaces).

Exterior differential forms are canonically attached to a differentiable manifold and a Riemannian metric also provides a Laplace operator on them. But more generally there are other kind of bundles one can look at, as well as suitable differential operators. Some are canonical, as in the case of spinors, while others are built up with various techniques e.g. twisting canonical ones, etc. In this context the heat equation method works and yields important results. Some are of interest in themselves; these will be described briefly in §14.2. Some are basic tools for Riemannian geometry; we will meet such applications twice in §§12.3.3.

Still thinking about heat, we mention the recent notion of heat content of a domain in a Riemannian manifold. This notion has various applications, even in the Euclidean case, and probably some future: see Savo 1998 [1099].

9.7.3 The Heat Kernel and Ricci Curvature

In §§9.4.2, we used the minimax principle to get upper bounds on the spectrum. Lower bounds are more difficult. The case of the first eigenvalue λ_1 is treated separately in §9.11. We will now address the question of a lower bound for every eigenvalue. An optimal result can be found in Bérard, Besson & Gallot 1985 [139]; see the book Bérard 1986 [135] for a detailed exposition. To formulate their result we introduce some notation.

Definition 169

$$Z(t) = \sum_k e^{-\lambda_k t}$$

which we will write as

$$Z_{M,g}(t)$$

when we need to specify which Riemannian manifold M and metric g is being invoked. Similar notation is used to specify the manifold and metric when discussing the heat kernel:

Definition 170

$$K_{M,g}(x,y,t) = K(x,y,t)$$

Then we can state:

Theorem 171 *There is a universal constant*

$$c = \mathrm{univ}(\inf\ \mathrm{Ricci}, \dim, \mathrm{diam})$$

(where inf Ricci *is the lower bound of the Ricci curvature,* dim *the dimension and* diam *the diameter of a Riemannian manifold M) such that for any time t*

$$Z_M(t) \leq \mathrm{Vol}(M) \sup_{x,y\in M} K_M(x,y,t) \leq Z_{S^d}(ct)$$

This is a very strong result since it is a bound for the whole heat kernel. Since the spectrum of the standard sphere S^d is known, one gets immediately:

Theorem 172 *There is universal constant such that all eigenvalues satisfy the lower bound*

$$\lambda_k \geq \mathrm{univ}(\inf\ \mathrm{Ricci}, d, \mathrm{diam}) k^{2/d} .$$

The term $k^{2/d}$ agrees with Weyl's asymptotic 9.19 on page 396 for the power of k but not for the volume. Moreover, simple examples show that the diameter, not only the volume, is really needed. Examples also show that these results are optimal as far as the ingredients (see how they enter more explicitly in Bérard, Besson & Gallot 1985 [139]. Finding optimal explicit values is an open problem. The authors' values are explicit but not optimal, especially in the case of negative Ricci curvature. This will be seen from the proof.

The proof is very geometrical. Look again carefully at the proof of the Faber–Krahn inequality 1.22 on page 81 for the fundamental tone of a plane vibrating membrane. There we used function symmetrization—a transplantation, going from the membrane D under study to the circular membrane D^* having same area. From any function on D, a function on D^* was constructed. Then the key ingredient (besides Fubini's theorem and a change of variable) was the isoperimetric inequality for plane curves.

Bérard, Besson & Gallot 1985 [139] enact a double generalization of the same ideas. First we symmetrize the whole heat kernel as a function (which depends on three variables). Second we use the result on the isoperimetric profile obtained in theorem 114 on page 319 which needs precisely a lower bound on Ricci curvature and diameter. The transplantation here goes from M to a sphere whose radius is precisely defined as a function of inf Ricci, the dimension and the diameter. It is then clear that on a manifold of negative Ricci curvature, the comparison sphere cannot be optimal.

The proof is then concluded by expensive and technical details. In particular it uses the maximum principle for parabolic partial differential equations (because the heat equation is parabolic). Time is taken in account as follows. The heat equation for the symmetrized kernel becomes an ordinary differential equation and one then applies Sturm–Liouville theory, in some sense as for Jacobi fields in §3.2. One can find this technique in Bandle 1980 [109]. Details of the above results can be found in chapter V of Bérard 1986 [135] or in Berger 1985 [163].

Brownian motion on Riemannian manifolds is very closely related to the heat equation. The "propagation speed" of Brownian motion "is the Ricci curvature." The reader will enjoy Stroock 1996 [1164], Elworthy 1988 [489], and Pinsky 1990 [1029, 1028].

9.8 The Wave Equation: The Gaps

Put together, the bounds from theorems 172 on the preceding page and 164 on page 386 frame the λ_k between two asymptotic curves. This is reasonable control, but does not say much about how the eigenvalues are distributed. Questions can be asked about the "jumps," about the evenness of the distribution, and more simply about the gaps. The formulas 9.15 on page 391 for the spectrum of the spheres and the $\mathbb{KP}^n$ show spectra which are not evenly distributed, since they are concentrated in intervals. The heat equation is not a deep enough tool to get information on the gaps—we need to analyze the wave equation on our Riemannian manifold.

This is like climbing Jacob's ladder. To get information on the manifolds "downstairs" we have to travel to the unit tangent bundle UM and to work

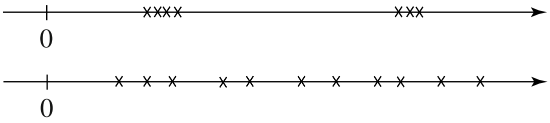

Fig. 9.7. Spectral gaps

with distributions. Downstairs, a function only has a gradient, but a distribution on UM has a wavefront, which is nothing but the set of its directional singularities. The first result is that under time evolution the wave front evolves exactly by the action of the geodesic flow: "the waves (the light) travel along geodesics." We cannot say more about the wave equation, it will need an entire book. Today this topic is called *microlocal analysis*. It involves subtle notions such as *Fourier integral operators* and *canonical transformations*. To our knowledge, there are no "popular" expositions of microlocal analysis; the most picturesque, and closest to Riemannian geometry, is that of Guillemin & Sternberg 1977 [670]. The four volumes of Hörmander 1983 [735, 736, 737, 738] are complete and encyclopedic (get the second edition of volume I); Tréves 1980-82 [1198, 1199] is also very informative. The wave kernel

$$\sum_k \cos\left(\sqrt{\lambda_k}t\right) \phi_k(x)\phi_k(y)$$

is no longer a function (only a distribution) but in exchange it carries much more information. It can also be remarked that microlocal analysis involves a lot a symplectic geometry, which takes place in T^*M, the cotangent bundle. It is better to ignore the fact that (thanks to the Riemannian structure) T^*M is canonically isomorphic to TM (see §15.2). One also works with the canonical contact structure on the unit tangent bundle UM (which is *Sasakian*): see page 56 of Sakai 1996 [1085].

The following result is part of a very general theory which applies to any elliptic operator on a compact manifold; we employ it here only to the Laplace operator.

Theorem 173 (Hörmander 1968 [734]) *The number $N(\lambda)$ of eigenvalues smaller than λ obeys the asymptotic law*

$$N(\lambda) = \frac{\mathrm{Vol}(M)\beta(d)}{(2\pi)^d}\lambda^{d/2} + O\left(\lambda^{(d-1)/2}\right)$$

It should be mentioned that such a result had been obtained in Avakumović 1956 [90] in three dimensions using a technical study of the parametrix. The immediate corollary (by the very definition of a "capital O" and elementary calculus) is the one we are after:

Theorem 174 *For any Riemannian manifold M there is a constant C_M such that for any real numbers a and b with $b-a$ large enough, the set of eigenvalues λ of the Laplacian in the interval $[a,b]$ satisfies*

$$\#\left\{\sqrt{\lambda} \in [a,b]\right\} > C_M(b-a)a^d \, .$$

Note that such gap results cannot be too general; think of theorem 186 on page 415 to the effect that there is always a Riemannian manifold whose spectrum is any chosen finite subset of the real numbers.

For the geometer there is major drawback in Hörmander's result. The way the constant is found in Hörmander's proof is not constructive; the geometry of the Riemannian manifold does not come in. But we would like to be able to estimate $C(M,g)$ as a function of the geometric invariants of (M,g). At the moment there is no such result obtained by working with the wave equation on a Riemannian manifold. But the following recent result is to be found in section $6\frac{9}{10}$ of Gromov 1996 [631]. The proof is extremely intricate, and uses the *Kac–Feynman–Kato inequality*. This formula bounds the spectrum of any elliptic operator on any bundle on a Riemannian manifold with the spectrum downstairs of the manifold itself, and was always used the other way around. But Gromov looks at suitable bundles over a compact Riemannian manifold and uses various tools from Vafa–Witten, Bochner–Lichnerowicz and Atiyah–Singer. See chapter §14.2 for a brief survey of those tools. Using that incredibly high climb up Jacob's ladder one has:

Theorem 175 (Gromov 1996 [631]) *In any odd dimensional Riemannian manifold whose sectional curvature satisfies*

$$|K| \leq 1$$

and whose injectivity radius is larger than 1, the spectral gaps are controlled:

$$\#\left\{\sqrt{\lambda} \in [a,b]\right\} > C_d(b-a)^d \operatorname{Vol}(M)$$

for any positive real numbers a, b such that with

$$b > a + C'_d$$

where C_d and C'_d are universal constants in the dimension d.

We leave the reader to use appropriate scaling to replace C_d by a constant depending on $\sup |K|$ and $\operatorname{Inj}(M)$. Let us remark that some geometric control is required in view of theorem 186 on page 415. It seems to be an interesting question to prove the above result by working only with the wave equation "down" on the manifold itself. Note also that one knows more (but not everything) about the distribution of the spectrum on the real line for certain special manifolds; see §9.13.

9.9 The Wave Equation: Spectrum and Geodesic Flow

In the pioneering paper Balian & Bloch 1972 [99], which we have discussed in §§1.8.6, the authors suspected a relation between the spectrum of a plane domain and its length spectrum.[4] The fact that compact plane domains have a boundary rendered this study difficult. This is one reason why people turned first to compact Riemannian manifolds (without boundary of course). We speak now about general Riemannian manifolds; the special case of space forms will be taken care of in §9.13. Relations between the spectrum (of functions) and the length spectrum will be met again in §9.12. For flat tori, we met a perfect link between the spectra furnished by the Poisson formula 9.14 on page 389 So the problem is to find, if possible, various generalizations of this formula.

For the general case, the first result was Colin de Verdière 1973 [389], but the proof was very tricky, using the heat kernel and the stationary phase technique. Soon after it was realized that the wave equation is the more powerful and elegant technique: Chazarain 1974 [327] and Duistermaat & Guillemin 1975 [465]. This yielded:

Theorem 176 *For any Riemannian manifold M the series*

$$\sum_i \cos\left(\sqrt{\lambda_i}t\right)$$

defines a distribution whose singular support is contained (besides the value 0) in the set of the lengths L of the periodic geodesics of M. For a generic Riemannian manifold, this singular support is a sum of distributions T_L, with L ranging over lengths of periodic geodesics, and where each T_L has support located in a small neighborhood of L. Moreover each T_L can be expressed with the sole help of the Poincaré return map (see the definition 10.4.3.2 on page 469) associated to the periodic geodesics of length equal to L and the holonomy map (the effect of parallel transport) along these geodesics.

Here is a very primitive explanation for theorem 176; it is not even an idea of a proof but just help for the reader who needs to visualize things to get some grasp of them. We look at a surface and, like throwing a stone in a pond, look for the wave generated by this action. The picture in figure 9.8 on the following page shows what is happening at the beginning: no problem occurs at small distances, but as in §§9.7.2 we might expect trouble at the cut locus. Two waves meeting transversally generate only nice interferences—this has been known for a long time. But the wave interferences are different when the two waves come one against the other in exactly opposite directions; this will be the case for any periodic geodesic. If moreover their common frequency is

[4] Recall that the *length spectrum* of a plane domain is the set of lengths of its periodic (billiard or light) trajectories.

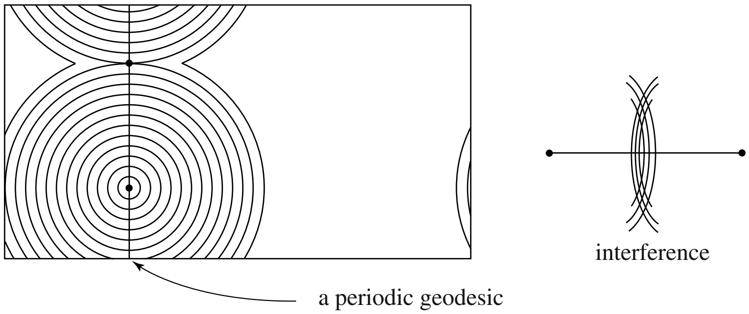

Fig. 9.8. **(a)** A periodic geodesic **(b)** Interference

of the form $2\pi nL$, where n is any integer and L is the length of the periodic geodesic under consideration, then we will have (probably) a resonance or, say, a tidal wave. This is the cause of singularities in the series above. Note that this does not happen for geodesic loops—they do not produce enough resonance.

We come back to more standard mathematical notation. First, there is a kind of reciprocal of the formula 9.15 on page 391:

Theorem 177 (Colin de Verdière 1979 [391] and Duistermaat & Guillemin 1975 [465]) If all the geodesics of a compact Riemannian manifold M are periodic with common length equal to L then for k large enough one has the inclusion

$$\mathrm{Spec}\,(M) \subset \bigcap_{k\in\mathbb{N}} \left[\frac{2\pi}{L}\left(k+\frac{\alpha}{4}\right)^2 - M, \frac{2\pi}{L}\left(k+\frac{\alpha}{4}\right)^2 + M\right]$$

and moreover the number of eigenvalues in every one of these intervals is polynomial in k.

We will see in §§10.10.2 the significance of the α which can only equal $0, 1, 3$, or 7 (the reader can—and should—think of the $\mathbb{KP}^n$).

The end of theorem 176 on the preceding page was very imprecise about the Poincaré and holonomy maps and in particular was only passing from the singularity of T_L to the Poincaré map. Recall that T_L was defined in theorem 176 on the previous page. Further recall that this Poincaré map has to be viewed in the unit tangent bundle UM (at some starting point) and is the differential at the origin of the return map after going once around a periodic geodesic. After various partial answers, strong results became available only recently in Guillemin 1993 [665] and 1996 [666]. In those works, the singularity of T_L is completely determined by the Poincaré map, this being done in terms of the so-called *Birkhoff canonical form*. This is moreover carried out for a general elliptic linear differential operator, with the periodic geodesics being replaced by the *periodic bicharacteristics*.

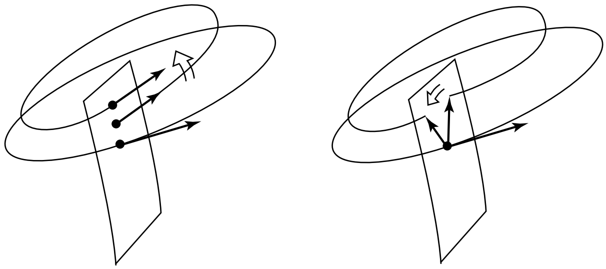

Fig. 9.9. (a) The Poincaré map works in $UM|_{\gamma^\perp}$ **(b)** The holonomy map works in $(\gamma')^\perp$

Our information about the gaps and relations with periodic geodesics (for general manifolds, see §9.13 for special manifolds) is still quite meager. Looking again at the picture in figure 9.6 on page 397, one can consider $N(\lambda)$ as a step function. Not only the repartition, but also the jumps are of interest. The common belief today is that those jumps are related in some way yet to be discovered to the length of the periodic geodesics (this set of lengths is called the *length spectrum*). This belief was initiated in Balian & Bloch 1972 [99]. But today we are still missing formal results. In exchange, there are many numerical computations, mainly done by theoretical physicists. This because they are extremely interested in the the semiclassical limit (see more about this on page 376). Recent numerical experiments and thoughts about them can be found in Sarnak 1995 [1095], Luo & Sarnak 1994 [885], Luo & Sarnak 1995 [886], Rudnick & Sarnak 1996 [1074] and the bibliographies of those articles.

The most baffling case will be seen in §§9.13.2; it is the case of negative curvature space forms. The idea is that we know that those forms are chaotic in the good sense: the geodesic flow is very ergodic, the behavior of periodic geodesics and of the geodesic flow are extremely well understood. Briefly speaking, the geodesic flow is extremely evenly distributed in the phase space UM. Because of theorem 176 on page 405, one would expect that the eigenvalues are evenly distributed as a subset of the reals. The answer should be that the spectrum looks like the eigenvalues of a random Gaussian symmetric matrix. This major question is almost completely open today; see §§9.13.2. There is a good result on the distribution of the eigenfunctions; see theorem 185 on page 412.

Question 178 *What is $N(\lambda)$ for a generic Riemannian manifold? Is it in*

$$o\left(\lambda^{(d-1)/2}\right)$$

instead of the extreme

$$O\left(\lambda^{(d-1)/2}\right)\ ?$$

We know only of the following intermediate result:

Theorem 179 (Bérard 1977 [132]) *If a compact Riemannian manifold has no conjugate points or has nonpositive sectional curvature then as $\lambda \to \infty$*

$$N(\lambda) = \frac{\mathrm{Vol}(M)\beta(d)}{(2\pi)^d}\lambda^{d/2} + O\left(\frac{\lambda^{(d-1)/2}}{\log \lambda}\right)$$

Note 9.9.0.1 (Quasimodes) An interesting link between the spectrum and the periodic geodesics is that of the *quasimodes*. The story started in Babich & Lazutkin 1967 [95] and is far from being finished today, remaining quite mysterious; see Colin de Verdière 1977 [390]. Briefly speaking, to one given periodic geodesic (satisfying certain conditions), one can associate a series of numbers which approach quite a few eigenvalues. The idea of the proof is to build up approximate solutions of the wave equation which will propagate along the geodesic.

Question 180 *Are they many cases for which one can obtain the whole spectrum in this fashion?*

The answer is that this possibility is exceptional and happens only when the geodesic flow is integrable. In general, the hyperbolic zones between the KAM tori will yield a contradiction. The entire book Lazutkin 1993 [852] is devoted to this topic. ♦

Note 9.9.0.2 For *scars*, see §§§ 9.13.2.1 on page 426. ♦

9.10 The First Eigenvalue

9.10.1 λ_1 and Ricci Curvature

The first nonzero eigenvalue λ_1 is of essential importance. It controls the Dirichlet quotient of functions of mean value zero, and it also controls resonances. Indirectly it controls even the pure metric geometry of the manifold—via the distance functions—as seen in Colding's formula 77 on page 264. Again lower bounds are the true prize; upper bounds can be useful but definitely are less useful and much easier to get. We now present results which are not simply a special case of theorems 164 on page 386 or 172 on page 401.

The first result on λ_1 to our knowledge is the following which is hidden on page 135 of Lichnerowicz 1958 [865] and used there to study transformation groups of Riemannian manifolds.

Theorem 181 (Lichnerowicz [865]) *If the Ricci curvature is larger than or equal to $d-1$ (that of the standard sphere of dimension d) then λ_1 is at least as large as the λ_1 of the sphere, namely d. Moreover equality happens only for manifolds isometric to the sphere.*

The proof is beautifully simple, based on Bochner's formula theorem 346 on page 595 (or equation 15.8 on page 707), applied to the 1-form which is the differential df of the first eigenfunction f. This df is not harmonic but

$$\Delta f = \lambda_1 f$$

is the trace of the Hessian

$$Ddf = \operatorname{Hess} f \ .$$

Bochner's formula as applied to df becomes, after integration over the manifold and using Stokes' theorem:

$$0 = \int_M \| \operatorname{Hess} f \|^2 - \lambda_1 \int_M \|df\|^2 + int_M \operatorname{Ricci}(df, df)$$

The proof is concluded by using Newton's inequality

$$\| \operatorname{Hess} f \|^2 \geq \frac{(\Delta f)^2}{d}$$

since after diagonalization at a point,

$$\| \operatorname{Hess} f \|^2 = a_1^2 + \cdots + a_d^2$$

and

$$(Df)^2 = (a_1 + \cdots + a_d)^2 \ .$$

The equality is obtained quite easily tracing back each inequality, and appeared first in Obata 1962 [971] (also see Cheng 1975 [362]). This result should be compared with Myers' theorem 63 on page 245. We will come back to this in §§12.2.5. The general result of theorem 172 on page 401 as applied only to λ_1 is an improvement of theorem 181 on the facing page since it involves moreover the diameter (think for example of real projective space). But its main source of interest is that it can be applied when the Ricci curvature is nonnegative or negative.

For those who love Riemannian pinching, we mention Croke 1982 [413] for pinching λ_1, and the recent Petersen 1999 [1020].

9.10.2 Cheeger's Constant

A somewhat intermediate result between Lichnerowicz's theorem 181 on the preceding page and theorem 172 on page 401 is based on Cheeger's constant h_c introduced on page 315.

Theorem 182 (Cheeger 1970 [329]) *On any compact Riemannian manifold*

$$\lambda_1 > \frac{1}{4} h_c^2 \ .$$

It was proved in Buser 1978 [291] that this inequality is optimal, but equality never occurs for a smooth metric. For more on this and the role of λ_1, see §§9.13.1. There is a huge literature on λ_1 but still it seems that there has never been any practical application to various questions concerning "vibrations of great structures," or "nondestructive and noninvasive tests." There is a relation obviously, but vibration today is largely an experimental area of mechanical engineering. Bell casters have always used tests of the sound of a bell to check for possible cracks; see Bourguignon 1986 [238].

9.10.3 λ_1 and Volume; Surfaces and Multiplicity

Despite theorem 164 on page 386 (which used Ricci curvature and volume), there cannot exist an upper bound involving only the volume. This was proven in Dodziuk 1993 [453], by simply building up suitable examples (of course of larger and larger diameter, and this only for dimension larger than or equal to 3). The question was raised because the case of surfaces is exceptional. In fact:

Theorem 183 (Hersch, 1970) *The first three eigenvalues of any Riemannian metric on the sphere S^2 obey the inequality*

$$\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} \geq \frac{3}{8\pi} \operatorname{Area}\left(S^2, g\right)$$

with equality only for the standard sphere. In particular

$$\lambda_1 < \frac{8\pi}{3} \frac{1}{\operatorname{Area}\left(S^2, g\right)} .$$

The proof is beautiful. It mixes three facts

1. the minimax principle of §§9.4.1,
2. the fact that the Dirichlet quotient of a surface is invariant under conformal change and finally
3. the fact that the conformal group of S^2 is large enough to transform any density on the sphere into a one whose center of mass is the origin.

For other compact surfaces, various authors found an upper bound involving only the area, with a constant depending on the genus. The optimal constant is still a pending problem. On this topic recent references can be found in the bibliographies of Dodziuk 1993 [453] and Nadirashvili 1996 [964].

The question of the highest possible multiplicity of λ_1 is also interesting for surfaces. Discard higher dimensions, thanks to the Colin de Verdière result 186 on page 415 to the effect that, starting in dimension three, any finite subset of the reals—including multiplicities—can always be realized as the beginning of the spectrum of a suitable Riemannian manifold. But for surfaces, the multiplicity of λ_1 is bounded with the genus of the surface. Results are optimal

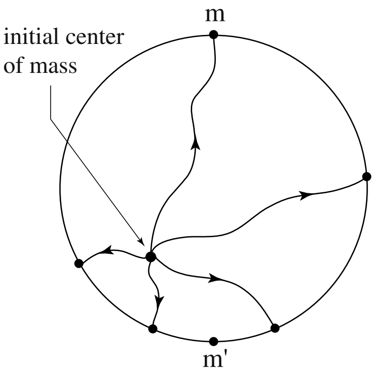

Fig. 9.10. Under radial conformal transformations sending all but one point into it, one can cover the whole ball with the initial center of mass

today for the sphere (triple) and the torus (sextuple). Optimal constants are still to be discovered for other surfaces. There is definitely a relation between this multiplicity and the chromatic number of the surface; see the definition on page 415.

To prove such an upper bound, one relies on the structure of the set of nodal lines, i.e. the set where an eigenfunction vanishes. Except at a finite set of singular points, the zero set is made up of regular curves. More important is that at singular points the curves meet with a set of tangents which are the directions of the diagonals of a regular polygon. Using this result of Bers the proof is concluded with arguments of algebraic topology; references are Besson 1980 [187], Yang & Yau 1980 [1289].

9.10.4 Kähler Manifolds

Mathematicians never stop asking questions. For example, can we have an upper bound on the first eigenvalue depending only on the volume when the manifold is "special"? Considering the geometric holonomy hierarchy introduced in chapter 13, the case to look at is that of Kähler manifolds (see §13.6). Indeed it is natural to wonder about the general spectrum of a Kähler manifold. One answer is the following extension to $\mathbb{CP}^n$ of Hersch's theorem 183 on the preceding page for S^2:

Theorem 184 (Bourguignon, Li & Yau 1994 [245]) *For any Riemannian metric g on $\mathbb{CP}^n$*

$$\lambda_1 \leq \frac{(n+1)\pi^n/n!}{\mathrm{Vol}\,(\mathbb{CP}^n, g)} .$$

We recall (see §§9.5.4 and §§§7.1.1.2) that the volume of the canonical metric of $\mathbb{CP}^n$ is

$$\text{Vol}\,(\mathbb{CP}^n, \text{Fubini–Study}) = \pi^n/n!$$

and its first eigenvalue is $n+1$. For the proof, the conformal group of the sphere is replaced here by the group of all biholomorphic transformations of $\mathbb{CP}^n$. In Bourguignon, Li & Yau 1994 [245] and Gromov 1992 [630] one will find generalizations of this result to various algebraic manifolds, and to the whole spectrum. For the big picture of the subject, it is important to remark that the spectrum is a robust invariant, while being Kähler is not: see note 9.4.1.1 on page 385 and §13.6 and §14.6.

The extremely important case of "Riemann surfaces", that is to say of constant curvature -1, whether or not compact, will be studied at large in §§9.13.1. In the spirit of §9.12, we are far from being able to recognize the spectra of Kähler manifolds.

9.11 Results on Eigenfunctions

9.11.1 Distribution of the Eigenfunctions

It seems hopeless to search for any general result valid for "any" Riemannian manifold. But one can hope for a regular distribution of the eigenfunctions when the manifold is generic (in any sense). A regular distribution would be one for which in any domain D of the manifold and for eigenfunctions with larger and larger eigenvalue, one finds the integral of the square of that function over that domain is in a proportion to the integral over the whole manifold which is closer and closer to the ratio of the volumes of D and M. To our knowledge there is not a single result in that direction; compare with the periodic geodesic result in §§10.3.5.

But if the manifold is "ergodic" (see §§10.5.1), then there are partial results. The conjecture is that ergodicity implies an even distribution of the eigenvalues and the eigenfunctions. Concerning the eigenfunctions one has only:

Theorem 185 *For an ergodic Riemannian manifold M, there is a sequence*

$$\{i(k)\}$$

of integers, of full density in the integers, such that for every $D \subset M$ with eigenfunctions $\phi_{i(k)}$ being normalized:

$$\lim_{k\to\infty} \int_M \phi_{i(k)}^2 = \frac{\text{Vol}(D)}{\text{Vol}(M)}\ .$$

Measure theory aficionados would prefer to write this as

$$\lim_{k\to\infty} \phi_{i(k)}^2 dV_M = dV_M$$

Full density means that the number of points in question in $[0, \lambda]$, compared to the whole spectrum, has a ratio closer and closer to one when λ goes to infinity. The latest general reference on this topic is Colin de Verdière 1985 [392] for our compact case, which completed the attempt of Shnirel′man 1973 [1137]. For the noncompact see Zelditch 1987 [1302] and Zelditch 1992 [1303]. The proofs involve a deep theorem of Yuri V. Egorov on Fourier integral operators and belong therefore to microlocal analysis. So again, the wave equation is used even if it disappears in the final statement. For the very special case of space forms of negative curvature, see further references in §9.13, but the results are still incomplete today.

9.11.2 Volume of the Nodal Hypersurfaces

Another way to look at regularity of eigenfunctions is to study their nodal hypersurfaces, namely the subsets of the manifold where they vanish. When the manifold is a surface, these subsets are curves. A reasonable behaviour to expect is that the volume of the $\phi_\lambda^{-1}(0)$ will grow as $\lambda \to \infty$, with some asymptotic order. The reader can check on examples (flat tori being the simplest) and also looking at spherical harmonics (see §§9.5.2 and §§9.5.3) that an eigenfunction with eigenvalue λ behaves like a polynomial of degree $\sqrt{\lambda}$. If this is more or less true for any compact Riemannian manifold, then one will have $\operatorname{Vol}\left(\phi_\lambda^{-1}(0)\right)$ roughly behaving like $\sqrt{\lambda}$. It was conjectured by Yau in 1982 that for every Riemannian manifold M with Riemannian metric g there are constants $c = c(g)$ and $c' = c'(g)$ such that

$$c\sqrt{\lambda} \leq \operatorname{Vol}\left(\phi_\lambda^{-1}(0)\right) \leq c'\sqrt{\lambda} \tag{9.25}$$

for every eigenvalue λ. The intuitive idea behind Yau's conjecture was that eigenfunctions for λ behave roughly like polynomials of degree $\sqrt{\lambda}$, which is the case for the standard sphere for which the eigenfunctions are the restrictions to the sphere of the harmonic polynomials of Euclidean space. After the partial result of Brüning 1978 [266], this was proven in Donnelly & Fefferman 1988 [461]. The volume is to be understood as the $(d-1)$ dimensional Hausdorff measure to be sure to make sense. The proof is extremely hard, and involves various results from analysis. One needs to know the local behaviour of the eigenfunctions, their local sup norm and the distribution of their singular zeroes. Another basic fact is the analyticity of the eigenfunctions of an elliptic operator (here the Laplacian). And the proof tells us even more about the eigenfunctions.

The story does not end here for at least two reasons. The first in that the proof we need the analyticity of both the manifold and the metric. But for the geometer the major drawback is that the two constants $c(g)$ and $c'(g)$ are unknown. They come from an atlas and its coordinate changes. The geometer would like to be able to express $c(g)$ and $c'(g)$ as functions of Riemannian invariants of (M, g) (and of course the cheapest possible ones). We know of no work on this. Let us mention a recent paper addressing noncompact manifolds: Donnelly & Fefferman 1992 [462]. Also see Savo 2001 [1101].

9.11.3 Distribution of the Nodal Hypersurfaces

Figure 1.98 on page 91 shows the extraordinary regularity of a nodal line. There is some reason to believe that when the geodesic flow of a Riemannian manifold is ergodic, the nodal sets are evenly distributed. In saying that nodal lines are evenly distributed, we mean something like asking that given any domain $D \subset M$

$$\lim_{\lambda\to\infty} \frac{\operatorname{Vol}\left(D \cap \phi_\lambda^{-1}(0)\right)}{\operatorname{Vol}\left(\phi_\lambda^{-1}(0)\right)} = \frac{\operatorname{Vol}(D)}{\operatorname{Vol}(M)} .$$

Today there are only numerical experiments. Nodal sets might also be connected to periodic geodesics by some mysterious phenomenon called *scarring*; see figure 1.100 on page 93. For a discussion of scars, we refer to Sarnak 1995 [1095], also see §§§9.13.2.1.

9.12 Inverse Problems

The general scheme is to try to understand the map

$$(M, g) \mapsto \operatorname{Spec}(M, g)$$

from Riemannian structures on a manifold M to the set of all discrete subsets of the positive real line:

$$\operatorname{Spec} : \mathcal{RS}(M) \to \left\{\text{discrete subsets of } \mathbb{R}^+\right\}$$

By a Riemannian structure, recall that we mean a point in the quotient set of the set of Riemannian metrics by all possible diffeomorphisms. We do not want to distinguish between two isometric Riemannian manifolds (metrics). The first question is to determine the image of this map, the second is about its inverse: is it one-to-one, and if not what can be said about the preimages of various points in the image?

9.12.1 The Nature of the Image

We are far today from being able to guess a sufficient condition for a subset of the reals to be realizable as the spectrum of some Riemannian manifold; we know only of Omori 1983 [973]. Of course all of the results above can be viewed as necessary conditions, the most typical one being Weyl's asymptotic as made precise in Hörmander's result 173 on page 403, as well as its gap corollary. But for a finite set to be realized as the beginning of a spectrum (including imposed multiplicities) there is no obstruction:

Theorem 186 (Colin de Verdière 1987 [393]) *For any compact manifold M of dimension larger than or equal to 3 and for any finite subset of the positive real numbers, indexed with finite multiplicities, there exists some Riemannian structure on M whose spectrum begins with that subset.*

The proof is nice. It consists in putting points on the manifold, considering them as oscillators with the desired frequency and multiplicity. Then one joins them by nonintersecting curves, building up a tubular neighborhood of that structure and controlling everything to keep this finite spectrum. One can also see this result as first finding a (finite) graph whose spectrum for its standard graph Laplacian is the desired finite piece under consideration, and then playing some kind of "tunnel effect" along the edges. Technically the multiplicities give troubles, which can finally be controlled by a subtle transversality argument. But for infinite subsets of the reals, the question of sufficient conditions seems completely open; however see note 9.12.1.1 on the following page.

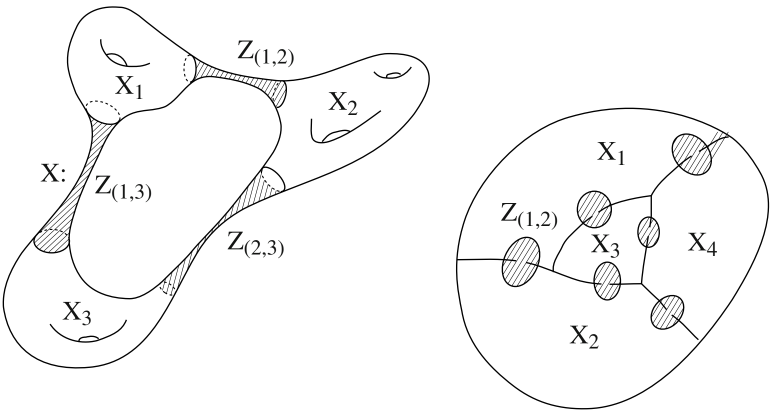

Fig. 9.11. Colin de Verdière's proof that one can choose any finite part of the spectrum of the Laplacian

It is when joining the points by nonintersecting curves that the condition on the dimension appears. This is of course impossible without extra conditions when the dimension is 2, since then some of those curves can be forced to meet. In fact this fits perfectly with the restriction on the multiplicity of λ_1 that we met in §§9.10.3. For the interested reader we mention here that pursuing this topic in the case of surfaces Colin de Verdière discovered recently a fascinating application to electrical circuits: see Colin de Verdière 1996 [396]. He was also led to make the following conjecture. For a compact surface M define its *chromatic number* Chrom (M) as the largest integer N so that there is an embedding into M of the complete graph with N vertices.

Conjecture 187 (Colin de Verdière) *For any surface M, the highest possible multiplicity of λ_1 of any Riemannian metric is equal to* $\mathrm{Chrom}\,(M) - 1$.

See page 601 of Colin de Verdière 1987 [393] for more on that.

Note 9.12.1.1 In theorem 167 on page 400 we saw a dramatic improvement of Colin de Verdière's results. The scheme for Lohkamp's proof is as follows: modify Colin de Verdière's construction by suitable "attachments of metrics". These constructions are hard and subtle—in particular the author uses Besicovitch's coverings and the technique of "crushed ice". ♦

9.12.2 Inverse Problems: Nonuniqueness

We have been studying direct problems: I know the manifold and some of its invariants. What can I say about the spectrum? Inverse problems have the form: I know various things about the spectrum, what can I recover of the metric? The first question is the uniqueness: are two isospectral manifolds necessarily isometric ?

The first time the author heard about this question was in letter written to him by Leon Green around 1960. In this letter, Green also remarked on an almost straightforward fact: if one knows not only the eigenvalues but also the eigenfunctions, then one knows the metric (two such manifolds can be called *homowave* or *homophonic*). This is because the completeness of the eigenfunctions (see §§9.3.3) implies knowledge of the Laplacian acting on functions, and then from the explicit formula of the Laplacian in coordinates, one recovers immediately the g_{ij}.

The isospectral question was a strong incentive in the sixties. In the case of Riemann surfaces, uniqueness was conjectured in Gel′fand 1962 [553]. For plane domains, we already met this question in §§1.8.4. The first counterexample came in Milnor 1964 [922]. It consists in two tori of dimension 16 with exactly the same spectra. By the results of §§9.5.2, we know the spectrum of a flat torus as soon as we know the lattice defining it. Then two lattices Λ and Λ' in $\mathbb{R}^d$ will yield isospectral tori if and only if the number N_m of points in them having a given norm m is always the same. The set of these numbers is completely encoded in the theta series of the lattices. Namely one defines the theta series of the lattice Λ by

$$\Theta_\Lambda(z) = \sum_{x\in\Lambda} q^{x\cdot x} = \sum_m N_m q^m \tag{9.26}$$

(where $q = exp(\pi i z)$) defined for suitable values of the complex variable z. These functions have been exhaustively studied for purposes of number theory. An excellent presentation is 2.3 (pages 44–47) of Conway & Sloane 1999 [403]. There one will found out how to compute the theta series of various lattices, depending how they are defined.

Milnor's examples were the two lattices called $E_8 \times E_8$ and E_{16}. The lattice E_8 is the famous lattice attached to the exceptional Lie group denoted also by E_8. It can be defined as the set of tuples $(n_1, \dots, n_8)$ where all n_i are integers or integers plus 1/2 and with the extra condition that $\sum_i n_i$ is even. The lattice E_{16} is then simple to construct. What is subtle is to compute their theta series and to show that they are identical; a very good exposition of this is to be found in Serre 1973 [1125]. Checking that they are not isometric (congruent) is the trivial part.

It is a easy exercise to show that isospectral 2-dimensional lattices are congruent (i.e. they can be rotated into one another in the Euclidean sense). Various people found isospectral flat tori of various dimensions. We refer the reader to Conway & Sloane 1999 [403] for them. The dimension can go as low as four. For this dimension one will find on page xxi of the preface (of the second edition) of Conway & Sloane unbelievably simple examples, depending moreover on four parameters. The case of dimension 3 was finally solved positively in Schiemann [1103]; indeed one only needs to know that the eigenvalues are not too large.

Then people got more and more examples of different types. Using number theory (quaternionic number fields) Gel′fand's 1962 conjecture of uniqueness for Riemann surfaces was disproven in Vignéras 1980 [1216]. Thereafter the field blew up so much that we just give few references permitting the reader to go back to all of them. The landmark Sunada 1985 [1168] put things in the right context, at least when considering space forms obtained by quotienting by discrete subgroups. One then finds a sufficient algebraic condition between two such groups to yield isospectral quotients. Then Gordon 1993 [575] gives very geometric methods (using transplantation techniques, see Bérard 1989 [137] which can also be used as a survey) to construct isospectral Riemann surfaces. It is interesting to note that the plane isospectral domains, mentioned in §§1.8.4, were found using isospectral abstract Riemann surfaces with no boundary (compare figures 9.12 on the following page and 1.94 on page 86).

It is a natural instinct to search for more and more general examples; the preceding ones were all space forms. People found locally homogeneous spaces, then nonhomogeneous ones and even one parameter deformations; see Bérard & Webb 1995 [141], Gordon & Webb 1994 [578], Gordon 1994 [575], Gordon 2000 [576], Gordon & Mao 1994 [577], Gornet 1998 [582], and Schueth 1999 [1112].

In Szabo 2001 [1174], very interesting pairs of isospectral metrics are constructed on spheres; they can be made as close to the canonical metric as you like.

We will see in §9.14 that there is a natural Laplacian for exterior differential forms of any degree, hence associated spectra for each degree. We naturally meet the question of obtaining isospectral, nonisometric metrics at the level of differential forms. Today one has examples of different types. For instance, the counterexamples with flat tori are always isospectral for differential forms of any degree, since the eigenvalues for differential forms coincide trivially with

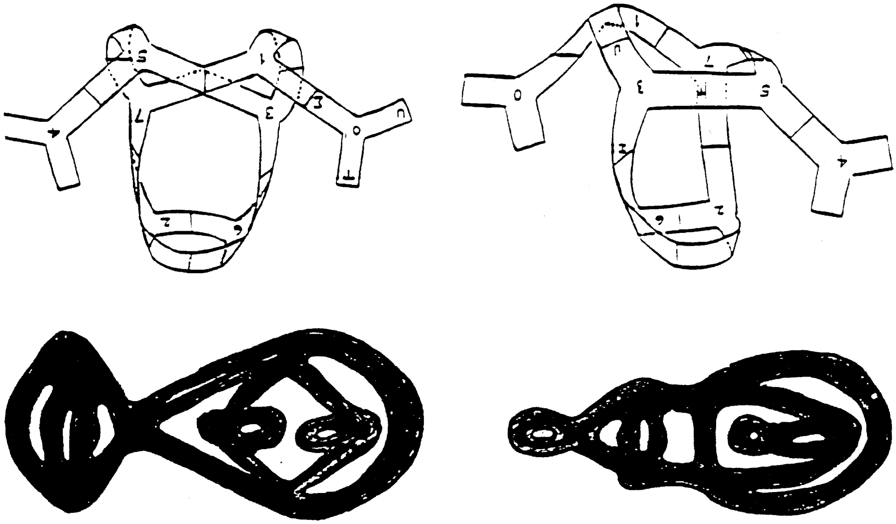

Fig. 9.12. Constructing isospectral plane domains out of isospectral surfaces of constant curvature

those for numerical functions, with only the multiplicities being multiplied by the fixed constant which is the binomial number $\binom{\dim M}{p}$. In the opposite direction, one will find in Gornet 1996 [581] examples distinguishing between isospectrality for functions and for differential forms.

9.12.3 Inverse Problems: Finiteness, Compactness

Since the geometry with a given spectrum is not unique, we can still try to have information on the possible geometries, i.e. sets of Riemannian structures having the same spectrum. How large can these sets be? Do they have any kind of structure, in particular are they "finite dimensional" in any reasonable sense or "compact"? To our knowledge, the finite dimensionality is a completely open problem (unlike the case of Einstein metrics as we will see in theorem 286 on page 531). The infinitesimal isospectral deformation equations in the space of metrics look hopeless; we will just meet a few exceptions.

But there is a nice result for surfaces:

Theorem 188 (Osgood, Phillips & Sarnak [979]) *For any choice of spectrum, the set of Riemannian structures (i.e. Riemannian metrics up to diffeomorphism) with that spectrum is compact.*

The proof is hard but two of its ingredients are of great importance for other purposes. The first is the collection of curvature terms which appear in the asymptotic expansion of the heat kernel; see theorem 165 on page 393. The second ingredient is new for us: it is the *determinant of the Laplacian*. Formally, it is defined as

$$\det \Delta = \prod_i \lambda_i \tag{9.27}$$

Approached naively, as it is written, this product is not convergent, but one can define it anyway, using various regularization tricks. The most common trick is to take its logarithm: consider the ζ *function* of the spectrum:

$$\zeta(s) = \sum_i \frac{1}{\lambda_i^s}$$

and compute formally $\zeta'(0)$. You will find the determinant. All the effort now focuses on rendering this analysis rigorous. We refer to Osgood, Phillips & Sarnak [979] for details of the proof and also for references on this determinant. Recall that we mentioned on page 86 a compactness result that was obtained for isospectral plane domains. This determinant is also used for an extraordinary proof of the conformal representation theorem 70 on page 254 using Ricci flow; see references and the current state of affairs in Chow [378].

The use of the determinant cannot be avoided. The heat invariants are certainly not enough. This can be seen simply because all of these invariants coincide in the case of constant curvature metrics, as already remarked in note 9.7.2.1 on page 399. On the other hand the Teichmüller space of all Riemann surfaces of a given genus is not compact. In particular, (at least for higher genus) one cannot prove compactness inside the set of metrics conformal to a given metric, a case which is much simpler since it only involves scalar functions instead of metric tensors; namely they involve only the Gauss curvature K and its various iterated Laplacians $\Delta^m K$. But in two dimensions, extensive study shows that the heat invariants are simple enough when they are controlled by the determinant of Δ. For this determinant as a functional of Riemannian metrics, see Sarnak [1096].

The above proof suggests the conjecture that, in two dimensions, there can be only a finite number of metrics isospectral to a given metric. This is certainly false in higher dimensions, since there are one-parameter isospectral deformations.

The compactness of higher dimensional isospectral sets is open. One reason is that the proof above involves controlling the nature of the heat invariants, which are so much simpler for surfaces. However there are good partial results, in dimensions 3 and 4: Osgood, Phillips & Sarnak [980], Anderson [41] and Brooks, Perry & Petersen [264]. For topological finiteness of isospectral sets, see Brooks, Perry, & Petersen [263, 262]. Also see Gordon 2000 [576].

9.12.4 Uniqueness and Rigidity Results

We already saw on page 399 the uniqueness of the spectra of the standard spheres up to dimension 6. The analogous question is open for higher dimensions. But there are good results:

Theorem 189 (Guillemin & Kazhdan 1980 [667] and Croke & Sharafutdinov [422]) On compact manifolds of negative curvature, there are no isospectral deformations.

The proof for surfaces is beautiful and we explain it in some detail because it seems to us that this technique could be used more widely. It is a kind of double Fourier analysis leading to a contradiction. There are three steps. One looks at the derivative of a deformation of metrics on the unit tangent bundle UM. This bundle has fibers which are circles, which leads to Fourier analysis for functions $UM \to \mathbb{R}$. If the Fourier subspaces are called H_i then the deformation function

$$t : UM \to \mathbb{R}$$

belongs to the direct sum

$$H_{-2} \oplus H_0 \oplus H_2$$

because Riemannian metrics are quadratic forms. Now one invokes theorem 176 on page 405 to the effect that the lengths of periodic geodesics are preserved under our deformation since it is isospectral. A periodic geodesic when lifted up to UM is now a periodic trajectory of the geodesic flow. An easy computation, the "first variation formula for changes of metric," shows that the integral of the deformation function t is zero along any periodic geodesic. But a manifold of negative curvature has a lot of periodic geodesics, dense in the best possible sense (see §10.6). This explains (although it is not a proof) a result of Livtsic to the effect that there exists a new function $s : UM \to \mathbb{R}$ such that t is the derivative of s along the geodesic flow. It remains now to look at how the geodesic vector field behaves with respect to the Fourier analysis above. The negativity of the curvature implies that differentiating in G lowers the rank in Fourier analysis. In particular $s' = t$ implies that

$$s \in H_1 \oplus H_0 \oplus H_{-1} \ .$$

But s also should be like t in $H_2 \oplus H_0 \oplus H_{-2}$ and this finishes the proof: t has to be constant along the fibers. The proof of Croke & Sharafutdinov 1997 [422] for higher dimensions is somewhat different.

9.12.4.1 Vignéras Surfaces

The Vignéras examples of surfaces with the same spectrum appeared in Vignéras 1980 [1216] The recent basic uniqueness and rigidity result Besson, Courtois & Gallot 1995 [189], which will be addressed in detail in theorem 251 on page 484, has already had so many applications that its authors are conjecturing (see 9.20, page 780) a result which would be in some sense the best possible:

Conjecture 190 *Isospectral, compact, negatively curved manifolds of dimension larger than 2 are isometric.*

Question 191 *Is isospectrality a nongeneric phenomena? Otherwise stated: are generic Riemannian manifolds spectrally isolated (solitude)?*

A third remark concerns the length spectrum, i.e. the set of length of periodic geodesics. From theorem 176 on page 405 one is certain that isospectrality implies coincidence of the length spectra; but Vignéras counterexamples in §§§9.12.4.1 show that different Riemann surfaces can have the same length spectrum. In §10.11 we will see that is not the case for the marked length spectrum. This is true for example for negative curvature manifolds of dimension higher than 2 and supports the conjecture just presented: see 9.14 in Besson, Courtois & Gallot 1995 [189].

9.13 Special Cases

9.13.1 Riemann Surfaces

By a Riemann surface we understand a compact orientable surface of constant curvature -1. In our hierarchy they are the negative space forms of dimension 2. This means we exclude the sphere and the torus.

Riemann surfaces have been studied since Riemann in great detail, for their intrinsic interest. They appeared originally in complex variable theory, in algebraic geometry and in number theory. Recently they became a favourite object for theoretical physicists, in particular in string theory. It is then not surprising that we have many strong results for them, including for their spectra. The book Buser 1992 [292] is a very complete exposition of the subject at that date. A more recent survey is Buser 1997 [293]. We just note that in Buser 1992 [292] the question of the regularity (randomness) of the spectrum and that of the eigenfunctions (compare with §9.9 and theorem 185 on page 412) are still not well understood, we will discuss them in §§9.13.2.

The first basic fact is that for Riemann surfaces theorem 176 on page 405 can be inverted. What theorem 176 says is that the function spectrum of the Laplacian determines the length spectrum (the set of lengths of the periodic geodesics). But the converse is false in general; one needs much more that the length spectrum, namely essentially the Poincaré map and the parallel transport of periodic geodesics. But in the case of Riemann surfaces, the parallel transport is always the identity since the dimension is two and we have orientability. The Poincaré map is also known because the curvature is constant. This explains (but of course does not prove):

Theorem 192 (Huber 1959 [745, 747]) *On a Riemann surface, the spectrum of the Laplace operator on functions determines the length spectrum and vice versa.*

The proof is based on a formula for Riemann surfaces which is a generalization of the Poisson formula 9.14 on page 389 which was valid for flat tori. The formula computes the heat kernel by a suitable summation formula involving the length spectrum. It is possible simply because there is an explicit formula for the heat kernel K^* of the total (noncompact!) hyperbolic space Hyp^d, and in particular for Hyp^2. Our surface is a quotient of Hyp^2 by a discrete group of hyperbolic isometries. It is enough to know the primitive elements of this group. Being without fixed points, they have to consist in a *gliding* along an hyperbolic line (called the *axis* and denoted by γ). The length of the gliding corresponds exactly to the length of a periodic geodesic downstairs. As a matrix of the group $\mathrm{Isom}\left(\mathrm{Hyp}^2\right)$ that length is exactly the trace of this matrix. This explains the name "trace formula." This formula of Huber is a particular case of Selberg's trace formula which we will meet below. The proof is finished by remarking that the heat kernel downstairs is a suitable summation of the type

$$K = \sum_{\gamma} K^* (x, \gamma y)$$

for the axis γ above; details are to be found in chapter 9 of Buser 1992 [292].

We now present to the reader a choice of results that we find especially appealing; most of them are in the book Buser 1992 [292]. The heuristic possibility of these results comes from Huber's theorem, as explained in the preface of the book:

> *This theorem does not show only that the eigenvalues contain a great deal of geometric information, it also indicates that spectral problems may be approached by geometric methods. . . .*
>
> Buser 1992 [292]

These geometric methods rest essentially on the fact that the set of all Riemannian surface structures on a given orientable surface of genus larger than 1 can be encoded in the lengths of the sides of the hexagonal pantaloon hyperbolic plane pieces and the twisting angles when one glues them together as was done in figure 4.10 on page 157. The study is still not too clear conceptually in Buser's book. But in Buser 1997 [293] the author made a decisive step. He succeeded, at least for a very large class of Riemann surfaces, to find the surface itself directly and explicitly from the spectrum. This means that the complete geometry is encoded in the spectrum. Those surfaces are called solitary because they don't have nonisometric isospectral companions.

We start with the eigenvalues called *small*.What is important for a Riemann surface is not only λ_1 and its position with respect to $1/4$, but also the set of λ's which are in $]0, 1/4]$ (called *small*). Why $1/4$ comes into the picture cannot be explained briefly; for details we refer the reader to Buser's book. From it we extract this. In writing the heat kernel as a summation, it is convenient to write the eigenvalues $\lambda = r^2 + 1/4$, so that the associated r are imaginary when λ is below $1/4$. A very heuristic reason is that in hyperbolic geometry, the modular domain is the one in figure 6.36 on page 255 and that

$$1/4 = (1/2)^2 \ .$$

Let us just recall that this modular domain is the quotient of the hyperbolic plane by the isometries whose matrix is integral. It might be the most important object of all mathematics, as its is connected with function analysis, complex variables, number theory, etc. Remember in this context the Riemann hypothesis for the zeros of the ζ function which "should" be all on the line $s = 1/2$.

Today the situation for small eigenvalues is satisfactory on one hand but on the other hand some conjectures are still open. Let us also mention that the small eigenvalues play a basic role in the refined version of the asymptotic expansion for the counting function of the length spectrum, as will be seen in theorem 205 on page 447. If we denote by $\mathcal{M}_\gamma$ the set of all Riemann surfaces of a given genus γ then

Theorem 193 (Buser 1992 [292] 8.1.1) *For any γ and any surface in $\mathcal{M}_\gamma$,*

$$\lambda_{4\gamma-2} > 1/4 .$$

Theorem 194 (Buser 1992 [292] 8.1.2) *For any γ and any integer n (think large) and for any $\varepsilon > 0$ (think of ε as small) there are elements of $\mathcal{M}_\gamma$ with*

$$\lambda_n \leq 1/4 + \varepsilon .$$

Together these two statements look surprising. There is a universal bound for the number of eigenvalues in $[0, 1/4]$ but not in any $[0, 1/4 + \varepsilon]$. A geometric reason is offered on page 211 of Buser's book; it mixes isoperimetric considerations for hyperbolic hexagons and the fact that $\mathcal{M}_\gamma$ is never compact—see just below.

Theorem 195 (Buser 1992 [292] 8.1.3) *For any $\varepsilon > 0$ there is a genus γ and a surface $\mathcal{M}_\gamma$ with*

$$\lambda_{2\gamma-3} < \varepsilon .$$

Theorem 196 (Buser 1992 [292] 8.1.4) *There is a universal constant $c > 0$ so that for any γ and any surface in $\mathcal{M}_\gamma$*

$$\lambda_{2\gamma-2} > c .$$

Although the conjectured value for c is in fact $1/4$, today the best known c is around 10^{-12}. There are many other results for small eigenvalues; see the Notes at the end of chapter 8 of Buser's book.

We turn now to the isospectral question. Recall that there are examples of isospectral but nonisometric Riemann surfaces: see §§9.12.2 and also that there is a general compactness result: see §§9.12.3. But in the present case we also have finiteness:

Theorem 197 (Buser 1992 [292] 13.1.1) *For a given genus γ there are at most* $\exp\left(720\gamma^2\right)$ *pairwise nonisometric isospectral Riemann surfaces.*

The last topic we will discuss in this section is *Wolpert's theorem* (1977-79). It says that for Riemann surfaces a certain finite part of the length spectrum determines the whole spectrum. In Buser's book the precise statement is theorem 10.1.4. Then Buser extends the theorem to the function spectrum as follows:

Theorem 198 (Buser 1992 [292] 14.10.1) *For any $\varepsilon > 0$ and any γ there is a universal constant* $\mathrm{univ}(\varepsilon, \gamma)$ *such that if two Riemann surfaces S and S' of the same genus γ both with injectivity radius larger than ε verify $\lambda_n(S) = \lambda_n(S')$ for every $n <$* $\mathrm{univ}(\varepsilon, \gamma)$ *then they are isospectral for their whole spectrum.*

Some remarks are now in order. First, the lower bound on the injectivity radius cannot be avoided. The noncompactness of $\mathcal{M}_\gamma$ is directly linked with the fact that the injectivity radius can go to zero. Conversely, compactness when there is lower bound on the injectivity radius is a very special case of the general compactness theorem which we will meet in complete detail in §§12.4.2 and also theorem 376 on page 621.

Second, the original proofs (both for the length and the function spectrum) were extremely expensive, using in particular the theory of real analytic varieties. Recently in Buser 1997 [293] the results on solitary surfaces (mentioned above) were used to give a much simpler proof of theorems like Wolpert's. Also see Schmutz 1996 [1106].

9.13.2 Space Forms

The preceding section concerned space forms of dimension two and of negative curvature. The case of zero or positive curvature was treated in section §§9.7.2 where we saw that the standard sphere and the standard $\mathbb{RP}^2$ are determined by their spectrum, as are flat tori. This was done using the asymptotic expansion of the heat kernel.

Looking now at higher dimensions, we saw in §§9.12.2 the state of affairs for flat tori and for spheres, circumstances being particularly unsatisfactory for spheres. Let us turn now to the compact manifolds of negative constant sectional curvature. This is very special case among manifolds of negative curvature. We saw at large in §9.11 that there are some results on the distribution of eigenvalues and of eigenfunctions for ergodic manifolds. But also that those results were very partial, the basic questions being completely open. Since negative curvature manifolds are ergodic in a very strong sense (see §10.6) and since we will see extremely satisfying results for them in §10.8 for the length spectrum with optimality for the space forms, it is then natural to expect for negative curvature space forms much stronger results than for the general ergodic or negatively curved ones. This was the case for Riemann surfaces as seen just above to some respect, in particular for the small eigenvalues.

There is a theoretical answer to every question concerning spectra of negatively curved Riemann surfaces, namely *Selberg's trace formula*, which in dimension 2 gives back part of Huber's theorem 192 on page 421. For higher dimensions, see Bunke & Olbrich [279].

These questions are under very intense study today. The hope is to use tools from number theory, since these space forms are mostly found by arithmetic means; see §§6.6.2. The tools are typically modular functions (for the flat tori in §§9.12.2 they were theta functions). Strong incentives come to this study from mathematical physics, in particular in what is called the semiclassical limit (see more on page 376) and in the present situation from *quantum chaos.*

There is no general picture arising from the various results obtained up to know. We already said in §9.9 that experts disagree, comparing mathematical results and numerical experiments (including dimension 2). We mention only references: Sarnak 1995 [1095], Luo & Sarnak 1994 [885], Luo & Sarnak 1995 [886], Rudnick & Sarnak 1996 [1074]. One should also of course look at the bibliographies of those. Today a conjecture is the following: there are numerical experiments from which it seems that the distribution of eigenvalues is not even for some arithmetic Riemann surfaces. That is, the distribution is not a *Gaussian orthogonal ensemble* (GOE), i.e. the set of the eigenvalues of a random $N \times N$ symmetric matrix as $N \to \infty$, with the whole business being rescaled to agree with Weyl's asymptotic. This negative statement was mathematically proven in Luo & Sarnak 1995 [886]. In figure 9.13 we see a picture taken from Sarnak 1995 [1095], comparing, for plane regions, arithmetic and the nonarithmetic spectra (see more on page 292 for the definition of arithmeticity in abstraction, but it is not really too much different for plane domains, and the plane domains are accessible to numerical computations).

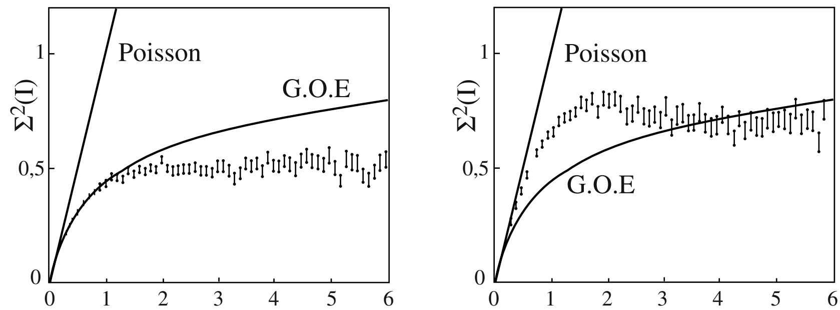

Fig. 9.13. (a) Σ^2 for a nonarithmetic triangle **(b)** $\Sigma^2(L)$ for an arithmetic triangle

However the geodesic flow is ergodic. Today it is believed that the distribution will be GOE for generic Riemann surfaces. And to explain the reason why arithmetic forms are exceptional, one should remember what was said in §9.9, namely that the jumps in the spectrum are linked with the structure of the length spectrum. But one knows that the length spectrum of an arithmetic form is very "degenerate" in the sense that the lengths are given by suitable integers—the reason for this is that we saw that the length shifts of gliding hyperbolic isometries are represented by the trace of an integral entry matrix. The asymptotic exponential behavior (see equation 9.19 on page 396) then forces all of these periodic geodesics to have very large multiplicities, hence huge jumps in the length spectrum.

In these results, precise descriptions of many quantities are studied for those space forms, not only the L^2 norms but also the sup norm. More: the behaviour of integrals like

$$\int_M P\left(\phi_{i_1}, \ldots, \phi_{i_k}\right) dV_M$$

for various polynomials P of degree k and their asymptotic behavior when one or more of the eigenvalues goes to infinity is related to possible scarring, which is the next problem we have to consider.

9.13.2.1 Scars

This is linked with the question of whether there are "scars." In some numerical experiments, people found that the nodal lines of some surfaces were, in some sense, accumulating along periodic geodesics. But in Sarnak 1995 [1095] it is proven that this can never happen for arithmetic space forms (for some suitable definition of what a scar is). A picture of a scar in a planar region is presented in figure 1.100 on page 93. This is an amusing paradox: the arithmetic case implies more regularity, and at the time it is a less common case (in the realm of space forms). The general state of affairs still divides experts, since scarring today is only purely experimental and because the definition of scars varies between authors; see Rudnick & Sarnak 1996 [1074], Shimizu & Shudo 1995 [1132] and the references there.

9.14 The Spectrum of Exterior Differential Forms

From equation 9.3 on page 379 we know that there is a sensible notion of Laplacian for exterior forms of any degree p from $p = 0$ (for functions) to the dimension $p = d = \dim M$. This time the kernel of Δ, i.e. the set of differential forms ω such that $\Delta\omega = 0$, is more subtle than for functions. From theorem 405 on page 665 we know that those forms, called *harmonic*, build up in degree p a real vector space isomorphic through the de Rham isomorphism 34 on page 171 to the cohomology space $H^p(M, \mathbb{R})$, hence of dimension equal to $b^p(M)$, the real p Betti number of M. This for the kernel of Δ. But Δ on

p-forms also has a spectrum, namely the set of its eigenvalues. We explain now how much information can be extracted, with our present state of knowledge, from the knowledge of the spectra for all degrees; we will denote by $\lambda_{p,k}$ the eigenvalues of exterior degree p.

There are today only two outcomes of spectral considerations for exterior forms which have a Riemannian geometry flavor; the Kähler case is richer and was briefly alluded to separately in §§9.10.4 above. First with McKean & Singer 1967 [910] a firework was ignited and in its brightness one could see with far greater clarity. We describe it briefly—a complete reference is Gilkey 1995 [564] (this second edition is very up to date). Roughly speaking what happens is the following. We look back at the asymptotic expansion in theorem 165 on page 393 for $\sum_k \exp(-\lambda_k t)$ with the U_k integrals, which are universal in the curvature tensor (this will mean always including its covariant derivatives). People were concerned that $U_{d/2}$ is not a topological invariant as soon as $d > 2$. Since differential forms also have a canonical Laplacian, we can do the same (it is not too much more expensive and appeared first in Gaffney 1958 [536]) with differential forms and get the pointwise invariants, denoted by $u_{p,k}(x)$ arising in the t^k term in the asymptotic expansions of the corresponding heat kernels. They are still universal in the curvature, but differ in general with various p. Their integrals over M will be denoted by capitals $U_{p,k}$ and the eigenvalues of the p-spectrum by $\{\lambda_{p,k}\}$. Now let us perform the alternate double sum

$$\sum_{p,k}(-1)^p \exp(-\lambda_{p,k} t) \ .$$

Because both of the operators d and d^* commute with the Laplacian Δ, they transform eigenfunctions into eigenfunctions. The Hodge decomposition theorem 406 on page 665 of any form into a harmonic part, a closed part and a coclosed part shows that into this alternate summation everything will disappear except at the harmonic level: there the zero eigenvalue $\lambda_{p,0}$ has a multiplicity equal to the pth Betti number $b^p(M)$. So in the alternating sum of the corresponding asymptotic expansions everything should also disappear for any k except when $k = d/2$. Hence the alternate pointwise sums

$$\sum(-1)^p u_{p,k}(x)\,,$$

when integrated on M and adding after multiplication by t_k, will yield identically the constant

$$\sum(-1)^p b_p = \chi(M).$$

This explains McKean and Singer's dream: a fantastic pointwise cancellation might well take place in the pointwise $u_{p,k}$ functions to yield the forced integrated cancellation. This was indeed proven in Patodi 1971 [1005].

The rebound was taken first in Gilkey 1973 [562] and then in Atiyah, Bott & Patodi 1973 [76, 77]. One studies Patodi's cancellation result, but puts it in successively more general bundles equipped with suitable elliptic operators, including the Dirac operator on spinors and uses Gilkey's results. It then turns out that those structures are plentiful enough to yield all elliptic operators, giving a new proof of the index theorem in §§14.2.3. It is important to use the theory of invariants "à la Gilkey" and the functorial behaviour of indices. The harvest is large: Hirzebruch's signature theorem 417 on page 717 can be obtained this way and of course this new insight yields many results in differential topology. This domain is still blooming; see the two books already mentioned. One point in this philosophy is that "pointwise cancellation" shows that local index theorems can exist. But Riemannian geometry is quite far away. However here comes the second byproduct of the rebound: the η invariant.

The main trick in the founding papers Atiyah, Patodi & Singer 1975–1976 [81, 82, 83] is to obtain the characteristic $\chi(M)$, not as the alternating sum of the zero eigenvalues of the various Laplacians on the exterior forms of a given degree on (M, g), but in one shot as the index of the first order operator $B = d - d^*$ acting on the total set of exterior forms on M (one just has to be careful to put the right signs in front of B). The eigenvalues of Δ are of the form λ^2 where λ is an eigenvalue of B but different signs are possible here. Hence the function

$$\eta(s) = \sum_{\lambda \neq 0} \operatorname{sign}(\lambda) |\lambda|^s$$

makes sense for suitable s. In an strict sense (as usual for this kind of function) $\eta(0)$ is not defined, but with some extra work one can still make sense out of it. It is then called the η invariant of (M, g) and measures the "spectral asymmetry." This invariant is especially interesting for manifolds with boundary. For a $4k$ dimensional manifold M' with a $4k-1$ dimensional boundary M (and provided that locally at the boundary the metric is a product) one can express the signature $\sigma(M')$ by the integral formula

$$\sigma(M') = \int_{M'} L(R) - \eta(M)$$

where L is the universal curvature integrand for the signature of Hirzebruch's theorem 417 on page 717. This invariant has many applications when looking at the subtle problem of the nonexistence of pointwise invariant integration formulas for the "signatures." Besides the original papers we refer the reader to Atiyah, Donnelly & Singer 1983 [78, 79] and Gilkey 1995 [564]. There are also relations with the secondary characteristic classes below, also with $\hat{A}$ genus when spinors are in view. The η invariant for 3-manifolds is applied in deriving the isolation result of Rong 1993 [1064] for the minimal volume in dimension 4 seen in equation 11.6 on page 518. The η invariant is also used in number theory: see Atiyah, Donnelly & Singer 1983 [78]. For η invariants of noncompact manifolds, see Hitchin 1996 [721]; for gluing and the η invariant see Bunke 1995 [278].

Another invariant based on the spectral analysis of differential forms is to be found in Ray & Singer 1971 [1052]. The result is that from the linear combination

$$\sum_{p=0}^{\dim M} (-1)^p p \zeta_p'(0)$$

of the $\zeta_p'(0)$ value of the ζ_p functions associated to the spectrum of the differential forms of all degrees p, one can recover a topological invariant. They conjectured that their invariant should coincide with the topological invariant called the *Reidemeister torsion* and gave some evidence for that. The conjecture was proven independently in Müller 1978 [952] and Cheeger 1979 [332]. The proof is very involved and was one of Cheeger's motivation for the study of the spectrum of certain singular manifolds, see Cheeger 1983 [333].

Do not hope that the knowledge of the differential form spectrum for all p from 0 to the dimension will determine the metric; in Milnor's examples discussed on page 417 all of those spectra coincide. For various questions concerning isospectrality of differential forms, see Gornet 1998 [582]. See Lott 2000 [881] for a subtle study of collapsing and the behaviour of differential forms.