

Hearing the Shape of a Triangle

Daniel Grieser and Svenja Maronna

Introduction

The question, “Can one hear the shape of a drum?” has attracted and inspired many mathematicians since it was asked by Mark Kac in 1966 [16]. The methods used to understand this problem draw on diverse areas: for example, partial differential equations, dynamical systems, group theory, number theory, and probability. In this article we will review some of the history and state of the art of the problem and add a new twist to the story that leads to a curious elementary geometric problem about triangles, which we then solve.

Let us state the problem precisely. For a domain (bounded open set) $\Omega \subset \mathbb{R}^2$ consider the problem of finding a function u on the closure of Ω , vanishing at the boundary $\partial\Omega$, and a number $\lambda \in \mathbb{R}$ satisfying

$$-\Delta u = \lambda u$$

in Ω , where $\Delta := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplace operator. We call λ a *Dirichlet eigenvalue* of Ω if there is a solution $u \neq 0$. Multiplying the equation by u and integrating by parts (i.e., using Green’s identity), one sees that any eigenvalue must be positive; and using basic techniques from PDEs and functional analysis, one can show (see [6]) that the set of eigenvalues is an infinite discrete subset of \mathbb{R} and that the eigenspace, i.e., the set of solutions u , corresponding to each eigenvalue is finite dimensional. Hence one may write the eigenvalues as a sequence $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \cdots \rightarrow \infty$, where each eigenvalue is repeated according to the dimension of its eigenspace.

In this way a sequence of numbers $\lambda_1, \lambda_2, \dots$ is associated with each domain Ω . This begs for

Daniel Grieser is professor of mathematics at Carl von Ossietzky Universität Oldenburg, Germany. His email address is daniel.grieser@uni-oldenburg.de.

Svenja Maronna is a student in the Master of Education program in mathematics at Carl von Ossietzky Universität Oldenburg, Germany. Her email address is svenja.maronna@uni-oldenburg.de.

DOI: <http://dx.doi.org/10.1090/noti1063>

mathematical investigation. Can we calculate the λ_k ? No, except in a very few cases, for example, rectangles, the disk,¹ and certain triangles. Can we say anything interesting about how the eigenvalues depend on the shape of Ω ? Yes. This is the subject of the mathematical discipline called spectral geometry (see [2] for a short introduction and more references, and also [6] and [17]). We can also pose the *inverse problem*: Is the domain Ω determined uniquely by its eigenvalue sequence? Of course, two congruent domains have the same eigenvalue sequence (we say they are *isospectral*), but do any two isospectral domains have to be congruent? This is Kac’s question, mentioned earlier, for the following reason: Think of Ω as a drum, i.e., a membrane that is stretched over a wire frame in the shape of $\partial\Omega$. The membrane can vibrate freely, except that it is fixed at the boundary. When the drum vibrates you will hear a sound, which is composed of tones of various frequencies. These frequencies are the numbers $\gamma\sqrt{\lambda_k}$, where γ is a constant depending on the material and tension of the drum.² So if you know γ , then in this sense you can “hear” the eigenvalues λ_k . Without that knowledge you can still hear the quotients $\sqrt{\lambda_k}/\lambda_1$, which correspond to the musical intervals between the overtones and the fundamental tone of the drum’s sound.

The problem may be easily generalized to higher dimensions and to compact Riemannian manifolds (with or without boundary). Already in Kac’s time it was known that the answer is “no” in the realm of Riemannian manifolds: Milnor had constructed two flat tori of dimension 16 which are isospectral but not isometric (the appropriate notion of congruence for Riemannian manifolds).

¹Here “calculate” is not to be taken literally: the eigenvalues are the squares of the zeroes of the Bessel functions.

²This is an idealized physical model; for real drums the frequencies are slightly different due to nonlinear effects and the influence of the resonance chamber.

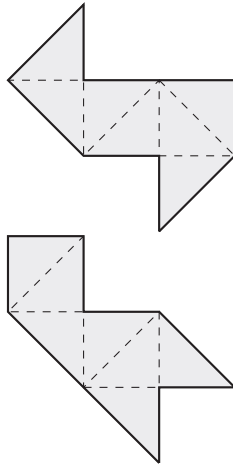


Figure 1. Two drums with the same overtones; see [13]. Isospectrality may be proved by transplantation; see [3], [5], and <http://www.geom.uiuc.edu/docs/research/drums/planar/planar.html>, and <http://www.math.udel.edu/~driscoll/research/drums.html> for pictures of eigenfunctions: For each triangle of the top drum one prescribes Euclidean motions to three triangles of the bottom drum. Then given any eigenfunction on the top drum, one transplants it to the bottom drum by moving the part of the eigenfunction on each top triangle to the bottom according to the given motions and adding (inserting suitable \pm signs) the functions obtained on each bottom triangle. The motions and signs can be chosen in such a way that the resulting function on the bottom drum is smooth across the dashed lines and hence an eigenfunction with the same eigenvalue.

So the question was whether there could also be a counterexample among domains in the plane.

It took twenty-six years to reduce the dimension of counterexamples and make them fit into the plane. The first planar counterexamples were given in 1992 by C. Gordon, D. Webb, and S. Wolpert [13]. Figure 1 shows one of the first examples that was found. Since then many more examples of isospectral Riemannian manifolds, among them continuous families, have been found. Recent surveys on these constructions are [11] and [12].

What Can You Hear?

Rather than focus on counterexamples to Kac's question, let's be positive and ask which geometric properties of a domain or Riemannian manifold *can* be determined from its eigenvalue sequence. The indirect way in which the eigenvalues arise makes this seem a tough question to attack. However, there is a wonderful idea that helps us. It is the idea of transforms and traces. The two most important

instances of this idea are the *heat trace* and the *wave trace*, corresponding to a sort of Laplace and Fourier transform of the eigenvalue sequence. More precisely, the heat trace is the function

$$(1) \quad h(t) = \sum_{k=1}^{\infty} e^{-\lambda_k t}, \quad t > 0,$$

and the wave trace is

$$(2) \quad w(t) = \sum_{k=1}^{\infty} \cos \sqrt{\lambda_k} t, \quad t \in \mathbb{R}.$$

The sum defining $h(t)$ converges for every $t > 0$, and h is a smooth function. The sum defining $w(t)$ never converges, but one can make sense of it in the sense of distributions, so w is a distribution on \mathbb{R} . For example, if $\lambda_k = k^2$ and we sum over $k \in \mathbb{Z}$, then the Poisson summation formula gives

$$(3) \quad w(t) = \sum_{k \in \mathbb{Z}} \cos kt = 2\pi \sum_{l \in \mathbb{Z}} \delta_{2\pi l}(t),$$

where $\delta_{2\pi l}$ is the delta distribution sitting at the point $2\pi l$ (to check this formally, simply calculate the Fourier series of the right-hand side). For this article we will be sloppy about the distinction between functions and distributions.

So why are the functions h , w useful for our problem? The reason is that there is a different way of understanding them, and this yields the desired link to the geometry of Ω . For h this involves the heat equation

$$(\partial_t - \Delta)v(t, x) = 0, \quad t > 0, \quad x \in \Omega,$$

where $\partial_t := \frac{\partial}{\partial t}$. This equation has a unique solution for any initial data $v(0, x) = f(x)$ if we impose the boundary condition that $v(t, x) = 0$ for all $t > 0$ and $x \in \partial\Omega$. By separation of variables we obtain $v(t, x) = \sum_{k=1}^{\infty} a_k e^{-\lambda_k t} u_k(x)$, where the u_k form an orthonormal basis of $L^2(\Omega)$ of real-valued eigenfunctions corresponding to the λ_k and $a_k = \int_{\Omega} f(y) u_k(y) dy$. In other words,

$$v(t, x) = \int_{\Omega} H(t, x, y) f(y) dy,$$

where $H(t, x, y) = \sum_{k=1}^{\infty} e^{-\lambda_k t} u_k(x) u_k(y)$. The function $H : (0, \infty) \times \Omega \times \Omega \rightarrow \mathbb{R}$ is called the heat kernel of Ω , and since the u_k are normalized in L^2 , one sees that

$$(4) \quad h(t) = \int_{\Omega} H(t, y, y) dy.$$

This is the trace of the operator $e^{t\Delta} : f \mapsto v(t, \cdot)$; hence the name heat trace for h . Now we observe that for any fixed $y \in \Omega$, the function $(t, x) \mapsto H(t, x, y)$ is the solution of the heat equation with initial data $f(x) = \delta_y(x)$; that is, it describes the distribution of heat after time t , when initially there is a single hot spot at y . Although heat spreads at infinite velocity (that is, $H(t, x, y) > 0$ for all x no matter how small $t > 0$), the value of

$H(t, x, y)$ at $x = y$ for t close to zero will be mostly influenced by the geometry of Ω near the point y . A precise analysis of the heat equation shows that for a Riemannian surface Ω without boundary we have the asymptotic behavior

$$H(t, y, y) \sim t^{-1} \sum_{j=0}^{\infty} a_j(y) t^j \quad \text{as } t \rightarrow 0,$$

where each $a_j(y)$ is a universal polynomial in derivatives of the Gauss curvature $K(y)$ of Ω at y . For example, $a_0(y) = \frac{1}{4\pi}$, $a_1(y) = \frac{1}{12\pi} K(y)$. If Ω has a boundary, then its influence is felt only when the distance of y to the boundary is of order at most \sqrt{t} , and in the integral (4) this contributes extra terms involving the curvature of the boundary and terms involving the powers $t^{-1/2+j}$. In the case of planar domains with polygonal boundary there is no curvature, but the corners give a contribution, and this leads to the formula

$$h(t) = a_0 t^{-1} + a_{1/2} t^{-\frac{1}{2}} + a_1 + O(e^{-\frac{c}{t}}) \quad \text{as } t \rightarrow 0$$

for some constant $c > 0$ where

$$a_0 = \frac{A}{4\pi}, \quad a_{1/2} = -\frac{P}{8\sqrt{\pi}}, \quad a_1 = \frac{1}{24} \sum_i \left(\frac{\pi}{\alpha_i} - \frac{\alpha_i}{\pi} \right),$$

where A is the area, P is the perimeter, and the α_i are the interior angles of the polygon. This formula was first mentioned in [18]; the first published proof was given in [20]. In the case of the triangle we have $\sum_i \alpha_i = \pi$, so $a_1 = \frac{\pi}{24} \sum_{i=1}^3 \frac{1}{\alpha_i} - \frac{1}{24}$. Therefore, if we know all the λ_k , then we know the function $h(t)$ and hence the coefficients $a_0, a_{1/2}, a_1$, hence the area, the perimeter, and the sum of the reciprocals of the angles of the triangle. So we can hear these quantities. This motivates the following theorem:

Theorem 1. *A triangle is determined uniquely up to congruence by its area A , its perimeter P , and the sum R of the reciprocals of its angles.*

Corollary 1. *One can hear the shape of a triangle among all triangles.*

That is, if we know that Ω is a triangle, then the spectrum of Ω determines which triangle it is.

Corollary 1 was first proved by C. Durso; see [10]. Durso used in her proof the *wave kernel*, which is a much more powerful tool in spectral geometry than the heat kernel—at the cost of harder technical issues in its analysis. The main idea, however, is beautiful and easy to understand. After telling this remarkable story we will return to the proof of Theorem 1. This yields a new proof of Corollary 1, which avoids the use of the wave kernel.

The Wave Kernel

The wave trace $w(t)$ can be obtained in the same way as the heat trace but starting with the wave equation

$$(\partial_t^2 - \Delta)u(t, x) = 0, \quad t \in \mathbb{R}, \quad x \in \Omega$$

with initial data $u(0, x) = f(x)$, $(\partial_t u)(0, x) = 0$ and boundary values $u(t, x) = 0$ for all $t \in \mathbb{R}$, $x \in \partial\Omega$. This equation has a unique solution for each f , and it describes vibrations of Ω , or propagation of waves on Ω , with initial shape f . Again the solution can be written in the form $u(t, x) = \int_{\Omega} W(t, x, y) f(y) dy$, where W is now a distribution, and

$$(5) \quad w(t) = \int_{\Omega} W(t, y, y) dy,$$

the trace of the operator $\cos t\sqrt{-\Delta} : f \mapsto u(t, \cdot)$.

How can we learn anything about the function $W(t, x, y)$? It may help to think of Ω as a lake. At time $t = 0$ we drop a stone into the lake at the place y —this corresponds to the initial condition $f(x) = \delta_y(x)$ —and observe the resulting waves. In a linear water wave model, $x \mapsto W(t, x, y)$ is the lake's surface at time t . Everyone knows what happens: A circular wave front centered at y will form, its radius increasing linearly with t . When it reaches the boundary of the lake, it will be reflected. In our simple model there is no loss of energy, and the wave will move on forever. The precise shape of the wave front can be described as follows: Starting at y walk in any direction at speed 1. Always walk straight, except when you hit the boundary. In this case reflect off the boundary according to the law “angle of reflection = angle of incidence”. The wave front at time t is the set of points that you can reach in this way when walking for time t .

This helps us understand the integrand $W(t, y, y)$ in (5): It will be large only for those times t for which the wave front returns to y after time t , i.e., for which there is a path³ from y to y of length t . A more careful analysis then shows that when integrated over y , many of these “large” contributions cancel with neighboring paths due to oscillation. Only contributions from *closed paths*—that is, those which return to y in the same direction in which they started—are not cancelled in this way. To summarize, we arrive at the conclusion that $w(t)$ is large only for $|t| \in T$, where

$$(6) \quad T = \{\text{lengths of closed paths in } \Omega\} \cup \{0\}.$$

³Here a “path” is a succession of straight lines, or geodesics, obeying the law of reflection when hitting the boundary of Ω . If Ω has corners, as for a triangle, then a path running into a corner can leave the corner in any direction.

Here we also need to count the “instant” path of length zero.

The precise mathematical statement of this involves the notion of singular support of a distribution, i.e., the set where the distribution is not given by a smooth function. The wave front at time t is precisely the singular support of the distribution $x \mapsto W(t, x, y)$. The result above translates into the statement

$$\text{singsupp } w \subset \text{clos}(T)$$

(see [14] for the rather technical proof and [7] for a survey of the history of this theorem). If there is precisely one path for each $t \in T$ (up to reversal of direction), then the singular support is equal to $\text{clos}(T)$. It is conjectured that this equality is always true, but this is an open problem. In the example where Ω is a circle of length 2π (or alternatively, the interval $[0, 2\pi]$ where we impose periodic boundary conditions), this can be seen explicitly: The eigenvalues are k^2 , $k \in \mathbb{Z}$, and the wave trace is (3); the numbers $2\pi l$ are precisely the lengths of closed paths in Ω .

So we see that essentially we can hear the set T of lengths of closed paths on Ω . This analysis can be refined substantially by analyzing the *kind* of singularities that the wave trace w has at points of T . It turns out that the singularity at $t = 0$ carries the same information as the full asymptotic expansion of the heat kernel at $t = 0$. The other singularities yield additional information, and using this, one can prove that one can hear generic convex domains with analytic boundary and certain symmetries; see [21], [15]. As a final remark on this, we would like to mention the remarkable recent work [8], in which for the first time the behavior of w at cluster points of T was analyzed in the special case of a disk. A recent survey on inverse spectral results obtained using trace formulae and related methods is [9].

To end this section, let us explain Durso’s proof that one can hear the shape of a triangle.

It is classical that in an acute triangle Ω there is a unique shortest closed path, and it is given by the triangle formed by the base points of the three altitudes of Ω ; see Figure 2. Therefore, one can hear the length of this path. Durso shows that in the case of an obtuse or right-angled triangle, the shortest closed path is the shortest altitude, traversed up and down, and that the wave trace w is singular at l_0 , the length of this path (this is the hard analytical part of the proof). So one can hear l_0 . Then she shows by an elementary geometric argument that any triangle is determined uniquely by area, perimeter, and the length of its shortest closed path.

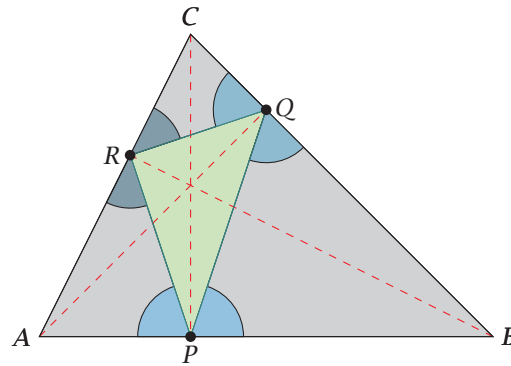


Figure 2. In an acute triangle $\triangle ABC$, how should one choose points P, Q, R on each side so that the triangle $\triangle PQR$ has minimal perimeter? The answer: Choose the base points of the altitudes of $\triangle ABC$. The resulting triangle is called the *Fagnano triangle*. There is a clever proof of this fact using reflections of the side AB across the sides AC and BC . Also, by a standard variational argument it follows that the circumference of the Fagnano triangle obeys the law of reflection at each of the points P, Q, R , so an ideal billiard ball on the billiard table $\triangle ABC$ will run forever along this line.

A Theorem about Triangles

We now return to the proof of Theorem 1. This is a rather peculiar statement: Have you ever heard of *reciprocals of angles*? There does not seem to be any geometric meaning to this, and our proof draws on classical analysis rather than geometry. Note that in contrast to Durso’s proof, our proof uses only tools known in the 1960s.

First, let us remark that it is quite clear that the three quantities A, P, R determine a triangle *up to finitely many choices*. This follows easily from the fact that the space of triangles \mathcal{T} is three dimensional (for example, it may be parameterized by the side lengths), that the functions A, P, R on \mathcal{T} are analytic and independent (in the sense that none of them can be expressed as a function of the other two; independence in a stronger sense will follow from Lemma 1 below), and that R is a proper function on $\mathcal{T}/\mathbb{R}_{>0}$, the quotient of \mathcal{T} by scalings: R tends to infinity when one of the angles tends to zero, which is the only way to leave all compact subsets of $\mathcal{T}/\mathbb{R}_{>0}$. However, just as prescribing the lengths of two sides and an angle not enclosed by them determines a triangle only up to two choices, it is not obvious why there should be only one triangle with any given A, P, R . Of course it is not hard to check this numerically, but it is far from obvious how to give a rigorous proof.

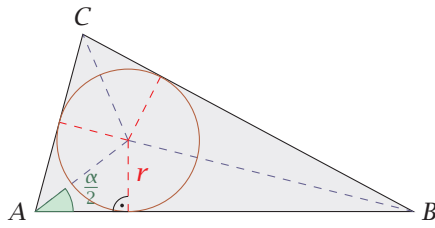


Figure 3. Proof of (7): $A = r\frac{P}{2}$ and $\cot \frac{\alpha}{2} + \cot \frac{\beta}{2} + \cot \frac{\gamma}{2} = \frac{1}{r}\frac{P}{2}$.

We denote the angles of the triangle by α, β, γ . We use the following formula from triangle geometry;⁴ see Figure 3:

$$(7) \quad \frac{P^2}{4A} = \cot \frac{\alpha}{2} + \cot \frac{\beta}{2} + \cot \frac{\gamma}{2}.$$

This allows us to work exclusively with angles. We will prove:

Proposition 1. *A triple (α, β, γ) of positive real numbers satisfying $\alpha + \beta + \gamma = \pi$ is uniquely determined, up to ordering, by the values of*

$$(8) \quad f(\alpha, \beta, \gamma) = \cot \frac{\alpha}{2} + \cot \frac{\beta}{2} + \cot \frac{\gamma}{2},$$

$$(9) \quad g(\alpha, \beta, \gamma) = \frac{1}{\alpha} + \frac{1}{\beta} + \frac{1}{\gamma}.$$

Theorem (1) follows directly from this: If area A , perimeter P and $R = g$ are given, then the angles are determined by equation (7) and the proposition, so the triangle is determined up to dilation. Then the given area fixes the dilation factor.

So it remains to prove the proposition. One way to proceed would be to eliminate one of the variables, say α , using the relation $\alpha = \pi - \beta - \gamma$; then eliminate another variable, say β , from the given value of g by solving a quadratic equation; then plug the expressions for α and β into f and investigate the resulting equation for γ . But this is horrible! Even if it works, it is ugly mathematics. If nothing else, the beautiful symmetry present in the statement of the proposition is lost.

Symmetry is a treasure. One should keep it and use it as long as possible. This is what we shall do.

Proof of Proposition 1. Let

$$D = \{(\alpha, \beta, \gamma) : \alpha, \beta, \gamma > 0, \\ \alpha + \beta + \gamma = \pi\} \subset \mathbb{R}_{>0}^3,$$

where $\mathbb{R}_{>0} = (0, \infty)$. We think of points of D as “marked triangles up to dilation”, where “marked” means that we have named the angles in a certain

⁴We are grateful to Richard Laugesen for pointing out this identity. Amazingly, both sides are also equal to the product $\cot \frac{\alpha}{2} \cot \frac{\beta}{2} \cot \frac{\gamma}{2}$. It's a nice little exercise in addition theorems to prove this.

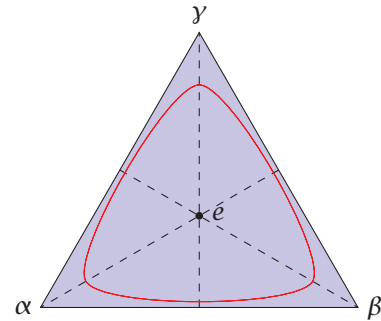
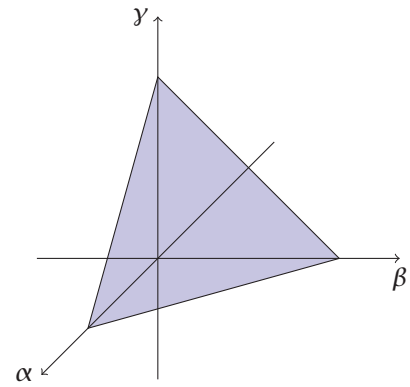


Figure 4. The space of angles of a triangle, and a level line of g .

order. The set D is (the interior of) a triangle itself—the triangle cut out of the plane $\alpha + \beta + \gamma = \pi$ by the positive octant; see Figure 4. Points on the dashed lines correspond to isosceles triangles; the center e corresponds to the equilateral triangle. Let us call a point that does not lie on a dashed line a *nonisosceles point*. The nonisosceles points form six connected subsets, which we call *chambers*. The dashed lines are also lines of symmetry: If we pick a nonisosceles point and reflect it step by step across all dashed lines, we obtain six points, one in each chamber. These six points correspond to the same triangle, with angles named in different orders. Each chamber corresponds to one ordering of the angles, for example, the lower left chamber to the ordering $\alpha > \beta > \gamma$ or $\alpha \geq \beta \geq \gamma$ when we include its dashed boundary parts.

The idea of the proof is to show that the level sets of the function g are convex curves (see Figure 4) and that f is strictly monotone along the part of any one of these curves lying in one chamber.

Lemma 1. *Consider the functions f, g and $h(\alpha, \beta, \gamma) = \alpha + \beta + \gamma$ on the positive octant $\mathbb{R}_{>0}^3$.*

- The function g is strictly convex on $\mathbb{R}_{>0}^3$.*
- The gradients $\nabla f, \nabla g, \nabla h$ are linearly independent at all nonisosceles points of D .*

Let us finish the proof of Proposition 1 and then return to prove Lemma 1. The strict convexity of

g implies that the sublevel set $G_{\leq s} = \{p \in \mathbb{R}_{>0}^3 : g(p) \leq s\}$ is strictly convex for any $s > 0$, with boundary the level surface $G_s = \{p \in \mathbb{R}_{>0}^3 : g(p) = s\}$. Furthermore, these sets are symmetric under all permutations of the coordinates. These properties then also hold for the intersections of the sublevel and level sets with the plane $\alpha + \beta + \gamma = \pi$. Since $g(p) \rightarrow \infty$ when p approaches the boundary of D (i.e., when at least one of the angles tends to zero), it follows that the sets $G_s \cap D$ are either closed curves in the interior of D which encircle the point e or are the point e or are empty. Since the equilateral triangle has $g(e) = \frac{9}{\pi}$, the first case corresponds to $s > \frac{9}{\pi}$.

In particular, we see that the point e is already determined by the value of g alone.⁵

Now consider any level curve $G_s \cap D$ with $s > \frac{9}{\pi}$. Consider the arc of the curve running inside one chamber, with endpoints p, q corresponding to isosceles triangles. Our proof will be complete if we can show that f is strictly monotone along this part of the curve.

Suppose f is not strictly monotone. Then there would be a point r on this arc, different from p and q , where f is stationary; that is, the derivative of f along the arc vanishes at r . By the Lagrange multiplier theorem this would mean that $\nabla f(r)$ is a linear combination of $\nabla g(r)$ and $\nabla h(r)$. But this would be a contradiction to part b of Lemma 1. This completes the proof of the proposition. \square

Proof of Lemma 1. a) The Hessian (matrix of second derivatives) of g is the diagonal matrix with entries $\frac{2}{\alpha^3}, \frac{2}{\beta^3}, \frac{2}{\gamma^3}$ on the diagonal. This is clearly positive definite for all $(\alpha, \beta, \gamma) \in \mathbb{R}_{>0}^3$, and this implies that g is strictly convex.

b) We have

$$\nabla f = -\frac{1}{2} \begin{pmatrix} \frac{1}{\sin^2 \frac{\alpha}{2}} \\ \frac{1}{\sin^2 \frac{\beta}{2}} \\ \frac{1}{\sin^2 \frac{\gamma}{2}} \end{pmatrix}, \quad \nabla g = - \begin{pmatrix} \frac{1}{\alpha^2} \\ \frac{1}{\beta^2} \\ \frac{1}{\gamma^2} \end{pmatrix}, \quad \nabla h = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Suppose there was a nonisosceles point (α, β, γ) (i.e., the numbers α, β, γ are pairwise different) and numbers R, S, T , not all zero, with $R\nabla f + S\nabla g + T\nabla h = 0$. Note that R, S cannot both be zero. Then the non-constant function

$$F(y) = -\frac{R}{2} \frac{1}{\sin^2 \frac{y}{2}} - S \frac{1}{y^2} + T$$

would have three different zeroes in the interval $(0, \pi)$, namely, $y = \alpha, y = \beta$ and $y = \gamma$. In order to show that this cannot happen, we prove that the function F is strictly monotone or strictly concave

⁵This can also be seen from the arithmetic-harmonic mean inequality $3 \left(\frac{1}{\alpha} + \frac{1}{\beta} + \frac{1}{\gamma} \right)^{-1} \leq \frac{\alpha + \beta + \gamma}{3}$ with equality iff $\alpha = \beta = \gamma$.

or convex on this interval, depending on the values R, S, T . We use the following fact, proved below:

Lemma 2. *The function $G(x) = \frac{1}{\sin^2 x} - \frac{1}{x^2}$ is strictly increasing and strictly convex on the interval $(0, \pi)$.*

This lemma implies that the function $G_C(x) = \frac{1}{\sin^2 x} - \frac{C}{x^2}$ is, on the interval $(0, \pi)$, strictly increasing for $C \geq 1$ and strictly convex for $C \leq 1$, since $G_C(x) = G(x) + \frac{1-C}{x^2}$ and the function $\frac{1-C}{x^2}$ is increasing for $C > 1$ and convex for $C < 1$. Now clearly we can write $F(y)$ as a non-zero constant multiple of $G_C(\frac{y}{2})$, for some C , plus a constant. Therefore, F cannot have three different zeroes on the interval $(0, 2\pi)$, and Lemma 1b) follows. \square

Proof of Lemma 2. First note that this is non-trivial: It is easy to check that both $\frac{1}{\sin^2 x}$ and $\frac{1}{x^2}$ have positive second derivative whenever they are defined, hence are convex, but it is not clear why their difference should be convex. However, things become very transparent when we use the series representation (partial fraction expansion)

$$\frac{1}{\sin^2 x} = \sum_{k=-\infty}^{\infty} \frac{1}{(x - k\pi)^2},$$

which follows from the well-known partial fraction expansion of the cotangent by differentiation. This yields $G(x) = \sum_{k \neq 0} \frac{1}{(x - k\pi)^2}$. Now every summand $\frac{1}{(x - k\pi)^2}$ is strictly convex on $(0, \pi)$ since the function $\frac{1}{x^2}$ is strictly convex on both half lines $x < 0$ and $x > 0$, so G is strictly convex. Furthermore, the series shows that G is regular at $x = 0$, and it is also even, so $G'(0) = 0$. Combined with strict convexity, this implies that G is strictly increasing on the interval $(0, \pi)$, which was to be shown. \square

Further Remarks

Let us take another look at the proof of Theorem 1 from a slightly different perspective. Proposition 1, which implies Theorem 1 by elementary triangle formulas, may be restated as saying that the map $\Phi = (f, g) : D \rightarrow \mathbb{R}^2$ is injective on the closure in D of each chamber. The proof of injectivity has two ingredients: First, Lemma 1b), which may be restated as saying that the differential of the map Φ is invertible in the chamber and hence, by the inverse mapping theorem, that Φ is locally injective everywhere; that is, every point of the chamber has a neighborhood on which Φ is injective. Second, the convexity of Lemma 1a) allows us to infer global injectivity from this local statement. Finally, the analytic core of the whole argument is Lemma 2, which is used in the proof of Lemma 1b). We now take another look at this.

A Different Proof of Lemma 2

While the given proof using the partial fraction representation is very elegant, you might wonder if there is a more pedestrian way to prove convexity of G . Indeed there is. Here is a sketch. It was our first proof of this result, and it is the result of the bachelor's thesis of the second author. A short calculation gives $\frac{1}{2}G''(x) = \frac{3}{\sin^4 x} - \frac{2}{\sin^2 x} - \frac{3}{x^4}$. We need to show that this is positive (here and in the sequel we always assume $x > 0$). This is equivalent to the inequality

$$(10) \quad 3 \sin^4 x + 2x^4 \sin^2 x < 3x^4.$$

How can one prove an inequality involving trigonometric functions and polynomials? Maybe your first idea is to use the well-known inequality $\sin x < x$ to get rid of the sines. But clearly this does not help, since $3x^4 + 2x^4 \cdot x^2 > 3x^4$. How can we do better?

Recall where the inequality $\sin x < x$ comes from: x is the first term in the Taylor series of $\sin x$; the next term is negative. Of course this is not a proof, but it is the core idea, which can be turned into a proof as follows: The function $f(x) = x - \sin x$ vanishes at $x = 0$ and has derivative $f'(x) = 1 - \cos x$, which is always nonnegative, and is positive for small positive x . Thus, $x - \sin x > 0$ for all positive x follows by integration: $f(x) = \int_0^x f'(t) dt > 0$.

So in order to prove (10) we can try to use a better estimate for $\sin x$ by using more terms from its Taylor series. We have the estimate

$$(11) \quad \text{associated with } \sin x < x - \frac{x^3}{6} + \frac{x^5}{120}.$$

This can be proved in the same way as $\sin x < x$: The function $f(x) = x - \frac{x^3}{6} + \frac{x^5}{120} - \sin x$ satisfies $f(0) = f'(0) = f''(0) = f'''(0) = f^{(4)}(0) = 0$ and $f^{(5)}(x) = 1 - \cos x \geq 0$, and > 0 for small positive x . Integrating, we obtain $f^{(4)}(x) = \int_0^x f^{(5)}(t) dt > 0$; then integrating again we get $f'''(x) > 0$, and so forth until we obtain $f(x) > 0$ for all $x > 0$.⁶

We now plug (11) into the left-hand side of (10). A rather tedious calculation shows that the result,

⁶Instead we could have used Taylor's formula with remainder for the function $g(x) = \sin x$:

$$g(x) = \sum_{k=0}^4 \frac{x^k}{k!} + \frac{1}{4!} \int_0^t (x-t)^4 g^{(5)}(t) dt,$$

which using $g^{(5)}(t) = \cos t \leq 1$ (and < 1 for small positive t) and $\int_0^t (x-t)^4 dt = \frac{1}{5}x^5$ yields the same result. Yet another proof uses Leibniz's criterion for the Taylor series $x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$ of $\sin x$, which is alternating. The terms after the fifth power are monotonically decreasing in absolute value if $\frac{x^{2n+1}}{(2n+1)!} < \frac{x^{2n-1}}{(2n-1)!}$ for $n \geq 4$, which is equivalent to $x^2 < 2n(2n+1)$, hence true for $x < \sqrt{72}$. Since the first omitted term after $\frac{x^5}{5!}$ is negative, we get that the sum of the

which starts as $3x^4 - \frac{1}{15}x^8 + \dots$, is less than $3x^4$ for $x < 4$. The main point is that the second term is negative. Among the higher terms some are positive, but they can easily be estimated against the negative ones.

A Few Open Problems

The way in which the Dirichlet eigenvalues determine the triangle is somewhat indirect: First one constructs the heat kernel h (see (1)), and then one considers the coefficients in its asymptotic expansion to prove the result. In particular, one needs to know (asymptotic information on) *all* the eigenvalues for this. It is natural to ask whether a finite number of eigenvalues, ideally only three, suffice to determine the triangle.

Problem. Do the first three Dirichlet eigenvalues, $\lambda_1, \lambda_2, \lambda_3$, determine a triangle?

Numerical evidence was provided in [1] that this is true—but that the corresponding statement for $\lambda_1, \lambda_2, \lambda_4$ is false. However, no proof of this is known. As a partial result in this direction, it is proved in [4] that for each $\varepsilon > 0$ there is a number N so that $\lambda_1, \dots, \lambda_N$ determine a triangle uniquely among all triangles whose angles are all greater than or equal to ε .

Problem. Is there a closed path (not hitting a corner) on every triangle?

For acute triangles the answer is “yes”; see Figure 2. The problem is open for general obtuse triangles.

Problem. Does the second Neumann eigenfunction on an obtuse triangle have its extrema on the boundary?

This is conjectured to be true and is a special case of the *hot spots conjecture*. See the recent discussion on polymath [19].

Let us mention two other open questions on the inverse spectral problem.

Problems. Can one hear the shape of a convex polygon? Can one hear the shape of a domain $\Omega \subset \mathbb{R}^2$ with smooth boundary?

We emphasize that the answer is “no” when convexity or smoothness is not required: All known counterexamples to “Can one hear the shape of drum?” are nonconvex polygons; cf. Figure 1.

series, which is $\sin x$, is less than $x - \frac{x^3}{3!} + \frac{x^5}{5!}$, at least for $x < \sqrt{72}$. Since $\sqrt{72} > \pi$, this is enough for our purpose.

References

- [1] PEDRO R. S. ANTUNES and PEDRO FREITAS, On the inverse spectral problem for Euclidean triangles, *Proc. R. Soc. Lond., Ser. A, Math. Phys. Eng. Sci.* **467**(2130):1546–1562, 2011.
- [2] RAFAEL D. BENGURIA, Dirichlet eigenvalue, in: *Encyclopedia of Mathematics*. http://www.encyclopediaofmath.org/index.php?title=Dirichlet_eigenvalue&oldid=14188. http://www.encyclopediaofmath.org/index.php/Dirichlet_eigenvalue.
- [3] PETER BUSER, JOHN CONWAY, PETER DOYLE, and KLAUS-DIETER SEMMLER, Some planar isospectral domains, *Int. Math. Res. Not.* **1994**(9):391–400, 1994.
- [4] PEI-KUN CHANG and DENNIS DETURCK, On hearing the shape of a triangle, *Proc. Amer. Math. Soc.* **105**(4):1033–1038, 1989.
- [5] S. J. CHAPMAN, Drums that sound the same, *Amer. Math. Monthly* **102**(2):124–138, 1995.
- [6] ISAAC CHAVEL, *Eigenvalues in Riemannian Geometry*, volume 115 of Pure and Applied Mathematics, Academic Press Inc., Orlando, FL, 1984. Including a chapter by Burton Randol, with an appendix by Jozef Dodziuk.
- [7] YVES COLIN DE VERDIÈRE, Spectrum of the Laplace operator and periodic geodesics: Thirty years after, *Ann. Inst. Fourier (Grenoble)* **57**(7):2429–2463, 2007. Festival Yves Colin de Verdière.
- [8] YVES COLIN DE VERDIÈRE, VICTOR GUILLEMIN, and DAVID JERISON, Singularities of the wave trace near cluster points of the length spectrum, preprint. <http://arxiv.org/abs/1101.0099>.
- [9] KIRIL DATCHEV and HAMID HEZARI, Inverse problems in spectral geometry. In *Inverse Problems and Applications. Inside Out II*, volume 60 of Mathematical Sciences Research Institute Publications, Cambridge Univ. Press, 2012, pp. 455–486.
- [10] CATHERINE DURSO, *Solution of the inverse spectral problem for triangles*, Ph.D. thesis, Massachusetts Institute of Technology, 1990.
- [11] OLIVIER GIRAUD and KOEN THAS, Hearing shapes of drums: Mathematical and physical aspects of isospectrality, *Reviews of Modern Physics* **82**(3):2213–2255, 2010.
- [12] CAROLYN GORDON, PETER PERRY, and DOROTHEE SCHUETH, Isospectral and isoscattering manifolds: A survey of techniques and examples. In *Geometry, Spectral Theory, Groups, and Dynamics*, volume 387 of Contemp. Math., Amer. Math. Soc., Providence, RI, 2005, pp. 157–179. Note: The claims of reference [59] in this paper have been refuted; see Arias-Marco, Schueth, *Geom. Funct. Anal.* **22** (2012), 1–21.
- [13] CAROLYN GORDON, DAVID L. WEBB, and SCOTT WOLPERT, One cannot hear the shape of a drum, *Bull. Amer. Math. Soc., New Ser.* **27**(1):134–138, 1992.
- [14] VICTOR GUILLEMIN and RICHARD MELROSE, The Poisson summation formula for manifolds with boundary, *Adv. in Math.* **32**(3):204–232, 1979.
- [15] HAMID HEZARI and STEVE ZELDITCH, Inverse spectral problem for analytic $(\mathbb{Z}/2\mathbb{Z})^n$ -symmetric domains in \mathbb{R}^n , *Geom. Funct. Anal.* **20**(1):160–191, 2010.
- [16] M. KAC, Can one hear the shape of a drum? Part II, *Amer. Math. Monthly*, **73**:1–23, April 1966.
- [17] R. S. LAUGESSEN, Tight frames and rotations: Sharp bounds on eigenvalues of the Laplacian. In *Proceedings of the AMSI International Conference on Harmonic Analysis and Applications (Macquarie University, 2011)*, 2012.
- [18] H. P. MCKEAN JR. and I. M. SINGER, Curvature and the eigenvalues of the Laplacian, *J. Differential Geometry* **1**(1):43–69, 1967.
- [19] Polymath project on hot spots conjecture, <http://polymathprojects.org/2012/09/10/polymath7-research-threads-4-the-hot-spots-conjecture/>.
- [20] M. VAN DEN BERG and S. Srisatkunarahaj, Heat equation for a region in \mathbb{R}^2 with a polygonal boundary, *J. London Math. Soc. (2)* **37**(1):119–127, 1988.
- [21] STEVE ZELDITCH, Inverse spectral problem for analytic domains. II. \mathbb{Z}_2 -symmetric domains, *Ann. of Math. (2)* **170**(1):205–269, 2009.