Isospectral hyperbolic surfaces have matching geodesics

Peter G. Doyle* Juan Pablo Rossetti†

Version 5.0 dated 28 March 2007

Abstract

We show that if two closed hyperbolic surfaces (not necessarily orientable or even connected) have the same Laplace spectrum, then for every length they have the same number of orientation-preserving geodesics and the same number of orientation-reversing geodesics. Restricted to orientable surfaces, this result reduces to Huber's theorem of 1959. Appropriately generalized, it extends to hyperbolic 2-orbifolds (possibly disconnected). We give examples showing that it fails for disconnected flat 2-orbifolds.

AMS Classification numbers Primary: 58J53 Secondary: 11F72, 20F67, 53C22

Keywords: isospectrality, closed geodesic, holonomy, almost conjugate, hyperbolic surface, Selberg trace formula, flat orbifold, prime geodesic theorem.

1 Introduction

We say that two hyperbolic surfaces (assumed closed but not necessarily orientable or even connected) are *almost conjugate* if their closed geodesics match, in the sense that for every length l they have the same number of

*Dartmouth College.

†FaMAF-CIEM, Univ. Nac. Córdoba. Partially supported by DFG Sonderforschungsbereich 647, Humboldt University, Berlin.

orientation-preserving geodesics and the same number of orientation-reversing geodesics. Looking beyond dimension 2, we say that two hyperbolic d-manifolds (assumed closed, but not necessarily orientable or even connected) are almost conjugate if their geodesics match with respect to length and 'twist'. The twist of a geodesic (also called its 'holonomy') is measured by the conjugacy class in $O(d-1)$ of the action of parallel translation around the geodesic. To say that geodesics have matching length and twist amounts to saying that the corresponding deck transformations are conjugate under the action of the full isometry group of hyperbolic d-space.

This usage of the term 'almost conjugate' accords with Sunada's well-known definition [13] that subgroups of a finite group are almost conjugate if they meet every conjugacy class equally. Here, two connected hyperbolic surfaces are almost conjugate just if their deck groups intersect equally every conjugacy class of the full group of isometries of the hyperbolic plane. Strictly speaking, it is the deck groups that are almost conjugate, but it is convenient to apply the term abusively to the surfaces themselves—the more so since we want to consider disconnected surfaces.

While we haven't specifically required it in the definition, the matching of geodesics between almost conjugate hyperbolic manifolds, whether surfaces or manifolds of higher dimension, can and should be taken to respect the imprimitivity index of the geodesics as well as their length and twist.

Please note that here and throughout, by 'geodesics' we mean oriented closed geodesics. Because our geodesics carry a designated orientation, the number of geodesics of length l will always be even, with each unoriented geodesic being counted twice, once for each orientation. So when we say, for example, that the number of geodesics of length at most l is asymptotically $\frac{e^l}{l}$, we're talking about oriented geodesics; the asymptotic number of unoriented geodesics would be $\frac{e^l}{2l}$.

According to the Selberg trace formula, almost conjugate hyperbolic manifolds are *isospectral*: They have the same Laplace eigenvalues with the same multiplicity. (Cf. Randol's chapter in [3]; Gangolli [8]; Bérard-Bergery [1].) Such manifolds can be constructed using the well-known method of Sunada [13]. Sunada's method is very flexible and powerful, and works in contexts that go well beyond the kind of locally homogeneous and isotropic case we encounter here in discussing hyperbolic manifolds. It readily yields examples of hyperbolic manifolds of dimension 2 and higher that are almost conjugate and hence isotropic: For an exposition, see Buser [2]. The examples Buser describes are all orientable, but with trivial modifications the constructions can be made to yield isospectral pairs of non-orientable surfaces. Since we don't know of any handy reference for this, we'll elaborate on this in Section 4.

We've said that according to the Selberg trace formula, almost conjugate manifolds are isospectral. All the examples coming from Sunada's method are automatically almost conjugate. This raises the obvious question:

Question 1 *If two hyperbolic manifolds are isospectral, must they be almost conjugate?*

In the case of orientable surfaces, where there is no twisting to contend with, the answer is yes: This is 'Huber's Theorem' [9, 10, 11], dating back to 1959. Nowadays we recognize this as a direct consequence of the Selberg trace formula, which shows how to 'read off' from the spectrum the lengths of the geodesics (and, it should go without saying, the associated multiplicities).

The purpose of this paper is to prove that the answer is still yes for surfaces, even without the orientability assumption.

Theorem 1 *If two hyperbolic surfaces (not necessarily orientable or even connected) are isospectral, then they are almost conjugate.*

Here, in contrast to the orientable case, we cannot use the Selberg trace formula to read off the lengths of geodesics directly from the spectrum. In fact, as we will see, there are 'scenarios' for constructing counter-examples consistent with the Selberg trace formula. But the Prime Geodesic Theorem comes to our rescue, because we can show that any scenario for constructing a counter-example requires the frequent participation of a large number of geodesics of length exactly l (specifically, at least a constant times $\frac{e^l}{l}$), and having this many geodesics of the same length is forbidden by the Prime Geodesic Theorem.

In higher dimensions Question 1 remains open, even in the case of connected orientable manifolds. The issues at stake in higher dimensions are well illustrated in the proof of Theorem 1—so you might be interested in this theorem even if you don't see why anyone would care about non-orientable surfaces.

To see that the possible existence of isospectral hyperbolic manifolds that are not almost conjugate is a question that must be taken seriously, we note that in dimension $d \geq 3$, there exist isospectral flat manifolds that are not almost conjugate. Our favorite example of this is the 3-manifold pair 'Tetra and Didi' [5]. In the flat case, some care is needed in defining almost conjugacy, because while in a hyperbolic manifold geodesics come only in isolation, in a flat manifold geodesics come in parallel families of varying dimension. So in the flat case, matching geodesics between manifolds involves measuring, rather than just counting. But Tetra and Didi will fail to be almost conjugate according to any definition.

Note. For further insight into the possible existence of isospectral spaces that are not almost conjugate, it is natural to expand the class of spaces we're considering from manifolds to orbifolds. (Cf. Dryden [6], Dryden and Strohmaier [7].) Of course we need to extend the definition of 'almost conjugacy' appropriately. We don't propose to discuss orbifolds in detail here, but for the benefit of those familiar with orbifolds, we have appended some comments in Section 5 below. Briefly, what we find is this: Theorem 1 extends to rule out examples among hyperbolic 2-orbifolds. However, there are examples of isospectral flat 2-orbifolds (necessarily disconnected) that are not almost conjugate. And we still don't know what happens in the hyperbolic case in dimension ≥ 3.

2 Outline

Let M be a hyperbolic surface, and γ a geodesic of length l and imprimitivity index ν (the number of times γ runs around a primitive ancestor). Define the *weight* $\mathrm{wt}(\gamma)$ as follows:

$$\mathrm{wt}_M(\gamma) = \begin{cases} \frac{1}{\nu} & \text{if } \gamma \text{ is orientation-preserving;} \\ \\ \frac{1}{\nu}\tanh(l/2) & \text{if } \gamma \text{ is orientation-reversing.} \end{cases} \tag{1}$$

Define the *total weight function*

$$W_M(l) = \sum_{l(\gamma)=l} \mathrm{wt}(\gamma). \tag{2}$$

From the Selberg trace formula, we have

Proposition 1 *Let M and N be hyperbolic surfaces, possibly non-orientable or disconnected. M and N are isospectral if and only if $W_M = W_N$.*

Sketch of proof. The weight $\mathrm{wt}_M(\gamma)$ tells the spectral contribution of γ, measured in units of the contribution of a primitive orientation-preserving geodesic of length $l(\gamma)$. Geodesics of different lengths make distinguishable contributions to the spectrum, but the contributions from geodesics of any given length get pooled together. (For details, see Randol's chapter in Chavel [3], specifically page 294; cf. also Gangolli [8] and Bérard-Bergery [1].) ■

To prove Theorem 1 above, it suffices to show

Theorem 2 *If M and N are hyperbolic surfaces and $W_M = W_N$, then M and N are almost conjugate.*

Observe that this is a purely geometrical statement: All reference to the Laplace spectrum has been laundered through the total weight function.

To prove Theorem 2, we will analyze how we might engineer agreement between W_M and W_N without having total agreement between the geodesics of M and N, and show that this is not possible without having infinitely many lengths l for which the number of geodesics of length exactly l is at least $C\frac{e^l}{l}$, for $C > 0$. This will contradict the following Proposition, which is a simple consequence of the so-called 'Prime Geodesic Theorem'.

Proposition 2 *For any compact hyperbolic surface, the number of geodesics of length exactly l is $o(\frac{e^l}{l})$.*

Proof. According to the Prime Geodesic Theorem (see [12]), for a connected hyperbolic surface (whether orientable or not) the number $F(l)$ of geodesics of length at most l is asymptotic to $\frac{e^l}{l}$. The number $f(l)$ of geodesics of length exactly l is given by the jump of F at l:

$$f(l) = \lim_{s \to l+} F(s) - \lim_{s \to l-} F(s). \tag{3}$$

But if F is any positive increasing function asymptotic to G, the jumps of F are $o(G)$. So $f(l) = o(\frac{e^l}{l})$. This establishes our claim for connected surfaces. The extension to the general case is immediate, because the $o(\frac{e^l}{l})$ estimate holds separately on each of the finitely many connected components. ■

3 Proof of Theorem 2

Let $\alpha_M(l)$ denote the number of primitive orientation-preserving geodesics in M of length exactly l, and $\beta_M(l)$ the number of primitive orientation-reversing geodesics.

Fix two surfaces M and N with $W_M = W_N$, and set

$$a(l) = \alpha_M(l) - \alpha_N(l); \tag{4}$$

$$b(l) = \beta_N(l) - \beta_M(l). \tag{5}$$

Note that, in the second definition, M and N have traded places.

Our job is to show that $a(l) = b(l) = 0$ for all l. The condition $W_M = W_N$ tells us that

$$\sum_k \frac{1}{k} a(\frac{l}{k}) = \sum_{k \text{ odd}} \frac{1}{k} b(\frac{l}{k}) \tanh(l/(2k)) + \sum_{k \text{ even}} \frac{1}{k} b(\frac{l}{k}).$$

Note how on the right-hand side we have had to distinguish between odd and even k, since going around an orientation-reversing geodesic an even number of times yields an orientation-preserving geodesic.

This system of constraints on the integer-valued functions a and b has solutions which at first blush look like they might permit the construction of a counter-example. To get the simplest solutions, fix an integer $q \geq 2$, and set

$$l_0 = \log q,$$

so that

$$\tanh(\frac{nl_0}{2}) = \frac{q^n - 1}{q^n + 1}.$$

Let

$$c_n = \frac{1}{n} \sum_{j|n} \mu(n/j) q^j.$$

We get a solution to our equations by taking $a(l) = b(l) = 0$ when l is not a multiple of l_0, and setting

$$a(l_0) = q - 1,$$

$$b(l_0) = q + 1,$$

$$a(2l_0) = 1,$$

$$b(2l_0) = 0,$$

$$a(nl_0) = b(nl_0) = c_n \ \ , n = 3, 5, 7, \ldots,$$

$$a(nl_0) = b(nl_0) = 0 \ \ , n = 4, 6, 8, \ldots.$$

The problem with these solutions is that they grow too fast: For n odd, $a(nl_0) \asymp q^n/n$. Our proof of Theorem 2 proceeds by showing that this kind of runaway growth is unavoidable.

Note. When $q = 2$, for the sequence $c_1, c_2, \ldots$ we get

$$2, 1, 2, 3, 6, 9, 18, 30, 56, 99, \ldots;$$

when $q = 3$ we get

$$3, 3, 8, 18, 48, 116, 312, 810, 2184, 5880, \ldots.$$

Looking these sequences up reveals that c_n tells the number of primitive length-n necklaces with beads of q colors, when turning over is not allowed. When q is a prime power, we have the alternative interpretation of c_n as the number of irreducible monic polynomials of degree n over the field with q elements. This seems suggestive: We can't use these solutions in the context of isospectral hyperbolic surfaces, but perhaps we could get some mileage out of them in a different context.... We leave that question for another day, and get back to the proof of Theorem 2.

Let

$$L = \{l : a(l) \neq 0 \text{ or } b(l) \neq 0\} \tag{6}$$

and

$$L_0 = \{l \in L : l \text{ is not a multiple of any other element of } L\}. \tag{7}$$

According to this definition, L is the set of lengths of geodesics where M and N exhibit different behavior, and L_0 consists of those lengths which are minimal with respect to the partial order where $l \preceq m$ means $m = kl$, $k \in \mathbf{N}^+$. Every element of L sits above some minimal element, i.e.

$$L \subseteq L_0\mathbf{N}^+. \tag{8}$$

Our job is to show that $L_0 = \emptyset$.

Lemma 1 $|L_0| < \infty$

Proof. Suppose $l \in L_0$. By assumption, $W_M(l) = W_N(l)$. Because l is minimal in L, any contributions by imprimitive geodesics to $W_M(l)$ are exactly matched by contributions to $W_N(l)$. This means that the contributions of primitive geodesics of length l must match:

$$a(l) = \tanh(l/2)b(l). \tag{9}$$

Assume for convenience that $a(l) > 0$, and hence $b(l) > a(l)$. Rewrite the equation above:

$$b(l) - a(l) = b(l)(1 - \tanh(l/2)); \tag{10}$$

$$b(l) = \frac{b(l) - a(l)}{1 - \tanh(l/2)}. \tag{11}$$

When l is large,

$$b(l) = \frac{b(l) - a(l)}{1 - \tanh(l/2)} \approx \frac{e^l}{2}(b(l) - a(l)) \geq \frac{e^l}{2}. \tag{12}$$

According to Proposition 2, the total number of geodesics of length exactly l is $o(\frac{e^l}{l})$. Here we have at least something on the order of $\frac{e^l}{2}$ geodesics of length l. This puts an upper bound on l, and thus forces $|L_0| < \infty$. ∎

Let P_{odd} denote the set of odd primes.

Lemma 2 *For any $l \in L_0$, only a finite number of the odd prime multiples of l are also multiples of an element of L_0 differing from l. Specifically,*

$$|lP_{\text{odd}} \cap (L_0 - \{l\})\mathbf{N}^+| \leq |L_0| - 1. \tag{13}$$

Proof. If $l_1 \in L_0$, $l_1 \neq l$, then $|lP_{\text{odd}} \cap l_1\mathbf{N}^+| \leq 1$. (This is a simple fact about divisibility: It has nothing special to do with lengths of geodesics!) ∎

Now fix any $l \in L_0$, and let p be an odd prime that avoids the finite set for which $pl \in (L_0 - \{l\})\mathbf{N}^+$. Since l is minimal in L, as above we have

$$a(l) = \tanh(l/2)b(l). \tag{14}$$

As above, assume for convenience that $a(l) > 0$, and hence $b(l) > a(l)$.

By assumption, $W_M(l) = W_N(l)$. The only geodesics that are 'in play' at length pl are those of length l or pl: That was the whole point of the restriction we've placed on p. So

$$a(pl) + \frac{1}{p}a(l) = \tanh(pl/2)\left(b(pl) + \frac{1}{p}b(l)\right). \tag{15}$$

Let's rework this:

$$a(pl) - \tanh(pl/2)b(pl) = \frac{1}{p}(\tanh(pl/2)b(l) - a(l)); \tag{16}$$

$$a(pl) - b(pl) + b(pl)(1 - \tanh(pl/2)) = \frac{1}{p}(\tanh(pl/2)b(l) - a(l)); \qquad (17)$$

$$b(pl) = \frac{\frac{1}{p}(\tanh(pl/2)b(l) - a(l)) + (b(pl) - a(pl))}{1 - \tanh(pl/2)}. \qquad (18)$$

Look at the numerator here. For p large, $\frac{1}{p}(\tanh(pl/2)b(l) - a(l))$ is close to $\frac{1}{p}(b(l) - a(l))$, and $b(l) - a(l)$ is a positive integer. And $b(pl) - a(pl)$ is always an integer: Not necessarily a positive integer, just some integer. As soon as p is larger than $2(b(l) - a(l))$, $\frac{1}{p}(b(l) - a(l))$ will be a positive fraction smaller than $1/2$, and adding an integer to it can only increase its absolute value. This means that for p large, the smallest the numerator can be in absolute value is something like $\frac{1}{p}(b(l) - a(l))$, which is at least $\frac{1}{p}$.

Meanwhile, the denominator is $1 - \tanh(pl/2) \approx 2e^{-pl}$. So $b(pl)$ is bigger than something like $\frac{e^{pl}}{2p}$. This contradicts Proposition 2—unless L_0 is empty! So $L_0 = \emptyset$, and M and N are almost conjugate. ■

4 Isospectral nonorientable surfaces

It is well known that there are many examples of isospectral closed hyperbolic surfaces. The first example goes back to Vigneras [14], who constructed arithmetic examples from quaternion algebras. More recent constructions have used Sunada's method. Sunada's method is very flexible, and can produce nonorientable examples as easily as orientable examples. But as we don't know of a reference for this, we briefly outline the procedure here. For necessary backgound, see Buser [2].

If you take any pair of Sunada isospectral closed hyperbolic surfaces without boundary, then they have a common quotient. Now just add what Conway calls a 'cross-handle' to this quotient, i.e., take the connected sum with a Klein bottle. Or more generally, take the connected sum with any closed non-orientable surface. Put the hyperbolic metric on this new quotient, and lift everything (the cross-handles and the metric) back up to the covers. The resulting surfaces are isospectral and nonorientable.

Another way of producing isospectral nonorientable pairs is to change some of the gluings in known orientable examples where isospectrality is proven using transplantation. To take a specific example, consider the surfaces described by Buser [2], Chapter 11, page 304. If you reinterpret Buser's gluing diagrams (Figures 11.5.1 and 11.5.2) so that the identifications on the

β geodesics are by translation, you get a non-orientable isospectral pair. The β identifications now add four cross-handles, rather than four handles. The transplantation method proving isospectrality in the orientable case continues to work here as well.

5 Comments on orbifolds

Here, as promised above, are some brief comments about orbifolds.

There are three independent examples of isospectral flat (disconnected) 2-orbifolds that are not almost conjugate, one involving quotients of a square torus, and two involving quotients of a hexagonal torus. We describe them using Conway's orbifold notation [4].

A standard square torus has as 2- and 4-fold quotients a 2222 orbifold and a 244 orbifold. If we call the torus S_1 and the quotients S_2 and S_4, spectrally

$$S_1 + 2S_4 = 3S_2, \tag{19}$$

i.e., you can't hear the difference between a torus with two 244s, and a trio of 2222s.

A standard hexagonal torus H_1 has as 2-, 3-, and 6-fold quotients a 2222 orbifold H_2 (this is a regular tetrahedron); a 333 orbifold H_3; and a 236 orbifold H_6. Spectrally,

$$H_2 + H_6 = 2H_3 \tag{20}$$

and

$$H_1 + H_3 + H_6 = 3H_2. \tag{21}$$

From these relations we can derive, for example:

$$H_1 + 3H_3 = 4H_2; \tag{22}$$

$$H_1 + 4H_6 = 5H_3; \tag{23}$$

$$2H_1 + 3H_6 = 5H_2. \tag{24}$$

These examples arise from a careful analysis of the contributions of rotations of various orders to the spectrum via the Selberg trace formula. To explain just how this works would take us too far afield. However, it is possible to verify isospectrality in these examples in a direct and elementary way by using Fourier series to represent explicitly the eigenfunctions of the component orbifolds, and checking that eigenvalues match up.

Among hyperbolic 2-orbifolds, no such examples exist, whether connected or not. This is a corollary of Theorem 1, together with the observation that, in contrast to the flat case, in the hyperbolic case elliptic elements of differing order make distinguishable contributions to the spectrum. Again, to go further into detail would take us too far afield.

Thanks

We're grateful to Emily Proctor, Carolyn Gordon, and Emily Dryden for stimulating discussions, and to Humboldt University (specifically the Sonderforschungsbereich 647 project) for hospitality and support during the writing of this paper. We're sorry to have to thank Peter Sarnak for pointing out that if a function is approaching a limit, its jumps must be getting smaller and smaller.

References

[1] Lionel Bérard-Bergery. Laplacien et géodésiques fermées sur les formes d'espace hyperbolique compactes. In *Séminaire Bourbaki, 24ème année (1971/1972), Exp. No. 406*, pages 107–122. Lecture Notes in Math., Vol. 317. Springer, 1973.

[2] Peter Buser. *Geometry and Spectra of Compact Riemann Surfaces.* Birkhäuser, 1992.

[3] Isaac Chavel. *Eigenvalues in Riemannian Geometry.* Academic Press, 1984.

[4] John Horton Conway. The orbifold notation for surface groups. In *Groups, Combinatorics and Geometry*, pages 438–447. Cambridge University Press, 1992.

[5] Peter G. Doyle and Juan Pablo Rossetti. Tetra and Didi, the cosmic spectral twins. *Geom. Topol.*, 8:1227–1242, 2004.

[6] Emily B. Dryden. Isospectral Finiteness of Hyperbolic Orbisurfaces, 2004.

[7] Emily B. Dryden and Alexander Strohmaier. Huber's theorem for hyperbolic orbisurfaces, 2005.

[8] Ramesh Gangolli. The length spectra of some compact manifolds of negative curvature. *J. Differential Geom.*, 12(3):403–424, 1977.

[9] Heinz Huber. Zur analytischen Theorie hyperbolischer Raumformen und Bewegungsgruppen, I. *Math. Ann.*, 138:1–26, 1959.

[10] Heinz Huber. Zur analytischen Theorie hyperbolischer Raumformen und Bewegungsgruppen, II. *Math. Ann.*, 142:385–398, 1961.

[11] Heinz Huber. Zur analytischen Theorie hyperbolischer Raumformen und Bewegungsgruppen, II: Nachtrag. *Math. Ann.*, 143:463–464, 1961.

[12] Ralph Phillips and Peter Sarnak. Geodesics in homology classes. *Duke Math. J.*, 55:287–297, 1987.

[13] T. Sunada. Riemannian coverings and isospectral manifolds. *Annals of Math.*, 121:169–186, 1985.

[14] M.-F. Vigneras. Varietés riemanniennes isospectrales et non isométriques. *Annals of Math.*, 112:21–32, 1980.