Mostow Rigidity Made Easier

Richard Evan Schwartz *

April 16, 2026

1 The Main Result

Mostow's rigidity theorem [**M**] is one of the most famous and spectacular results about hyperbolic manifolds. The traditional proofs (and also Gromov's proof [**T**]) rely on a fair amount of real analysis. The student who wants to learn the result *all the way to the bottom* is in for an arduous journey. In this article I give a proof of Mostow rigidity that is self-contained modulo undergraduate real analysis. The proof should be accessible to first-year graduate students interested in geometry and topology. My approach has a lot in common with existing proofs, but it is analytically lighter. The ideas here also overlap with [**S**].

Given metric spaces (X_1, d_1) and (X_2, d_2), a map $H : X_1 \to X_2$ is *BL* (bi-Lipschitz) if H is a bijection and if there is some $K \geq 1$ such that

$$K^{-1} d_1(x, y) \leq d_2(H(x), H(y)) \leq K d_1(x, y), \qquad \forall x, y \in X_1. \qquad (1)$$

When $K = 1$, the map H is an *isometry*. I will prove the following theorem.

Theorem 1.1 (Mostow) *If M_1 and M_2 are compact hyperbolic 3-manifolds and $f : M_1 \to M_2$ is BL, then there is an isometry $g : M_1 \to M_2$. So, diffeomorphic compact hyperbolic 3-manifolds are isometric.*

This is a somewhat limited version of Mostow's original Theorem. In §6, I discuss how the same arguments prove the full-blown theorem.

*Supported by N.S.F. Grant DMS-2505281

Paper Overview: The paper is organized in a top-down manner, so that the big result, Mostow's Theorem, is presented first and proved modulo three medium-sized results. The medium-sized results are then proved modulo some small results, and then finally the small results are proved. Here is a chapter-by-chapter account.

- In §1 (this chapter) we introduce background material, and then prove Mostow Rigidity modulo Theorem 1.2, Lemma 1.3, and Theorem 1.7.
- In §2 we prove Theorem 1.2 and Lemma 1.3 using classical hyperbolic geometric arguments, including the Morse Lemma. There is nothing new here, but I try to present it well.
- In §3 we prove Theorem 1.7 modulo several auxiliary results that have an analytic flavor.
- In §4 we review some basic measure-theoretic tools that are needed for the auxiliary results left over from §3.
- In §5 we use the tools from §4 to prove the auxiliary results left over from §3.

Acknowledgements: I thank Sujung Jo, Dan Margalit, and ChatGPT for helpful conversations. I would also like to thank the anonymous referee for making helpful expository suggestions.

1.1 Hyperbolic Geometry Background

Hyperbolic Space: Let $\boldsymbol{C}$ denote the complex numbers. *Hyperbolic* 3-*space*, denoted $\boldsymbol{H}^3$, is a metric space modeled on the upper half space in $\boldsymbol{C} \times (0, \infty)$. Technically, $\boldsymbol{H}^3$ is a complete Riemannian manifold of constant negative sectional curvature. You don't need to know this to read the proof of Mostow Rigidity.

Here are some useful facts about the hyperbolic metric.

F1. When $a, b \in \boldsymbol{C}$ and $a \neq 0$, the map $(z, t) \to (az + b, |a|t)$ acts isometrically on $\boldsymbol{H}^3$

F2. The geodesics in $\boldsymbol{H}^3$, which are the length minimizing paths, are either vertical rays or semicircles that meet $\boldsymbol{C} \times \{0\}$ at right angles.

F3. Letting ℓ and $\ell_{\boldsymbol{E}}$ denote hyperbolic and Euclidean arc-length respectively, we have $a\ell \leq \ell_E \leq b\ell$ on $\boldsymbol{C} \times [a, b]$.

F4. The map $\phi(p) = (0, \|p\|)$ carries $p \in \boldsymbol{H}^3$ to the point on $\gamma = \{0\} \times (0, \infty)$ nearest to p. If $t \in (0, 1)$, the distance from $(1, t)$ to γ is less than $\ln(1/t) + 1$.

Hyperbolic Manifolds: All manifolds are assumed to be compact. Let $\boldsymbol{I}$ be the group of isometries of $\boldsymbol{H}^3$. A *hyperbolic manifold* is any quotient of the form $M = \boldsymbol{H}^3/\Gamma$, where Γ is a subgroup of $\boldsymbol{I}$. We only make this definition when M is actually a manifold. What this means is that we have a universal covering map $\pi : \boldsymbol{H}^3 \to M$, and Γ is the *deck group*. Γ is usually called a *co-compact torsion-free lattice*, but we will call Γ a *nice lattice*. The metric on M is such that π is a local isometry.

Conformal Transformations: Let $\mathbf{S} = \boldsymbol{C} \cup \infty$ be the Riemann sphere. A *generalized circle* is either a round circle in $\boldsymbol{C}$ or else a set of the form $L \cup \infty$ where L is a straight line in $\boldsymbol{C}$. A *conformal transformation* is a homeomorphism of $\mathbf{S}$ which maps generalized circles to generalized circles. When such a map is orientation preserving, it has the form

$$z \to \frac{az + b}{cz + d}, \qquad a, b, c, d \in \boldsymbol{C}, \qquad ad - bc = 1. \tag{2}$$

As a special case, a *homothety* is a map of the form $z \to az + b$ where $a \in (0, \infty)$. Such maps preserve the directions of lines in $\boldsymbol{C}$.

Every isometry of $\boldsymbol{H}^3$ extends to give a conformal transformation of $\mathbf{S}$ and every conformal transformation of $\mathbf{S}$ arises this way. For this reason, we will sometimes abuse the notation and speak of a conformal transformation h as being a member of $\boldsymbol{I}$; technically we are referring here to the isometry that extends h.

Affine Transformations: An affine map of $\boldsymbol{R}$ is a map $x \to ax + b$ with $a \neq 0$. Interpreting $\boldsymbol{C}$ as $\boldsymbol{R}^2$, we say that a *real affine transformation* of $\boldsymbol{C}$ is a map of the form $A(v) = T(v) + w$ where T is an invertible real linear transformation and $w \in \boldsymbol{C}$. Most real affine transformations are not conformal. However, if there are round circles C_1, C_2 such that $A(C_1) = C_2$ then A is in fact conformal.

Equivariance and Lifting: A BL map $H : \boldsymbol{H}^3 \to \boldsymbol{H}^3$ is *equivariant* if there is a pair of nice lattices Γ_1, Γ_2 such that

$$H\Gamma_1 H^{-1} = \Gamma_2. \tag{3}$$

Here is another way to express this condition. For all $\gamma_1 \in \Gamma_1$, there exists a $\gamma_2 \in \Gamma_2$ such that $H \circ \gamma_1 = \gamma_2 \circ H$.

Here is how this arises in Mostow Rigidity. We have $M_j = \boldsymbol{H}^3/\Gamma_j$ where Γ_j is a nice lattice. Suppose $f : M_1 \to M_2$ is BL. Then f has an equivariant BL *lift* $H : \boldsymbol{H}^3 \to \boldsymbol{H}^3$ which conjugates Γ_1 to Γ_2 and is such that

$$f \circ \pi_1 = \pi_2 \circ H.$$

Note that if H is K-BL then so is fHg for any $f, g \in \boldsymbol{I}$.

We make similar definitions for a homeomorphism $h : \mathbf{S} \to \mathbf{S}$ (which we usually call a *homeo*) The homeo h is equivariant if $h\Gamma_1 h^{-1} = \Gamma_2$. We will often work with pairs (H, h) where h is a continuous extension of H to $\mathbf{S}$. (We prove the existence of such extensions below.) In this case, we call the pair (H, h) *equivariant* if both maps are simultaneously equivariant with respect to the same pair of nice lattices.

Tame Sequences: We call a sequence $\{g_n\}$ in $\boldsymbol{I}$ *tame* if $\{g_n(p)\}$ is bounded for each $p \in \boldsymbol{H}^3$. In this situation we can pass to a subsequence and arrange that there are 4 points $p_1, p_2, p_3, p_4 \in \boldsymbol{H}^3$, say the vertices of a regular tetrahedron, so that all 4 sequences $\{g_n(p_j)\}$ converge. But then $\{g_n\}$ converges in $\boldsymbol{I}$. In short, a tame sequence converges on a subsequence.

Here is a criterion for tameness. Suppose $a, b, c \in \mathbf{S}$ are 3 distinct points, and $\{g_n(a)\}$, $\{g_n(b)\}$, $\{g_n(c)\}$ all converge to 3 distinct points of $\mathbf{S}$. Then $\{g_n\}$ is tame. This derives from the fact that (up to specifying whether the map preserves or reverses orientation) an element of $\boldsymbol{I}$ is determined by what it does on 3 distinct points of $\mathbf{S}$. The takeaway here is that we can tell that a sequence is tame by looking at how it acts on the ideal boundary.

Below we will generalize these ideas for sequences of K-BL maps.

1.2 A Lesson in Calculus

Suppose that $h : \boldsymbol{R} \to \boldsymbol{R}$ is a function that is differentiable at some point $a \in \boldsymbol{R}$. After translating we normalize so that $a = 0$ and $h(a) = 0$. Then what we are saying is that there is some constant A such that

$$\lim_{x \to 0} \frac{h(x)}{x} = A. \tag{4}$$

It follows that, for any nonzero real number u, we have

$$\lim_{n \to \infty} \frac{h(u/n)}{u/n} = A. \tag{5}$$

Put another way

$$\lim_{n \to \infty} nh(u/n) = Au. \tag{6}$$

We introduce auxiliary functions

$$f_n(x) = nx, \qquad g_n(x) = x/n. \tag{7}$$

We define

$$h_n = f_n \circ h \circ g_n. \tag{8}$$

Then we have

$$\lim_{n \to \infty} h_n(u) = Au. \tag{9}$$

This last equation holds for all $u \in \boldsymbol{R}$, even when $u = 0$. Hence $h_n \to h'$ where h' is the linear function $h'(u) = Au$. Given the way that limits work – and I invite you to think about this – the convergence is uniform over compact sets of $\boldsymbol{R}$. In short $h_n \to h'$ uniformly on compact sets. When $A = 0$ the map h' is the zero map. Otherwise h' is a linear map. This is how we want to think about the derivative: If you zoom in to a point of differentiability, you get a linear map in the limit.

Before moving on, let us extend the lesson a bit. Suppose that $h : \boldsymbol{C} \to \boldsymbol{C}$ is a map normalized so that $h(0) = 0$ and the directional derivative of h exists in the direction of a line L through the origin. Then we can define h_n as above. If we know that the limit h' exists, then we can say that $h'|_L$ is a linear map from L into $\boldsymbol{C}$. The whole map h' might be a mystery but the restriction to L is not a mystery. Our constructions below are all built around this idea.

1.3 Extensions of Equivariant BL Maps

In §2 we prove the following well-known result.

Theorem 1.2 *A BL map H of $\boldsymbol{H}^3$ extends continuously to a homeo h of* $\mathbf{S}$. *If H is equivariant with respect to a pair Γ_1, Γ_2 of nice lattices then so is h.*

For the subsequent results in this section we work exclusively with equivariant pairs (H, h). We will show in the equivariant case that the map h from Theorem 1.2 is a conformal transformation. This means that h extends to an equivariant isometry $h^* : \boldsymbol{H}^3 \to \boldsymbol{H}^3$ which in turn defines an isometry $g : M_1 \to M_2$. So, proving Mostow Rigidity boils down to proving that h is a conformal transformation. We will move h around by isometries until it spills its secret.

We write $(H, h) \sim (H', h')$ if there are $f, g \in \boldsymbol{I}$ so that $H' = fHg$ and $h' = fhg$. A sequence $\{(H_n, h_n)\}$ is *derived from* (H, h) if we have $(H_n, h_n) \sim (H, h)$ for all n. Compare Equation 8. Generalizing the definition given in §1.1, we call a sequence $\{\phi_n\}$ of K-BL maps *tame* if $\{\phi_n(p)\}$ is bounded for each $p \in \boldsymbol{H}^3$. We call $\{(H_n, h_n)\}$ tame if $\{H_n\}$ is tame. In §2 we prove the following generalization of the tameness criterion in §1.1.

Lemma 1.3 *Suppose that $\{(H_n, h_n)\}$ is a sequence derived from (H, h) and there are 3 distinct points $a, b, c \in \mathbf{S}$ such that $\{h_n(a)\}, \{h_n(b)\}, \{h_n(c)\}$ converge to 3 distinct points in $\mathbf{S}$. Then $\{(H_n, h_n)\}$ is tame.*

Before we implement our calculus lesson we need one more result, a subtle and powerful one.

Lemma 1.4 *Let $\{(H_n, h_n)\}$ be a tame sequence derived from (H, h). Then $(H, h) \sim (H', h')$ where, on a subsequence, H_n converges to H' uniformly on compact subsets of $\boldsymbol{H}^3$ and h_n converges to h' uniformly on $\mathbf{S}$.*

Proof: Let $H_n = f_n H g_n$ and $h_n = f_n h g_n$. Note that $\{f_n\}$ and $\{g_n\}$ might not be tame. For any $g \in \Gamma_1$ we have $f_g \in \Gamma_2$ so that $H = f_g H g$ and $h = f_g h g$. Combining these equations with the fact that $\boldsymbol{H}^3/\Gamma_1$ is compact, we can <u>also</u> write $H_n = f_n' H g_n'$ and $h_n = f_n' h g_n'$, where $\{g_n'\}$ is tame. Here

$$f_n' = H_n \circ (g_n')^{-1} \circ H^{-1}.$$

Since $\{H_n\}$ and $\{(g_n')^{-1}\}$ are tame so is $\{f_n'\}$. As noted in §1.1, we have $f_n' \to f' \in \boldsymbol{I}$ and $g_n' \to g' \in \boldsymbol{I}$ on a subsequence. So, we get the desired convergence by setting $h' = f' h g'$ and $H' = f' H g'$. ♠

1.4 Zooming In

Let (H, h) be an equivariant pair. Our goal is to show that $h \in \boldsymbol{I}$. We normalize so that $h(\infty) = \infty$ and interpret h as a homeo of $\boldsymbol{C}$.

Lemma 1.5 *If h is a real affine transformation then $h \in \boldsymbol{I}$.*

Proof: The map h is equivariant with respect to nice lattices Γ_1 and Γ_2. For any $f \in \Gamma_1$ there is some $g \in \Gamma_2$ such that $gh = hf$. We can choose a line L and an element $f \in \boldsymbol{I}$ so that $C_1 = f(L)$ is a round circle. But then $C_2 = h(f(L))$ is a (possibly non-circular) ellipse. On the other hand $h(L)$ is a line and $C_2 = g(h(L))$ is either a line or a round circle. The only way the two descriptions match is if C_2 is a round circle. But then h maps the circle C_1 to the circle C_2. As we mentioned in §1.1, this forces $h \in \boldsymbol{I}$. ♠

A line $L \subset \boldsymbol{C}$ is *good* for h if the restriction $h|_L$ is an affine map. This means that $h(L)$ is a line and h scales lengths by a constant factor on L. We say that the direction D is *good* for h if every line of direction D is good for h. The affine maps here may depend on the line but, since h is a homeo, the image under h of this parallel line family is another parallel line family.

Lemma 1.6 *If two directions are good for h then $h \in \boldsymbol{I}$.*

Proof: There are real affine transformations A, B so that the horizontal and vertical directions for $\phi = AhB$ are good, and ϕ is the identity on the union of the two coordinate axes. But $\phi(x, y)$ is on the same horizontal line as $(0, y)$ and on the same vertical line as $(x, 0)$. Hence $\phi(x, y) = (x, y)$. In short, ϕ is the identity. Since AhB is the identity, $h = A^{-1}B^{-1}$. Hence h is a real affine map. Hence $h \in \boldsymbol{I}$ by Lemma 1.5. ♠

Choose $z \in \boldsymbol{C}$. Specializing the constructions from §1.3, we call $\{(H_n, h_n)\}$ the *zoom sequence based on* (h, z) if $H_n = f_n H g_n$ and $h_n = f_n h g_n$, where

- f_n is the homothety that fixes $h(z)$ and scales distances by n.
- g_n is the homothety that fixes z and scales distances by $1/n$.

We write $h \to h'$ if $\{(H_n, h_n)\}$ is tame, and (H_n, h_n) converges on a subsequence to (H', h') in the sense of Lemma 1.4. For ease of notation we omit mention of H, even though H is lurking in the background at every step.

By Lemma 1.4, we have $h \sim h'$ if $h \to h'$. Since $h' \sim h$, the map h' is also equivariant. We write

$$h \Rightarrow h^{(n)} \qquad \text{if} \qquad h \to h' \to h'' \to \ldots \to h^{(n)}. \tag{10}$$

By several applications of Lemma 1.4 we have $h \sim h^{(n)}$ if $h \Rightarrow h^{(n)}$. So, if $h \Rightarrow h^{(n)}$ then $h^{(n)}$ is equivariant. Also, if $h \Rightarrow h^{(n)} \in \boldsymbol{I}$ then $h \in \boldsymbol{I}$.

The *directional derivative* of h at z in the direction v, when it exists, is given by

$$D_v h(z) = \lim_{t \to 0} \frac{h(z + vt) - h(z)}{t}. \tag{11}$$

We call this a *rational* directional derivative if the real and imaginary parts of v are rational. A point $z \in \boldsymbol{C}$ is an *asterisk* for h if every rational directional derivative exists at z and if $D_1 h(z) \neq 0$. In §3-5 we prove the following result.

Theorem 1.7 *h has an asterisk.*

Corollary 1.8 *Let D be a direction. Then $h \Rightarrow h''$ where D is good for h''.*

Proof: Let z be an asterisk for h. Let $\{h_n\}$ be the zoom sequence based on (h, z). Let $\mathcal{Q}$ denote the set of lines of rational slope through z and let $\mathcal{R}$ denote the set of all lines through z. Remembering the calculus lesson, we see that the restriction of h_n to each line of $\mathcal{Q}$ converges to an affine map. In particular $h_n(z+m) \to h(z) + mD_1h(z)$ for $m = 0, 1, 2$, and these points are distinct. So $\{h_n\}$ is tame by Lemma 1.3. Hence $h \to h'$ and the restriction of the homeo h' to every line of $\mathcal{Q}$ is a nontrivial affine map. Hence every line of $\mathcal{Q}$ is good for h'. Since h' is a homeo, and $\mathcal{Q}$ is dense in $\mathcal{R}$, all lines of $\mathcal{R}$ are good for h'.

Choose $z' \in \boldsymbol{C}$ so that the line through z and z', which we denote as $\overline{zz'}$, has direction D. Let $h'_n = f'_n h' g'_n$. The sequence $\{h'_n\}$ is the zoom sequence based on (h', z'). The restriction of h'_n to $\overline{zz'}$ is independent of n. So, by Lemma 1.3, $\{h'_n\}$ is tame. Every line in $(g'_n)^{-1}(\mathcal{R})$ is good for h'_n, and as $n \to \infty$ this set converges to the set of lines in direction D. (The point $(g'_n)^{-1}(z)$ common to all these lines moves to ∞ along the line $\overline{zz'}$.) Hence D is good for h''. ♠

If D is good for h and $h \Rightarrow h''$ then D is also good for h''. So, applying Corollary 1.8 twice, for different directions, we get $h \Rightarrow h^{(4)}$ where two directions are good for $h^{(4)}$, and $h^{(4)}$ is equivariant. But then $h^{(4)} \in \boldsymbol{I}$ by Lemma 1.6. By Lemma 1.4 again, $h \in \boldsymbol{I}$. Our proof of Mostow Rigidity is done.

1.5 Discussion

The idea of the asterisk is the most unusual ingredient in our proof. I cannot remember how I thought of it, but let me explain how it functions in the proof, and what considerations might lead to it. The essential idea behind the above kind of proof of Mostow Rigidity is that zooming in to a point that has some regularity reveals some extra structure of the map. Let's discuss what the phrase *some regularity* might mean.

At one extreme, *some regularity* might mean that the partial derivatives exist at the point. If the partial derivatives are zero, this information is completely useless to us. If the partial derivatives are nonzero, then when we zoom in and take a suitable limit, we find (after conjugating) that our map is actually a linear transformation when restricted to two lines. This is nice to know, but does not move us much towards a proof of Mostow Rigidity. On the positive side, this amount of regularity is not too hard to establish.

At the other extreme, *some regularity* might mean that the map is actually differentiable at a point. As is well known, the map h from Theorem 1.2 is *quasi-conformal*. Such maps are known to be differentiable almost everywhere, and the derivative is non-singular almost everywhere. See [**LV**]. Given a point of nonsingular differentiability at z, we get a swifter proof. We consider the zoom sequence $\{h_n\}$ based on (h, z) and we get $h \to h'$ where h' is real affine and equivariant – and hence in $\boldsymbol{I}$. The difficulty with this alternate proof is that the analysis behind the differentiability result just quoted is (to me) rather formidable.

The asterisk idea is a happy compromise between these two extremes. On the one hand, the amount of work needed to show the existence of an asterisk is not much more than the amount needed to show that the partial derivatives exist and are nonzero at a point. On the other hand, when we zoom into an asterisk we get control over what happens in all rational directions. This is an intermediate amount of information, compared to the two extremes, but it is enough to push through the proof. The rest of this section will give some variations on a theme, showing other ways to convert the asterisk idea into a proof.

Three Fair Directions: Let us say that h is *fair* on a line L if $h(L)$ is a straight line. We say that a direction D is *fair* for h if h is fair for every line in the direction D. Corollary 1.8 says in particular that for any direction D we have $h \Rightarrow h''$, where D is fair for h''. Applying Corollary 1.8 three times we get $h \Rightarrow h^{(6)}$, where 3 different directions are fair for $h^{(6)}$. It is a fun exercise to show that this forces the equivariant $h^{(6)}$ to be real affine – and hence in $\boldsymbol{I}$.

Zoom Ad Infinitum: We could apply Corollary 1.8 an infinite number of times and then take a limit (guaranteed by a result much like Lemma 1.4) and conclude that $h \sim h^*$ where h^* is fair on every line. But then the equivariance implies that h preserves generalized circles and hence is in $\boldsymbol{I}$.

The Minimal Proof: The first half of the proof of Corollary 1.8 shows that $h \to h'$ where all lines through a point $z \in \boldsymbol{C}$ are good for h'. Since $h \sim h'$ we see that there are two points $a, b \in \mathbf{S}$ such that the restriction of h to each generalized circle through a, b is a conformal map. Since h is equivariant, there is a second pair a', b' of points, both distinct from a and b, such that the restriction of h to each generalized circle through a', b' is a conformal transformation. It is a fun exercise to show that these conditions force $h \in \boldsymbol{I}$. This proof is a bit harder to generalize to higher dimensions.

Challenge: I was hoping to figure out a proof which combines equivariance with a differentiability result even weaker than Theorem 1.7 but couldn't get this to work. Can you?

2 The Extension

2.1 Proofs modulo the Morse Lemma

In this chapter we prove Theorem 1.2 and Lemma 1.3. In this first section we reduce these results to a well-known geometric result called the Morse Lemma. See [**T**, Prop. 5.9.2]. We work with the K-BL map $H : \boldsymbol{H}^3 \to \boldsymbol{H}^3$.

Given a geodesic $\gamma \subset \boldsymbol{H}^3$ let $N_r(\gamma)$ denote the set of points in $\boldsymbol{H}^3$ at most r from γ. The set $N_r(\gamma)$ is called the *r-tubular neighborhood* of γ.

Lemma 2.1 (Morse) *Let γ be a geodesic. Then there is a unique geodesic γ' such that $H(\gamma) \subset N_{K'}(\gamma')$. The constant K' only depends on K.*

Our proof gets the constant $K' = 4K^3 + 2K + 1$, but this is not optimal. The choice of K', however, has no impact on the proof of Theorem 1.2. The Morse Lemma holds in great generality, with good control on K'. See [**GS**].

Proof of Theorem 1.2: The Morse Lemma induces a map of the set of oriented geodesics. If γ is an oriented geodesic then $\gamma \to \gamma'$ where γ' is the unique geodesic such that $H(\gamma) \subset N_{K'}(\gamma')$. The orientation of γ combines with H to determine the orientation of γ'. If τ is another geodesic and $\tau \to \tau'$ and γ, τ have a common endpoint, then $H(\gamma)$ and $H(\tau)$ are asymptotic to each other in the corresponding direction. Hence γ' and τ' also have a common endpoint. This property lets us define h as the unique map of the Riemann sphere $\mathbf{S}$ with the following property: if $\gamma \to \gamma'$ and γ connects p and q, then γ' connects $h(p)$ to $h(q)$. If we apply the construction to H^{-1} we get h^{-1}.

It remains to show that h is a homeo. By construction, h is a bijection. If h is not a homeo, there are H-corresponding sequences of geodesics $\{\gamma_n\} \leftrightarrow \{\gamma_n'\}$ such that the endpoints of the geodesics in the one sequence come together on $\mathbf{S}$ and the endpoints of the geodesics in the other sequence do not. But then the one sequence exits every compact subset of $\boldsymbol{H}^3$ and the other sequence does not. This contradicts the BL nature of H. Given the way that the extension only depends on H we see that h is equivariant if H is equivariant. ♠

Before proving Lemma 1.3 we need one preliminary lemma. An *ideal triangle* is a union of 3 geodesics defined by 3 distinct points on $\mathbf{S}$. Any two ideal triangles are equivalent under the action of $\boldsymbol{I}$.

Lemma 2.2 *Let Δ be an ideal geodesic triangle in $\boldsymbol{H}^3$. Then for any R, the set of points in $\boldsymbol{H}^3$ within R of all three geodesics of Δ is compact.*

Proof: Let $\gamma_{a,b}$ be the geodesic in $\boldsymbol{H}^3$ whose endpoints are $a, b \in \mathbf{S}$. We can normalize so that Δ has vertices $0, 1, \infty$. Let $p = (z, t) \in \boldsymbol{H}^3$. If $t \to 0$ then the distance from p to one of $\gamma_{0,\infty}$ or $\gamma_{1,\infty}$ tends to ∞. If $t \to \infty$ or $|z| \to \infty$ then the distance from p to $\gamma_{0,1}$ tends to ∞. So, p is close to all these geodesics when $t \sim 1$ and $|z|$ is fairly small. ♠

Proof of Lemma 1.3: We can normalize by a pair of elements of $\boldsymbol{I}$ so that $\{(H_n, h_n)\}$ is such that $h_n(0) \to 0$ and $h_n(1) \to 1$ and $h_n(\infty) \to \infty$. Further composing with convergent sequences in $\boldsymbol{I}$ we can assume that h_n fixes each of $0, 1, \infty$. Let $p \in \boldsymbol{H}^3$ be arbitrary. By the Morse Lemma, $H_n(p)$ is uniformly close to all three of the geodesics of the ideal triangle with vertices $0, 1, \infty$. But then, by Lemma 2.2, $\{H_n(p)\}$ remains within a bounded subset of $\boldsymbol{H}^3$. Hence $\{(H_n, h_n)\}$ is tame. ♠

2.2 The Morse Lemma Modulo the Tube Lemma

We keep the notation from the previous section. Let d, ℓ respectively denote hyperbolic distance, hyperbolic arc length. As in **F4** from §1.1, we define

$$\gamma = \{0\} \times (0, \infty), \qquad \phi(p) = (0, \|p\|) \in \gamma. \tag{12}$$

Again, ϕ is the map such that $\phi(p)$ is the point of γ closest to p.

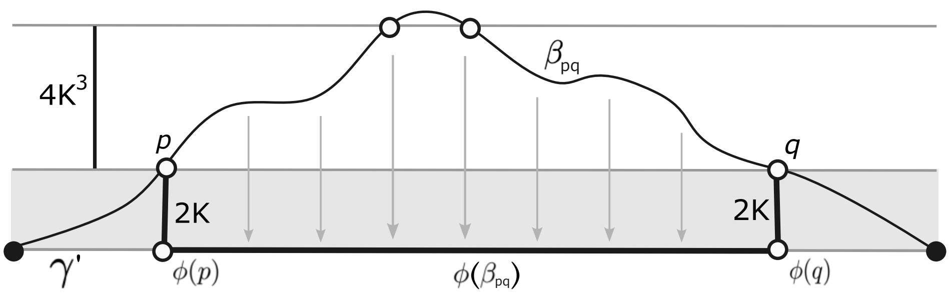
In the next section we prove the following result.

Lemma 2.3 (Tube) *If $r > 1$ then $\ell \circ \phi \le e^{-r+1}\ell$ on $\boldsymbol{H}^3 - N_r(\gamma)$.*

Now we use the Tube Lemma to prove the Morse Lemma. The proof comes in two steps. The first step is showing that an arc of $H(\gamma)$ cannot wander too far from the geodesic connecting its endpoints. If the arc does wander far away, we can use ϕ to produce a much shorter competing path which connects the endpoints. This will give a contradiction. After we have this result about finite arcs of $H(\gamma)$, the second step involves considering a sequence of increasingly long arcs of $H(\gamma)$ and taking a suitable limit. The key point is that the geodesics connecting the endpoints of these arcs all have to line up.

Lemma 2.4 *Let $C = 4K^3 + 2K$. Let α be a geodesic segment in $\boldsymbol{H}^3$. Then $H(\alpha) \subset N_C(\gamma')$ for the geodesic γ' through the endpoints of $H(\alpha)$.*

Proof: We normalize so that $\gamma' = \{0\} \times (0, \infty)$, as in the Tube Lemma. We show a schematic picture of the construction.



Let $\beta = H(\alpha)$. Suppose $\beta \not\subset N_C(\gamma')$. Then there are $p, q \in \partial N_{2K}(\gamma')$ and an arc β_{pq} of β joining p to q that remains outside $N_{2K}(\gamma')$ and exits $N_C(\gamma')$. Note that $\ell(\beta_{pq}) \geq 8K^3$. The path $\phi(\beta_{pq})$ connects $\phi(p)$ to $\phi(q)$. By the Tube Lemma

$$d(\phi(p), \phi(q)) \leq \ell(\phi(\beta_{pq})) \leq \ell(\beta_{pq}) e^{-2K+1}. \tag{13}$$

By **F4** from §1.1, we have $d(p, \phi(p)) = d(q, \phi(q)) = 2K$. By Equation 13 and the triangle inequality

$$d(p, q) \leq d(p, \phi(p)) + d(\phi(p), \phi(q)) + d(\phi(q), q) \leq e^{-2K+1} \ell(\beta_{pq}) + 4K. \tag{14}$$

Since β_{pq} is the image of a geodesic segment under a K-BL map,

$$\ell(\beta_{pq}) \leq K \ell(H^{-1}(\beta_{pq})) = K d(H^{-1}(p), H^{-1}(q)) \leq K^2 \, d(p, q). \tag{15}$$

Combining Equations 14 and 15, and using $\ell(\beta_{pq}) \geq 8K^3$, we have

$$\left(e^{-2K+1} + \frac{1}{2K^2}\right) \ell(\beta_{pq}) \geq e^{-2K+1} \ell(\beta_{pq}) + 4K \geq d(p, q) \geq \frac{\ell(\beta_{pq})}{K^2}. \tag{16}$$

Multiplying through by $2K^2 / \ell(\beta_{pq})$ and subtracting 1, we get

$$2K^2 e^{-2K+1} \geq 1.$$

This is false for all $K \geq 1$, a contradiction. ♠

Proof of the Morse Lemma: Let $C = 4K^3 + 2K$ be as in Lemma 2.4. We fix an origin O of γ and let $\alpha_n \subset \gamma$ be a geodesic segment of length at least $2n$ centered at O. Let the endpoints be a_n and b_n. Each ray of γ emanating from O exits every compact subset of $\boldsymbol{H}^3$, so we can choose our segments so that $\{H(a_n)\}$ and $\{H(b_n)\}$ both converge to points $A, B \in \mathbf{S}$. Let γ'_n be the geodesic through the endpoints of $H(a_n)$ and $H(b_n)$. Given our description of hyperbolic geodesics, and the fact (from Lemma 2.4) that γ'_n comes within C of $H(O)$ for all n, we cannot have $A = B$.

We normalize so that $A = (-1, 0)$ and $B = (1, 0)$. Let $\Delta_r \subset \boldsymbol{H}^3$ be the ball of radius r about O. Let γ' be the geodesic connecting $(-1, 0)$ and $(1, 0)$. The endpoints of γ'_n converge to the endpoints of γ'. By compactness and Lemma 2.4, we have the following for sufficiently large n:

$$H(\gamma) \cap \Delta_r = H(\alpha_n) \cap \Delta_r \subset N_C(\gamma'_n) \cap \Delta_r \subset N_{C+1}(\gamma') \cap \Delta_r \subset N_{C+1}(\gamma').$$

Letting $r \to \infty$ we see that $H(\gamma) \subset N_{C+1}(\gamma')$. Now we set $K' = C + 1$. ♠

2.3 Proof of the Tube Lemma

Let ϕ and γ be as in Equation 12. Let $\ell_{\boldsymbol{E}}$ denote Euclidean arc length. Again, ℓ denotes hyperbolic arc length. We use the facts **F1** – **F4** from §1.1. Looking at the formula for ϕ we can see that $\ell_{\boldsymbol{E}} \circ \phi \leq \ell_{\boldsymbol{E}}$. That is, ϕ does not increase Euclidean arc length. Define

$$\tau = \{1\} \times (0, \infty), \qquad U = (\boldsymbol{C} \times (0, e^{-r+1})) \cap \phi^{-1}(\boldsymbol{C} \times (1, \infty)) \tag{17}$$

The set U is open and $\phi(U) \subset \boldsymbol{C} \times (1, \infty)$. By **F3**, we have $\ell \leq \ell_{\boldsymbol{E}}$ on $\phi(U)$ and $\ell_{\boldsymbol{E}} \leq e^{-r+1}\ell$ on U. Hence, on U, we have

$$\ell \circ \phi \leq \ell_{\boldsymbol{E}} \circ \phi \leq \ell_{\boldsymbol{E}} \leq e^{-r+1}\ell.$$

Combining **F3** and **F4** we see that $(1, t)$ is less than r units from γ when $t \in [e^{-r+1}, \infty)$. Therefore

$$\tau - N_r(\gamma) \subset U. \tag{18}$$

Let $\Gamma \subset \boldsymbol{I}$ denote the stabilizer subgroup of γ. Elements of Γ have the form $(z, t) \to (az, |a|t)$ with $a \in \boldsymbol{C} - \{0\}$. Every point of $\boldsymbol{H}^3 - N_r(\gamma)$ can be mapped into $\tau - N_r(\gamma)$ by an element of Γ. Hence $\boldsymbol{H}^3 - N_r(\gamma) \subset \Gamma(U)$, the orbit of U under Γ. But elements of Γ commute with ϕ. For this reason, our inequality $\ell \circ \phi \leq e^{-r+1}\ell$ on U holds on all of $\boldsymbol{H}^3 - N_r(\gamma)$. This proves the Tube Lemma.

3 Existence of Asterisks

3.1 Reduction to a Technical Lemma

A set $S \subset \boldsymbol{C}$ is *null* if, for every $\epsilon > 0$, there is a countable collection $\{D_j\}$ of disks such that $S \subset \bigcup D_j$ and $\sum \text{area}(D_j) < \epsilon$. Otherwise, S is *fat*. For someone who knows about measure theory, we remark that a null set is one with zero outer measure and a fat set is one with positive outer measure.

Let (H, h) be an equivariant pair, normalized so that $h(\infty) = \infty$. The purpose of this chapter is to prove that h has an asterisk, a point where all the rational directional derivatives exist and are nonzero.

Lemma 3.1 *The subset of* $[0,1]^2$ *where* $\partial h/\partial x$ *does not exist is null. The subset of* $[0,1]^2$ *where* $\partial h/\partial x \neq 0$ *is fat.*

Proof of Theorem 1.7: Tile $\boldsymbol{C}$ by unit squares. By symmetry and Lemma 3.1 the derivative $\partial h/\partial x$ exists in each square outside of a null set. The countable union of null sets is null, so $\partial h/\partial x$ exists outside of a null set $S_0 \subset \boldsymbol{C}$. By rotational symmetry $D_r h$ exists outside a null set for all rational vectors. The countable union of these null sets is null, and outside $\bigcup S_r$ all the rational directional derivatives of h exist. A subset of a null set is null, so the fat set where $\partial h/\partial x \neq 0$ intersects the set where all the rational directional derivatives of h exist. This gives us an asterisk, and in fact many. ♠

3.2 Analytic Preliminaries

The rest of the chapter is devoted to the proof of Lemma 3.1. We first gather together some analytic results.

The *Borel σ-algebra* is the smallest collection of subsets of $[0,1]^d$ that contains all closed subsets and is closed under the operations of taking complements, countable intersections, and countable unions. We only care about the cases $d = 1, 2$. A *Borel set* is a member of the Borel σ-algebra.

Lemma 3.2 *Let* $\phi : [0,1]^2 \to \boldsymbol{R}$ *be a continuous function. Then the set of points in* $(0,1)^2$ *where* $\partial\phi/\partial x$ *exists is a Borel set.*

We give a self-contained proof of this result in §5.1. It is a special case of an extremely general result, [**Z**, Prop. 3.3].

We identify the set of horizontal lines in $[0,1]^2$ with the interval $[0,1]$ in the obvious way.

Theorem 3.3 *Let $S \subset [0,1]^2$ be a set. Let $F_S \subset [0,1]$ be the set of horizontal lines L such that $S \cap L$ is fat. If S is null then F_S is null. If S is a Borel set and F_S is null then S is null.*

This result is a special case of the Fubini-Tonelli Theorem [**F**, p. 65]. We give a self-contained proof in §5.2.

Let $\mathcal{J}$ denote the set of closed intervals of $[0,1]$. Suppose $A : \mathcal{J} \to (0,\infty)$ has the property that $A([0,1]) \geq \sum_j A(J_j)$ when $\{J_j\}$ is a set of disjoint intervals in $[0,1]$. An interval $J \subset [0,1]$ is N-*stretched* if $|A(J)| \geq N|J|$. A point $p \in [0,1]$ is *stretchy* if for any N there is an interval J centered at p that is N-stretched. Otherwise we call p *stiff*.

Theorem 3.4 *The set of stretchy points in $[0,1]$ relative to A is null.*

This result is a bit too idiosyncratic to have a direct proof in the literature, but it is a quick consequence of Lemma 4.6 below, which is a special case of the kind of covering result used in Folland's proof of the Maximal Theorem. See [**F**, §3.4]. We give a self-contained proof of Theorem 3.4 in §5.3.

Now we come to the crucial analytic definition, the definition of an absolutely continuous function.

Definition (AC Functions): Suppose $f : [0,1] \to \boldsymbol{R}$ is continuous. Let $I = \{I_1, ..., I_n\}$ denote a finite list of intervals of $[0,1]$ having pairwise disjoint interiors. We call I a *partial partition*. Let $|I| = \sum |I_k|$. We define I'_k to be the interval bounded by the two points of $f(\partial I_k)$. We define $I' = \{I'_1, ..., I'_n\}$ and $|I'| = \sum |I'_k|$. The function f is *AC* (absolutely continuous) if, for each $\epsilon > 0$, there is some $\delta > 0$ such that $|I| < \delta$ implies that $|I'| < \epsilon$.

Theorem 3.5 *Suppose $f : [0,1] \to \boldsymbol{R}$ is AC. Then*

1. *f is differentiable on the complement of a null set.*
2. *If $f(0) \neq f(1)$ then $f' \neq 0$ on a fat set.*

Theorem 3.5 is an immediate consequence of what Folland calls *The Fundamental Theorem of Calculus for Lebesgue Integrals*. See [**F**, p. 102]. We give a self-contained proof in §5.4.

3.3 Proof of Lemma 3.1

We work with closed intervals and disks, though sometimes consider their interiors. Here is a well-known property of our homeo h. See [**T**, §5.9].

Theorem 3.6 (Disk) *There is a constant K, depending only on h, with the following property. Let $\Delta \subset \boldsymbol{R}^2$ be a disk. Then there are disks D_1, D_2 so that $D_1 \subset h(\Delta) \subset D_2$ and* $\text{diam}(D_2)/\text{diam}(D_1) < K$.

Proof: Suppose $\{\Delta_n\}$ is a sequence where the best ratio for $h(\Delta_n)$ tends to ∞. Let Δ be the unit disk. Composing with homotheties, and passing to a subsequence we get a derived sequence $\{(H_n, h_n)\}$ such that $h_n(0) = 0$ and $\{h_n(1)\}$ converges. The best ratio for $h_n(\Delta)$ tends to ∞. This sequence is tame by Lemma 1.3. So, passing to a subsequence again, we arrange that h_n converges uniformly to a homeo h'. Since h' is a homeo, there are concentric disks D_1, D_2 with $D_1^o \subset h'(\Delta) \subset D_2^o$. Here D_j^o is the interior of D_j. But then $D_1 \subset h_n(\Delta) \subset D_2$ for large n. This is a contradiction. ♠

Now we come to the key geometric idea. I learned this idea from [**LV**]. When $S \subset \boldsymbol{C}$ we define $\alpha(S)$ to be the supremum of all finite sums $\sum \text{area}(D_i)$ where $\{D_i\}$ is a collection of disjoint disks contained in S. For each interval $J \subset [0,1]$ let

$$A(J) = \alpha(h([0,1] \times J)). \tag{19}$$

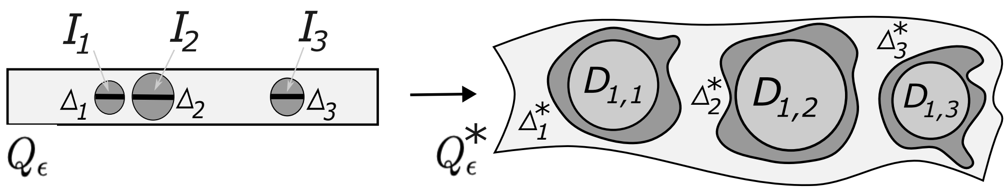
By construction, A satisfies the hypothesis of Theorem 3.4.

Theorem 3.7 *Suppose $y \in (0,1)$ is a stiff point for A. Let $\pi : \boldsymbol{C} \to \boldsymbol{R}$ be any linear projection. Then $\pi \circ h$ is AC on $L_y = [0,1] \times \{y\}$.*

Proof of Lemma 3.1: Let $h_1 = \text{Re}(h)$ and $h_2 = \text{Im}(h)$. Let $D_j \subset (0,1)^2$ be the set where $\partial h_j/\partial x$ exists. By Lemma 3.2, the set D_j is a Borel set. Let L_y be some horizontal line where y is a stiff point. By Theorems 3.7 and 3.5, we see that $D_j^c \cap L_y$ is null. So, $D_j^c \cap L_y$ is null unless y is stretchy, and the stretchy set is null by Theorem 3.4. By Theorem 3.3, D_j^c is null. Hence $\partial h/\partial x$ exists outside of the null set $D_1^c \cup D_2^c$. Let $T \subset [0,1]^2$ be the set where $\partial h/\partial x$ exists and is nonzero. There is a fat set $F \subset [0,1]$ and a linear projection π with the following property. For each $y \in F$, the map $\pi \circ h$ does not identify the endpoints of $[0,1] \times y$. By Theorem 3.5, the set $T \cap L_y$ is a fat subset of L_y for any $y \in F$. By Theorem 3.3, the set T is fat. ♠

Proof of Theorem 3.7: Let $Q = [0,1]^2$. Let $f = \pi \circ h|_{L_y}$. If f is not AC, we can scale so that there is a sequence of partial partitions $\{I^n\}$ with $|I^n| < 1/n$ and $|(I^n)'| \geq K$, the constant from the Disk Theorem. We can subdivide so that the intervals in each partition have the same size up to a factor of 2.

For any set Y let $Y^* = h(Y)$. Fix n and let $I^n = \{I_1, ..., I_k\}$. Let $\epsilon = \max|I_j|$. Since $\epsilon < 2\min|I_j|$, we have $k\epsilon < 2|I|$. Hence $\epsilon < 2/(kn)$. Let $Q_\epsilon = [0,1] \times [y-\epsilon, y+\epsilon]$. Since y is stiff, there is some Ω such that $\alpha(Q_\epsilon^*) < \Omega\epsilon$. Let Δ_j be the open disk having I_j as a diameter. These disks are disjoint. The figure shows the situation when $k = 3$.



Since h is a homeo, the image sets $\{\Delta_j^*\} \subset Q_\epsilon^*$ are also disjoint. Hence

$$\sum_{j=1}^{k} \alpha(\Delta_j^*) \leq \Omega\epsilon. \tag{20}$$

Let $D_{1,j} \subset \Delta_j^* \subset D_{2,j}$ be as in the Disk Theorem. We have

$$\sum_{j=1}^{k} \text{diam}(D_{2,j}) \geq K, \quad \sum_{j=1}^{k} \text{diam}(D_{1,j}) \geq 1, \quad \sum_{j=1}^{k} \text{area}(D_{1,j}) \geq \frac{1}{4k}. \tag{21}$$

The first equation, which comes from $|I'| \geq K$, implies the second equation, and the second equation implies the third. Hence

$$\frac{2\Omega}{kn} = \Omega\epsilon > \sum_{j=1}^{k} \alpha(\Delta_j^*) \geq \sum_{j=1}^{k} \text{area}(D_{1,j}) \geq \frac{1}{4k}. \tag{22}$$

So $n < 8\Omega$, a contradiction when n is large enough. ♠

4 A Primer on Measure Theory

4.1 Measurability

This chapter proves some standard measure-theoretic results. The material in this section can be found in [**F**, §1].

We work in $[0,1]^d$. We only care about $d = 1, 2$. A *dyadic cube* in $[0,1]^d$ is a cube we get by starting with $[0,1]^d$ and recursively subdividing cubes into 2^d equal-sized sub-cubes and selecting one of them. Two dyadic cubes are either nested or else have disjoint interiors. Say that a *carpet* is a countable union of dyadic cubes having pairwise disjoint interiors. When $S \subset [0,1]^d$ we define the *outer measure*

$$\mu(S) = \inf_{S \subset \bigcup Q_n} \sum \text{volume}(Q_j) \tag{23}$$

The infimum is taken over all carpets containing S.

A subset $S \subset [0,1]^d$ is *open* iff the complement $S^c = [0,1]^d - S$ is closed. Every open subset of $[0,1]^d$ is the countable union of dyadic cubes. In particular, this is true for open balls. Using this fact it is an easy exercise to show that S is null iff $\mu(S) = 0$, and fat iff $\mu(S) > 0$. We also have:

- *monotonicity:* $\mu(A) \leq \mu(B)$ for any $A \subset B$.
- *subadditivity* If $S = \bigcup T_n$ then $\mu(S) \leq \sum \mu(T_n)$.

A subset $S \subset [0,1]^d$ is *measurable* if

$$\mu(E \cap S) + \mu(E \cap S^c) = \mu(E) \tag{24}$$

for all subsets $E \subset [0,1]^d$. Note that S is measurable iff S^c is measurable. It follows directly from monotonicity and subadditivity that null sets are measurable.

We defined Borel sets in §3.2.

Theorem 4.1 *Let S be a Borel set. Then the following is true.*

1. *S is measurable.*
2. *If $S = \bigcup S_n$, an increasing union of Borel sets, then $\mu(S) = \lim \mu(S_n)$.*
3. *For any $\epsilon > 0$ we have a compact set K and an open set U such that $K \subset S \subset U$ and $\mu(U - K) < \epsilon$.*

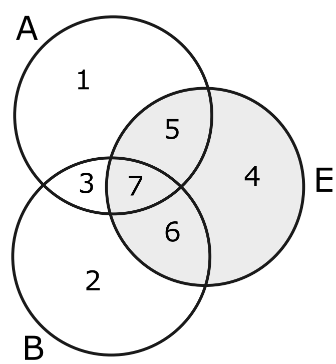
We prove this result through a series of lemmas.

Lemma 4.2 *Dyadic cubes are measurable.*

Proof: Let A be a dyadic cube. Let $E \subset [0,1]^d$ be an arbitrary set. Any carpet containing E can be further subdivided so that each of its cubes is contained in either A or A^c. The union of the former gives a sub-carpet covering $E \cap A$ and the union of the latter gives a sub-carpet covering $E \cap A^c$. Since this holds for all covers of E we have $\mu(E) = \mu(E \cap A) + \mu(E \cap A^c)$. ♠

Lemma 4.3 *If A, B are measurable then so are $A \cap B$ and $A \cup B$ and $A - B$.*

Proof: Note that A^c and B^c are also measurable. Since $(A \cup B)^c = A^c \cap B^c$ and $A - B = A \cap B^c$, it suffices to prove that $A \cap B$ is measurable. Let R_k be the region shown in the Venn diagram for A, B, E as in the figure. Let $[k_1 ... k_m] = \mu(R_{k_1}) + ... + \mu(R_{k_m})$. We want to prove $[4567] = [456] + [7]$.



Since B is measurable, $[4567] = [45] + [67]$. Since A is measurable, we have $[67] = [6] + [7]$. Using these equations and sub-additivity, we have

$$[4567] \leq [456] + [7] \leq [45] + [6] + [7] = [45] + [67] = [4567].$$

So, we have equalities everywhere and we have proved what we want. ♠

We note in particular that when A and B are measurable and $A \subset B$ we have $\mu(B \cap A) + \mu(B - A) = \mu(B)$, or $\mu(B - A) = \mu(B) - \mu(A)$. Also, by induction, finite unions and intersections of measurable sets are measurable.

Lemma 4.4 *The countable union of measurable sets is measurable.*

Proof: Suppose $S = \bigcup S_n$ is a countable union of measurable sets. Since finite unions of measurable sets are measurable, we can suppose that we have an increasing union. Now we define $T_n = S_n - S_{n-1}$ and $T_1 = S_1$. Since these sets are all measurable we have $\mu(T_n) = \mu(S_n) - \mu(S_{n-1})$. Hence

$$\sum_{k=1}^{n} \mu(T_k) = \mu(S_n) \leq \mu(S) \leq \sum_{k=1}^{\infty} \mu(T_k). \tag{25}$$

The first inequality is monotonicity and the last is subadditivity. Letting $n \to \infty$ and using the fact that a bounded monotone sequence converges,

$$\sum_{k=1}^{\infty} \mu(T_k) = \lim \mu(S_n) \leq \mu(S) \leq \sum_{k=1}^{\infty} \mu(T_k). \tag{26}$$

But then, in fact, we have equality everywhere in Equation 26. In particular the series converges. Hence we can choose n so large that

$$\mu(S_n^c - S^c) = \mu(S - S_n) = \mu\bigg(\bigcup_{k=n}^{\infty} T_k\bigg) \leq \sum_{k=n}^{\infty} \mu(T_k) < \epsilon, \tag{27}$$

Combining Equation 27 with the measurability of S_n and sub-additivity,

$$\mu(E) \leq \mu(E \cap S) + \mu(E \cap S^c) \leq \mu(E \cap S_n) + \mu(E \cap S_n^c) +$$
$$\mu(S - S_n) + \mu(S_n^c - S^c) \leq \mu(E) + 2\epsilon.$$

Since ϵ is arbitrary, we have $\mu(E) = \mu(E \cap S) + \mu(E \cap S^c)$. ♠

Proof of Statement 1: Any open set is a countable union of dyadic cubes. So, Lemma 4.2 and Lemma 4.4 say that open sets are measurable. Hence closed sets are measurable. This fact combines with Lemma 4.4 to show that all Borel sets are measurable. ♠

Proof of Statement 2: This follows from Equations 25 and 26. ♠

Proof of Statement 3: We get open $U \supset S$ with $\mu(U) - \mu(S) < \epsilon/2$ by taking a suitable cover of S by open cubes and then taking their union. We get a compact $K \subset S$ by applying the same construction to S^c and then taking complements. The measurability of all sets involved guarantees that $\mu(S) - \mu(K) < \epsilon/2$. Combining these, and again using measurability, we have $\mu(U - K) = \mu(U) - \mu(K) < \epsilon$. ♠

4.2 Baby Fubini

Here we prove a weak special case of the Fubini-Tonelli Theorem [**F**, p. 65] which is sufficient for Theorem 3.3. We identify the set of horizontal lines of $[0,1]^2$ with $[0,1]$.

Lemma 4.5 *Suppose S is measurable. Let $F_{t,S}$ denote the set of horizontal lines L such that $\mu(L\cap S)>t$. If $\mu(S)<t^2$ then $\mu(F_{t,S})\leq t$.*

Proof: In view of Statement 3 of Theorem 4.1, and the corresponding result for null sets, it suffices to prove this result when S is open. If S is a finite union of dyadic squares, we subdivide and reduce to the case when they all have the same size and belong to a common grid of $[0,1]^2$. In this case, we get the result by counting. In general $S=\bigcup Q_n$ is a carpet. Also, the intersection $L\cap S$ is a carpet for each horizontal line L. Let Q^n be the union of the first n squares. By the finite case, $\mu(F_{t,Q^n})\leq t$. By Statement 2 of Theorem 4.1, $\mu(\bigcup F_{t,Q^n})\leq t$ and $\mu(L\cap S)=\lim\mu(L\cap Q^n)$ for each horizontal line L. Hence $F_{t,S}\subset\bigcup F_{t,Q^n}$. Hence $\mu(F_{t,S})\leq t$. ♠

4.3 Covering Results

Our proofs here take their ideas from [**F**, §3.4]. We work with closed intervals. A *Besicovich cover* $\mathcal{B}$ of $S\subset[0,1]$ is a union of intervals, such that each point of S is centered on some interval of $\mathcal{B}$ and each interval of $\mathcal{B}$ is centered on some point of S. Let $|I|$ denote the length of an interval I.

Lemma 4.6 *Suppose $\mathcal{B}$ is a Besicovich cover of S. Then there exists a subset of $\mathcal{B}$ consisting of disjoint intervals whose total length-sum is at least $\mu(S)/3$.*

Proof: Greedily choose intervals, always picking a largest one that is disjoint from the previous ones picked. Let $\{I_j\}$ be this collection. Let $S'=\bigcup I_j$. Let $3I_j$ be the interval obtained by dilating I_j by a factor of 3 about its midpoint. Any $x\in S-S'$ is the center of some J of $\mathcal{B}$ not picked by the algorithm. But then J intersects some I_j with $|I_j|\geq|J|$. This forces $x\in 3I_j$. Hence $\{3I_j\}$ covers S. Hence $\sum|3I_j|\geq\mu(S)$. Hence $\sum|I_j|\geq\mu(S)/3$. ♠

An interval J is *δ-porous* if $\mu(J\cap K)<(1-\delta)|J|$. A point $p\in[0,1]$ is *δ-porous* if p is the center of arbitrarily small δ-porous intervals. Finally, K is *porous* if for some $\delta>0$ every point of K is δ-porous.

Lemma 4.7 (Porous) *If S is measurable and porous then $\mu(S) = 0$.*

Proof: Suppose $\mu(S) > 0$. By Statement 3 of Theorem 4.1 we can find $K \subset S \subset U$ with K compact, U open, and $\mu(U - K) < \epsilon$. If we take ϵ small, we have $\mu(K) > 0$. The set K remains porous.

Since K is compact, there is some $\lambda > 0$ such that every interval of length λ centered at a point of K lies in U. We take a Besicovich covering $\mathcal{B}$ of K by δ-porous intervals all of length less than λ. Letting $\{I_j\}$ be as in Lemma 4.6, we have $\delta_j := \mu(I_j \cap U) - \mu(I_j \cap K) \geq |I_j| - (1-\delta)|I_j| = \delta|I_j|$. From this calculation, we see that the set $U - K$ intersects $\bigcup I_j$ in a set of size at least $\sum \delta_j \geq \delta\mu(K)/3$. This is a contradiction for small enough ϵ. ♠

We say that a covering $\mathcal{B}$ by intervals is *renewable* if, for every $\epsilon > 0$, every point of S is the (left or right) endpoint of an interval in $\mathcal{B}$ having length less than ϵ. Let $S\Delta T = (S - T) \cup (T - S)$.

Theorem 4.8 (Vitali) *Suppose $\mathcal{B}$ is a renewable cover for a Borel set S. For any $\epsilon > 0$ there is a disjoint collection of intervals $\{I_j\}$ of $\mathcal{B}$ such that $\mu(S\Delta T) < \epsilon$ where $T = \bigcup I_j$.*

Proof: If S is null, the result is trivially true. So, assume $\mu(S) > 0$. By Statement 3 of Theorem 4.1 we have $K \subset S \subset U$ where K is compact and U is open and $\mu(U - K) < \epsilon/2$. There is some $\lambda > 0$ so that every interval of length λ having an endpoint in K belongs to U. We make $\{I_j\}$ by length-greedily picking disjoint intervals of $\mathcal{B}$, having endpoints in K and length less than λ. Let $T = \bigcup I_j$. By monotonicity, $\mu(T - S) < \epsilon/2$.

Let $p \in K' = K - T$. Let I be an interval in $\mathcal{B}$ having length less than λ and p as an endpoint. Let J be the interval centered at p having length $8|I|$. The interval I was not picked in the algorithm, so there is some interval I_j with $|I_j| \geq |I|$, such that p is at most $|I_j|$ away from an endpoint of I_j. But then $\mu(I_j \cap J) \geq |I| = |J|/8$. Hence $\mu(K' \cap J) \leq (7/8)|J|$. Hence p is $(1/8)$-porous with respect to K'. Hence K' is $(1/8)$-porous. Also, K' is a Borel set and hence measurable. We conclude from the Porous Lemma that $\mu(K') = 0$. Hence $\mu(S - T) \leq \mu(S - K) + \mu(K') < \epsilon/2$. This combines with $\mu(T - S) < \epsilon/2$ to give $\mu(S\Delta T) < \epsilon$. ♠

5 Proofs of the Analytic Results

5.1 Proof of Lemma 3.2

Let $\phi : [0,1]^2 \to \boldsymbol{R}$ be a continuous map. We extend ϕ to be continuous on $\boldsymbol{R}^2$. Let $D \subset (0,1)^2$ be the set where $\partial\phi/\partial x$ exists. Let

$$\phi_\epsilon(x,y) = \frac{\phi(x+\epsilon,y)-\phi(x,y)}{\epsilon} \tag{28}$$

Let's call ϕ *rationally differentiable* at (x,y) if $\{\phi_\epsilon(x,y)\}$ is a Cauchy sequence relative to any sequence of rational ϵ with $\epsilon \to 0$. Because ϕ is continuous, ϕ is differentiable at (x,y) if and only if ϕ is rationally differentiable at (x,y).

In other terms, this is true iff for each $n \geq 1$ there is an integer $m \geq 1$ such that if $0 < |a|, |b| < 1/m$ are rational then $(x,y) \in U_{m,a,b,n}$, the set such that $|\phi_a(x,y) - \phi_b(x,y)| < 1/n$. Since ϕ_a and ϕ_b are continuous, $U_{m,a,b,n}$ is an open set, and hence a Borel set. So, D is a Borel set because

$$D = \bigcap_{n=1}^{\infty}\bigcup_{m=1}^{\infty} U_{m,n}, \qquad U_{m,n} = \bigcap_{0<|a|,|b|<1/m} U_{m,a,b,n}.$$

5.2 Proof of Theorem 3.3

For any set Y we let Y^c denote its complement in the relevant domain. Given $S \subset [0,1]^2$ we let F_S and $F_{t,S}$ respectively denote the set of horizontal lines L such that $\mu(L \cap S) > 0$ and $\mu(L \cap S) > t$.

Suppose $\mu(F_S) > 0$. Then there is some $t > 0$ such that $\mu(F_{t,S}) > t$. If $\mu(S) = 0$ then also $\mu(S) < t^2$ and we contradict Lemma 4.5.

Suppose S is a Borel set and $\mu(S) > 0$. Then $\mu(S^c) < t^2$ for some $t < 1$. By Lemma 4.5, we have $\mu(F_{t,S^c}) \leq t$. But if $L \notin F_{t,S^c}$ then $\mu(L \cap S^c) \leq t$, and hence $\mu(L \cap S) \geq 1-t$. So $(F_{t,S^c})^c \subset F_{1-t,S}$. Therefore $\mu(F_{1-t,S}) \geq 1-t$. Since $F_{1-t,S} \subset F_S$, we have $\mu(F_S) > 0$. Hence, if $\mu(F_S) = 0$ then $\mu(S) = 0$.

5.3 Proof of Theorem 3.4

Let S be the stretchy set for $A : \mathcal{J} \to (0,\infty)$. If S is fat then $\mu(S) > 1/N$ for some N. Let $\mathcal{B}$ be a Besicovich cover of S by N^2-stretched intervals. Let $\{J_j\}$ be the set of disjoint intervals produced by Lemma 4.6. For large N we have a contradiction: $A([0,1]) \geq \sum A(J_j) \geq N^2\mu(S)/3 \geq N/3$. So, S is null.

5.4 Proof of Theorem 3.5

Proof Outline: First we give the proof modulo four auxiliary lemmas. Lemma 5.2 below shows that an AC function is the difference between two monotone AC functions. Lemma 5.3 below shows that a monotone AC function is differentiable except on a null set. These two results combine to show that an arbitrary AC function is differentiable away from a null set.

Let f be an AC function and let A be the set where $f' = 0$. Let $B = [0,1] - A$. Suppose B is null. Lemma 5.1 below says in particular that the image of a null set under an AC map is null. Hence $f(B)$ is null. By Lemma 5.4 below, the set $f(A)$ is null. But then $f([0,1]) = f(A) \cup f(B)$ is both null and connected. Hence $f([0,1])$ is a single point. In particular, $f(0) = f(1)$. Hence, if $f(0) \neq f(1)$ then B is fat.

Now we will take care of the details of these lemmas. For convenience we repeat the definition of an AC function. Suppose that $f : [0,1] \to \boldsymbol{R}$ is a continuous function. Let $I = \{I_1, ..., I_n\}$ denote a finite list of intervals of $[0,1]$ having pairwise disjoint interiors. We call I a *partial partition*. Let $|I| = \sum |I_k|$. We define I'_k to be the interval bounded by the two points of $f(\partial I_k)$. We define $I' = \{I'_1, ..., I'_n\}$ and $|I'| = \sum |I'_k|$. The function f is *AC* (absolutely continuous) if, for each $\epsilon > 0$, there is some $\delta > 0$ such that $|I| < \delta$ implies that $|I'| < \epsilon$.

Lemma 5.1 *Let f be an AC function. Let $S \subset [0,1]$. For each $\epsilon > 0$ there is some $\delta > 0$ such that $\mu(S) < \delta$ implies that $\mu(f(S)) < \epsilon$. Hence, if S is null then $f(S)$ is null.*

Proof: Suppose $\mu(S) < \delta$. We can find an open set U such that $S \subset U$ and $\mu(U) < \delta$. Note that U is a countable union of intervals with disjoint interiors. Let U^n denote the union of the first n intervals of U. We construct a partial partition $I = I^n$ as follows. For each compact connected component C of closure(U^n), we include in I^n an interval connecting points of C where f respectively achieves its min and max. By construction, $|I| < \delta$ and $f(U^n) \subset V^n := \bigcup I'_k$. Choosing δ small enough, we have $|I'| < \epsilon$. Hence $\mu(V^n) < \epsilon$. Also by construction $V^n \subset V^{n+1}$ for all n. Since $V = \bigcup V^n$ is an increasing union of sets, each a finite union of intervals, $\mu(V) \leq \epsilon$. But $f(S) \subset V$. Hence $\mu(f(S)) < \epsilon$. ♠

Our next lemma is usually stated for functions of *bounded variation*. In the notation of the lemma below, this means that $v(f,[0,1]) < \infty$. Here we have stronger hypotheses and a stronger conclusion.

Lemma 5.2 *If f is AC then $f = f_+ - f_-$ where $f_\pm$ is monotone and AC.*

Proof: Given an interval $Y \subset [0,1]$ we define the *variation* $v(f,Y)$ to be the supremum of $|I'|$ taken over all partial partitions I of Y.

We claim first that $v(f,[0,1]) < \infty$. If $v(f,[0,1]) = \infty$ then for any $\delta > 0$ we can find an interval Y such that $|Y| < \delta$ and $v(f,Y) \geq 1$. This contradicts the fact that f is AC. Hence $v(f,[0,1]) < \infty$.

Given the monotonicity properties of the variation, we conclude that $v(f,Y) < \infty$ for every interval $Y \subset [0,1]$. We also note an additivity property: For $0 \leq a < b$ we have

$$v(f,[0,b]) = v(f,[0,a]) + v(f,[a,b]). \tag{29}$$

Let $f_+(x) = v(f,[0,x])$. By construction, f_+ is increasing. By Equation 29,

$$f_+(b) - f_+(a) = v(f,[a,b]) \geq f(b) - f(a).$$

Hence $f_- = f_+ - f$ is also increasing.

It remains to show that f_+ and f_- are AC. Since f is AC, and sums and differences of AC functions are AC, it suffices to prove that f_+ is AC. Let us reformulate the definition of AC first. Instead of defining $|I'_k|$ as the distance between the two points of $f(\partial I_k)$ we might have defined $|I'_k| = v(f,I_k)$. Since the AC criterion is already defined in terms of taking a supremum over all partial partitions, the inclusion of the further supremum in the definition of $|I'_k|$ changes nothing. This alternate definition picks out the same class of functions as AC. But this definition is formulated entirely in terms of the variation function. Now, by Equation 29,

$$v(f_+,[a,b]) = f_+(b) - f_+(a) = v(f,[a,b]).$$

Since f and f_+ have the same variation function and f is AC, so is f_+. ♠

The next result is the workhorse in our proof of Theorem 3.5. It actually holds for all monotone functions. See [**Fa**]. The case of monotone AC functions is much easier, though the core idea is the same as in the general case.

Lemma 5.3 *If f is monotone and AC then f' exists outside a null set.*

Proof: Adding a linear function to f, we reduce to the case when f is strictly monotone. Let $E_{a,b}$ denote the set of points $p \in [0,1]$ such that

1. For all $n \in \boldsymbol{N}$ the point p is the endpoint of an interval I such that $\mu(I) < 1/n$ and $\mu(f(I)) < a\mu(I)$.
2. For all $n \in \boldsymbol{N}$ the point p is the endpoint of an interval I such that $\mu(I) < 1/n$ and $\mu(f(I)) > b\mu(I)$.

To prove this lemma, it suffices to prove that $\mu(E_{a,b}) = 0$ for all $0 \leq a < b$. Suppose some $S = E_{a,b}$ has $\mu(S) > 0$. Since f is continuous, $S = E_{a,b}$ is defined by a countable collection of open conditions. Hence S is a Borel set.

Let $\mathcal{B}$ be the renewable cover of S made from the intervals in Item 1. For any $\delta > 0$ let $\{I_j\}$ and $T = \bigcup I_j$ be as in Theorem 4.8 so that $\mu(S\Delta T) < \delta$. By construction $\mu(f(T)) \leq a\mu(T)$. By Lemma 5.1, we can make $\mu(S\Delta T)$ and $\mu(f(S\Delta T))$ as small as we like by shrinking δ. Hence $\mu(S) \leq a\mu(S)$.

Running the same argument with Item 2 in place of Item 1, we get $\mu(S) \geq b\mu(S)$. Since $b\mu(S) \leq a\mu(S)$ and $a < b$ we have $\mu(S) = 0$. ♠

Our final result does not need the AC hypothesis.

Lemma 5.4 *Suppose $f : [0,1] \to \boldsymbol{R}$ is continuous. Suppose also that f' exists and equals 0 on a set $A \subset [0,1]$. Then $f(A)$ is null.*

Proof: Each $x \in A$ is the midpoint of an interval I_x such that $|f(I_x)| < \epsilon |I_x|$. Let $\mathcal{B}$ be the Besicovich cover made from these intervals. We introduce an auxiliary Besicovich cover $\mathcal{B}^*$ of A, obtained by shrinking all the intervals in $\mathcal{B}$ by a factor of 3 about their midpoints. Let $\{I_j^*\}$ be the disjoint collection of intervals of $\mathcal{B}^*$ produced by the greedy algorithm. Our proof of Lemma 4.6 shows that the corresponding union of dilated intervals $\{I_j\}$ is a cover of A. By construction, $\sum |I_j| \leq 3$. But now we can say $\mu(f(A)) < 3\epsilon$. Since ϵ is arbitrary, $f(A)$ is null. ♠

6 Discussion

Usually one defines a compact hyperbolic manifold to be a compact Riemannian manifold that is locally isometric to $\boldsymbol{H}^3$. By the Cartan-Hadamard Theorem, this more typical definition agrees with the definition in §1.1.

Our proof showed that the extension map h lies in $\boldsymbol{I}$. Hence the isometry from M_1 to M_2 we get induces the same isomorphism from $\pi_1(M_1)$ to $\pi_1(M_2)$ that the BL map f does.

If we just assume that $f : M_1 \to M_2$ is a homotopy equivalence, then the lifted map H is still a quasi-isometry. Theorem 1.2 still works in this case. So two homotopy equivalent compact hyperbolic 3-manifolds are isometric.

The same proof works for compact hyperbolic n-manifolds when $n \geq 4$. The key change is in Theorem 3.7. When $n = 4$ we would let $\mathcal{J}$ denote the set of axis-aligned squares of $[0, 1]^2$ instead of the set of intervals of $[0, 1]$ and then Equation 19 would be in terms of volume rather than area.

The proof also works for finite volume hyperbolic n-manifolds, when we have $n \geq 3$. In the finite volume case, you have to avoid zooming into cusps.

7 References

[**F**] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, John Wiley & Sons, New York (1984).

[**Fa**] C. Faure, *An elementary proof of the differentiability almost everywhere of monotone functions*, Real Analysis Exchange **15** (1989–90)

[**GS**] S. Gouëzel and V. Shchur, *A corrected quantitative version of the Morse lemma*, Journal of Functional Analysis **277**, pp. 1258–1268 (2019).

[**LV**] O. Lehto and K. I. Virtanen, *Quasiconformal Mappings in the Plane*, Second Edition, Springer–Verlag, New York (1973).

[**M**] G. D. Mostow, *Quasi-conformal mappings in n-space and the rigidity of hyperbolic space forms*, Inst. Hautes Études Sci. Publ. Math. No. 34, pp. 53–104 (1968).

[**S**] R. E. Schwartz, *The quasi-isometry classification of rank one lattices*, Inst. Hautes Études Sci. Publ. Math. No. 82, pp. 133–168 (1995).

[**T**] W. P. Thurston, *The Geometry and Topology of Three-Manifolds*, Princeton University Lecture Notes (1978).

[**Z**] L. Zajíček, *Fréchet differentiability via partial Fréchet differentiability*, Comment. Math. Univ. Carolin. **64**, No. 2, pp. 185–207 (2023).