Hyperbolic surfaces

A **hyperbolic surface** is a metric space with some additional properties: it has the **shortest length property** and every point has an open neighborhood that is isometric to an open neighborhood of H^2. These are of interest to us because they allow us to characterize the topologies of surfaces which locally look like H^2.

Rectifiable curves

For a curve $\gamma : [a, b] \to X$ in a metric space X, the curve is rectifiable if γ is continuous and sums of the form

$$\sum_{i=1}^{n} d(\gamma(a_{i-1}), \gamma(a_i))$$

are bounded for the subdivision $a = a_0 \leq a_1 ... \leq a_n = b$. If the curve is rectifiable, then it has a finite length.

Shortest length property

The length of a rectifiable curve, $l(\gamma)$, is defined to be the supremum of these sums over all subdivisions of the interval. A metric space with shortest length property is one in which any two points can be joined by a rectifiable curve and

$$d(x, y) = inf_\gamma \ell(\gamma)$$

. For H^2, the shortest length is the arc length of the geodesic.

Lemma

Let X be a metric space with shortest length property.
Let D be an open disc in H^2 with center a and radius r. We will subsequently denote such hyperbolic discs as $D(a; r)$. For any isometry $\sigma : D \to X$ of D onto an open neighborhood of $\sigma(a)$ in X, $d(\sigma(a), x) \geq r$ for all points $x \in X - \sigma(D)$.

Since X is a metric space with shortest length property, we can find a rectifiable curve γ from $\sigma(a)$ to x. In D, find a $\rho \in (0, r)$ such that we have a compact, closed disc in D, which we will call K. K has a compact, closed image in $\sigma(D)$, and so its complement, K^c is open. Since X is connected, γ meets $\sigma(K - \partial K)$, $\sigma(\partial K)$, and $X - \sigma K$. γ must meet $\sigma(K - \partial K)$ at some y on the boundary, so the length of this curve from $\sigma(a)$ to y must be greater than the length of the entire curve. So,

$$l(\gamma) \geq d(\sigma(a), y) = \rho$$

$$sup_\rho \ell(\gamma) \geq sup_\rho d(\sigma(a), y) = r$$

$$inf_\gamma \ell(\gamma) \geq r$$

So, we have our result by the definition of the shortest length property.

□

Proposition

Let X be a connected Hausdorff space, Y a hyperbolic surface and f:X→ Y a local homeomorphism. Then, $d(u, v) = inf_\gamma \ell(f\gamma)$ defines a metric on X such that f is a local isometry and X is a hyperbolic surface.

A local homeomorphism is a continuous map such that each point $x \in X$ has an open neighborhood which is mapped to an open neighborhood around f(x) in Y. If X and Y are both metric spaces, then this is a local isometry.

First, we need to show that any u and v in X can indeed be joined by a curve whose image in Y is rectifiable. We can do this by fixing a point u in X and dividing X into two subsets X_1 such that the points $v \in X_1$ can be joined to u in X by a curve γ such that $f\gamma$ is rectifiable in Y, and X_2 of points which cannot be joined by such a curve. We will show that both of these sets are open, and that since X is a connected space, one of the sets must be empty. X_1 is not empty, since it contains u (which can trivially be joined to itself) and an open neighborhood around u whose image behaves like H^2. Suppose there exists a point v in X_2. Then, there exists an open neighborhood around v in which points can be joined to v but cannot ultimately be joined to u such that the image of the curve is rectifiable in Y. Since X_2 is open, it must be empty, as its complement is also open but nonempty. Therefore, all points u and v in X can be joined by such curves.

To show that $d(u,v) = inf_\gamma \ell(f\gamma)$ defines a metric on X, we first need to show that distances between distinct points in X are not 0. We can do this using the preceding lemma. Fixing a point x in X, we can find an open neighborhood W around x which is homeomorphic to a hyperbolic disc $D(f(x);r)$ in Y, such that points in W can be joined to x by a geodesic and thus the curves have non-zero length. For points outside W, we know from the lemma that they are at least distance r from x.
We also need to show that $d(u,v) = inf_\gamma \ell(f\gamma)$. Choosing u and v in W, we can find a curve λ such that its image is the geodesic joining $f(u)$ to $f(v)$, which we will call $f\lambda$. We already know that

$$inf_\lambda \ell(f\lambda) \geq inf_{f\lambda} \ell(f\lambda) = d(f(u), f(v))$$

We need only to prove the opposite inequality. This is simple, since $d(f(u), f(v)) = \ell(f\lambda) \geq inf_\lambda \ell(f\lambda)$.
Finally, to show that X is also a hyperbolic space, we still need to show that it has the short length property. For a curve $\gamma : [a,b] \to X$ in X, we can use the Heine-Borel theorem to subdivide [a,b] such that the curve in each interval is in a hyperbolic disc like W. Since in each of the W, which are isometric to H^2, we can find the length of the curve equivalent to the length in their image in Y, we conclude by summing these curve lengths that $\ell(\gamma) = \ell(f\gamma)$. □

The Hopf-Rinow Theorem

The Hopf-Rinow theorem essentially allows us to join two points x and y on a complete hyperbolic surfaces with a geodesic curve of length equal to the distance between x and y. We need two notions here: First, we will be thinking about curves primarily as the parametrized path of a point on a surface. Intuitively, then, the "time" it takes to travel from one endpoint to the other will be at least as long as the distance between the endpoints:

$$d(\gamma(s), \gamma(t)) \leq |t - s|$$

Let us make an analogy: Suppose our surface is a circle. On small intervals, the distance between the endpoints is equal to the distance between the parameters, but as our point travels farther, the distance between the endpoints (defined, of course, as the length of the shortest curve between the endpoints) may now be less than the length of the path.

The second notion we will be dealing with is that of completeness. A complete hyperbolic space is one on which Cauchy sequences converge. This means that we would be able to find the exact length of a curve on our surface – consider the Poincaré disk which is missing the center point. This is still a hyperbolic space, but without the center point, if we were to attempt to find a distance between $-\frac{1}{2}$ and $\frac{1}{2}$, we can only find a Cauchy sequence of geodesic curves which have lengths that approximate the "distance" more and more closely but never actually reaches the "distance" – in fact, there would not be any distance to speak of. On a complete surface, the distance between x and y is the length of the geodesic connecting x to y. Therefore, we will see that it is important when we get to proving the Hopf-Rinow theorem that we are working on complete hyperbolic spaces.

The Hopf-Rinow Theorem continued on next page. . .

Proposition 1

Let X be a hyperbolic surface, not necessarily complete. Let $\alpha, \beta : \Re \to X$ agree on some nonempty open interval $J \subseteq \Re$. Then, α and β agree on all of $\Re$. Returning to the parametrization perspective, this is equivalent to saying that any two particles that have a segment of their paths in common have the same pasts and the same futures.

Suppose without loss of generality that 0 in in our interval J. Our strategy will require us to show that the set of points $S := \{s|\alpha, \beta \text{ agree on } [0, s]\}$ is unbounded, and therefore $S = \Re$. Let us assume S is bounded. Since α and β agree on a non-empty open interval, we can find a small open disc D with radius r around s in the interval on which α and β agree. Now, α and β agree on [0,s+r]. We can continue this process to $+\infty$ and similarly to $-\infty$, and so S is unbounded. □

Proposition 2

Let X be a complete hyperbolic surface. Then, the geodesic curve $\gamma : J \to X$ can be extended to $\Re \to X$.

Suppose again without loss of generality that $0 \in J$, and let E be the set of points $e \in [0, +\infty)$ such that γ extends to [0,e). Since $d(\gamma(s), \gamma(t)) = |t - s|$ on small intervals and Cauchy sequences converge on X, we know that the curve is uniformly continuous and e is actually in E's interval. Then, we can find a disc around $\gamma(e)$ which is hyperbolic, and in which we can find a geodesic that extends the curve. In this way, we can extend the interval to $\Re$. □

Lemma

In order to prove the Hopf-Rinow Theorem, we will also need a distance preservation lemma. This lemma says that if a curve $\gamma[a, b] \to X$ and $\gamma[b, c] \to X$ are distance preserving, then $\gamma[a, c] \to X$ is also distance preserving.

Suppose we find an s in [a,b] and t in [b,c]. Then, by the Triangle inequality, we have

$$d(\gamma(s), \gamma(t)) \leq d(\gamma(s), \gamma(b)) + d(\gamma(b), \gamma(t)) = (b - s) + (t - b) = t - s$$

$$c - a = (c - b) + (b - a) = d(\gamma(a), \gamma(b)) + d(\gamma(b), \gamma(c)) \leq d(\gamma(a), \gamma(s)) + d(\gamma(s), \gamma(t)) + d(\gamma(t), \gamma(c))$$

$$c - a \leq (s - a) + d(\gamma(s), \gamma(t)) + (c - t)$$

$$t - s \leq d(\gamma(s), \gamma(t))$$

which gives us the equality. □

continued on next page...

Hopf-Rinow Theorem

Let X be a complete hyperbolic surface. For two points x,y in X with d(x,y) = d, there exists a geodesic curve σ such that $\sigma(0) = x$ and $\sigma(d) = y$

For our proof, we will be using the notation $HR(x, z, y)$ to indicate that x and z satisfy the Hopf-Rinow theorem and that d(x,y) + d(y,z) = d(x,z). The general scheme of the proof is to show that we can "build" a sequence of geodesic curves from x to z such that curve with the minimum total distance, does in fact coincide with the geodesic curve from x to z.

Our first goal is to show that for any x and y in X, we can indeed find a z ≠ x such that $HR(x, z, y)$. In other words, Choosing an open hyperbolic disc around x, we will find a compact circle S of radius r around x, which we will denote $S(x; r)$. On S, choose a point z that minimizes $d(z, y)$. Of course, $d(x, z) = r$.

Consider the rectifiable curve $\gamma : [a, c] \to X$, with $\gamma(a) = x$ and $\gamma(c) = y$. Choose $b \in [a, c]$ such that $d(\gamma(a), \gamma(b)) = r$. Then,

$$\ell(\gamma) = \ell(\gamma_{(a,b)}) + \ell(\gamma_{(b,c)})$$

Since distance is the infimum of the rectifiable curves,

$$\ell(\gamma_{(a,b)}) + \ell(\gamma_{(b,c)}) \geq d(\gamma(a), \gamma(b)) + d(\gamma(b), \gamma(c))) \geq r + d(z, y)$$

This shows that we cannot find any curves from x to y that are shorter than $r + d(z, y)$.

Now we need to further extend our curve into [y,z]. To accomplish this, we will show

$$HR(x, z, y) \,\&\, HR(z, w, y) \Rightarrow HR(x, w, y)$$

By the Triangle Inequality, and using both of our initial statements,

$$d(x, w) + d(w, y) \leq d(x, z) + d(z, w) + d(w, y) = d(x, y)$$

However, again applying the Triangle inequality yields

$$d(x, w) + d(w, y) \geq d(x, y)$$

and since by our lemma, this kind of sum preserves distance, we have an equality and $HR(x, w, y)$. Finally, let's pick a point $v \neq x$ such that $HR(x, v, y)$, and let $e = d(x, v)$. Then, using proposition 2, we can construct a unique geodesic such that $\sigma : \Re \to X$ with $\sigma(0) = x$ and $\sigma(e) = v$. If we can show that this geodesic is in fact the same one on which our other points lie, then we have proven the theorem. In other words, we need to show that the set S of points in [0,d] such that $d(x, \sigma(s)) = s$ and $d(\sigma(s), y) = d - s$ is all of [0,d].

We know already that 0 is in S. We also know that z is in S from the first part of our proof (such that the s solving $\sigma(s) = z$ is the b∈[a,c]). Now we see that any point w ≠ z such that $HR(z, w, y)$ is also on the geodesic, and S is the whole interval [0,d].

□

Uniformization

The Uniformization Theorem shows us that any complete hyperbolic surface can be "constructed" by means of a unique local isometry from H^2 to X. The theorem uses the notion that plane figures cannot be isometric without also being congruent, which will allow us to cover X using congruent triangles from H^2.

Uniformization theorem

Let D we an open disc in H^2. Any isometry $\phi : D \to X$, a complete hyperbolic space, can be uniquely extended to a local isometry $H^2 \to X$.

First, let O be the center of D. By extending all of the geodesic curves in D through O to the real line, as we have shown is possible in the previous section, we will have a map $\phi : H^2 \to X$. What we have in X, then, essentially is the image of D such that the geodesics have been extended to curves beyond the image of D (and so are not necessarily geodesics).

Now, choosing another point A in H^2, we will find an open disc around A such that ϕ maps this disc to X isometrically. To describe this disc, we will look at the covering of [O,A] and pick r such that every point on the curve from $\phi(O)$ to $\phi(A)$ has a hyperbolic disc of radius r. This basically creates a "tube" of coverings from O to A. For our disc around A, we will choose a radius small than all of these radii; have radius $\frac{r}{2}$. We ultimately want to show that in $D(A, \frac{r}{2}), d(\phi(B), \phi(C) = d(B, C)$. This would lead to the general conclusion we want – that ϕ is locally a geodesic on all such small discs in X.

First, let's show that $d(A, B) = d(\phi(A), \phi(B))$. Let's divide OA in H^2 into a sequence of segments such that each segment is less than $\frac{r}{2}$ long.

$$O = A_0, A_1, A_2, ..., A_n = A$$

Then, we will project these perpendicularly onto the geodesic through OB, creating the points $B_{0,1,2,...n}$. This divides the triangle OAB into one triangle and many quadrilaterals. In $D(A_i, r)$, we see that A_{i-1} and A_i have images in $D(\phi(A_i); r)$, since our segments are no longer than $\frac{r}{2}$. So, we have

$$d(A_{i-1}, A_i) = d(\phi(A_{i-1}), \phi(A_i))$$

We will further divide our quadrilaterals into two triangles each, and then use hyperbolic triangle congruence theorems in induction to show that for all i, the corresponding triangles are congruent.

$$\Delta A_{i-1}B_{i-1}A_i \approx \Delta\phi(A_{i-1})\phi(B_{i-1})\phi(A_i)$$

$$\Delta B_{i-1}A_iB_i \approx \Delta\phi(B_{i-1})\phi(A_i)\phi(B_i)$$

Keep in mind that, even though our triangles are nicely arranged on H^2, this may not be the case on X.

Starting with $D(A_i, r)$ and $\hat{D}(\phi(A_i), r)$, we see that $\phi : [A_i, A_{i+1}]X$ is in $\hat{D}(\phi(A_i), r)$. So,

$$d(A_i, A_{i+1}) = d(\phi(A_i), \phi(A_{i+1}))$$

And, since we are working on a hyperbolic triangle in H^2 currently, $d(A_i, B_i)$ can be no longer than $d(A_n, B_n) \leq \frac{r}{2}$, so A_i, B_i also has an image in $\hat{D}_i$, and so

$$d(A_i, B_i) = d(\phi(A_i), \phi(B_i))$$

And, since B_{i+1} can be at most $\frac{r}{2}$ from B_i and therefore strictly less than r from A_i, we know that $[B_i, B_{i+1}]$ also has image under ϕ in $\hat{D}_i$. A_i, B_i, and B_{i-1} form a triangle. (This is why B_{i+1}, A_i cannot be exactly r apart – if they are, then A_i, B_i, and B_{i+1} are on the same geodesic. And so, we have that A_{i+1}, B_{i+1} and A_{i-1}, B_{i-1} are also in D_i and their images respectively in $\hat{D}_i$)

So, we can conclude using congruence theorems that the angle $\angle A_{i-1}A_iB_{i-1}$ is equal to its image in $\hat{D}_i$ and likewise for $\angle B_{i-1}A_iB_i$ and its image, and this gives us the total angle $\angle A_{i-1}A_iB_i$ and its image, as well as our induction result, that $\angle A_{i+1}A_iB_i$ is equal in measure to the one in its image. SAS gives us that the next triangle, $\Delta A_{i+1}A_iB_i$ is congruent to its image. Now, shifting our attention to the disc centered at A_{i+1}, we can achieve the same result for the following triangle, and therefore this is true for all i. Similarly, we can show this for the lower triangles of the form $\Delta A_{i+1}B_iB_{i+1}$.
Finally, notice that B is not necessarily on B_n. To show that this last triangle ΔAB_nB is congruent it its

image in $\hat{D}_n$, use the fact that $\angle B_n$ and its image are right angles, and that B cannot be more than $\frac{r}{2}$ from A to conclude that corresponding sides are congruent and that $d(A, B) = d(\phi(A), \phi(B))$

To prove that $d(C, B) = d(\phi(C), \phi(B))$, use similar arguments to show that d(A,C) is preserved. Then, we can construct ΔABC and show that the corresponding sides have equal lengths by the hyperbolic sine or cosine rules. QED