Math 181 Handout 9

Rich Schwartz

October 30, 2005

The purpose of this handout is to prove two results:

- Any path connected hyperbolic surface has a universal cover.
- The universal cover of a complete hyperbolic surface is isometric to the hyperbolic plane.

Combining these two results we see that every complete hyperbolic surface is universally covered by the hyperbolic plane. Both of these results hold true in much greater generality, and I will prove the first one in a somewhat more general setting.

1 Nearby Paths

Let X be a path connected metric space which is also a manifold. (So, every point of X has a neighborhood which is homeomorphic to Euclidean space.). Given two paths $f_0, f_1 : [0,1] \to X$ we define

$$D(f_0, f_1) = \sup_{t \in [0,1]} d(f_0(t), f_1(t)).$$

Say that the path f_0 is *good* if there is some $\epsilon > 0$ with the following property: Suppose that $D(f_0, f_1) < \epsilon$ and $f_0(0) = f_1(0)$. Then there is a homotopy F from f_0 to f_1 such that $F_t(0) = f_0(0)$ for all t. In other words, the two paths can be homotoped to each other without moving the initial point. Say that X is *good* if every path in X is good. (The value of ϵ might depend on the path.)

I'll prove that any good manifold has a universal cover. It turns out that any Riemannian manifold is good.

Lemma 1.1 *A hyperbolic surface is good.*

Proof: Let f_0 be a path. For every $t \in [0,1]$ we can find a supremal $\epsilon(t)$ such that the ball of radius $\epsilon(t)$ about $f_0(t)$ is isometric to a disk in the hyperbolic plane. There is some $\epsilon > 0$ such that $\epsilon(t) > \epsilon$ for all t. Otherwise we could find a sequence $\{t_n\}$ such that $\epsilon(t_n) \to 0$. But then there would be some limit point t^* with $\epsilon(t^*) = 0$ and this is a contradiction. So, now we know that there is some $\epsilon > 0$ such that every point $f_0(t)$ has an ϵ ball neighborhood which is isometric to an ϵ ball in the hyperbolic plane. If $D(f_0, f_1) < \epsilon/2$ then we can join $f_0(t)$ to $f_1(t)$ by a geodesic γ_t which remains within the $\epsilon/2$ ball about $f_0(t)$. As t varies the endpoints of γ_t vary continuously. Also, for t small, everything takes place in a set which is isometric to a ball in the hyperbolic plane. Hence, γ_t varies continuously as well.

We set things up so that $\gamma_t(0) = f_0(t)$ and $\gamma_t(1) = f_1(t)$ and γ_t is constant speed. Then the map

$$F(s,t) = \gamma_t(s)$$

is the desired homotopy from f_0 to f_1. Basically, we are just doing the most obvious thing, pushing points of f_0 towards points of f_1 along geodesics. ♠

Exercise 1: Draw a careful picture of the construction from the lemma.

Exercise 2: Give an example of a metric space which has no non-trivial good paths. (Hint: swiss cheese.)

2 Proof of the First Result

2.1 Definition of the Universal Cover

Let X be a good space. Let $x \in X$ be a basepoint. We define $\widetilde{X}$ to be the set of pairs $(y, [f])$ where $y \in X$ is a point and f is a path which joins x to y. Here $[f]$ denotes the path homotopy equivalence class of f.

So far $\widetilde{X}$ is just a set. We define

$$D([f_0],[f_1]) = \inf D(f_0, f_1).$$

The inf is taken over all paths f_0 which represent $[f_0]$ and all paths f_1 which represent $[f_1]$. Finally, we define

$$\widetilde{d}((y_0,[f_0]),(y_1,[f_1])) = d(y_0,y_1) + D([f_0],[f_1]).$$

Exercise 3: Prove that $\widetilde{d}$ is a metric on $\widetilde{X}$. Hint: The only hard part of this exercise is showing that $\widetilde{d}(p,q) = 0$ implies $p = q$. Here $p, q \in \widetilde{X}$. This amounts to showing that $D([f_0],[f_1]) = 0$ implies that $[f_0] = [f_1]$. Deduce this from the goodness of X.

There is an obvious map $E : \widetilde{X} \to X$, given by $E(y,[f]) = y$. This map is distance non-increasing, and hence continuous. Note also that E is onto because X is path connected.

2.2 Evenly Covered Neighborhoods

Let $y \in X$ be a point and let U be a neighborhood of y which is homeomorphic to $\boldsymbol{R}^n$. Each point $z \in U$ can be joined to y by a canonical path $\gamma(z,y) \subset U$: The image of $\gamma(z,y)$ under the homeomorphism to $\boldsymbol{R}^n$ is a straight line segment. In the hyperbolic surface case we can just say simply that U is a small ϵ ball about y and $\gamma(z,y)$ is the geodesic connecting the two points.

We want to show that $E^{-1}(U)$ is a disjoint union of open sets, and the restriction of E to each one of them is a homeomorphism. Let H denote the set of path homotopy classes of curves joining x to y. We are going to first produce a homeomorphism Ψ from $E^{-1}(U)$ to $U \times H$. This is a formal way of saying that $E^{-1}(U)$ is a disjoint union of copies of U.

2.2.1 Well Definedness

Let $(z,[f]) \in E^{-1}(U)$ be a point. Let f be any representative of f and let $g = f * \gamma(z,y)$. We are just extending f so that it connects x to y. We define

$$\Psi((z,[f])) = (z, [f * \gamma(z,y)]).$$

If f_0 and f_1 are both representatives of $[f]$ then a path homotopy from f_0 to f_1 extends to a path homotopy from $f_0 * \gamma$ to $f_1 * \gamma$. Hence $[f_0 * \gamma] = [f_1 * \gamma]$. Hence, our map Ψ is well defined.

2.2.2 Continuity

To show that Ψ is continuous, suppose that $(z_0,[f_0])$ and $(z_1,[f_1])$ are very close. Then $f_0 * \gamma(z_0,y)$ and $f_1 * \gamma(z_1,y)$ are very close. Since X is good, we have $[f_0 * \gamma(z_0,y)] = [f_1 * \gamma(z_1,y)]$ once these paths are sufficiently close. Also z_0 and z_1 are very close. So, the second coordinates of $\Psi(z_0,[f_0])$ and

$\Psi(z_1, [f_1])$ agree and the first coordinates are very close. This shows (a bit sketchily) that Ψ is continuous.

2.2.3 Surjectivity

Given any pair $(z, [g]) \in U \times H$ we can consider the path $f = g * \gamma(z, y)^{-1}$. In other words, we first take our path g from x to y and then take the inverse of $\gamma(z, y)$ from y to z. This gives us a path from x to z. The two paths g

$$f * \gamma(z, y) = g * \gamma(z, y)^{-1} * \gamma(z, y)$$

are clearly homotopic. Hence $\Psi(z, f) = (z, g)$. This shows that Ψ is onto.

Exercise 4: Draw a careful picture of this construction, and define a homotopy from $f * \gamma$ to g.

2.2.4 Injectivity

Suppose that $\Psi(z_0, [f_0]) = (z_1, [f_1])$. Then, first of all, $z_0 = z_1$. So, we can write $z = z_0 = z_1$. Second of all, we know that $[f_0 * \gamma(z, y)] = [f_1 * \gamma(z, y)]$. Writing $\gamma = \gamma(z, y)$ we have $[f_0 * \gamma] = [f_1 * \gamma]$ but then

$$[f_0 * \gamma * \gamma^{-1}] = [f_1 * \gamma * \gamma^{-1}].$$

(We can just use our homotopy from $f_0 * \gamma$ to $f_1 * \gamma$ and extend it trivially to γ^{-1}.) Finally, we have $[f_0 * \gamma * \gamma^{-1}] = [f_0]$, as in Exercise 4. Likewise $[f_1 * \gamma * \gamma^{-1}] = [f_1]$. Hence $[f_0] = [f_1]$, as desired. This shows that Ψ is injective.

2.2.5 Continuity of the Inverse

Using the notation from the surjectivity proof, we have

$$\Psi^{-1}(z, [g]) = (z, [f]),$$

where $f = g \circ \gamma^{-1}$. To show that Ψ^{-1} is continuous, we argue as follows: If $(z_0, [g_0])$ and $(z_1, [g_1])$ are sufficiently close−i.e. less than 1 unit apart−then $[g_0] = [g_1]$. The idea here is that points in different components of $U \times H$ are declared to be 1 away from each other. Since then $[g_0] = [g_1]$, we can use the same path g to represent both $[g_0]$ and $[g_1]$. But then $f_0 = g * \gamma(z_0, y)^{-1}$ and $f_1 = g * \gamma(z_1, y)^{-1}$ are also close. This shows that Ψ^{-1} is continuous.

2.3 The Covering Property

Now we know that Ψ is a homeomorhism from $E^{-1}(U)$ to $U \times H$. Let $\pi : U \times H \to U$ be projection onto U. Then the restriction of π to each component of $U \times H$ is clearly a homeomorphism. These components are of the form $U \times \{h\}$ where $h \in H$.

Finally, note that

$$E = \pi \circ \Psi.$$

For each component $\widetilde{U}$ of $E^{-1}(U)$ there is some $h \in H$ so that $\Psi(\widetilde{U}) = U \times \{h\}$ and Ψ is a homeomorphism from $\widetilde{U}$ to $U \times \{h\}$. But then the restriction to $\widetilde{U}$ of $E = \pi \circ \Psi$ is the composition of two homeomorphisms, and hence a homeomorphism. This completes the proof that E is a covering map. It remains to show that X is simply connected and path connected.

2.4 Simple Connectivity

We still need to show that $\widetilde{X}$ is simply connected. Warning: This proof is pretty tough. I couldn't find a way to boil it down to something simpler. You might want to skip this part on the first reading.

We take the basepoint $\widetilde{x} \in \widetilde{X}$ to be the pair $(x, *)$ where $*$ is the trivial loop connecting x to x. So, suppose that

$$f : [0,1] \to \widetilde{X}$$

is a loop. This means that $f(0) = f(1) = \widetilde{x}$ and

$$f(t) = (x_t, [\gamma_t]),$$

where $x_t \in X$ and γ_t is a path connecting x to x_t. We need to show that f is homotopic to the trivial loop.

Define β by the formula $\beta(s) = x_s$. So, β is a path which traces out "the first coordinate" of f. Now define the path β^t by the formula

$$\beta^t(s) = \beta(st).$$

Note that β^t is a path which joins x to $\beta^t(1) = \beta(t) = x_t$. So, β^t and γ_t are both paths which join x to x_t.

Lemma 2.1 $[\beta^t] = [\gamma_t]$. *In other words,* β^t *is a representative for* $[\gamma_t]$.

Proof: Since γ_0 represents the trivial element in $\pi_1(X,x)$ and β_0 is the trivial element in $\pi_1(X,x)$ we have $[\beta_0]=[\gamma_0]$. Let J be the set of parameter values for which $[\beta^t]=[\gamma_t]$. We have just seen that $0\in J$. To show that $J\in[0,1]$ it suffices to prove that J is both open and closed.

Openness: If $t\in J$ then $[\beta^t]=[\gamma_t]$. For s close to t we have $D([\gamma_s],[\gamma_t])$ small. This means that we can take γ_t and γ_s so that $D(\gamma_s,\gamma_t)$ is small. Consider the path $\delta_s=\gamma_t*\beta[s,t]$. In other words, we take γ_t from x to x_t and then use β to get from x_t to x_s. By continuity $\beta[s,t]$ is short for s close to t. Hence $D(\delta_s,\gamma_t)$ is small. Hence $D(\delta_s,\gamma_s)$ is small. Since X is good we have $[\delta_s]=[\gamma_s]$ for s sufficiently near t. Since $[\gamma_t]=[\beta^t]$ we have $[\delta_s]=[\beta^t*\beta(s,t)]=[\beta_s]$. Hence $[\gamma_s]=[\beta_s]$.

Closedness: Suppose that $[\beta^t]=[\gamma_t]$ for a sequence of t values converging to s. For t very close to s we can again define $[\delta_s]$ as above. The same argument as above shows that $[\delta_s]=[\gamma_s]$ if t is sufficiently close to s. But again $[\delta_s]=[\beta_s]$. Hence $[\gamma_s]=[\beta_s]$.

Being nonempty, open, and closed, $J=[0,1]$. ♠

Now we know that we can take $f(t)=(x_t,[\beta^t])$. Since f is a loop, we know that $[\beta]$ is the trivial element in $\pi_1(X,x)$. Let B be a homotopy from β to the trivial loop. Let B_u be the uth level of the homotopy, so that $B_1=\beta$ and B_0 is the trivial loop. Define

$$B_u^t(s)=B_u(st).$$

Finally define

$$F_u(t)=(B_u^t(1),B_u^t).$$

Here B_u^t is a path connecting x to $B_u^t(1)$. Everything in this crazy definition varies continuously and F_0 is the trivial loop in $\widetilde{X}$.

2.5 Path Connectedness

Now let's show that $\widetilde{X}$ is path connected. Let $(x,[f])$ be a point in $\widetilde{X}$. Then $x=f(1)$. So, we can write $(x,[f])=(f(1),[f])$. Let f^t be the path defined by the equation $f^t(s)=f(st)$. Then $t\to(f(t),[f^t])$ is a path in $\widetilde{X}$ from $\widetilde{x}$ to $(x,[f])$.

3 Proof of the Second Result

3.1 Replacing the Metric

So far we have defined a fairly wierd metric on $\widetilde{X}$. However, when X is a hyperbolic surface, we can define a much better metric on $\widetilde{X}$. Given a path $\gamma \subset \widetilde{X}$ we define $L(\gamma)$ to be the hyperbolic length of $E(\gamma)$, using the hyperbolic metric on X. We define the distance between $\widetilde{p}, \widetilde{q} \in \widetilde{X}$ to be the infimal length of curves joining these two points.

Lemma 3.1 *When $\widetilde{X}$ is equipped with the new metric, the map E is still distance non-increasing and also a universal covering map.*

Proof: If $d(\widetilde{p}, \widetilde{q}) = D$ then there is a path $\widetilde{\gamma}$ joining $\widetilde{p}$ to $\widetilde{q}$ which has length $D + \epsilon$ for any $\epsilon > 0$. But then $\gamma = E(\widetilde{\gamma})$ has length $D + \epsilon$ and joins p to q. Hence the distance on X between p and q is at most $D + \epsilon$. Since ϵ is arbitrary, the distance on X from p to q is at most D.

Let U be a neighborhood on X as in §2. In fact, we take U to be isometric to a ball in the hyperbolic plane. Then the map $E : \pi^{-1}(U) \to U$ is actually an isometry on each component. To show this, let $\widetilde{p}, \widetilde{q}$ be two points in the same component $\widetilde{U}$ of $E^{-1}(U)$. Then the shortest path $\widetilde{\gamma}$ joining $\widetilde{p}$ to $\widetilde{q}$ is the one which projects to the geodesic from p to q in U. This path always exists, because we can always lift the geodesic connecting p to q. Being an isometry from $\widetilde{U}$ to U, the map E is clearly a homeomorphism. ♠

3.2 Completeness

Recall that a metric space is *complete* if every Cauchy sequence converges.

Lemma 3.2 *The universal cover of a complete metric space is complete.*

Proof: Let $\{\widetilde{x}_n\}$ be a Cauchy sequence in X. We have constructed things in such a way that the map $E : \widetilde{X} \to X$ is distance non-increasing. Setting $x_n = E(\widetilde{x}_n)$ we now know that $\{x_n\}$ is a Cauchy sequence in X. Since X is complete, there is some limit point x_*. There is an evenly covered neighborhood U of x_* which contains x_n for n large. But then all the points $\widetilde{x}_n$ lie in the same component of $\widetilde{E}^{-1}(U)$ for n large. But $E : \widetilde{U} \to U$ is a homeomorphism. In particular E maps convergent sequences to convergent sequences and so does E^{-1}. Since $\{x_n\}$ is a convergent sequence in U, the sequence $\{\widetilde{x}_n\}$ is a convergent sequence in $\widetilde{U}$. ♠

Combining the result from this section with the results from the last section, we see that the universal cover of a complete hyperbolic surface is a path connected, simply connected, complete hyperbolic surface.

3.3 Uniqueness

Now we show that any complete, connected, simply connected hyperbolic surface H is isometric to the hyperbolic plane. Let $\boldsymbol{H}$ denote the hyperbolic plane.

Let $h \in H$ be a point and let $\mathbf{h} \in \boldsymbol{H}$ be a point. Both points have neighborhoods which are isometric to disks in the hyperbolic plane. Thus we can find an isometry I between a neighborhood U of h and a neigborhood $\boldsymbol{U}$ of $\mathbf{h}$ in $\boldsymbol{H}$. Let $x \in H$ be any point. We can find a path γ from h to x.

Lemma 3.3 *I can be defined uniquely on γ so that I is a local isometry at every point along γ.*

Proof: We have $\gamma : [0,1] \to H$. Note that $0 \in J$. Hence J is nonempty. If I is defined on $\gamma(t)$ for all $t < s$ then we can define $I(\gamma(s)) = \lim_{s \to t} I(\gamma(t))$. This works because $\boldsymbol{H}$ is complete. If I is defined for $\gamma(t)$ for all $t \leq s$ then we can use the fact that both spaces are locally isometric to the hyperbolic plane to extend I uniquely to a neighborhood of $\gamma(s)$. These two properties show that we can extend I uniquely to $\gamma(t)$ for all $t \in [0,1]$. ♠

Recall that $\gamma(1) = x$. We define $I(x) = I(\gamma(1))$, using the extension from the previous lemma. We need to see that our map is well defined. Here is where simple connectivity comes in. Suppose that γ_0 and γ_1 are two paths connecting h to x. We want to show that $I(\gamma_0(1)) = I(\gamma_1(1))$. That is, we want to show that the extension based on γ_0 is the same as the extension based on γ_1.

Let γ_t, for $t \in [0,1]$ be a path homotopy from γ_0 to γ_1. (We know that such a path homotopy exists from handout 5.) The point $\mathbf{x_t} = I(\gamma_t(1))$ varies continuously with t. On the other hand, note that the same extension in the above lemma works for both γ_s and γ_t as long as s and t are close together. Hence $\mathbf{x_s} = \mathbf{x_t}$ for s and t close. But this shows that $\mathbf{x_t}$ does not move at all.

Now we know that our isometry I extends to a local isometry $I : H \to \boldsymbol{H}$. But the existence of our extension just used completeness of $\boldsymbol{H}$, path connectivity of H and simple connectivity of H. Reversing the roles of H and $\boldsymbol{H}$ we construct the inverse map I^{-1} using the same method. Hence both I and I^{-1} are homeomorphisms and local isometries.

Since I is a local isometry and also a homeomorphism, and distances in H are defined using lengths of paths, we see that I is a global isometry.