Lecture 10: The hyperbolic plane.

October 2, 2014

Let X be a Riemann surface. Recall from last time the notion of a *compatible (Riemannian) metric* on X: it is a metric $\{\langle -, -\rangle_x\}_{x \in X}$ on the smooth surface underlying X, satisfying the condition that the notion of angle provided by the metric agrees with the notion of angle coming from the holomorphic charts of the Riemann surface.

In such a situation, every oriented isometry of the metrized smooth surface X is automatically an automorphism of the Riemann surface X:

$$Isom^+(X) \subset Aut(X).$$

This is because a biholomorphic map on open subsets of $\mathbb{C}$ is exactly a smooth map with smooth inverse whose derivatives preserves oriented angles (see the first problem on your problem set).

Now, in general one would expect the inclusion to be strict, because a biholomorphism is only expected to preserve angles, not distances. But we made the claim in the last lecture, that as a consequence of uniformization, almost every Riemann surface X carries a canonical compatible metric for which $Aut(X) = Isom^+(X)$.

The key to understanding this is to explore the canonical metrics on the three model spaces $\mathbb{P}^1$, $\mathbb{C}$, and $\mathbb{D}$.

We talked about $\mathbb{P}^1$ in the last lecture. We defined its canonical metric to be the spherical metric, coming from the identification of $\mathbb{P}^1$ with S^2. We saw that the inclusion $Isom^+(\mathbb{P}^1) \subset Aut(\mathbb{P}^1)$ was strict, and that the difference is accounted for by the automorphisms of the form

$$z \mapsto \lambda z + b$$

with $\lambda \in \mathbb{R}_{>0}$ and $b \in \mathbb{C}$, in the sense that an aribitrary automorphism of $\mathbb{P}^1$ can be uniquely written as the composition of one of these guys with an oriented isometry.

The case of $\mathbb{C}$ falls out from your problem set this week. You saw that the automorphisms of $\mathbb{C}$ are the maps of the form $z \mapsto az + b$ with $a \neq 0$ and b complex numbers. This can uniquely be written as the composition of a $z \mapsto \lambda z$ (for $\lambda \in \mathbb{R}_{>0}$ again) with an oriented Euclidean isometry, namely some $z \mapsto az + b$ with $|a| = 1$.

(Note: it actually follows that every oriented isometry of the plane is of this form, since oriented isometries surely give automorphisms of the Riemann surface, and when $|a| \neq 1$ the map $z \mapsto az + b$ is not an isometry!)

Thus, in this lecture we'll finish things up by discussing $\mathbb{D}$. Here the situation is different: it turns out that every automorphism of $\mathbb{D}$ actually *is* an oriented isometry:

$$Isom^+(\mathbb{D}) = Aut(\mathbb{D}).$$

But we're getting ahead of ourselves, since we still need to say what the hyperbolic metric on $\mathbb{D}$ is.

Recall that every metric compatible with the complex structure on an open subset U of $\mathbb{C}$ is given by a formula of the type

$$\langle -, - \rangle_z = \lambda(z)^2 \cdot \langle -, - \rangle_{std},$$

where $\lambda : U \to \mathbb{R}_{>0}$ is a smooth function, measuring the amount distances are scaled in going from the standard Euclidean metric to the new metric. For example, the formula

$$\lambda(z) = \frac{2}{1 + |z|^2}$$

describes the spherical metric on $\mathbb{C} \subset \mathbb{P}^1$.

The hyperbolic metric on $\mathbb{D}$ is gotten by just changing a sign: it's given by

$$\lambda(z) = \frac{2}{1 - |z|^2}.$$

Before making a more detailed study, let us explore this in a vague way. At $z = 0$, the center of the disk, the scale is 2, so hyperbolic distances are twice as big as euclidean ones. But as z approaches the boundary of the disk, the scale tends to ∞. So, near the boundary, hyperbolic distances are much bigger than they appear to our Euclidean eyes.

There is another model for this same geometry. Namely, in your problem set last week you found that $\mathbb{D}$ and $\mathbb{H}$ are biholomorphic, via the biholomorphism

$$z \mapsto f(z) = (-i) \cdot \frac{z + i}{z - i}.$$

Using this biholomorphism we can transport the hyperbolic metric on $\mathbb{D}$ over to $\mathbb{H}$. There the formula for it is simpler: the scale factor becomes

$$\lambda(z) = \frac{1}{Im(z)}$$

instead.

(The claim is that the above f gives an isometry between $\mathbb{D}$ with the $\frac{2}{1-|z|^2}$ metric and $\mathbb{H}$ with the $\frac{1}{Im(z)}$ metric. This follows from the formula

$$\frac{1}{Im(f(z))} \cdot |f'(z)| = \frac{2}{1 - |z|^2}.$$

Indeed, since f is a holomorphism (with nonzero derivative), it stretches the Euclidean distances by factor $|f'(z)|$.)

So we have two models for hyperbolic geometry: one, the unit disk $\mathbb{D}$ with the $\lambda(z) = \frac{2}{1-|z|^2}$ scale, and another, the upper-half plane $\mathbb{H}$ with the $\lambda(z) = \frac{1}{Im(z)}$ scale. We can transport between them using the biholomorphism f. It's actually handy to consider both of them, because some aspects of hyperbolic geometry are easier to see in the unit disk model, and some are easier to see in the upper half plane model. (In other words, some aspects look nice to our Euclidean eyes on the unit disk, and some look nice to our Euclidean eyes on the upper half plane.)

To illustrate this, let us describe three classes of oriented hyperbolic isometries. Two will look nice on $\mathbb{H}$, and one will look nice on $\mathbb{D}$.

1. For every $a \in \mathbb{R}$, the map $z \mapsto z + a$ is a hyperbolic isometry of $\mathbb{H}$. [This example gets the name "Parabolic"]

2. For every $\lambda \in \mathbb{R}_{>0}$, the map $z \mapsto \lambda z$ is a hyperbolic isometry of $\mathbb{H}$. [This example gets the name "Hyperbolic", confusingly enough]

3. For every $u \in \mathbb{C}$ with $|u| = 1$, the map $z \mapsto u \cdot z$ is a hyperbolic isometry of $\mathbb{D}$. [This example gets the name "Elliptic".]

In all three cases, the verification that the given map is an isometry is trivial. By the way, if you want to see what the first two kinds of isometries look like in the unit disk model, you can see these YouTube videos, respectively: https://www.youtube.com/watch?v=P0s7wGkdEDM and https://www.youtube.com/watch?v=p7HB2cfZ4mw. (Also tilt your head to the left if you want to match our normalizations.) Of course, the third example is just a rotation around the origin.

Now we can state the main theorems. The first two theorems express, in two different ways, that the above examples explain all the automorphisms of $\mathbb{D}$.

Theorem 0.1. *Every automorphism of the Riemann surface $\mathbb{D}$ is uniquely the composition of a map as in 1. above, followed by a map as in 2. above, followed by a map as in 3. above.*

Corollary 0.2. *Every automorphism of $\mathbb{D}$ is an oriented isometry: $Aut(\mathbb{D}) = Isom^+(\mathbb{D})$.*

The corollary follows immediately, since all of the three types of maps above are oriented isometries, and the composition of oriented isometries is an oriented isometry.

Theorem 0.3. *Every automorphism of $\mathbb{D}$ is conjugate to one of the maps listed above in 1,2,3.*

This means that for every γ in $Aut(\mathbb{D})$, there is an $A \in Aut(\mathbb{D})$ such that $A\gamma A^{-1}$ is equal to one of those maps φ listed above. In other words, γ and the model φ are equivalent via the "coordinate change" A. So from the perspective of hyperbolic geometry, γ looks like φ: the picture for γ is the same as the picture for φ, if you "tilt your head" using the isometry A.

Theorem 0.4. *The automorphisms of $\mathbb{H}$ are exactly the maps*

$$z \mapsto \frac{az+b}{cz+d}$$

with $a, b, c, d \in \mathbb{R}$ and $ad - bc > 0$. (In symbols, $Aut(\mathbb{H}) = PGL_2^+(\mathbb{R})$.)

Corollary 0.5. *Every automorphism of $\mathbb{H}$ (or $\mathbb{D}$) extends to an automorphism of $\mathbb{P}^1$.*

In the case of $\mathbb{H}$, the corollary follows immediately from the theorem, since our description of the automorphism group of $\mathbb{P}^1$ showed that all the maps as in the theorem are automorphisms of $\mathbb{P}^1$. As for $\mathbb{D}$, it follows from the case of $\mathbb{H}$ since the map f and its inverse are also automorphisms of $\mathbb{P}^1$ (they were originally produced as rotations of the sphere). The corollary also follows easily from the first theorem, for essentially the same reason.

There are at least two approaches to proving the above theorems, both relying on results due to Schwarz. One approach actually starts with the last corollary, which can be proved using Schwarz's reflection principle. We won't follow that approach, but I encourage you to look up the reflection principle, because it's very fun.

Instead we'll follow an approach based on Schwarz's lemma, which is the following:

Lemma 0.6. *Let $h : \mathbb{D} \to \mathbb{D}$ be a holomorphic map with $h(0) = 0$. Then $|h'(0)| \le 1$, with equality if and only if there is a $c \in \mathbb{C}$ of norm one such that $h(z) = c \cdot z$.*

Before we give the proof I have to remind you of the *maximal modulus principle*. It says the following. Let $F : U \to \mathbb{C}$ be any non-constant holomorphic function on a connected open subset of $\mathbb{C}$, and suppose $\overline{D} \subset U$ is a closed disc contained in U. Since F is in particular continuous, it follows that there is a point $z_0 \in \overline{D}$ on which f has maximum modulus, i.e.

$$|F(z_0)| \geq |F(z)| \text{ for all } z \in \overline{D}.$$

(Recall the reason: $|F|$ is continuous and $\overline{D}$ is compact, so $|F(\overline{D})|$ is compact, hence closed and bounded, hence attains a finite supremum.)

Then the maximum modulus principle says that any such z_0 of maximum modulus necessarily lies on the *boundary* $\partial\overline{D}$. One proof uses the mean value property of holomorphic maps, but here is another proof. Suppose z_0 lies in the interior of $\overline{D}$. By the local structure theorem for holomorphic maps, we know that there is a small disk N around z_0 such that $F(N)$ contains a small disk around $F(z_0)$. But a small disk around a point always contains points of larger modulus than that point, so this is a contradiction to $F(z_0)$ having maximum modulus.

Okay, now let's prove Schwarz's lemma.

Proof. Let $H(z) = h(z)/z$. This H is holomorphic on $\mathbb{D}$, since $h(0) = 0$ (so the power series expansion of h starts with a z term). Now, the key claim is that $|H(z)| \leq 1$ for all $z \in \mathbb{D}$. Taking $z = 0$ will give the desired $|h'(0)| \leq 1$.

To prove the claim, note that on the circle of radius $r < 1$ centered at 0, we have the bound

$$|H(z)| = |h(z)|/r \leq 1/r.$$

By the maximum modulus principle applied to H, it follows that $|H(z)| \leq 1/r$ for all $z \in \overline{D}(0, r)$. Fixing z and letting $r \to 1$, we deduce the claim.

Now suppose we have equality $|h'(0)| = 1$. Thus $H(0) = h'(0)$ has modulus 1. But we just proved that $|H(z)| \leq 1$ for all $z \in \mathbb{D}$. This is a contradiction to the maximum modulus principle, unless H is constant with value $h'(0)$, in which case $h(z) = h'(0) \cdot z$, as claimed. $\square$

Corollary 0.7. *Every automorphism of $\mathbb{D}$ fixing 0 is of the form $z \mapsto c \cdot z$ for some $c \in \mathbb{C}$ of norm 1, i.e. it is an isometry of type 3 above.*

Proof. The Schwarz lemma says $|f'(0)| \leq 1$. But the Schwarz lemma applied to f^{-1} gives the opposite inequality. Thus we have equality, so the Schwarz lemma gives the claim. $\square$

Now we can prove Theorem 0.1.

Proof. First let us prove existence, i.e. that every holomorphic automorphism of $\mathbb{D}$ is the composition of a map of type 1, followed by 2, followed by 3.

Let $h : \mathbb{D} \to \mathbb{D}$ be a holomorphic automorphism of $\mathbb{D}$. Suppose that h sends z_0 to 0. First, I claim there is a unique composition $h_2 \circ h_1$ of a map of type 1 followed by a map of type 2 which also takes z_0 to 0.

This is easiest to verify in the upper half plane model. The claim is that for any $w \in \mathbb{H}$, there is a unique such composition taking w to i. That is clear: we have to translate over to the imaginary axis, then scale appropriately to get to i.

Now consider $h_3 := h \circ (h_2 \circ h_1)^{-1}$. This is an automorphism of $\mathbb{D}$ which takes 0 to 0. By the corollary to the Schwarz lemma, it is of type 3, which proves the claim.

Now for uniqueness. If we have any such decomposition $h = h_3 \circ h_2 \circ h_1$, then since h_3 fixes 0 it must be that $(h_2 \circ h_1)(z_0) = 0$. Then the above argument actually shows that h_1 and h_2, and hence h_3, are uniquely determined. $\square$

Let us skip Theorem 0.3 for now, and prove Theorem 0.4. This could be done by pure calculation, granting Theorem 0.1. But instead let's give a more abstract argument.

First, we claim that the class of transformations as in Theorem 0.4 is exactly the class of automorphisms of $\mathbb{P}^1$ which send $\mathbb{H}$ to itself. Indeed, first suppose that $\phi \in Aut(\mathbb{P}^1)$ sends $\mathbb{H}$ to itself. By continuity, it must also send the boundary circle $\mathbb{R} \cup \{\infty\}$ to itself. In particular, applying this argument to h^{-1}, we see that the values $h^{-1}(0), h^{-1}(1)$, and $h^{-1}(\infty)$ lie in $\mathbb{R} \cup \{\infty\}$. Looking over the proof of the classifications of the automorphisms of $\mathbb{P}^1$, we only needed these three numbers to write down the formula for h in terms of a, b, c and d. Thus it follows that the a, b, c and d representing h can be chosen to be real. Then the only thing we need to see is that the condition $ad - bc > 0$ holds. For this we can just note that $h'(0) = \frac{ad-bc}{d^2}$ must be positive if $h(i\epsilon)$ is to lie in $\mathbb{H}$ for small ϵ.

Next suppose that h is given as stated in Theorem 0.4; we need to see that h sends $\mathbb{H}$ to itself. Such an h must preserve the circle $\mathbb{R} \cup \infty$ again, this time for obvious reasons. But if we remove this circle from the sphere $\mathbb{P}^1$ we get two connected components; hence our automorphism h must either preserve the components or swap them. Then the same derivative argument as above shows it must preserve them since $ad - bc > 0$. Hence h restricts to an automorphism of $\mathbb{H}$, as claimed.

Now the proof of Theorem 0.4 is trivial. Indeed, all of the above listed three types of automorphisms are evidently restricted from automorphisms of $\mathbb{P}^1$ (recall that $f : \mathbb{D} \leftrightarrow \mathbb{H}$ also came from an automorphism of $\mathbb{P}^1$). Hence so are any of their compositions. Thus, by Theorem 0.1, the same holds for any automorphism of $\mathbb{H}$.

Now we can prove Theorem 0.3, to finish our work.

Suppose that $h : \mathbb{H} \to \mathbb{H}$ is an arbitrary automorphism. As we saw above, h extends to an automorphism of $\mathbb{P}^1$ which preserves the boundary circle $\mathbb{R} \cup \{\infty\}$. In fact, more generally $h(\overline{z}) = \overline{h(z)}$, since the a, b, c and d of Theorem 0.4 are real. Now, we analyze the fixed points of h on $\mathbb{P}^1$. These correspond to complex eigenlines of a 2x2 real matrix, so there are four cases:

1. h has a unique fixed point r, lying in $\mathbb{R} \cup \infty$.

2. h has exactly two fixed points $r_1 \neq r_2$, both of which lie in $\mathbb{R} \cup \infty$.

3. h has exactly two fixed points, one of which z lies in $\mathbb{H}$ and the other is $\overline{z}$.

4. h fixes every point.

In the fourth case $h = id$. Then h actually lies in all three of our model classes, so the claim is certainly true. But now we'll see that the other three cases exactly match up with the 3 model classes.

Suppose we are in the first case. The type 3 maps just rotate the boundary circle, so we can find one of them, call it A, such that $A(r) = \infty$. Then $A \circ h \circ A^{-1}$ has the unique fixed point ∞. In particular, just because it fixes ∞, it follows that $A \circ h \circ A^{-1}$ must be of the form $z \mapsto az + b$ with $a \neq 0$ and b complex numbers. But to avoid more fixed points we need $a = 1$. Then we need b real to preserve $\mathbb{H}$. It follows that $A \circ h \circ A^{-1}$ is of type 1. This proves the claim in the first case.

Now suppose we are in the second case. First suppose $r_1 < r_2$ and neither is ∞. By applying the automorphism $A(z) = z \mapsto \frac{z-r_1}{z-r_2}$, we can move r_1 to 0 and r_2 to ∞. We can accomplish the same thing with no condition on r_1 and r_2 if we throw in some rotations of the boundary circle as well.

Then $A \circ h \circ A^{-1}$ fixes both 0 and ∞. It follows easily, either by algebraic or geometric arguments, that $A \circ h \circ A^{-1}$ is of type 2.

Now suppose we are in the third case. We already checked above that there is an automorphism A (composition of type 1 and type 2) which sends z to i. Thus $A \circ h \circ A^{-1}$ fixes i. If we move to the unit disk model and apply Schwarz's lemma, we see that $A \circ h \circ A^{-1}$ is of type 3. This finishes the proof.

To summarize, we analyze the fixed points of our automorphism, not just inside the space $\mathbb{D}$, but also on its boundary. We apply automorphisms to move those fixed points into standard position, after which point it becomes routine to identify the resulting conjugated automorphism with one of the 3 types above. The nature of the fixed points determines which of the three cases we're in.

Let me finish by giving a geometric reinterpretation of Schwarz's lemma. Schwarz's lemma says that every holomorphic map $\mathbb{D} \to \mathbb{D}$ is distance non-increasing in the hyperbolic metric, and that if the distances are preserved at any point then it must be an isometry. Indeed, Schwarz's lemma says that such a map, if it sends 0 to 0, must be distance non-increasing at 0, and if it preserves distances there then it is an isometry. But by composing on both sides with a hyperbolic isometry we can reduce the case of any arbitrary point of $\mathbb{D}$ with arbitrary image to this case.