Lecture 8: The Riemann sphere, take two: geometry.

September 25, 2014

Last time we studied the Riemann sphere $\mathbb{P}^1$ from a holomorphic/algebraic perspective. In particular, we saw that the field of meromorphic functions on $\mathbb{P}^1$ is $\mathbb{C}(z)$, and that its automorphism groups is $PGL_2(\mathbb{C})$.

Now we will study $\mathbb{P}^1$ from a more geometric perspective. The main result is the following. Consider the unit 2-sphere

$$S^2 = \{v \in \mathbb{R}^3 \ | \ |v| = 1\} = \{(x, y, z) \in \mathbb{R}^3 \ | \ x^2 + y^2 + z^2 = 1\}.$$

Theorem 0.1. *There is a canonical* conformal isomorphism *between S^2 and $\mathbb{P}^1$.*

Let us explain the meaning of the words "conformal isomorphism", first imprecisely, then later in more detail. Roughly, a conformal isomorphism is a bijection which preserves oriented angles.

To tease this out a little bit, recall that any Riemann surface (such as $\mathbb{P}^1$) is obtained by gluing open subsets of $\mathbb{C}$ along biholomorphic identifications. Recall also that biholomorphisms preserve oriented angles. Thus, the notion of oriented angle (say, between two smooth curves meeting at a given point) makes sense on an arbitrary Riemann surface. More specifically, we can choose a chart around the point in question and then calculate the angle in the chart. The result is independent of the chart, since any two charts differ by a biholomorphism, which preserves angles.

On the other hand, S^2 as a geometric object inside $\mathbb{R}^3$ carries its own notion of oriented angle, which we know well from living on a sphere ourselves (approximately). Actually, we need to specify the orientation, but let us take it to be the clockwise one if you're walking on the outside of the surface and looking down. And on $\mathbb{C}$ we'll take it to be the counterclockwise one, in the usual representation of $\mathbb{C}$ as a plane, with i pointing up and 1 to the right.

Thus we have notions of oriented angle both on S^2 and $\mathbb{P}^1$, and the claim is that there's a bijection $\pi : S^2 \leftrightarrow \mathbb{P}^1$ which matches them up. (Again, we'll be more precise later.) There are some immediate corollaries of this:

Corollary 0.2. *The group of conformal automorphisms of S^2 is $PGL_2(\mathbb{C})$.*

That's because via our bijeciton π, conformal automorphisms of S^2 are the same as conformal automorphisms of $\mathbb{P}^1$. But we know that a biholomorphism is the same thing as a conformal isomorphism (see lecture 1). So this is the same as automorphisms of $\mathbb{P}^1$ as a Riemann surface.

So if you've ever wondered how many ways there are to map a sphere bijectively to itself while preserving all angles, now you know the answer: it's equivalent to a problem in complex linear algebra.

Now, among these conformal automorphisms are evidently the *isometries* of S^2, meaning those bijections from S^2 to itself which preserve not only angle, but also distance. We'll see in the beginning of the next lecture that every such isometry is given by rotation around some axis, and that the group of these identifies with the group $SO(3)$ of orthogonal 3x3 real matrices with determinant one.

In particular, combining the above we obtain an inclusion

$$SO(3) \subset PGL_2(\mathbb{C}):$$

these are the isometries sitting inside the conformal automorphisms. It turns out that the inclusion is proper, but that one can completely understand the discrepancy between the two groups, and it's not difficult to manage. We'll also talk about this at the beginning of next lecture.

For now, though, let us turn to our task, which is the construction of the conformal isomorphism $\pi : S^2 \leftrightarrow \mathbb{P}^1$. It will be given by *stereographic projection*.

Namely, let $N = (0,0,1) \in S^2$ be the North Pole, and consider the complex numbers as sitting inside $\mathbb{R}^3$ via the embedding $x + iy \mapsto (x, y, 0)$. Then we say that points $p \in S^2 \setminus \{N\}$ and $z \in \mathbb{C}$ *correspond via stereographic projection (from the north pole)* if:

- The points p, z and N all lie on a single straight line in $\mathbb{R}^3$.

We claim that this correspondence defines a bijection π between $S^2 \setminus \{N\}$ and $\mathbb{C}$. The meaning of this is that given any $z \in \mathbb{C}$ there is a unique $p \in S^2 \setminus \{N\}$ which corresponds to it, and vice-versa. This is supposed to be geometrically obvious: given a z we can draw the line connecting z and N, and this line will cut the sphere in exactly one other point, which will be p. And vice-versa for going from p to z.

A more picturesque way of thinking about the resulting map $\pi : S^2 \setminus \{N\} \to \mathbb{C}$ is the following. Imagine putting a light source at the north pole, and imagine also that the sphere is made out of some transparent material, but the (x, y)-plane $\mathbb{C} \subset \mathbb{R}^3$ acts as a solid ground. Then start to draw on the sphere with black paint. This black paint will cast a shadow on $\mathbb{C}$ consisting exactly of the stereographic projections of the points you've painted. (This doesn't quite work for points below the equator, but you get the idea.)

We can extend this bijection $\pi : S^2 \setminus \{N\} \leftrightarrow \mathbb{C}$ to a bijection $\pi : S^2 \leftrightarrow \mathbb{P}^1$ by letting N correspond to ∞. Note that this does make sense from a continuity standpoint: if we take a point which approaches the north pole, its stereographic projection wanders out to ∞ (fairly quickly). Again, picturesquely, the inverse to stereographic projection wraps the plane around the sphere. But although near the origin this wrapping only distorts distances by a factor of $1/2$, near ∞ it sends huge swaths of land to relatively small parcels near the north pole.

In particular, the stereographic projection certainly doesn't preserve distances. But as we will see, it does preserve angles.

First, however, I want to derive a formula for π and its inverse.

Proposition 0.3. *In coordinates, $\pi : S^2 \setminus \{N\} \to \mathbb{C}$ is given by*

$$(x, y, z) \mapsto \frac{x}{1-z} + \frac{y}{1-z} \cdot i,$$

and its inverse is given by

$$a + bi \mapsto \left(\frac{2a}{1+a^2+b^2}, \frac{2b}{1+a^2+b^2}, \frac{-1+a^2+b^2}{1+a^2+b^2}\right).$$

Note that these expressions are all rational functions of the coordinates. It might seem surprising that we can parametrize a round object like the sphere using only rational functions — no trig functions, for example. But let us explain why, without doing any calculations, one should expect that the formulas involve only rational functions.

Let's think of the inverse to π, which is the more complicated case. What we are doing is looking for the intersection of a line and a sphere. But a line is described by a degree one equation, and a sphere by a degree two equation. So the intersection is described by a degree two (quadratic) equation. There are two intersection points, corresponding to the roots of this quadratic equation. However, one is known a priori: the north pole. Then we can find the other, since the second root of a quadratic equation is a rational function of the coefficients and the first root.

But anyway, now let's implement this idea to derive the above formulas.

Proof. Suppose given a point $(x, y, z) \in S^2 \setminus \{N\}$. The line joining N to (x, y, z) can be parametrized as

$$t \cdot (x, y, z) + (1 - t) \cdot N = (tx, ty, tz + 1 - t).$$

(This is linear in t, and equals N when $t = 0$ and (x, y, z) when $t = 1$. We need to see where this intersects the (x, y)-plane, i.e. where

$$tz + 1 - t = 0.$$

Solving for t, we find

$$t = \frac{1}{1 - z}.$$

Plugging back in, we find the corresponding x and y values, namely they are $\frac{x}{1-z}$ and $\frac{y}{1-z}$. This verifies the first formula.

As for the second one, suppose given a point $(a, b, 0)$ on the (x, y)-plane. Again, the line joining N to $(a, b, 0)$ can be parametrized as

$$t \cdot (a, b, 0) + (1 - t) \cdot N = (ta, tb, 1 - t).$$

We need to see where this intersects the sphere, so we solve

$$(ta)^2 + (tb)^2 + (1 - t)^2 = 1.$$

Rewriting this as a quadratic equation in t, we get

$$(1 + a^2 + b^2) \cdot t^2 - 2t = 0.$$

We don't want the solution $t = 0$, since that corresponds to the north pole. So we divide by t, which lets us solve

$$t = \frac{2}{a^2 + b^2}.$$

Plugging back in gives the displayed result. □

The fact that these formulas are given by rational functions is sort of fun. For example, if you plug in rational numbers for a and b, you get a bunch of points on S^2 with rational coordinates. One dimension lower, this is what gives rise to Pythagorean triples.

But anyway, back to geometry. We need to see that π preserves angles. There will be two steps: first we will check this away from the north pole, this we will check it at the north pole, by looking in a different chart.

The first step is the major one. There are two possible approaches (at least):

1. Use standard Euclidean geometry reasoning from high school.

2. Check it by a coordinate calculation.

In class we used the first approach, but here we'll use the second one. But, as in class, I want to first explain why it's possible in principle to check such a thing using calculation in coordinates.

This involves talking about another major idea of Riemann's, which now goes under the name *Riemannian Geometry*. We'll revisit this later in more detail, so for now I will be a little imprecise.

Suppose we want to define geometry on a given surface X. (Say, X is a $\mathcal{G}$-structured surface for $\mathcal{G}$ the class of smooth bijections with smooth inverse: so X is a *smooth surface*.) That is, we want to be able to talk about concepts like the length of a path on X, the area of a region on X, the angle between two curves intersecting on X, etc. Riemann's idea for doing this is actually a combination of two ideas:

1. To specify the global geometry of X, it's enough to specify the first-order geometry at every point $x \in X$.
2. The first-order geometry at the point $x \in X$ can be specified by giving an *inner product* on the tangent space $T_x(X)$.

The tangent space $T_x(X)$ is a two-dimensional vector space which corresponds to possible first-order movements away from the point X. If we choose a smooth identification of X with an open subset of $\mathbb{R}^2$ locally around x, then we get an identification of $T_x(X)$ with the vector space $\mathbb{R}^2$ of displacements in the plane. If we change the chart by a smooth map with smooth inverse, the identification of $T_x(X)$ with $\mathbb{R}^2$ changes by the linear isomorphism given by the derivative of our smooth map at the point in question.

Note that a smooth diffeomomorphism of the plane can induce an arbitrary invertible linear map on derivatives, so there is a priori no more structure on $T_x(X)$ than just that of a vector space. This is in contrast to the case of a Riemann surface, where the 2-dimensional real vector space $T_x(X)$ actually had the extra structure of a one-dimensional complex vector space, this structure being respected by all biholomorphisms.

Now we should say what an inner product of a vector space is, and why it encodes first-order geometry.

Definition 0.4. *Let V be a finite-dimensional real vector space. An* inner product *on V is a function*

$$V \times V \to \mathbb{R},$$

usually denoted $(v, w) \mapsto \langle v, w\rangle \in \mathbb{R}$, satisfying the following properties:

1. *$\langle v, w\rangle = \langle w, v\rangle$ for all $v, w \in V$. (Symmetry)*
2. *$\langle \lambda v, w\rangle = \lambda\langle v, w\rangle$ and $\langle v_1 + v_2, w\rangle = \langle v_1, w\rangle + \langle v_2, w\rangle$ for all $v, w, v_1, v_2 \in V$ and $\lambda \in \mathbb{R}$. (Linearity)*
3. *$\langle v, v\rangle \geq 0$ for all $v \in V$, with equality if and only if $v = 0$. (Positive-definiteness)*

Note that we only required linearity in the first variable, but by symmetry this implies linearity in the second variable as well.

The first example should be $V = \mathbb{R}^n$, with the so-called *standard inner product*, namely the dot product:

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n)\rangle = x_1y_1 + \ldots + x_ny_n.$$

This is supposed to encode the standard Euclidean geometry on $\mathbb{R}^n$.

A finite dimensional real vector space together with an inner product is called an *inner product space*. Given an inner product space, we can define various geometric notions, such as the *length of a vector*:

$$|v| := \sqrt{\langle v, v \rangle}$$

and the *(unoriented) angle between two vectors v and w*, determined by

$$cos(\theta) = \frac{\langle v, w \rangle}{|v| \cdot |w|}.$$

Note that it's unoriented, because it's symmetric in v and w. Or, again, $cos(\theta) = cos(-\theta)$. An important special case is the notion of *orthogonality*: two vectors v, w are said to be orthogonal if

$$\langle v, w \rangle = 0.$$

These definitions are motivated by the case of the standard inner product on $\mathbb{R}^n$, where they correspond to the usual notions from Euclidean geometry.

We remark that the notion of length of a vector actually uniquely determines the inner product, by the identity

$$\langle v, w \rangle = \frac{1}{2} \left(|v + w|^2 - |v|^2 - |w|^2 \right),$$

so we could have axiomatized our structure based on a notion of length rather than inner product. Sometimes one or the other is more convenient; we used the inner product because it only involves linear algebra, not quadratic algebra.

Now let us give a very simple example of a *Riemannian surface*, meaning a surface X equipped with an inner product on each tangent space $T_x(X)$ which "varies smoothly" with the point $x \in X$ (such a structure is sometimes called a Riemannian *metric* on X).

Let X be our old friend $\mathbb{C} \setminus \{0\}$. Each tangent space $T_x(X)$ identifies with $\mathbb{C} = \mathbb{R}^2$ since X is already an open subset of $\mathbb{C}$. So we could just take the standard inner product on $T_x(X)$. But instead we *scale* the inner product as follows: on $T_z(\mathbb{C} \setminus \{0\})$ we take the inner product given by

$$\langle v, w \rangle = \frac{1}{|z|^2} \langle v, w \rangle_{standard}.$$

What we're doing here is artificially scaling distances up as we approach the origin, and scaling them down as we approach ∞ (but the angles stay the same!). For example, in this Riemannian structure, all of the circles centered at the origin have the same length, namely 2π. This is an open infinite cylinder again.

Note that the notion of a *Riemannian surface* is a priori quite different from that of a Riemann surface. This is confusing, but we can't help it that Riemann had so many great ideas. But actually the coincidence of terminology is not as unlucky as it seems: as we have hinted at, and will see in more detail later, the uniformization theorem shows that every Riemann surface carries a (essentially canonical) Riemannian structure as well. E.g. the above metric on $\mathbb{C} \setminus \{0\}$ is the same as the one coming from its uniformization via the exponential map.

Moving back to our present discussion, another example of a Riemannian surface is S^2, the so-called *round sphere*. Indeed, the tangent space to S^2 at a point p is a vector subspace of the space of all vectors in 3-space based at p, which is $\mathbb{R}^3$. Then we can restrict the standard inner product on $\mathbb{R}^3$ to $T_p(S^2)$ in order to get a Riemannian structure on S^2.

Thus, e.g., to measure the angle between two vectors tangent to S^2 we just measure their angle in the ambient 3-space.

If you run through the machine of Riemannian geometry in this case, you'll find yourself recovering the usual spherical geometry: straight lines are great circles, the total area is 4π, etc.

Now we can explain how to check that π preserves angles. It's a matter of taking the derivative of π, and seeing how it plays with the two inner products in question. Namely, the derivative of π at a point $p \in S^2$ is a linear map

$$D = d\pi|_p : T_p(S^2) \to T_{\pi(p)}\mathbb{C} = \mathbb{R}^2.$$

What we will see is that there is a scalar λ (depending on the point p) such that

$$\langle Dv, Dw \rangle = \lambda^2 \cdot \langle v, w \rangle$$

for $v, w \in T_p(S^2)$. (This λ will then be the "conformal factor", or "scale" of the map π at p.) Here, as discussed above, the inner product on the right is restricted from the dot product on the ambient $\mathbb{R}^3$, and the inner product on the left is the dot product in $\mathbb{R}^2$.

Such an equation evidently implies that unoriented angles are preserved by D, by the formula for angle in terms of inner product given above. It then follows that unoriented angles are preserved by π, since this is a first-order question, more-or-less by definition. Finally, we can easily see that the orientations match up as well, and this finishes the claim.

Now let's get to calculating to verify the above equation. But actually, we should hold our horses and be a little bit clever first. I claim that it suffices to verify the above equation when p has x-coordinate zero. This is because the whole picture inside $\mathbb{R}^3$ is invariant under rotation about the north-south axis. Such a rotation is an isometry of $\mathbb{R}^3$, hence restricts to an isometry of both S^2 and $\mathbb{C}$. So stereographic projection will preserve angles at a given point p if and only it preserves angles at one (or every) point on the same latitude line as p. That's why we can assume our point p is of the form $(0, y, z)$.

Now, here we go. the formula for π was

$$(x, y, z) \mapsto (\frac{x}{1-z}, \frac{y}{1-z}).$$

Thus its derivative, viewed as a linear map from $\mathbb{R}^3$ to $\mathbb{R}^2$, is given by the matrix of partial derivatives: the first row is

$$(\frac{1}{1-z}, 0, \frac{x}{(1-z)^2})$$

and the second row is

$$(0, \frac{1}{1-z}, \frac{y}{(1-z)^2}).$$

This is the derivative at the point (x, y, z).

In other words, inputting our assumption that $x = 0$, we find that D is the linear map

$$(a, b, c) \mapsto \frac{1}{1-z} \cdot (a, b + c \cdot \frac{y}{1-z}).$$

Now, what happens when we restrict this to $T_p(S^2)$? Since S^2 is a level-set of the function $f(x, y, z) = x^2 + y^2 + z^2$, the subspace $T_p(S^2)$ is described by the equation $df|_p = 0$: a vector is tangent if a first-order movement in that direction keeps you on the level set, i.e. doesn't change the value under f. Computing partial derivatives we see that $(a, b, c) \in \mathbb{R}^3$ lies in $T_p(S^2)$ if and only if $ax + by + cz = 0$. (This is just an algebraic restatement of the fact that the tangent plane to the sphere is orthogonal to the vector joining the center of the sphere to the point in consideration.)

Now, again, we're trying to find a constant $\lambda = \lambda(x, y, z)$ such that

$$\langle Dv, Dw \rangle = \lambda^2 \cdot \langle v, w \rangle$$

for all $v, w \in T_p(S^2)$. But let's be clever again. Since both sides are linear in both v and w, it suffices to verify the identity as v and w run over some chosen basis of $T_p(S^2)$. If we chose this basis carefully, we can simplify our calculations.

A geometrically motivated choice of basis would be unit tangent vectors along the latitudinal and longitudinal lines. These also have nice algebraic expressions, namely

$$e = (1, 0, 0)$$

and

$$f = (0, z, -y).$$

It is trivial to check that are orthogonal and have length 1 and both lie in $T_p(S^2)$ by the above description of $T_p(S^2)$. Thus they do form a basis, and it will suffice to see that

$$\langle De, Df \rangle = 0$$

and that

$$|De| = |Df|.$$

(This common value will then give us the constant λ.)

The rest is just algebra. Plugging in, we have

$$De = \frac{1}{1-z} \cdot (1, 0)$$

and

$$Df = \frac{1}{1-z} \cdot (0, z - \frac{y^2}{1-z}) = \frac{1}{1-z}(0, 1),$$

where the last manipulation is because $y^2 + z^2 = 1$. These vectors are certainly orthogonal and have the same common length $\lambda = \frac{1}{1-z}$, so that finishes the proof.

Exercise: run a similar argument with the inverse of π to reach the same conclusion. Which is the easier argument, in your opinion? I couldn't decide.

There's only on thing left to do, and that's check that stereographic projection preserves angles at ∞. For this we need to look in the other chart of $\mathbb{P}^1$. The other chart can be accessed by (a slight modification of) stereographic projection to the south pole.

Namely, I claim that if you take a point $p \in S^2$ different from both the north and south poles, then if its stereographic projection from the north pole is z, its stereographic projection from the south pole is $\overline{z}^{-1}$. In class we saw this with a geometric argument, but it can also just be calculated from the formulas for stereographic projection and its inverse, since it's equivalent to saying that if we switch the sign of the last coordinate of $\pi^{-1}(z)$ and then reapply π, we get to $\overline{z}^{-1}$. It may help to note that since $\overline{z}^{-1}$ and z have the same angle and have inverse norms, this claim is also symmetric by rotation along the north-south axis, so we can assume $z \in \mathbb{R} \setminus \{0\}$ if we want.

Anyway, granting that claim, we see that another description of our bijection $\pi : \mathbb{P}^1 \leftrightarrow S^2$ is that in the chart around ∞ of $\mathbb{P}^1$, namely the bijection $\mathbb{C} \leftrightarrow \mathbb{P}^1 \setminus \{0\}$ given by $z \mapsto 1/z$, our bijection π

sends a point z in the chart $\mathbb{C}$ to the stereographic projection of $\overline{z}$ to the south pole. (Thus, e.g. $z = 0$, which corresponds to $\infty \in \mathbb{P}^1$, goes to the north pole.) Now we see by the same argument as above that π preserves angles on all of this chart, and in particular preserves angles at ∞ too. This finishes the proof.

In other words, again, here are the two descriptions of $\pi : \mathbb{P}^1 \to S^2$. One is that it is stereographic projection to the north pole. The other is that it is complex conjugation, followed by multiplicative inversion, followed by stereographic projection to the south pole.

Exercise: What does the automorphism $z \mapsto z^{-1}$ of $\mathbb{P}^1$ look like geometrically on the sphere S^2?