Nov 2017

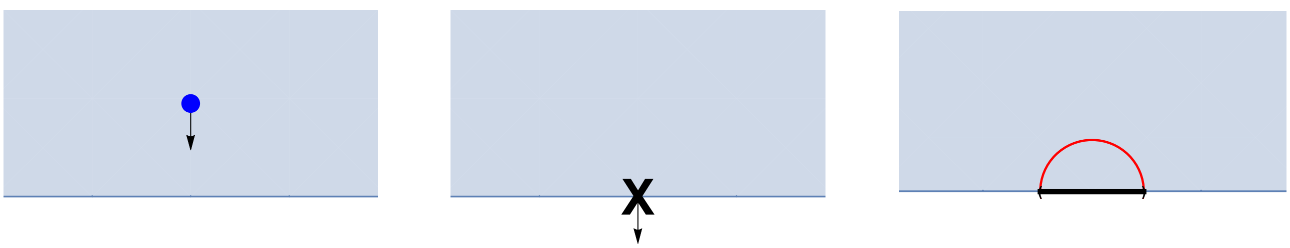
Letter to Bill Duke on "Inversive Coordinates"

from Alex Kontorovich

Dear Bill,

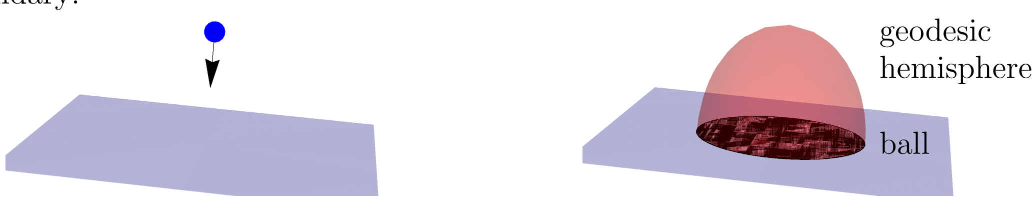
Let me make more precise my comment about using so-called "inversive coordinates" to treat points in the upper half plane $\mathbb{H}$ in the same vein as geodesics in $\mathbb{H}$. My discussion below is basically "well-known" and some initial ideas can be traced back to Clifford and Darboux; I include at the end just a few references. (I learned much of this point of view from discussions with Kei Nakamura.) My claim is the following: imagine being a point in $\mathbb{H}$ and moving straight down to the boundary $\partial\mathbb{H} = \widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. After an infinite amount of time, you arrive at the boundary (x-axis).

Claim: when you keep moving past the x-axis, what develops in the upper half-plane is actually a geodesic above a (Euclidean) interval in the boundary. Here is what I have in mind:



Point moving down | Arriving at $\partial\mathbb{H}$, moving past | Turning into geodesic above the black interval

Before explaining why this is a reasonable thing to say, let me point out that it generalizes to arbitrary dimensional upper half-space $\mathbb{H}^n$. When a point moves "down" past the boundary, what appears in $\mathbb{H}^n$ is a co-dim-1 geodesic hemisphere above a Euclidean ball in the boundary:



I'll return to $\mathbb{H}$ to keep the original discussion, but everything below basically generalizes on replacing "interval" with "ball," x with $\mathbf{x} = (x_1, \dots, x_{n-1})$, and x^2 with $|\mathbf{x}|^2 = x_1^2 + \cdots + x_{n-1}^2$.

To a geodesic above the Euclidean interval $|x - x_0| < r$ in the boundary $\partial\mathbb{H}$, we attach the following "inversive coordinates":

$$\mathbf{v} = \mathbf{v}(x_0, r) := \left(\frac{1}{\widehat{r}}, \frac{1}{r}, \frac{x_0}{r} \right),$$

where $\widehat{r}$ is the "co-radius", defined as the radius of the inversion of the interval through the unit interval. It is clear that the interval $[x-r, x+r]$ inverted through the unit interval becomes $[1/(x+r), 1/(x-r)]$, so the co-radius is:

$$\widehat{r} = \frac{1}{2}\left[\frac{1}{x-r} - \frac{1}{x+r}\right] = \frac{r}{|x|^2 - r^2}.$$

(We write $|x|^2$ here instead of just x^2 to emphasize that this formula for the co-radius is valid in all dimensions.) Rearranging terms, we may write this as

$$\frac{1}{\widehat{r}}\frac{1}{r} - \frac{|x|^2}{r^2} = -1, \qquad \text{or} \qquad Q(\mathbf{v}) = -1,$$

where Q is the "discriminant" (quadratic) form with half-Hessian

$$Q = \begin{pmatrix} 0 & 1/2 & \\ 1/2 & 0 & \\ & & -1 \end{pmatrix}.$$

In general, the "-1" above is replaced by $-I_{n-1}$. Unsurprisingly, this form Q has signature $(1,2)$, so the quadric $Q=-1$ is a one-sheeted hyperboloid.

In a similar way, we may attach the following inversive coordinates to a point $(x,y) \in \mathbb{H}$:

$$\mathbf{w} = \mathbf{w}(x,y) := \left(\frac{1}{\widehat{y}}, \frac{1}{y}, \frac{x}{y}\right),$$

where $\widehat{y}$ now is the "co-height", defined to be

$$\widehat{y} := \frac{y}{|x|^2 + y^2}. \tag{1}$$

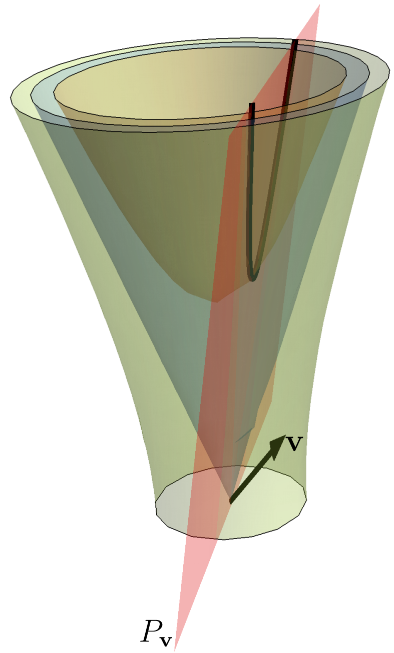
Again this is the "y-part" of the image of (x,y) reflected through the (Euclidean) unit circle. From (1), it follows that

$$Q(\mathbf{w}) = +1,$$

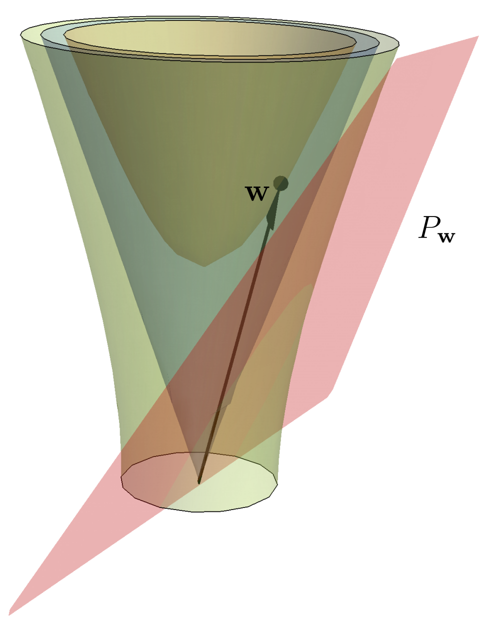
so $\mathbf{w}$ lies on one sheet (the "top" one, the one with $y>0$) of the two-sheeted hyperboloid $Q=1$. This top sheet is of course itself a model for $\mathbb{H}$, and is what you "see". So once the point $\mathbf{w}$ has traveled infinitely far, arrived at the boundary cone $Q=0$, and moved past the cone to become $\mathbf{v}$ on $Q=-1$, what happens is this. The point $\mathbf{v}$ is Q-orthogonal to a plane

$$P_{\mathbf{v}} = \{\mathbf{t} \in \mathbb{R}^{n+1} \ : \ \mathbf{v}Q\mathbf{t}^{\dagger} = 0\}, \qquad (\dagger\text{=transpose})$$

and the intersection of this plane with the top sheet is (pointwise, under the map $(x,y) \mapsto \mathbf{w}$) the corresponding geodesic in $\mathbb{H}$! Here's the picture I have in mind.



What happens to this plane picture when we use the point $\mathbf{w}$ on $Q = 1$ instead? The orthogonal plane $P_{\mathbf{w}}$ is still "there," but we simply don't "see" it because it doesn't intersect the top sheet; all we see instead is the point $\mathbf{w}$.



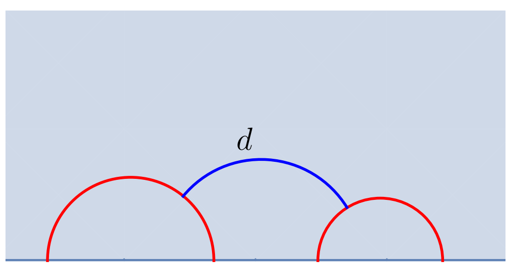
A lot of nice things happen in these coordinates. For one, conformal maps are now just elements of $SO_Q^{\circ}(\mathbb{R})$, the connected component of the identity of the real special orthogonal group preserving Q. This is simply a restatement of the standard fact that $\mathrm{SL}_2(\mathbb{R})$ is a double cover of $\mathrm{SO}_{2,1}$ (and in higher dimensions, one uses Clifford algebras). Another is that it's very easy to write down the action of reflection of a point $\mathbf{w}$ through a geodesic expressed in inversive coordinates as $\mathbf{v}$. Indeed, one has the standard formula for reflection through the plane $P_{\mathbf{v}}$ orthogonal to $\mathbf{v}$:

$$\mathbf{w} \mapsto \mathbf{w} - 2\frac{\mathbf{w}Q\mathbf{v}^\dagger}{\mathbf{v}Q\mathbf{v}^\dagger}\mathbf{v} \;=\; \mathbf{w}\cdot(I + 2Q\cdot\mathbf{v}^\dagger\cdot\mathbf{v}).$$

The matrix in parentheses above is the (anti-conformal) Möbius transformation in $O_Q(\mathbb{R})$ representing reflection through the geodesic corresponding to $\mathbf{v}$. (And there's no typo here: $\mathbf{v}^\dagger\cdot\mathbf{v}$ is a rank one matrix; note also that we used $\mathbf{v}Q\mathbf{v}^\dagger = -1$.)

What happens if the desired geodesic is vertical, so is not expressible as being above an interval $|x| < r$, i.e. the one connecting x_0 to ∞? Simply take the limit as $r \to \infty$ of the inversive coordinates corresponding to the intervals $|x_0 + r| < r$; it is easy to compute that one obtains in the limit $\mathbf{v} = (2x_0, 0, 1)$. By the way, inversive coordinates actually give a geodesic an orientation: Imagine taking the geodesic above an interval $[a, b]$, that is, going "from" a to b, and sending b to the right all the way to infinity, and past infinity to negative infinity around until it comes up from below, becoming the interval $[c, a]$. The corresponding geodesic still goes from a to c, so has the opposite orientation; thus the interval is really $(-\infty, c]\cup[a,\infty)$, that is, the exterior of $[c, a]$. One way of seeing this is to follow the geodesic flow forwards and backwards.

Yet another nice property is the **Claim** that the Q-product of two geodesics computes their "generalized" hyperbolic distance:

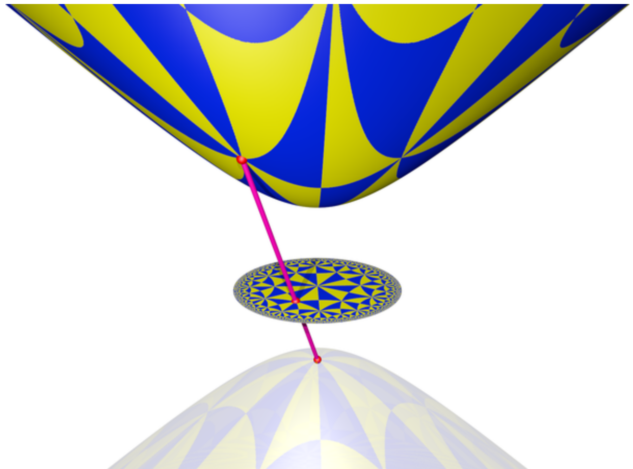


$$\mathbf{v}_1 \star \mathbf{v}_2 := \mathbf{v}_1 Q \mathbf{v}_2^\dagger = \cosh d$$

(For but one proof, send one geodesic to the upper y-axis and compute.)

If the geodesic on the left in the above image moves to the right until after the two intersect (at angle θ, say), the product becomes $\mathbf{v}_1 \star \mathbf{v}_2 = \cos\theta$ (that is, it is less than 1), which one can think of as being the same as above but with $d = i\theta$ imaginary. If it keeps moving to the right until the two intervals are nested (keeping in mind the orientation discussion above), the product becomes less than -1, so it is $-\cosh d$. This is what is meant by "generalized" distance in the Claim above.

These inversive coordinates are really nice for the reasons described above (i.e., the calculations are particularly simple), but this whole discussion becomes completely standard (at the expense of slightly more complicated formulae) on making a linear change of variables from the discriminant form Q to the "Pythagorean" (or "Lorenzian") form $Q_1(a, b, c) = a^2+b^2-c^2$. (This phenomenon is also familiar – compare the spin representation of the discriminant form (3.31) of [Kon13] (granted, in higher dimension) to that of the Pythagorean form (4.6) there.) The top sheet of $Q_1 = -1$ (note that Q_1 has signature $(2, 1)$ while the discriminant form Q has opposite signature; hence the sign change) is identified with the unit disk $A^2 + B^2 < 1$ in the plane $c = 0$ (which then is standard to identify with $\mathbb{H}$) under projection to the point $(0, 0, -1)$. Here is a nice picture I found online [Sta] (so didn't have to make).



Best wishes,

Alex

PS Let me make the last paragraph above completely explicit. The standard map $\mathbb{H} \to \mathbb{D}$ is given by:

$$\mathbb{H} \to \mathbb{D} : z \mapsto \frac{z-i}{z+i}, \qquad x + iy \mapsto \frac{x^2+y^2-1}{x^2+(y+1)^2} + i\frac{-2x}{x^2+(y+1)^2}.$$

For $(A, B) = A + iB \in \mathbb{D}$ in the disk, the map (projection through $(0, 0, -1)$) to the top Lorenzian sheet $\mathcal{T}_1 : a^2 + b^2 - c^2 = -1, c > 0$ is given by:

$$\mathbb{D} \to \mathcal{T}_1 : (A, B) \mapsto (a, b, c) = \left(\frac{2A}{1-(A^2+B^2)}, \frac{2B}{1-(A^2+B^2)}, \frac{1+A^2+B^2}{1-(A^2+B^2)}\right).$$

It is trivial to concatenate the two, getting a map $\mathbb{H} \to \mathcal{T}_1 : (\mathbb{H} \to \mathbb{D}) \circ (\mathbb{D} \to \mathcal{T}_1)$ given by

$$\mathbb{H} \to \mathcal{T}_1 : x + iy \mapsto \left(\frac{x^2+y^2-1}{2y}, -\frac{x}{y}, \frac{x^2+y^2+1}{2y}\right).$$

I think this much is basically everywhere in the literature. What is (perhaps) not completely trivial is to recognize the above as

$$\left(\frac{x^2+y^2-1}{2y}, -\frac{x}{y}, \frac{x^2+y^2+1}{2y}\right) = \left(\frac{1}{2}\left(\frac{1}{\widehat{y}} - \frac{1}{y}\right), -\frac{x}{y}, \frac{1}{2}\left(\frac{1}{\widehat{y}} + \frac{1}{y}\right)\right).$$

Hence concatenating this with the linear map $\mathcal{T}_1 \to \mathcal{T}$ (where $\mathcal{T} : Q(\frac{1}{\widehat{y}}, \frac{1}{y}, \frac{x}{y}) = 1, y > 0$ is the "discriminant" top sheet) given by

$$\mathcal{T}_1 \to \mathcal{T} : (a, b, c) \mapsto (\frac{1}{\widehat{y}}, \frac{1}{y}, \frac{x}{y}) = (c + a, -b, c - a),$$

furnishes explicitly the inversive coordinates map $\mathbb{H} \to \mathcal{T} : x + iy \to (\frac{1}{\widehat{y}}, \frac{1}{y}, \frac{x}{y})$.

PPS Here are some references to the above material:

[Cli82] W. H. Clifford. On the powers of spheres (1868). In Mathematical Papers. Macmillan (London), 1882

[Dar72] G. Darboux. Sur les relations entre les groupes de points de cercles et de spheres dans le plan et dans l'espace. Ann. Ecole Norm Sup., 1:323–392, 1872

[Kon13] A. Kontorovich. From Apollonius to Zaremba: local-global phenomena in thin orbits. *Bull. Amer. Math. Soc.* (N.S.), 50(2):187–228, 2013.

[LMW02] Jeffrey C. Lagarias, Colin L. Mallows, and Allan R. Wilks. Beyond the Descartes circle theorem. Amer. Math. Monthly, 109(4):338–361, 2002

[Sta] https://math.stackexchange.com/questions/2372877/is-projection-an-isometry-from-the-poincare-disk-to-a-two-sheeted-hyperboloid

[Wil69] J. B. Wilker, Four proofs of a generalization of the Descartes circle theorem, Amer Math. Monthly 76 (1969) 278-282.