The disc model of hyperbolic geometry

In (2-dimensional) Euclidean geometry, the points are given by ordered pair of real numbers. In the disc model of hyperbolic geometry, the points are pairs (x, y) of real numbers such that $x^2 + y^2 < 1$. It will often be convenient to consider the point (x, y) as the complex number $x + iy$.

In Euclidean geometry, the length of a path γ is given by integrating the distance form $\sqrt{dx^2 + dy^2}$ along γ. In the disc model of hyperbolic geometry, we instead integrate the distance form $\frac{2\sqrt{dx^2+dy^2}}{1-x^2-y^2}$ along γ. In effect distances are scaled by the factor $\frac{2}{1-x^2-y^2}$, so distances are longer further from the origin, and we can see that the distance to points on the unit circle becomes infinite (so lines can be extended to an arbitrary length in either direction).

Like in Euclidean geometry, we define the straight lines in hyperbolic geometry to be the shortest curves between their endpoints. From this definition, we can see that the unique straight line segment between $(0, 0)$ and $(a, 0)$ is the part of the real axis between those two points:

Lemma 1. *The unique hyperbolic straight line between the points* $(0, 0)$ *and* $(a, 0)$ *in the disc model, is the Euclidean straight line between them.*

Proof. (non-examinable). Let γ be a curve from $(0, 0)$ to $(a, 0)$. We get that the length of γ is $\int_\gamma \frac{2\sqrt{dx^2+dy^2}}{1-x^2-y^2} \geqslant \int_\gamma \frac{2\sqrt{dx^2}}{1-x^2}$, which is the length of the straight line segment from $(0, 0)$ to $(a, 0)$. Equality holds in the above inequality if and only if y is constantly 0 along the path. Therefore, the part of the real axis between $(0, 0)$ and $(a, 0)$ is the unique hyperbolic straight line between those points. □

Having identified what one hyperbolic straight line looks like in the disc model, we can determine what all the others look like by studying the isometries of the disc model of hyperbolic geometry.

Firstly, we note that the hyperbolic distance $\frac{2\sqrt{dx^2+dy^2}}{1-x^2-y^2}$ is rotationally symmetric about the origin - i.e., rotation about the origin is an hyperbolic isometry. Similarly, reflection in a line through the origin is also an hyperbolic isometry.

From this we can deduce:

Lemma 2. *Hyperbolic straight lines through the origin are exactly Euclidean straight lines.*

Proof. The image of an hyperbolic straight line through the origin under a rotation about the origin is another hyperbolic straight line, and in particular, if the hyperbolic straight line passes through the point z, we can choose a rotation about 0 that sends z to a real number. Then we know that the unique hyperbolic line from 0 to z is the Euclidean straight line. □

We will show:

Lemma 3. *The only hyperbolic isometries that fix the origin are rotations about the origin and reflections in lines through the origin.*

Proof. (non-examinable). Let f be an hyperbolic isometry that fixes the origin - i.e. $f(0) = 0$. We know that the distance of a point from the origin is entirely determined by its distance from the origin in the disc model. Furthermore, if two points are the same hyperbolic distance from the origin, then they must be the same Euclidean distance from the origin in the disc model. (The hyperbolic line to a point passes through points nearer to the origin in the Euclidean sense, so the hyperbolic distance from the origin increases with Euclidean distance.) Therefore, since hyperbolic isometries preserve hyperbolic distances, f must send every point to a point the same Euclidean distance from the origin. i.e. $|z| = |f(z)|$ (recall that for a complex number $z = x + iy$, $|z| = \sqrt{x^2 + y^2}$) We also know that f sends hyperbolic straight lines to other hyperbolic straight lines. In particular, we know that the hyperbolic straight lines through the origin are exactly Euclidean straight lines through the origin. Therefore, we know that f sends Euclidean straight lines through the origin to other Euclidean straight lines through the origin. We just need to show that f preserves the angle between any two of these lines. For some $0 < r < 1$, let z_1 and z_2 be the points on two lines through the origin, at distance r from the origin. Let γ be the arc of a circle of radius r, centred at the origin, between z_1 and z_2. Since f preserves distance from the origin, we know that $f(\gamma)$ is also an arc of a circle centre 0, radius r. Since f is an isometry, and therefore preserves the length of curves, the curve $f(\gamma)$ must have the same length as γ. Since they are both arcs of circles of radius r, they must subtend the same angle at the origin. This is the angle between the lines l_1 and l_2. Therefore, f preserves distance from the origin, preserves lines through the origin, and preserves the angle between two such lines. It must therefore be either a rotation about the origin, or a reflection in a line through the origin. □

Now that we know all the isometries that fix the origin, once we have an isometry that sends a point a to the origin for every point a, we will be able to determine all hyperbolic isometries - first determine where the origin gets sent; then after we compose with an isometry that sends that point to the origin, we get an isometry that fixes the origin, so by the preceding lemma, we know that this isometry is either a rotation or a reflection.

First, we need to find an hyperbolic isometry sending the point a to the origin.

Proposition 1. *The map $z \mapsto \frac{z-a}{\overline{a}z-1}$ is an hyperbolic isometry that sends the point a to the origin. It is its own inverse - i.e. applying it twice gives the identity.*

Proof. (non-examinable). First we need to show it is an hyperbolic isometry. Let $f(z) = \frac{z-a}{\overline{a}z-1}$, and let $f(z) = g(z) + ih(z)$. By differentiating, we get $\frac{df(z)}{dz} = \frac{\overline{a}z-1-\overline{a}(z-a)}{(\overline{a}z-1)^2} = \frac{a\overline{a}-1}{(\overline{a}z-1)^2}$. Observe that $dg(z)^2 + dh(z)^2 = (dg(z) + idh(z))(dg(z) - idh(z)) = df(z)\overline{df(z)}$. Therefore, $\frac{2\sqrt{dg(z)^2+dh(z)^2}}{1-g(z)^2-h(z)^2} = \frac{2\sqrt{df(z)\overline{df(z)}}}{1-|f(z)|^2} = \frac{2\sqrt{\frac{df(z)}{dz}dz\overline{\frac{df(z)}{dz}}d\overline{z}}}{1-|f(z)|^2} = \frac{2\sqrt{\frac{(a\overline{a}-1)^2}{(\overline{a}z-1)^2(a\overline{z}-1)^2}}}{1-|f(z)|^2}\sqrt{dx^2 + dy^2}$.

Now we note that $1-|f(z)|^2 = 1-\left(\frac{z-a}{\overline{a}z-1}\right)\left(\frac{\overline{z-a}}{\overline{\overline{a}z-1}}\right) = \frac{(\overline{a}z-1)(a\overline{z}-1)-(z-a)(\overline{z}-\overline{a})}{(\overline{a}z-1)(a\overline{z}-1)} = \frac{\overline{a}a\overline{z}z+1-\overline{a}a-\overline{z}z}{(\overline{a}z-1)(a\overline{z}-1)}$. This gives us that $\frac{2\sqrt{\frac{(a\overline{a}-1)^2}{(\overline{a}z-1)^2(a\overline{z}-1)^2}}}{1-|f(z)|^2} = \frac{2(a\overline{a}-1)}{(a\overline{a}-1)(z\overline{z}-1)} = \frac{2}{1-z\overline{z}}$. This is exactly the factor we get from the usual hyperbolic metric. Therefore f is an hyperbolic isometry.

We check that indeed $f(a) = 0$. We want to check that f is self-inverse. To do this, we plug the result of f back into f to get $f(f(z)) = \frac{\frac{z-a}{\overline{a}z-1}-a}{\overline{a}\frac{z-a}{\overline{a}z-1}-1} = \frac{(z-a)-a(\overline{a}z-1)}{\overline{a}(z-a)-(\overline{a}z-1)} = \frac{z(1-a\overline{a})}{1-a\overline{a}} = z$.

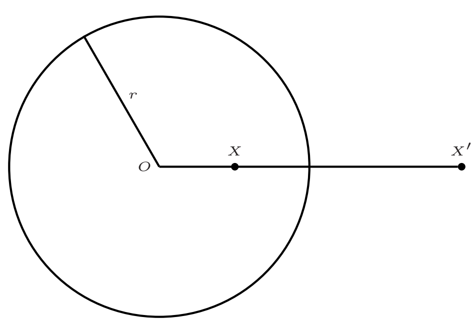
Finally, we want to check that f sends the hyperbolic plane (the unit disc) to itself. Suppose that $|z| < 1$. We want to show that $\left|\frac{z-a}{\overline{a}z-1}\right| < 1$. This is equivalent to showing that $|z-a| < |\overline{a}z-1|$. Squaring both sides (which is valid because both sides are positive) reduces this to showing $(z-a)(\overline{z}-\overline{a}) < (\overline{a}z-1)(a\overline{z}-1)$, or equivalently, $z\overline{z}+a\overline{a} < z\overline{z}a\overline{a}+1$, or $(1-z\overline{z})(1-a\overline{a}) > 0$, which is true since $1-z\overline{z}$ and $1-a\overline{a}$ are both > 0. $\square$

We can look at the isometry $z \mapsto \frac{z-a}{\overline{a}z-1}$ in more detail. It can be expressed as the composite of the three maps:

$$z \mapsto z - \frac{1}{\overline{a}}$$
$$z \mapsto \frac{\frac{1}{\overline{a}} - a}{\overline{a}z}$$
$$z \mapsto z + \frac{1}{\overline{a}}$$

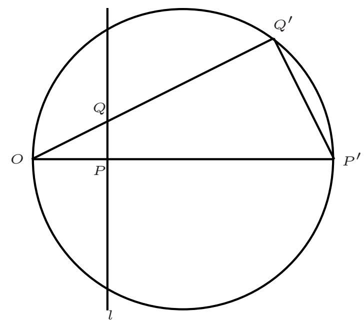
The composite of the first two is $z \mapsto \frac{\frac{1}{\overline{a}}-a}{\overline{a}z-1}$. When we add $\frac{1}{\overline{a}}$ to this, we get $\frac{z-a}{\overline{a}z-1}$. The first and third of these maps are translations. The second is the composite of an inversion and a reflection in the real axis.

Inversion



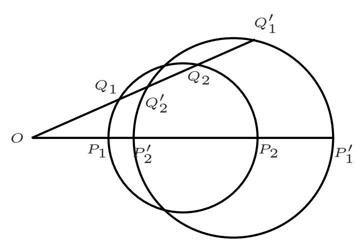
Consider a circle γ, with centre O and radius r. Consider the transformation that sends a point X other than the origin to the point X' on the line OX, on the same side of O as X, such that $OX'.OX = r^2$. This transformation is called inversion in γ.

Inversion clearly sends straight lines through O to themselves. It sends straight lines not through O to circles through O − let l be a straight line not through O. Drop the perpendicular from O to l, and let the foot of this perpendicular be P. Let its image under the inversion be P'. Let Q be another point on l, and let Q' be its image.



We know that $OP.OP' = OQ.OQ' = r^2$, so P, P', Q, and Q' are concyclic. Since $\angle OPQ = 90°$, we get that $\angle OQ'P = 90°$. Therefore, Q' lies on the circle with diameter OP'. This circle is therefore the image of l. (We think of O as the image of ∞ under the inversion.)

Finally, given a circle S that does not pass through O, the inversion of S in γ is another circle. Let l_1 be the line from O to the centre of S, and let l_2 be another line through O that meets S. Let P_1 and P_2 be the points where l_1 meets S, with P_1 nearer to O, and let Q_1 and Q_2 be the points where l_2 meets S, with Q_1 nearer to O. Let the images of these points be P_1', P_2', Q_1' and Q_2' respectively.



Since $OP_1.OP_1' = OP_2.OP_2' = r^2$, we get that $\frac{OP_2'}{OP_1} = \frac{OP_1'}{OP_2} = \frac{r^2}{OP_1.OP_2}$. This means that the image of S is an enlargement of S about O. This sends circles to circles, so the image of S is a circle.

We can now deduce that hyperbolic lines in the disc model are Euclidean circles, since they are the images of straight lines through the origin under the transformation $z \mapsto \frac{z-a}{\overline{a}z-1}$, which is the composite of a translation, an inversion, a reflection, and another translation. (In fact it can be expressed as the composite of just a reflection and an inversion - inversion in the circle with centre $\frac{1}{\overline{a}}$.) This means that it sends straight lines to circles (or other straight lines). We know that it sends 0 to a and we can see where it sends infinity (which we think of as being a point on the straight line): one way to do this is to observe that as $z \to \infty$, $\frac{1}{z} \to 0$. Now $\frac{z-a}{\overline{a}z-1} = \frac{1-\frac{a}{z}}{\overline{a}-\frac{1}{z}} \to \frac{1}{\overline{a}}$ as $\frac{1}{z} \to 0$. Therefore, the image of a line through the origin will be a circle passing through the points a and $\frac{1}{\overline{a}}$. Observe that $\frac{a}{\frac{1}{\overline{a}}} = a\overline{a}$ is real. This means that the line through a and $\frac{1}{\overline{a}}$ passes through the origin. This means that the power of the origin with respect to this circle is 1. Therefore, the tangent to the circle from the origin must have length 1, so this circle will meet the unit circle (the circle centred at the origin, of radius 1) at right angles.

We therefore have that general hyperbolic lines in the disc model are circles that meet the unit circle at right angles.

When performing calculations in the disc model, it is often easier to apply an isometry of the form $z \mapsto \frac{z-a}{\overline{a}z-1}$ to send a relevant point to the origin, then use the same isometry on the resulting point if necessary.

Area

The area of a region in the disc model is obtained by integrating $\frac{4dxdy}{(1-x^2-y^2)^2}$ over the region. Recall that in Euclidean geometry we find the area of a region by integrating $dxdy$ over that region, and that in hyperbolic geometry, lengths are obtained by multiplying the infinitesimal Euclidean lengths by the scale factor $\frac{2}{1-x^2-y^2}$. To get areas, we need to multiply by the square of this factor, since a Euclidean area is the product of two lengths.

In particular, as we shall see, there is a simple formula for the area of an hyperbolic triangle. First, we will need a few lemmas about asymptotic hyperbolic triangles (triangles with a vertex on the boundary of the disc).

Lemma 4. *Any two triply asymptotic triangles (triangles with all 3 vertices on the boundary of the disc) are congruent.*

Proof. It is enough to show that any triply asymptotic triangle is congruent to the triangle with vertices at 1, -1 and i. Let ABC be a triply asymptotic triangle. Let l be the (hyperbolic) line AB. Pick a point w on l, and apply the isometry $z \mapsto \frac{z-w}{\overline{w}z-1}$, to send w to 0. Now applying a rotation about 0, we can send l to the real line. Let the image of C under this isometry be C'. Note that C' will be on the boundary of the disc because the hyperbolic isometry sends the boundary of the disc to itself. There is an hyperbolic line from C' perpendicular to the real axis – if C' is the point $x + iy$, then this hyperbolic line is the Euclidean circle with centre at $\frac{1}{x}$. Let d be the point where this hyperbolic line meets the real axis. Now we apply the isometry $z \mapsto \frac{z-d}{dz-1}$ (since d is real, $\overline{d} = d$). This isometry preserves the real line - if z is real, then $z - d$

and $dz - 1$ are both real. Since it is an isometry, it preserves angles, so there is an hyperbolic line from the image C'' of C', perpendicular to the real axis at 0. This must be the imaginary axis. Therefore, C'' must be either i or $-i$. If it is i, we have sent ABC to the triangle with vertices at 1, -1 and i. If it is $-i$, we can reflect in the real axis. Therefore, we have shown that ABC is congruent to the triangle with vertices at 1, -1 and i. $\square$

Lemma 5. *Any two doubly asymptotic triangles with the same angle at the non-asymptotic vertex are congruent.*

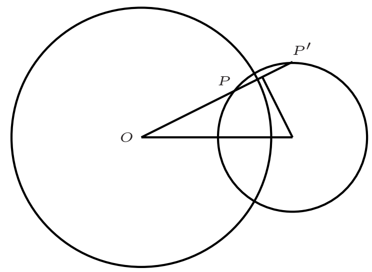
Proof. It is enough to show that any doubly asymptotic triangle with angle θ is congruent to the triangle with vertices at 0, $\cos\frac{\theta}{2} + i\sin\frac{\theta}{2}$ and $\cos\frac{\theta}{2} - i\sin\frac{\theta}{2}$. Suppose the non-asymptotic vertex is at a. Apply the isometry $z \mapsto \frac{z-a}{\overline{a}z-1}$. This sends the triangle to an asymptotic triangle with one vertex at 0, and angle θ at 0. Now if we apply the rotation to send one of the vertices to $\cos\frac{\theta}{2} - i\sin\frac{\theta}{2}$, the other vertex must be sent to either $\cos\frac{\theta}{2} + i\sin\frac{\theta}{2}$ or $\cos\frac{3\theta}{2} - i\sin\frac{3\theta}{2}$. In the first case, we are done. In the second case, we can apply a reflection in the line through 0 and $\cos\frac{\theta}{2} - i\sin\frac{\theta}{2}$ to get to the triangle we want. $\square$

Proposition 2. *The area of a doubly asymptotic hyperbolic triangle with angle ϕ radians is $\pi - \phi$.*

Proof. (non-examinable). First we apply an isometry to send the triangle to the doubly asymptotic triangle with one vertex at the origin, and the two sides meeting there being the lines at angles $\pm\frac{\phi}{2}$ from the positive real axis.

Now if we let γ be the circle forming the third side of the hyperbolic triangle, and let A be the interior of the triangle. We find the hyperbolic area by integrating $\int\int_A \frac{4dxdy}{(1-x^2-y^2)^2}$. Converting to polar coordinates, we get $\int\int_A \frac{4rdrd\theta}{(1-r^2)^2}$. Let $f(\theta)$ be the length of the line segment in A at angle θ from the positive real axis. The area is then $\int_{-\frac{\phi}{2}}^{\frac{\phi}{2}}\int_0^{f(\theta)} \frac{4r}{(1-r^2)^2}drd\theta$. For the inner integral, note that $\frac{d}{dr}\left(\frac{1}{1-r^2}\right) = \frac{2r}{(1-r^2)^2}$. Therefore, $\int_0^{f(\theta)} \frac{4r}{(1-r^2)^2}dr = \left[\frac{2}{1-r^2}\right]_0^{f(\theta)} = 2\left(\frac{1}{1-f(\theta)^2} - 1\right) = 2\frac{f(\theta)^2}{1-f(\theta)^2}$.

We can calculate $f(\theta)$ as a function of θ. Firstly, we let L be the distance from the origin to the centre of γ. Since the tangent from the origin to γ has length 1, and is at an angle $\frac{\phi}{2}$ from the line between the origin and the centre of γ, we get that $L = \frac{1}{\cos\frac{\phi}{2}}$. Now we consider the line l through the origin at angle θ to the positive real axis (which goes through the centre of γ).



Since the origin has power 1 with respect to γ, the distances from the origin to the points where l meets γ are $f(\theta)$ and $\frac{1}{f(\theta)}$. We can drop the perpendicular from the centre of γ to l: it will meet l at the midpoint of the points of intersection of l and γ. Therefore, we get that $\frac{f(\theta)+\frac{1}{f(\theta)}}{2} = L\cos\theta$. Multiplying by $f(\theta)$, we get the quadratic equation $f(\theta)^2 - 2L\cos\theta f(\theta) + 1 = 0$. We can solve this to get $f(\theta) = L\cos\theta - \sqrt{L^2\cos^2\theta - 1}$ (we take this solution because we know that $f(\theta) \leqslant 1$). From this we get $f(\theta)^2 = 2L^2\cos^2\theta - 1 - 2L\cos\theta\sqrt{L^2\cos^2\theta - 1}$, and $1 - f(\theta)^2 = 2 - 2L^2\cos^2\theta + 2L\cos\theta\sqrt{L^2\cos^2\theta - 1} = 2\sqrt{L^2\cos^2\theta - 1}(L\cos\theta - \sqrt{L^2\cos^2\theta - 1}) = 2f(\theta)\sqrt{L^2\cos^2\theta - 1}$. We therefore get $\frac{f(\theta)^2}{1-f(\theta)^2} = \frac{L\cos\theta - \sqrt{L^2\cos^2\theta-1}}{2\sqrt{L^2\cos^2\theta-1}} = \frac{L\cos\theta}{2\sqrt{L^2\cos^2\theta-1}} - \frac{1}{2}$

The area is therefore $\int_{-\frac{\phi}{2}}^{\frac{\phi}{2}} \left(\frac{L\cos\theta}{\sqrt{L^2\cos^2\theta-1}} - 1\right) d\theta$. This equals $2\int_0^{\frac{\phi}{2}} \frac{L\cos\theta}{\sqrt{L^2\cos^2\theta-1}} d\theta - \phi$.

We now make the substitution $\alpha = \sqrt{L^2\cos^2\theta - 1}$. We get that $\frac{d\alpha}{d\theta} = -\frac{L^2\sin\theta\cos\theta}{\sqrt{L^2\cos^2\theta-1}}$. Therefore, the integral becomes

$$-\int_{\sqrt{L^2-1}}^{\sqrt{L^2\cos^2\frac{\phi}{2}-1}} \frac{L}{\sin\theta} d\alpha = \int_{\sqrt{L^2\cos^2\frac{\phi}{2}-1}}^{\sqrt{L^2-1}} \frac{L}{\sin\theta} d\alpha$$

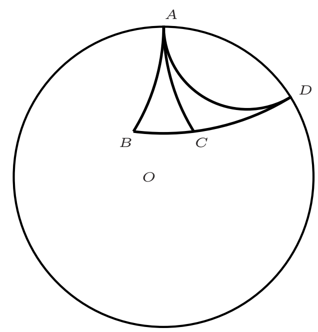
We need to express $\sin\theta$ in terms of α. Now $\sin\theta = \sqrt{1-\cos^2\theta} = \sqrt{1 - \frac{\alpha^2+1}{L}}$, so the integral is $\int_{\sqrt{L^2\cos^2\frac{\phi}{2}-1}}^{\sqrt{L^2-1}} \frac{1}{\sqrt{L^2-1-\alpha^2}} d\alpha$. We note that since $L = \frac{1}{\cos\frac{\phi}{2}}$, $L^2\cos^2\frac{\phi}{2} - 1 = 0$.

Now we make the substitution $\beta = \frac{\alpha}{\sqrt{L^2-1}}$ to get $\int_0^1 \frac{\sqrt{L^2-1}}{\sqrt{L^2-1}\sqrt{1-\beta^2}} d\beta = \int_0^1 \frac{1}{\sqrt{1-\beta^2}} d\beta = \left[\sin^{-1}\beta\right]_0^1 = \frac{\pi}{2} - 0 = \frac{\pi}{2}$. Therefore, the area of the hyperbolic triangle is $\pi - \phi$. □

From this, we can deduce the formula for the area of a singly asymptotic triangle, and a non-asymptotic triangle:

Proposition 3. *The area of a singly asymptotic triangle with angles α and β radians is $\pi - \alpha - \beta$.*

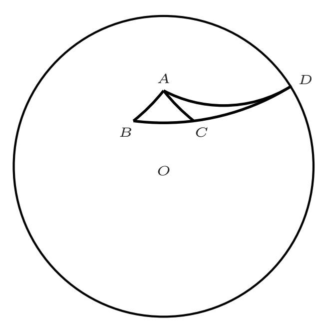
Proof. Let ABC be a singly asymptotic triangle, with A a point on the boundary of the disc. Extend BC past C to meet the boundary of the disc at D.



Now the area of the singly asymptotic triangle ABC is the area of the doubly asymptotic triangle ABD minus the area of the doubly asymptotic triangle ACD. The triangle ABD has angle α, while the triangle ACD has angle $\pi - \beta$. Therefore the area of triangle ABD is $\pi - \alpha$, and the area of triangle ACD is $\pi - (\pi - \beta) = \beta$. The area of triangle ABC is therefore $\pi - \alpha - \beta$. □

Proposition 4. *The area of an hyperbolic triangle with angles α, β and γ is $\pi - \alpha - \beta - \gamma$.*

Proof. Let ABC be an hyperbolic triangle with angles α at A, β at B and γ at C. Extend BC past C to meet the boundary of the disc at D.



The area of ABC is the area of ABD minus the area of ACD. Let $\angle CAD = \delta$. Then the area of $\triangle ABD$ is $\pi - \alpha - \beta - \delta$, and the area of $\triangle ACD$ is $\pi - (\pi - \gamma) - \delta = \gamma - \delta$, so the area of $\triangle ABC$ is $\pi - \alpha - \beta - \gamma$. □