EXPLORING HYPERBOLIC GROUPS WITH ELEMENTARY GEOMETRY AND GENETIC ALGORITHMS

ASHANI DASGUPTA

Abstract. This article presents an elementary geometric journey into hyperbolic geometry and group dynamics. Beginning with classical ideas such as cross ratio and inversion in a circle, we intuitively explore the hyperbolic metric on the unit disk using transformation geometry. We then study distance-preserving maps of the disk and use a ping–pong argument to construct a free group of isometries. By iterating these transformations, we observe the emergence of a fractal limit set structure resembling a Cantor set. Finally, we illustrate how simple evolutionary algorithms can computationally explore this limit set, offering a modern experimental perspective on classical geometric group theory.

1. A bit of Transformation Geometry

Geometric transformations are at the heart of modern geometry. Recall that in we study properties of shapes such as circles and triangles in school geometry. Now we begin the study of functions that keep a certain shape unchanged. At the moment it is okay to imagine that the shape is drawn in some space X such as the $x-y$-coordinate plane. Then the functions that we study are maps from $X \to X$.

Consider the shape of a flower drawn in the coordinate plane $\mathbb{R}^2$. What functions from $\mathbb{R}^2 \to \mathbb{R}^2$ keeps the flower unchanged?

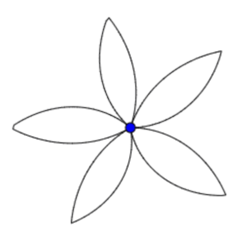

Figure 1. Flower with Five Petals

Clearly rotations by $k \times \frac{360^\circ}{5}$ about the center of the flower will keep the flower visually undisturbed. Here k can be any integer. This type of functions are known as geometric transformation. The contention being that they

Date: February 28, 2026.

transform the shape or position (or both) of a geometric object. One might ask: what is a geometric object? This is a much deeper question and the answer depends on who you ask. We will avoid this question at the moment.

There are many geometric transformations that we all are familiar with. Here are three well-known transformations.

(1) rotation
(2) translation
(3) reflection

There are plenty of other functions from $\mathbb{R}^2 \to \mathbb{R}^2$. It is a daunting task to study all of them in a systematic manner. Hence at times we take a step back and ask something for fundamental: is there something else that these functions are keeping unchanged?

Let us be precise about this. Suppose R_{60° is rotation by 60° counter-clockwise about the point $(0,0)$. Some geometric shapes drawn on the plane might be left unchanged when this function is applied on the plane. *However there is something far more fundamental that is left unchanged or invariant.*

Turns out that for rotation, the **distance** between pairs of points is such an invariant quantity. More precisely, suppose $A = (x_1, y_1)$ and $B = (x_2, y_2)$ are any two points in the plane. We may compute distance between them using the familiar Pythagorean formula from coordinate geometry.

$$d(A,B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Suppose after rotation $A \to A_1$ and $B \to B_2$. Then it is easy to prove that $d(A,B) = d(A_1, B_1)$. Try to prove it as an **exercise**.

Distance is not the only quantity preserved by rotational motion. Rotations also preserve angles. Given a function $f : X \to X$, it may be difficult to determine all the quantities that are preserved by f. There may be more than one such invariant. These invariant quantities are often objects of human discovery. **Some of these quantities can also be discovered using artificial intelligence and computational tools such as genetic algorithms.** In this exploration, we examine one such toy example. We uncover a fractal-like object that is invariant under a class of functions on a disc. The word *fractal* may sound mysterious; informally, it refers to a geometric object that exhibits similar structure at every scale.

At the moment, let us explore another geometric transformation. It is known as a perspective map. Suppose P is any point in the plane and L is a line that works as a screen for projection. Then the perspective map sends any point T on the plane to a point on the line L in the following manner:

- join P to T and possibly extend it.
- find the intersection of this line with the screen L.
- the point of intersection T_1 is the image of point T

This map is strangely simple. It requires two fixed items: point P and the line L. What measurable quantity is left unchanged by this perspective map? Human civilization required few hundred years to figure that out. It is known as **cross ratio**.

Suppose A, B, C, D are four points. Assume that they are either on a straight line or they are on a circle (concyclic with P). Under the perspective map, their images are A_1, B_1, C_1, D_1.

Then the following ratio of ratios is preserved under the perspective function. Try the experiment.

$$\frac{\frac{AC}{BC}}{\frac{AD}{BD}} = \frac{\frac{A_1C_1}{B_1C_1}}{\frac{A_1D_1}{B_1D_1}}$$

This was initially observed in a special case by Pappus of Alexandria. He worked on the case where the ratio of ratios is 1. About a thousand years later mathematicians and artists explored this concept in much greater detail.

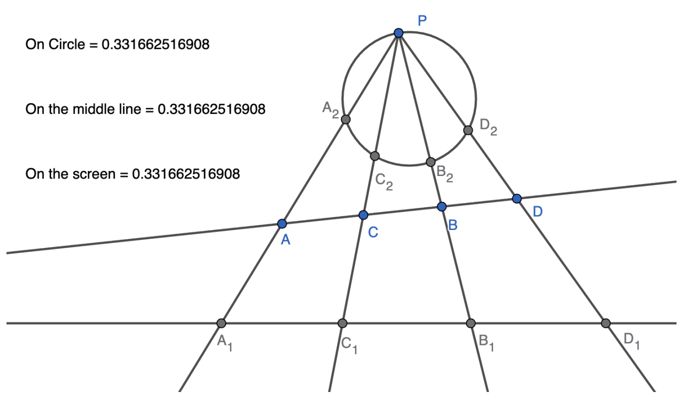

FIGURE 2. Cross Ratio of Points on a Circle and a Line

The cross ratio revealed something remarkable about geometry. In some sense, a circle and a straight line have a lot in common. After all the perspective map does not care if the points are linear or there are placed in a circular manner. The cross ratio is preserved in either case!

It is natural to investigate other maps, apart from the perspective maps, that preserve the cross ratio. Since segment lengths are preserved by distance preserving maps such as rotations, translations and reflections, therefore they trivially preserve cross ratio. Are there any other weird maps that preserves cross ratio as well? Turns out there is. It is known as **inversion**. One may think of it as a reflection about a circle.

Suppose c is a circle centered at $O = (a, b)$ and radius R. If A is any point on the plane then its inverse A' is a special point with the following properties.

- A' is on the ray OA.
- $OA \times OA' = R^2$

This map is well defined on the punctured plane; that is any point other than the center O is a valid input for this function.

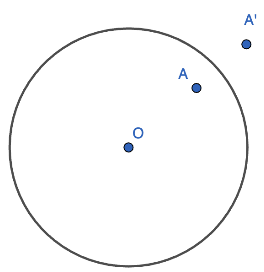

FIGURE 3. Inversion is reflection about a circle

Let us choose four points A, B, C, D on a line. Invert them about the circle. Suppose the images are A', B', C', D'. Cross ratio of the input points would be same as cross ratio of the out points. Notice that the output points are circularly spread out. Try to prove that they are actually on a circle. Also try this experiment.

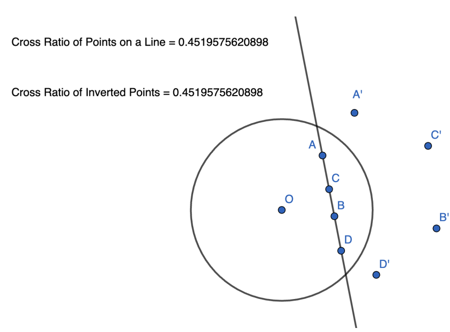

FIGURE 4. Inversion preserves cross ratio

Recall that perspective map preserves cross ratio of four points lying on a circle. Will inversion work the same way? Consider the set up in Figure 5. The green circle is the mirror (about which we invert points). Suppose H, D, C, G are four points on a black circle. Their inverted images are H_1, D_1, C_1, G_1. You may check that the cross ratio is preserved in this case. Try this experiment.

Notice that we have suspiciously added a red circle in this diagram that passes through G, G_1, H_1, H. As an exercise show that this red circle is orthogonal to the mirror (green circle). That is tangents drawn to each other at the point of intersection of red and green circles make $90°$.

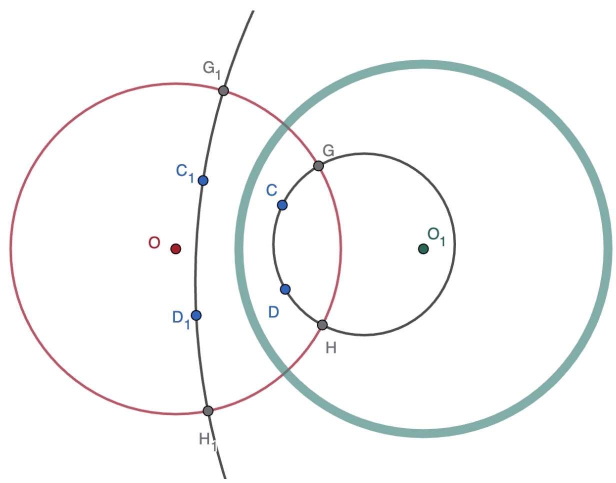

FIGURE 5. Cross Ratio of Points on Circle

Proposition 1.1. *Suppose* Γ *is a circle (in the figure drawn in green color and imagined as a mirror). Let* A *and* B *be any two points that are reflected (inverted) about* Γ *to* A' *and* B' *respectively. Then there exists a circle* G *(drawn in red in the figure) that passes through* A, B, A', B'. *Moreover* Γ *and* G *are orthogonal to each other.*

Proof. The proof is a simple application of power of a point. Suppose the radius of mirror Γ is R_{green}. Clearly $PA \times PA' = R_{green}^2 = PB \times PB'$. Consider the triangles ΔPAB and $\Delta PA'B'$. Notice that

$$\frac{PA}{PB} = \frac{PB'}{PA'}$$

Moreover $\angle P$ is a common angle of both the triangles included in the proportional sides. Hence the two triangles are similar. This implies corresponding angles are equal.

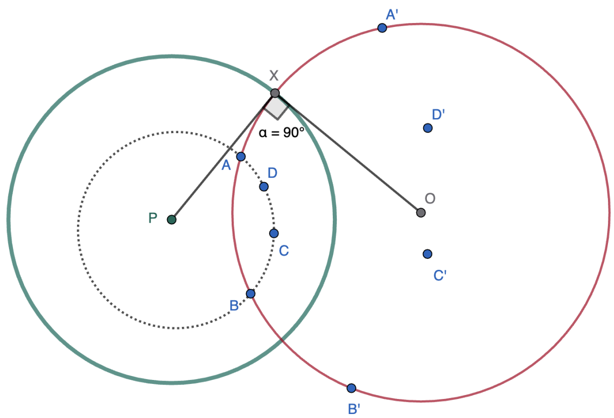

FIGURE 6. Orthogonal Red Circle

$$\angle PAB = \angle PB'A'$$

Since $\angle BAA' = 180° - \angle PAB$ therefore $\angle BAA' + \angle BB'A' = 180°$. Since opposite angles of quadrilateral $BB'A'A$ adds up to $180°$ hence it is cyclic. We draw this circle G in red color and mark its center as O.

Suppose the radius of the red circle is R_{red}. Again using the power of a point technique we notice that $Power(P) = OP^2 - R_{red}^2$. Moreover using inversion we know that $Power(P) = PA \times PA' = R_{green}^2$. Hence we have,

$$OP^2 - R_{red}^2 = R_{green}^2$$

Notice that, this implies ΔPXO is right angled. The conclusion is immediate from here. □

Let us put together what we have learned thus far.

(1) Perspective map is a very simple geometric motion. It preserves a surprising quantity known as cross ratio.
(2) Cross ratio of four points on a circle (passing through the point of projection) or four points lying on a straight line are preserved under perspective map.
(3) Another map that preserves cross ratio is inversion (relection) about a circular mirror.
(4) Again the four points can be on a straight line or a circle.
(5) There is a natural red circle that is orthogonal to the green mirror.

We again flip (dualise) our line of thought. We begin thinking like this: points inside the red circle are being reflected (inverted) by green mirrors that are orthogonal to the red circle. There are of-course infinitely many green mirrors and they produce a collection of reflections that move around points inside the red circle. Each green mirror represents a unique inversion, hence a function whose domain and co-domain are points inside the disc.

In order make calculations easier, we fix a red circle with radius 1. Next we start drawing a bunch of green mirrors (see Figure 7. Points inside the red circle, that is on the disc, are moved around by the inversions. As we observed earlier, this process preserves the cross ratio if input points are chosen from a straight line arrangement or a circle arrangement. One may naturally wonder if the cross ratio can be considered as measure of some kind of a distance and if inversions can be regarded as maps that preserve *that* special kind of distance.

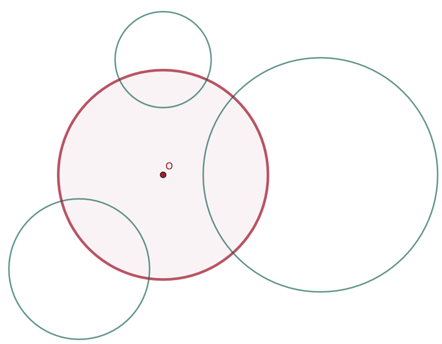

FIGURE 7. Orthogonal Mirrors on a Red Circle

First, we need to clarify what we mean by the word **distance**.

Definition 1.2. Suppose X is a set. Let d be a function that takes pairs of elements of X as input and produces a non-negative real number as output. Moreover the function d satisfies the following properties.

- $d(a, a) = 0$ for all $a \in X$
- $d(a, b) > 0$ if and only if $a \neq b$
- (symmetric condition) $d(a, b) = d(b, a)$ for all $a, b \in X$
- (triangular inequality) $d(a, b) + d(b, c) \geq d(a, c)$ for all $a, b, c \in X$

Such a function d is known as a distance or a metric on the set X.

Thinking carefully, we notice that this definition captures the essence of the concept of distance. Next we wish to understand if cross ratio can be used to construct a new type of distance function.

Let us experiment with the cross ratio. Here is the set up of that experiment.

- Fix the red circle with radius one unit.
- Fix a circular green mirror.
- Choose four points placed on an arc of a circle. Two of the four points, A, B are placed on the circumference of the red circle. Two other points C, D are placed in the interior of the red disc.
- Fix the position of C at the moment.
- Move D along the dotted circle toward C.
- What happens to the cross ratio?

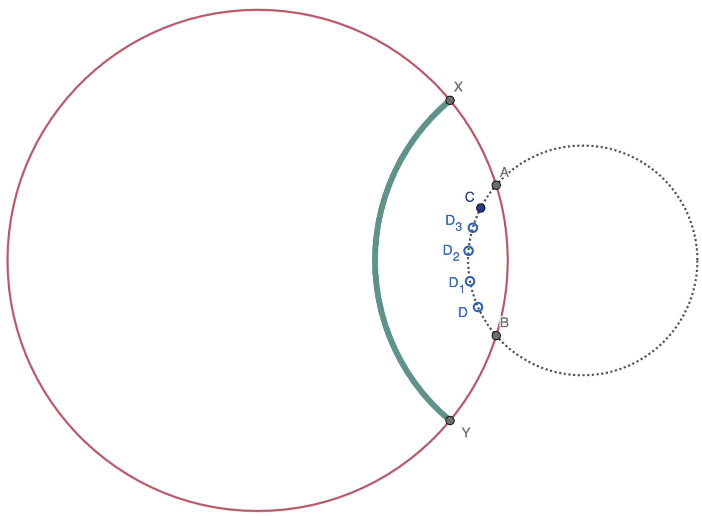

Figure 8. Experiment with Cross Ratio

We computed the values (see experiment). They are recorded in this table. The cross ratio formula that we used is as follows:

$$\frac{\frac{DA}{DB}}{\frac{CA}{CB}}$$

The good news is as D moves toward C, the value of the cross ratio gets diminished. Hence we may hope to consider the cross ratio as a measure of how far D is from C. The bad news is, while computing cross ratio we put $\frac{DA}{DB}$ at the top and $\frac{CA}{CB}$ at the denominator. This choice seems ambiguous.

Note that cross-ratio is itself monotonic (as D moves toward C it is consistently decreasing) as evident in the table. Try proving that using elementary

Position of D	**Cross Ratio**
D	17.01
D_1	7.72
D_2	3.72
D_3	2.05

TABLE 1. Change in Cross Ratio

geometric arguments. We wish induce symmetric behavior in the quantity. This means, we wish to remove the ambiguity of putting D-segments at the numerator and C-segments at the denominator.

One way mathematicians fix issues like this is to wrap the quantity with absolute value of natural logarithm function. Logarithm function itself is monotonic. Hence as cross ratio diminishes, log of cross ratio will also diminish. Similarly if cross-ratio is increasing, log of cross ratio will increase as well. Moreover we have the following beautiful property of logarithm function.

$$\log \frac{M}{N} = \log M - \log N$$

If we wrap the logarithm inside absolute value function, then we completely resolve "which one is at the top" ambiguity. In otherwords we have,

$$\left|\log \frac{\frac{DA}{DB}}{\frac{CA}{CB}}\right| = \left|\log \frac{DA}{DB} - \log \frac{CA}{CB}\right| = \left|\log \frac{CA}{CB} - \log \frac{DA}{DB}\right|$$

There is one more awesome observation. If we move D toward the circumference, suppose toward B, then DA becomes large and DB becomes very small. Hence $\frac{DA}{DB}$'s value explodes to infinity. This implies $|\log \frac{CA}{CB} - \log \frac{DA}{DB}|$ also explodes to infinity. In simple terms, the distance from C becomes very very large as D approaches the circumference at B.

Though the disc appears finite, this formula allows for infinite travel in this apparently finite space. The only thing we need to verify is whether, the formula that we created is a true distance function. At the moment, the formula is as follows.

- Suppose X, Y are any two points inside the red circle G. Choose any circle G_1 through X, Y that meets G at A, B.
- The proposed distance $d_{new}(X, Y) = |\log \frac{XA}{XB} - \log \frac{YA}{YB}|$

There is another slight ambiguity in the manner this new distance formula has been treated thus far. The ambiguity lies in the fact that we have not specified anything about G_1. For different choices of G_1, the distance value would change. For the sake of precision, we may demand that G_1 itself is orthogonal to G. Then we are able to draw the circle G_1 uniquely.

We can use inversion to determine G_1. Invert X about G to reach X'. The circle through X, X', Y is G_1. Try proving that G_1 is orthogonal to G.

There is a slightly more interesting reason to consider G_1 to be orthogonal to G. Try this experiment to notice that this arrangement maximizes the value of:

$$\left|\log \frac{XA}{XB} - \log \frac{YA}{YB}\right|$$

Lemma 1.3. *Suppose X, Y be any two points inside the unit disc and A, B are two points on the circumference G. Then $\left|\log \frac{XA}{XB} - \log \frac{YA}{YB}\right|$ is maximum when A, X, Y, B are on a circle orthogonal to G.*

Proof. Let Γ be a green circle of inversion such that the images X', Y' and the center O are in the same straight line. Can you explicitly construct such a circle of inversion using compass and straight edge? Try this computationally.

We draw coordinate axes to make calculations. Assume that the diameter through X', Y', O is the Y-axis. Suppose this diameter intersects the circle at $A' = (0, 1), B' = (0, -1)$. Let $X = (0, x), Y = (0, y)$ and pick any point $B'' = (\cos\theta, \sin\theta)$.

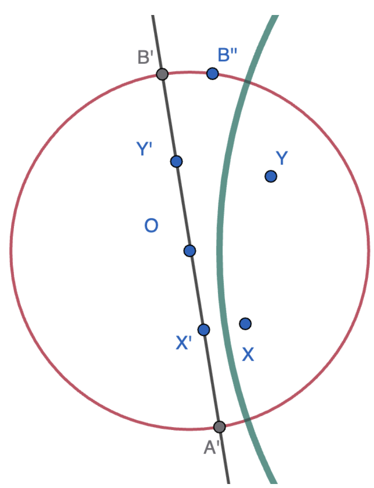

FIGURE 9. Inverting into Diameter

Let us compute the cross ratio as follows. We want to show that the cross ratio is maximized if $B'' = B'$.

$$\frac{\frac{X'A'}{X'B''}}{\frac{Y'A'}{Y'B''}} = \frac{\frac{x+1}{\sqrt{\cos^2\theta+(x-\sin\theta)^2}}}{\frac{y+1}{\sqrt{\cos^2\theta+(y-\sin\theta)^2}}}$$

This simplifies to the following expression.

$$\frac{\frac{X'A'}{X'B''}}{\frac{Y'A'}{Y'B''}} = \frac{\frac{x+1}{\sqrt{1+x^2-2x\sin\theta}}}{\frac{y+1}{\sqrt{1+y^2-2y\sin\theta}}} = \frac{x+1}{y+1}\cdot\frac{\sqrt{1+y^2-2y\sin\theta}}{\sqrt{1+x^2-2x\sin\theta}}$$

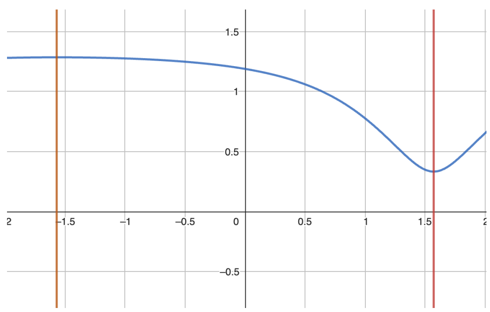

Figure 10. Monotonic Expression

Now, by differentiating with respect to θ or otherwise observe that the quantity $\dfrac{\sqrt{1+y^2-2y\sin\theta}}{\sqrt{1+x^2-2x\sin\theta}}$ extremizes at $\theta = \dfrac{\pi}{2}$. Try this experiment. Therefore $B'' = B'$ if we wish to maximize the cross-ratio type quantity.

Inverting the diameter back, we get an orthogonal circular arc passing through X, Y. Therefore absolute value of the logarithm of cross ratio is maximized in the desired configuration. □

In order to prove that d_{new} is a true distance function we need to verify the four conditions presented in Definition 1.2. The first three conditions are easy to verify. Let us explicitly show that the triangular inequality holds as well.

Proposition 1.4. *Suppose X, Y, Z are three points inside the red circle G. Draw three circles orthogonal to G as follows:*

- *Circle G_1 through X, Y meeting G at A_1, B_1*
- *Circle G_2 through Z, X meeting G at A_2, B_2*
- *Circle G_3 through Y, Z meeting G at A_3, B_3*

Show that $d_{new}(X, Y) + d_{new}(Y, Z) \geq d_{new}(Z, X)$

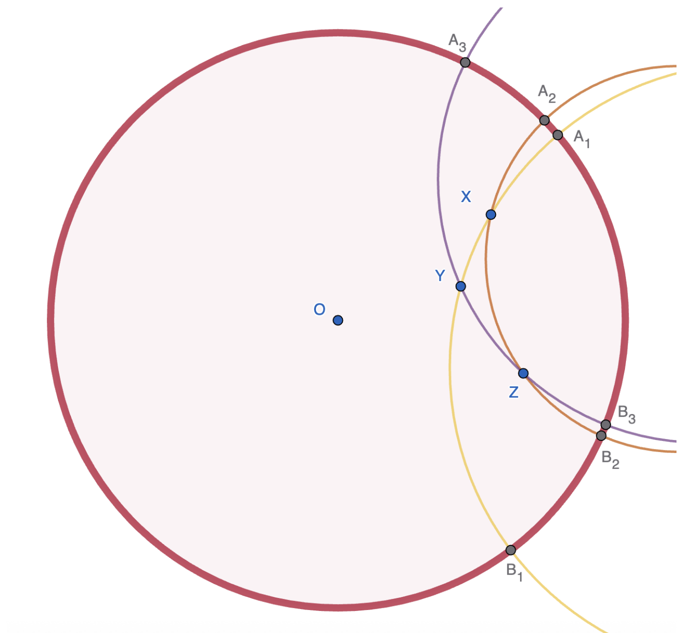

FIGURE 11. Triangular Inequality

Proof. $d_{new}(X,Y) + d_{new}(Y,Z)$ is equal to,

$$\left|\log\frac{XA_1}{XB_1} - \log\frac{YA_1}{YB_1}\right| + \left|\log\frac{YA_3}{YB_3} - \log\frac{ZA_3}{ZB_3}\right|$$

$$\geq \left|\log\frac{XA_1}{XB_1} - \log\frac{YA_1}{YB_1} + \log\frac{YA_3}{YB_3} - \log\frac{ZA_3}{ZB_3}\right|$$

Notice that if X, Y, Z are on the same orthogonal circle then we have $A_1 = A_2 = A_3 = A$ and $B_1 = B_2 = B_3 = B$

$$\begin{aligned}
&\geq \left|\log\frac{XA}{XB} - \log\frac{YA}{YB} + \log\frac{YA}{YB} - \log\frac{ZA}{ZB}\right| \\
&\geq \left|\log\frac{XA}{XB} - \log\frac{ZA}{ZB}\right| \\
&= d_{new}(X,Z)
\end{aligned}$$

Now suppose X, Y, Z are not on the same orthogonal circle. By Lemma 1.3, logarithm of cross ratio maximizes when the points in the circumference are end points of the orthogonal circle. Therefore we have

$$\left|\log\frac{XA_1}{XB_1} - \log\frac{YA_1}{YB_1}\right| \geq \left|\log\frac{XA_2}{XB_2} - \log\frac{YA_2}{YB_2}\right| \tag{1}$$

$$\left|\log\frac{YA_3}{YB_3} - \log\frac{ZA_3}{ZB_3}\right| \geq \left|\log\frac{YA_2}{YB_2} - \log\frac{ZA_2}{ZB_2}\right| \tag{2}$$

Adding (4) and (5) we have

$$\begin{aligned}
&\left|\log\frac{XA_1}{XB_1} - \log\frac{YA_1}{YB_1}\right| + \left|\log\frac{YA_3}{YB_3} - \log\frac{ZA_3}{ZB_3}\right| \\
&\geq \left|\log\frac{XA_2}{XB_2} - \log\frac{YA_2}{YB_2}\right| + \left|\log\frac{YA_2}{YB_2} - \log\frac{ZA_2}{ZB_2}\right| \\
&\geq \left|\log\frac{XA_2}{XB_2} - \log\frac{YA_2}{YB_2} + \log\frac{YA_2}{YB_2} - \log\frac{ZA_2}{ZB_2}\right| \\
&= \left|\log\frac{XA_2}{XB_2} - \log\frac{ZA_2}{ZB_2}\right| \\
&= d_{new}(X, Z)
\end{aligned}$$

□

2. Distance Preserving Maps and Their Actions

We have uncovered three beautiful objects of investigation. They are closely tied to each other. It is a new world enclosed in a circle and you may travel infinitely far away from each other inside that enclosed world. This is the classical Poincare disc model of hyperbolic geometry. The distance function that we have observed is indeed the Poincare metric. However for pedagogical reasons we avoid the word “hyperbolic" in this exposition.

The three objects are as follows:

(1) **Space** D.

    It is a disc of radius 1. We have been working with points inside the disc. The circumference of this disc is G

(2) **Distance** function d_{new}.

    Given any two points X, Y inside the disc, draw a circle Γ, passing through X, Y orthogonal to G. Notice that if X, Y and the center of G are on a straight line then the *orthogonal circle* is a diameter (imagined as an arc of circle with center very far away!) Suppose Γ intersects G at A, B then,

$$d_{new} = \left|\log\frac{XA}{XB} - \log\frac{YA}{YB}\right|$$

(3) Collection of **functions** H

These functions preserve d_{new} between pairs of points chosen from the disc. This is true because these functions (from $D \to D$) preserve cross-ratio and the d_{new} is the absolute value of the logarithm of the cross ratio. The functions that we studied so far are precisely inversions about orthogonal circles to G or reflections about diameters or composition of these two fundamental type of functions.

The distance preserving maps are often christened as isometries. The collection of isometries form a *group*. For the uninitiated, a group H is a set of elements with the following properties:

- There is a combination rule $\circ$. Combining any two elements $a, b \in H$, we may produce a third element $c \in H$ This element is denoted as $a \circ b$. In our example, the elements of H are distance preserving functions such as inversion about orthogonal circles and reflection about diameters. The combination rule is function composition.
- There is a do-nothing element $e \in H$ such that $e * h = h$ for any element $h \in H$. In our example, the do-nothing element is the identity map $f(x) = x$.
- Each element h has an inverse function h^{-1} such that $h * h^{-1} = e$. In our example, if f is a reflection about some diameter, then f is its own inverse. This is because if you reflect twice about the same line (or circle) then you are back in your own position.

There are a few other rules that elements of H must satisfy. We won't go into the detail in this exposition. Please refer to any textbook on group theory.

We have already noticed that each inversion and each reflection is a self-inverse. That is if we apply an inversion twice to any point in the disc (about some orthogonal circle), that point returns to its initial position.

$$I \circ I(x) = x$$

Though this is a composition of functions (same function composed with itself), we sometimes write this in a somewhat misleading short hand: I^2. Since this map I^2 returns all points to their initial position, therefore it is an identity map. We write this relation as:

$$I^2 = e$$

So far not much interesting is happening in the disc. However if we combine a reflection with an inversion, then immediately beautiful motion of points begin to happen inside the disc.

Suppose I_1 is inversion about a circle centered at $(1.2, 0)$ and radius $\sqrt{1.2^2 - 1^2}$. Check that I_1 is orthogonal to G. Suppose R_1 is the reflection about the diameter connecting the points $(0, -1)$ and $(0, 1)$. We wish to examine the composition function $T = R_1 \circ I_1$. Lets investigate what happens to the points on the disc when $T = R_1 \circ I_1$ is applied on them repeatedly. In order to simplify things, we target the point $(0, 0)$.

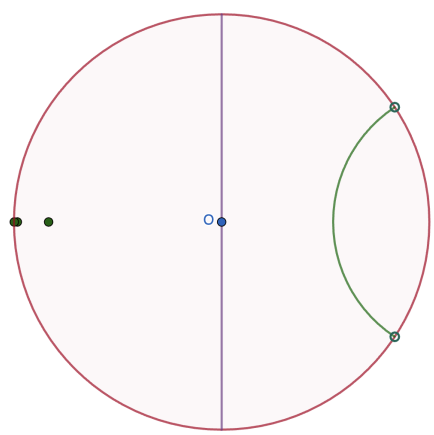

FIGURE 12. Pushing (0,0) in the disc

Iteration	Output Point
$(R_1 \circ I_1)^0$	$(0, 0)$
$(R_1 \circ I_1)^1$	$(-0.83333333333333, 0)$
$(R_1 \circ I_1)^2$	$(-0.983606557377, 0)$
$(R_1 \circ I_1)^3$	$(-0.9984984984985, 0)$
$(R_1 \circ I_1)^4$	$(-0.9998634066384, 0)$
$(R_1 \circ I_1)^5$	$(-0.9999875816506, 0)$
$(R_1 \circ I_1)^6$	$(-0.9999988710528, 0)$
$(R_1 \circ I_1)^7$	$(-0.9999998973684, 0)$

Why is this an expected behavior? Note that inverting $(0, 0)$ about the green circle would send it to the *right side* of the green circle. Reflecting the image about purple line bounces the point back to the left side of green circle. But this makes it go further away from green mirror. This back-and-forth march pushes the input point $(0, 0)$. It marches toward the circumference, but never reaches it. We say $(-1, 0)$ is a limit point of the collection of output points that we produce in this process.

Let us continue the experiment with a different starting point, say $(0.2, 0.3)$. We notice something very similar is happening. The point $(-1, 0)$ acts as an attractor point toward which points keep on marching

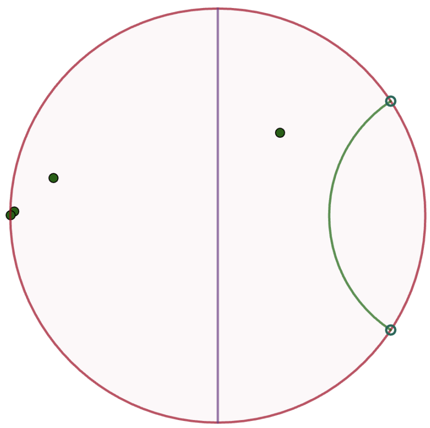

FIGURE 13. (0.2, 0.3) marches toward (-1,0)

Iteration	Point
$(R_1 \circ I_1)^0$	$(0.3, 0.4)$
$(R_1 \circ I_1)^1$	$(-0.7917525773196, 0.1814432989691)$
$(R_1 \circ I_1)^2$	$(-0.9809072164948, 0.0199587628866)$
$(R_1 \circ I_1)^3$	$(-0.9982659926676, 0.0018461863889)$
$(R_1 \circ I_1)^4$	$(-0.9998423798012, 0.0001680098911)$

Problem 2.1. Create an algebraic formula for $(R_1 \circ I_1)^k$. Prove that $(R_1 \circ I_1)^k(P) \to (-1, 0)$ as $k \to \infty$ for any point P inside the disc.

For a given function T, we observed an attractor point on the circle. We can do this for each member of H and mark the attractor and repeller points (defined later; technically they are attractor points of some group element in our set-up) on the circumference of the disc. These points on the circumference may further have accumulation points on the circumference. This collection of all attractor and repeller points and their accumulation points on the circumference is known as the *limit set* of the group H.

At the moment we exclusively work with two such functions as indicated in the diagram. We have marked the corresponding reflectors by the name of the functions.

- $T = R_1 \circ I_1$

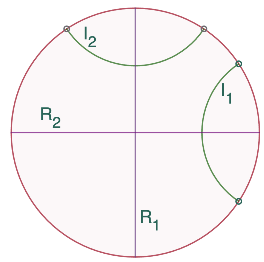

FIGURE 14. Two maps

- $S = R_2 \circ I_2$

We wish to understand the limit set of all the functions produced by these two functions T, S. We investigate how they move points inside the disc. Earlier we observed that T is pushing points toward $(-1, 0)$. Similarly we can observe that S pushes points toward $(0, -1)$. Consider the following function.

$$T \circ S \circ T \circ T \circ S$$

What does it do to the points in the disc? As an exercise you may try to find the attractor point of this function.

Notice that the inverse of T is $I_1 \circ R_1$. After all,

$$T \circ I_1 \circ R_1 = R_1 \circ I_1 \circ I_1 \circ R_1 = e$$

We denote the inverse function of T as T^{-1}. Notice that the attractor point of T^{-1} is $(1, 0)$. Similarly the attractor point of S^{-1} is $(0, 1)$. We say that *repelling* point of T is the attracting point of T^{-1}.

There is another crucial observation. The function T preserves the diameter connecting $(1, 0)$ and $(-1, 0)$. In our mind, we have a very clear picture of the dynamics of T. Points on the diameter are pushed from $(1, 0)$ toward $(-1, 0)$. As a set, the diameter is not moved by this map. All other points in the disc are flowing toward $(-1, 0)$.

One can show that each of these functions have the following following properties.

(1) They push points inside the disc toward a specific point on the circumference. This point is known as *attracting point* of that map.
(2) The attracting point of inverse of a map f, is known as *repelling point* of f.

(3) Suppose X is the attracting point and Y is the repelling point of one such function f. We observe that the function f stabilizes (does not move) the diameter or a orthogonal circular arc connecting X and Y.

The last claim needs a bit of justification. We observed this property for the function f. It stabilized the diameter connecting $(1,0)$ and $(-1,0)$. Explicitly, this means, that any point on the diameter is mapped to some other point of the diameter by the map T. What about some other transformation such as $T^2 \circ S$? How can we find its attracting point X and repelling point Y? How can we show that the map stabilizes the diameter or orthogonal circular arc connecting X and Y? One may approach this problem analytically using complex numbers. This leads to the derivation of the normal form of this type of transformations (often known as Mobius Transformation). We will leave this exploration of Mobius Transformations using complex numbers for the reader.

2.1. **Extension to boundary.** Since T pushes points toward $(-1,0)$, it is should be perfectly reasonable to deduce that it *fixes* the point $(-1,0)$. We can establish this geometrically. Notice we are now "extending" the action of T on the disc to an action of T on the circumference.

Problem 2.2. Show that T fixes $(-1,0)$

Proof. Suppose $P' = (x,0)$ is the inverse of $P = (-1,0)$ about the circle with center $O_1 = (1.2,0)$ and radius $\sqrt{1.2^2 - 1}$. Then we have,

$$OP \times OP' = 1.2^2 - 1$$

This implies,

$$\begin{aligned}
&(1.2 - x) \times (1.2 + 1) = 1.2^2 - 1 \\
&1.2^2 + 1.2 - 1.2x - x = 1.2^2 - 1 \\
&2.2 = 2.2x \\
&x = 1
\end{aligned}$$

Hence the inverse of $(-1,0)$ is $(1,0)$. If we reflect that about vertical diameter we are back to $(-1,0)$. Hence $T(-1,0) = (-1,0)$. □

Let us continue the investigation on the boundary of the disc. We will do it with a very specific function and a very useful picture.

$$T \circ S \circ T$$

The picture has two colored regions of the circumference: blue and green. The blue regions are marked about the attracting and repelling points of T.

Similarly the green regions are marked about attracting and repelling points of S. One may appreciate the following lemma using an experiment with a computational tool like Geogebra (see here) or prove it using calculations.

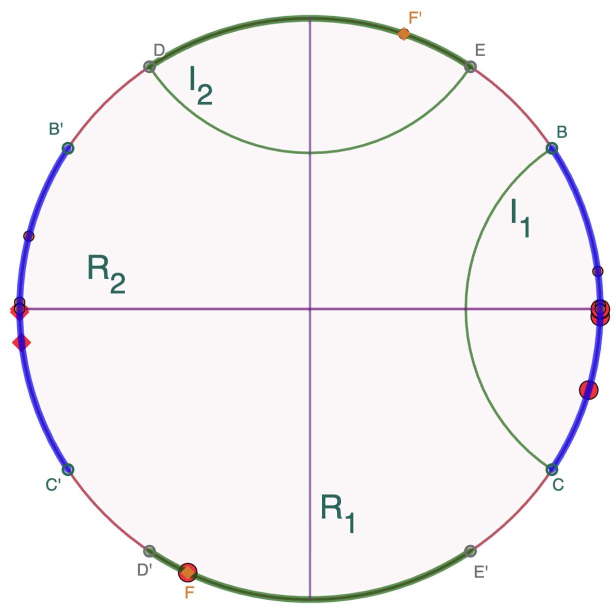

FIGURE 15. Ping Pong

Lemma 2.3. *Repeated application of T or T^{-1} on points in green region, puts them inside the blue region. Similarly repeated application of S or S^{-1} on points in blue region, puts them in green region.*

The claim is believable. We observed earlier that a point located anywhere in the disc, is pulled toward $(-1, 0)$ by the repeated application of T. Hence points very close to the green regions travel toward $(-1, 0)$ inside the disc. By extension, points on the green arcs travel toward $(-1, 0)$ along the circumference by repeated application of T.

Now we play ping pong with a point F in the green region. Let us apply $T \circ S \circ T$ on F. Clearly $T(F) \in Blue$, $S \circ T(F) \in Green$ and finally $T \circ S \circ T(F) \in Blue$. Since we started with a green point and ended up with a blue point, clearly $T \circ S \circ T$ is not an identity map. In fact the same strategy shows that $T^* \circ S^\circ \circ \cdots \circ S^* \circ T^*$ is not an identity element. Here $*$ can any non-zero positive or negative integer.

We just illustrated with an example that the group generated by T and S is free! This roughly means that any sequence of composition of functions

T, T^{-1}, S, S^{-1} is not an identity function (unless that happens trivially). In particular we investigate the *words* produced by T, T^{-1}, S, S^{-1}. For example, $T^2 \circ S^3 \circ T^{-5}$ is a word. We illustrated with an example that no such word can be an identity function. The group generated by T and S is therefore a *free* group.

This is an algebraic fact about the group generated by T, S. We derived it using a geometric argument. This argument is also known as a ping pong lemma. We have not provided explicit proofs but essentially the argument is complete. A *slight* check is needed to complete it. We must show all elements of the group generated by T and S are conjugates of the elements of the form $T^* \circ S^\circ \circ \cdots \circ S^* \circ T^*$. Try that as an exercise.

3. Discovering the Limit Set of the Free Group

The group F, generated by T, T^{-1}, S, S^{-1}, has infinitely many words. We wish to know the limit set of the entire group in the boundary circle. Recall that when we hit $(0, 0)$ by $T, T^2, T^3, \ldots$ we noticed the $(0, 0)$ approached (-1.0). Hence $(-1, 0)$ is a point in the limit set. Similarly $(0, -1)$ is another point in the limit set (found by hitting $(0, 0)$ by powers of S). We wish to list all members of the group and find the limit set of all the functions in the group.

There is a classical way of approaching this problem using human intuition. In this exposition we attempt to understand it using the power of genetic algorithms.

The first step would be to define the genes (internal instructions) or the symbols that the genetic algorithm needs to manipulate. In our example, the genes are T, T^{-1}, S, S^{-1}. The genotype would be a word in these genes. For this experiment we write Python code. It is a bit awkward to write T^{-1}. We use t instead. Similarly we use s for S^{-1}.

We use a simple Python code to generate 1000 random words using the genes with fixed length 20 . Here are the first 5 words.

Word Count	Individual
Word 1	$sSTSSTStsStstsSStSTS$
Word 2	$SsTTTSssTtsssSTsttts$
Word 3	$tSSStTstTtsSTtstsStT$
Word 4	$StsttTtTsTstSsSTSsTt$
Word 5	$SsssSSSssTttSststTtt$

Next we define the phenotype; that is the observable characteristic of each individual. In our example, it is the geometric location of the point $(0, 0)$ in the disc when a particular individual (a word made up of genes) is applied on $(0, 0)$.

Table 2. Phenotypes of Generation 0 Random Words

Individual	**Genotype (Word)**	**Final Position** (x, y)	d_{Euc}
Word 1	sSTSSTStsStstsSStSTS	$(-0.146064, -0.989275)$	1.000000
Word 2	SsTTTSssTtsssSTsttts	$(0.180018, 0.983663)$	1.000000
Word 3	tSSStTstTtsSTtstsStT	$(0.989360, 0.145367)$	0.999983
Word 4	StsttTtTsTstSsSTSsTt	$(-0.974380, -0.224908)$	1.000000
Word 5	SsssSSSssTttSststTtt	$(0.999897, 0.013814)$	0.999993

We notice that the Euclidean distance of the final point from $0, 0$ is not very informative. Hence we will use the d_{new} that we produced earlier. This fixes a fitness measure on the data. We apply an individual word (function) on $(0, 0)$ and calculate the *new* distance of the output from $(0, 0)$. Words that sends $(0, 0)$ furthest and closer to the circumference G are rewarded with a higher fitness score.

Table 3. Genotype to Phenotype Mapping and Hyperbolic Fitness (d_{new})

ID	**Genotype (Word)**	**Final Position** (x, y)	d_{Eucl}	**Fitness** (d_{new})
1	sSTSSTStsStstsSStSTS	$(-0.146064, -0.989275)$	1.000000	26.371672
2	SsTTTSssTtsssSTsttts	$(0.180018, 0.983663)$	1.000000	25.336624
3	tSSStTstTtsSTtstsStT	$(0.989360, 0.145367)$	0.999983	11.663637
4	StsttTtTsTstSsSTSsTt	$(-0.974380, -0.224908)$	1.000000	17.347494
5	SsssSSSssTttSststTtt	$(0.999897, 0.013814)$	0.999993	12.503735

After calculating the *new* fitness ranks for Generation 0, we isolate the top 100 ‘elite’ individuals—those words whose phenotypes lie furthest from the origin in the d_{new} metric. To produce Generation 1, we employ a mutation strategy: each elite word generates 10 variants through random single-gene alterations. This process creates a new batch of 1,000 individuals that are genetically predisposed to high fitness but possess enough variation to discover new regions of the limit set.

In fact, if we plot these phenotypes of generation 1 in the disc, we immediately observe an emergent pattern!

We notice that the points are clustering in some spots on the circle but they are not everywhere. This pattern is emerging in generation 1. We can (and should) run the selection and mutation experiment for several generations. If we zoom in further, closer to $(-1, 0)$ we start noticing a self-similar fractal

TABLE 4. Top 5 Elite Parents from Generation 0 (Random Population)

Rank	**Genotype (Word)**	**Final Position** (x, y)	**Fitness** (d_{new})
0	ststsTTsTSTStSTSSstt	$(1.0000, -0.0015)$	37.429948
1	ttSSstttttsssTStsstt	$(0.9999, 0.0162)$	37.429948
2	tstsstStttttSTsstSTs	$(-0.2253, 0.9743)$	36.736801
3	sTTSSSSSSTStttStSss	$(0.2093, 0.9779)$	36.331335
4	STSSTSsSTssTSTTTTTss	$(-0.0165, 0.9999)$	36.331335

TABLE 5. Top 5 Mutated Variants from Generation 1

Rank	**Genotype (Word)**	**Final Position** (x, y)	**Fitness** (d_{new})
0	stSTStSSSTTTSSSTSSstss	$(-0.0198, 0.9998)$	inf
1	ststsTTsTSTStSTSSstt	$(1.0000, -0.0015)$	37.429948
2	ststsTTsTSTStSTSSstt	$(1.0000, -0.0015)$	37.429948
3	ststsTTsTSTStSTSSstt	$(1.0000, -0.0015)$	37.429948
4	ststsTTsTSTStSTSSstt	$(1.0000, -0.0015)$	37.429948

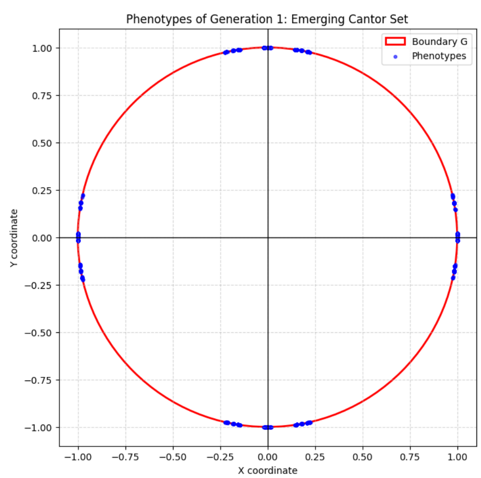

FIGURE 16. Scatter Plot of Phenotypes

like pattern. Rather than a continuous arc, we observe a hierarchy of satellite clusters separated by distinct 'forbidden' gaps.

The fact that it looks like a fractal (and in fact a Cantor set) is not sufficient to *prove* it is a Cantor set. There is a rigorous topological proof of that fact. However the example is useful for illustration purposes. Instead of intuitively guessing the limit set, we have a robust "first guess" aided by genetic algorithms.

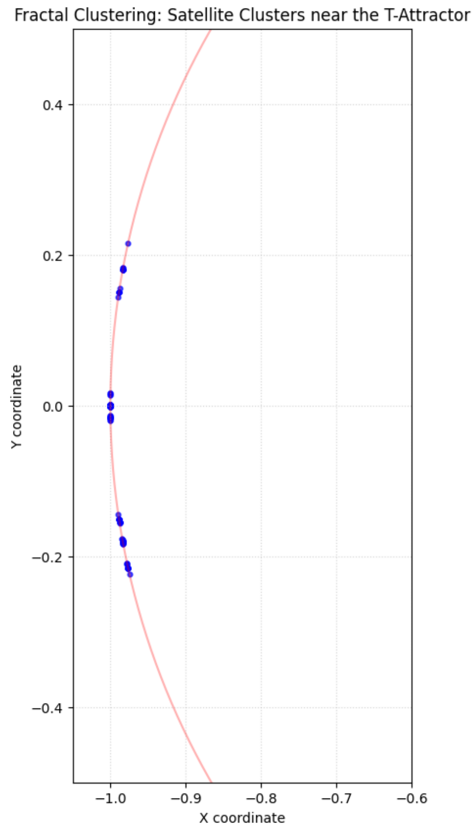

FIGURE 17. Zoomed in at (-1, 0)

4. A Gentle Walk toward Bowditch Boundary

One of the key methods of doing mathematics is trying a bunch of examples and noticing patterns. The picture that we produced earlier is quite intriguing.

We notice that there are big and small gaps in the clusters. No red points are landing in those gaps. Suppose A_1, B_1 are end-points of one such gap. Is there a word that fixes A_1, B_1? It would stabilizes the orthogonal arc connecting A_1, B_1. How can we discover such a word?

If we pinch the endpoints A_1, B_1 of the blue arc, we collapse the corresponding complementary gap of the Cantor limit set. Because the limit set is self-similar, zooming in reveals infinitely many such gaps. If we collapse every complementary gap arc to a point, the resulting quotient space has no gaps and is therefore homeomorphic to a circle. In this sense, we have modified the boundary: the loxodromic words that stabilized these gap endpoints are now "neutralized" at the boundary level. The resulting circle may be

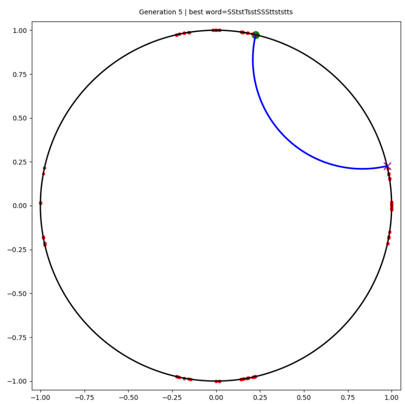

FIGURE 18. Gap Stabilizer

viewed as the boundary of F relative to the directions determined by these gap-stabilizing elements.

Can we computationally discover a gap gluing word? Indeed we can try. Here are the conceptual steps of the experiment.

(1) Enumerate all reduced non-trivial words in $\{T, S, t, s\}$ of length at most 12.
(2) For each word w, compute the two fixed points of the corresponding Möbius transformation on the boundary circle.
(3) Compare the unordered pair of fixed points of w with $\{A_1, B_1\}$ using a suitable distance metric.
(4) Among those words whose fixed points lie within a prescribed tolerance of $\{A_1, B_1\}$, select the shortest ones.

Here are the first 10 in the rank-list. Notice that in rank 2 we have,

$$TstS = TS^{-1}T^{-1}S = [T, S^{-1}],$$

the commutator of T and S^{-1}.

The word that ranked 1 is simply the inverse of $TstS$. This immediately encourages us to explore the commutator elements and gap-gluing words. Moreover, we are forced to ask a more careful question: is it possible to have $A_1 = B_1$? Is it possible to change T, S and make that happen?

Problem 4.1. Suppose F is a free group generated by T, S, T^{-1}, S^{-1} acting on the unit disc as discussed above. Let ΛF be the limit set marked in red color. Show that if we collapse each gap arc in $S^1 \backslash \Lambda(F)$ (whose endpoints may be detected as fixed points of suitable words, e.g. commutator-type elements), then the resulting quotient of S^1 is a circle.

Rank	**Word**	**Length**	Max. Dist.
1	sTSt	4	0.0170579
2	TstS	4	0.0170579
3	TststS	6	0.00525793
4	sTSTSt	6	0.00525793
5	sTSSTSt	7	0.000514774
6	TstsstS	7	0.000514774
7	TsttstS	7	0.00636153
8	sTSTTSt	7	0.00636153
9	TstssstS	8	0.000109728
10	sTSSSTSt	8	0.000109728

TABLE 6. Shortest reduced words (within tolerance) whose boundary fixed points are closest to $\{A_1, B_1\}$.

We now take the experiment one step further. Can we choose T and S such that the big gap is extinguished in the boundary circle (or we have $A_1 = B_1$). Let us draw orthogonal circles of reflection that creates a curvy looking square inside the disc. The *arc* that is stabilized by the gap-gluing word is no longer an arc! Its end-points are also glued into a yellow circle that is tangential to the red circle.

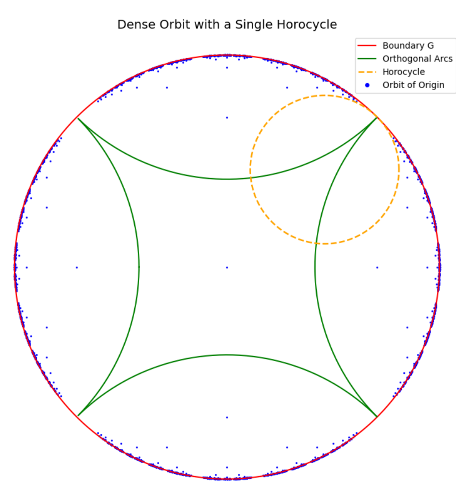

FIGURE 19. Horocycle

In this set-up limit set of the orbit of $(0, 0)$ in the boundary circle is the entire boundary circle. The gaps are closed. The stabilizer of each collapsed gap P is called a parabolic subgroup. Most orbit points of $(0, 0)$ stay outside the yellow circle stabilized by parabolic subgroup. This type of yellow circle is often known as a horocycle. In-fact if we hit a horo-cycle by some other group element (not in the parabolic group) then we get other yellow circles which are also horocycles. They are stabilized by conjugates of P. Outside horocycles points are pushed around isometrically (and properly-discontinuously).

The disc that we created is known as the Poincare model of the hyperbolic plane. It naturally leads to a concept of negative curvature which we won't discuss in this exposition.

The group generated by T, S acts in this hyperbolic plane. For a point $(0, 0)$ the orbits mostly avoid regions bounded by horocycles stabilized by parabolic subgroups and their conjugates. Hence we say the group generated by T, S is hyperbolic relative to the parabolic subgroups.

Cheenta Academy, Kolkata 700029, India
Email address: ashani.dasgupta@cheenta.com