The Modular Group
A Finitely Generated Group with Interesting Subgroups

Paul R. McCreary
Teri Jo Murphy
Christian Carter

The action of Möbius transformations with real coefficients preserves the hyperbolic metric in the upper half-plane model of the hyperbolic plane. The modular group is an interesting group of hyperbolic isometries generated by two Möbius transformations, namely, an order-two element \( g_2(z) = -1/z \) and an element of infinite order \( g_\infty(z) = z + 1 \). Viewing the action of the group elements on a model of the hyperbolic plane provides insight into the structure of hyperbolic 2-space. Animations provide dynamic illustrations of this action.

This article updates an earlier article [1].

1. Introduction

Transformations of spaces have long been objects of study. Many of the early examples of formal group theory were the transformations of spaces. Among the most important transformations are the isometries, those transformations that preserve lengths. Euclidean isometries are translations, rotations and reflections. The groups and subgroups of Euclidean isometries of the plane are so familiar to us that we may not think of them as revealing much about the space they transform. In hyperbolic space, however, light traveling or even a person traveling on a hyperbolic shortest-distance path tends to veer away from the boundary. Thus, the geometry is unusual enough so that viewing the actions of isometries of hyperbolic 2-space reveals some of the shape of that space. Two-dimensional hyperbolic space is referred to as the hyperbolic plane.
Here are graphic building blocks used for all of the animations.

\[
\text{RightTriangle} = \text{Module}\left[\{\text{arc, center, start, end}\}, \right.
\text{arc} = \text{Table}\left[\text{center + Cos[start + (end - start) n / 21.]} \right.
\text{I Sin[start + (end - start) n / 21.], \{n, 21\}}; \right.
\text{Join[}
\text{arc} /. \{\text{center \rightarrow 1, start \rightarrow Pi, end \rightarrow 2 Pi / 3}, \right.
\text{arc} /. \{\text{center \rightarrow 0, start \rightarrow Pi / 3, end \rightarrow Pi / 2}, \right.
\text{Table}[\text{I (1 - n / 21), \{n, 21 - 1}\}, \right.
\text{\{.001 I\}
\text{]}]; \right.
\\]
\]
\[
\text{VerticalStrip} = \frac{-1}{\text{RightTriangle}};
\]
\[
\text{LeftTriangle} = -\text{Conjugate}\@\text{RightTriangle};
\]
\[
\text{RightVerticalStrip} = \frac{-1}{\text{LeftTriangle}};
\]

Figure 1 shows four cyan and white regions, each bounded by some combination of three arcs or rays. Any two adjacent regions make up a fundamental region. The two fundamental regions shown on either side of the \(y\) axis (each with one white and one cyan half) are related by the function \(-1/z\), which is an inversion over the unit circle composed with a reflection in the \(y\) axis.

\text{Graphics[}\left[\left[\right.
\text{Circle[\{}\right.\right., \right.
\text{EdgeForm[Black]}, \right.
\text{White, Polygon@*ReIm@LeftTriangle,} \right.
\text{Polygon@*ReIm@RightVerticalStrip,} \right.
\text{Cyan, Polygon@*ReIm@RightTriangle,} \right.
\text{Polygon@*ReIm@VerticalStrip} \right.\right.\left.\right], \right.
\text{PlotRange \rightarrow \{\{-1, 1\}, \{0, 2\}\}, ImageSize \rightarrow 200 \{1, 1\},} \right.
\text{Background \rightarrow LightGray}\right.\right.\right]}
This article examines how elements of the modular group rearrange the triangular-shaped regions shown in Figure 2. The curved paths are arcs of circles orthogonal to the x axis. Arcs on these circles are hyperbolic geodesics, that is, shortest-distance paths in hyperbolic 2-space. In Euclidean space, the shortest-distance paths lie on straight lines. In hyperbolic space, shortest paths lie on circles that intersect the boundary of the space at right angles. Hyperbolic distances are computed as if there is a penalty to pay for traveling near to the plane’s boundary. Thus, the shortest-distance paths between two points must bend away from the boundary.

```
Graphics[
  
Cyan, EdgeForm[Black],
  Polygon@ReIm@Flatten[
    Table[j - \frac{1}{k - \frac{1}{n+VerticalStrip}}, {j, -2, 2}, {k, -4, 4},
    {n, -2, 3}], 2
  ],

PlotRange -> {{-1.5, 1.505}, {-2, 2}}, ImageSize -> 450]
```

▲ Figure 2. The upper half-plane model of the hyperbolic plane.
In the animations that follow, it is instructive to focus on the action that a transformation takes on the family of circles that meet the $x$ axis at right angles. The transformations that we consider, namely members of the modular group, preserve this family of circles. The circles in the family are shuffled onto different members in the family, but no new circles are created and none are taken away. One could say that in the context of hyperbolic geometry, the transformations preserve the family of all shortest-distance paths. Indeed this is an excellent thing for an isometry to do!

The context of this article can be found described in Chapter 2 of [2]. In this small text, one can find illustrations that inspired our animations. The formulas, which made coding the animations much simpler than one might expect, are given and justified in detail.

### 2. Möbius Transformations

First consider a class of functions known as Möbius transformations. These transformations are named after the same mathematician with whom we associate the one-sided, half-twisted Möbius band. Möbius transformations are defined by

$$f(z) = \frac{az + b}{cz + d}, \quad a, b, c, d \in \mathbb{C}, \quad ad - bc \neq 0.$$ 

Here $\mathbb{C}$ stands for the complex numbers. Over the reals, a Möbius transformation $f(z) = \frac{a z + b}{c z + d}$ with real coefficients falls into one of two categories: either $c = 0$, and the graph is a straight line, or $c \neq 0$, and the graph is a hyperbola. A representation of this latter type of function is shown in Figure 3.

```
Show[
  Plot[(2 x + 1) / (x + 1), {x, -10, 10},
    PlotRange -> {{-5, 5}, {-4, 6}}],
  Graphics[{Dashing[.02, .05]], Line[{{-50, 2}, {50, 2}}],
    Line[{{-1, -50}, {-1, 50}}]},
  PlotRange -> {{-5, 5}, {-4, 6}}]
```

▲ Figure 3. Graph of $f(x) = \frac{2x+1}{x+1}$ shown with dashed asymptotes.
Our purpose is to investigate how Möbius transformations stretch and twist regions in the \textit{extended complex plane}. The \textit{complex plane} is the usual Euclidean plane with each point identified as a complex number—namely, \((x, y) = x + iy\). The \textit{extended complex plane} is formed from the complex plane by adding the point at infinity. A Möbius transformation is one-to-one (injective) on the extended complex plane \(\mathbb{C}^* = \mathbb{C} \cup \{\infty\}\).

When a Möbius transformation \(F\) acts on a complex number, \(F(x + iy) = u + iv\), we may view the action as moving the point \((x, y)\) to the point \((u, v)\). Importantly, a Möbius transformation maps the set \(S\) of circles and lines in \(\mathbb{C}^*\) back to \(S\) in \(\mathbb{C}^*\). A comprehensive proof of this fact may be found in most elementary texts on complex variables, for example, in [3], p. 158.

The figures of our animations live in the extended complex plane. Each point of a figure, taken as a complex number, is acted on by the Möbius transformations. These transformations spin hyperbolic 2-space about a fixed point or shift the space in one direction or another.

\section{3. The Modular Group and Hyperbolic Space}

The \textit{modular group} \(M\) is a special class of Möbius transformations:

\[
M := \left\{ f(z) = \frac{az + b}{cz + d} \mid a, b, c, d \in \mathbb{Z}, ad - bc = 1 \right\}.
\]

That is, if \(f \in M\), the coefficients of \(f\) are integers and the coefficient matrix \(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\) has determinant equal to one.

What is a group? Recall that a \textit{group} is a set \(G\) together with a binary operation satisfying certain properties: (1) the set \(G\) must be closed under the operation; (2) the operation must be associative; and (3) there must be an identity element for the operation, and all inverses of elements in \(G\) must themselves be elements of \(G\). The proof that the modular group is, in fact, a group under the operation of function composition is a standard exercise in a course on complex analysis. (See, for example, [3], p. 277–278.)

We take as established that the elements of the modular group do indeed form a group and investigate some of the interesting subgroups.

\section{Fundamental Regions}

One of our main goals is to investigate how the elements of the modular group act on fundamental regions. That is to say, how the regions are stretched and bent when we view them as Euclidean objects. As hyperbolic objects, the regions are all carbon copies of each other, in much the same way that the squares on a checkerboard are all identical in ordinary Euclidean geometry.
In general, a group of one-to-one transformations acting on a topological space partitions that space into \textit{fundamental regions}. For a collection of sets \{\(U_i\)\} to be a collection of fundamental regions, certain properties must hold. First and foremost, the \(U_i\) must be pairwise disjoint. Second, given any transformation \(f\) in the group other than the identity, \(U_i\) and \(f(U_i)\) are disjoint. Finally, given any two regions \(U_i\) and \(U_j\), there exists some transformation \(f\) such that \(f(U_i) = U_j\).

Generally, in order to cover the entire space without overlapping, each fundamental region must contain some but not all of its boundary points. This technicality is set aside for the purposes of this article.

In fact, in this article we relax the definition to include \textit{all} of the boundary points for a particular fundamental region. Thus, adjacent fundamental regions can only overlap on their boundaries. The essential feature remains that there is no \textit{area} in the intersection of adjacent regions.

A group of transformations does not necessarily yield a unique partition of the space into fundamental regions. Thus, the fundamental regions we view are merely representative fundamental regions.

Figure 4 shows a fundamental region of the modular group with some parts highlighted.
A fundamental region with vertices marked and a pair of tangents.

Each fundamental region contains four vertices that can be fixed by elements of the modular group. (A point $f$ is fixed by $f$ if $f(p) = p$.) Tangents are drawn at one vertex; the angle is 60 degrees. The vertex at the top has a straight, 180 degree angle. The vertex at the bottom has a zero degree angle because the tangents to the intersecting arcs coincide there. Any hyperbolic polygon with a vertex on the boundary of the space (the $x$ axis in this case of the upper half-plane) has a zero degree angle at that vertex. The corresponding four angles in each fundamental region have the same measures as those indicated here. Each vertex can be fixed by some element in the modular group. Further, each fundamental region can be mapped onto any other fundamental region by an element of the modular group.

A classic view of the matter is to see the upper half-plane as tessellated (or tiled) by triangular-shaped regions, as in Figure 2. A checkerboard tessellation of the Euclidean plane can be constructed by sliding copies of a square to the left, right, up and down. Eventually, the plane is covered with square tiles. The modular group tessellates the hyperbolic plane in an analogous way. The elements of the group move copies of a fundamental region until triangular-shaped tiles cover the upper half-plane model of the hyperbolic plane. Of course, these tiles do not appear to be identical to our eyes, trained to match shapes and lengths in Euclidean geometry. However, the triangular-shaped tiles are all identical if measured using the hyperbolic metric. In the tiling process, all areas in the upper half-plane are covered by tiles and no two tiles have any overlapping area. In fact, this procedure is precisely how the hyperbolic plane illustration was constructed. The boundary points for a single fundamental region were acted on by function elements of the modular group, and the resulting points were drawn as a boundary line in the illustration.

It helps to note that each transformation in $\mathcal{M}$ has at least one fixed point. Some transformations in $\mathcal{M}$ have two fixed points. Only the identity map has more than two. In the illustrations that follow, we observe the placement of fixed points and the way transformations map fundamental regions near them.
Hyperbolic Lengths

The hyperbolic metric is a rather curious metric that challenges our notion of distance. Under the hyperbolic metric in the upper half-plane, the shortest distance between two points is along a vertical line or an arc of a circle perpendicular to the boundary (the real axis). For example, the shortest hyperbolic path between the points \( z = i \) and \( z = 1 + i \) is the top arc of the circle \( \{ z : | z - \frac{1}{2} | = \frac{\sqrt{5}}{2} \} \), which passes through both points and is perpendicular to the real axis (Figure 5).

```
Show[
  Graphics[
    
    Circle[.5, 0], Sqrt[5/4],
    Text[Style["i", Italic, 14], (0, 1), (-7, 2)],
    Text[Style[Row[{1, " + ", Style["i", Italic]}], 14],
         {1, 1}, {1.5, 2}],
    Red, Circle[.5, 0], Sqrt[5/4],
          {ArcTan[2], Pi - ArcTan[2]}],
    RGBColor[0, 0, 0], PointSize[.02], Point[{(0, 1), {1, 1}}],
    PlotRange -> {{-7, 1.7}, {-1.2, 1.2}}
    ], Axes -> True
  ]
```

\[ \text{Figure 5.} \text{ The shortest hyperbolic path between the points } i \text{ and } 1 + i. \]
Without discussing precisely how hyperbolic lengths and areas are measured, we state that every image under a transformation in the modular group is congruent to every other image under the hyperbolic metric. Thus, all of our fundamental regions shown in the animations are actually the same size in the hyperbolic metric. For a discussion on hyperbolic metrics, [4] is a good place to start.

4. Structure of the Modular Group

We structure our investigation of the modular group by considering four cyclic subgroups. Recall that a cyclic subgroup can be generated by computing all powers of a single group element. The four cyclic subgroups we present are representative of the four possible types of subgroups found in the modular group.

- Four Examples of Cyclic Subgroups

- \( \langle g_2 \rangle \)

For the first subgroup, consider the function \( g_2(z) = -1/z \); it is a Möbius transformation with coefficients \( a = 0, b = -1, c = 1, d = 0 \) and coefficient matrix \( \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \). The subscript indicates that \( g_2 \) is of order two in \( M \); that is, \( g_2(g_2(z)) = z \), and so \( g_2 \) is its own inverse. In this case, \( g_2 \) generates a subgroup with only two elements, namely \( M_2 = \langle g_2 \rangle = \{z, -1/z\} \).

In this article, we adopt the standard notation that angular parentheses \( \langle \rangle \) indicate the set of elements generated by taking products from the elements enclosed by the parentheses. Curly braces \( \{\} \), on the other hand, enclose the delineated list of elements in a set.

\[
g_2[z_] := -1/z
\]

The Möbius transformation \( \rho \), its inverse \( \sigma \) and the function \( R2 \) are used for the motion in Figure 6.

\[
\rho[z_] := \frac{z - 1}{z + 1}
\]

\[
\sigma[z_] := \frac{z 1 + 1}{-z + 1}
\]

\[\]
\begin{verbatim}
R2[F_, t_] :=
    Table[
        ReIm@Table[σ[Exp[2 Pi I t] ρ[F[[j]]], {j, Length[F]}][[i]],
            {i, 2}]
    Manipulate[
        Graphics[
            
            
            EdgeForm[Black],
            LightGray,
            Line@Flatten[Join[
                Table[ReIm[n - \frac{1}{RightTriangle}], {n, -2, 2}],
                Table[ReIm[m - \frac{1}{Table[n - \frac{1}{RightTriangle}, {n, -2, 2}]}],
                    {m, -2, 2}]
            ], 1],
            (* background figures *)
            Table[{
                Lighter@Red, Darker@Green}[[n]],
                Line[R2[{LeftTriangle, \frac{1}{LeftTriangle}}, t6][[n]]],
                Polygon[R2[{RightTriangle, \frac{1}{RightTriangle}}, t6][[n]]
            ], {n, 2}]
            (* pinweel arms *)
            ],
            PlotRange -> {{-1.5, 1.5}, {-1, 3}}, Axes -> True,
            Ticks -> None, ImageSize -> 400],
            {{t6, 0, "rotation ", 0, 1, Appearance -> "Labeled"},
            SaveDefinitions -> True
        ]
    
\end{verbatim}
The Manipulate depicts the way in which $g_2$ maps the two fundamental regions shown in Figure 2 onto one another. In fact, the action of $g_2$ on the fundamental regions is to hyperbolically rotate them $180^\circ$ onto each other about the central fixed point $z = i$. The actual mapping is performed instantaneously without rotation. In particular, only the first, middle and final frames contain illustrations of fundamental regions. However, the sequence of intermediate mappings illustrates through animation the mapping properties of $g_2$. In a later section, we discuss how the functions illustrated were broken into a composition of functions so that the hyperbolic nature of their motion was made continuous.

This example highlights the fact that vertical lines are paths of least distance in the upper half-plane model of the hyperbolic plane. Indeed, it is usual to view straight lines as circles that have radii with infinite length and that pass through the point at infinity. With this bending of the definition of a circle, a vertical line has all the characteristics required of a geodesic in the hyperbolic plane. Like the circles, a vertical line is perpendicular to the $x$ axis, which is the boundary of the upper half-plane model of the hyperbolic plane. A vertical line is the limit of a sequence of geodesic circles.
The second example (see Figure 7) is a subgroup of infinite order generated by the linear shift (or translation) \( g_\infty(z) = z + 1 \); it is a Möbius transformation with \( a = 1, b = 1, c = 0, d = 1 \) and matrix \( \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \). The function \( g_\infty \) has infinite order because \( g_\infty^n(z) = z + n \), and no point of \( \mathbb{C} \) ever returns to its original position no matter how many times \( g_\infty \) is applied, though \( g_\infty(\infty) = \infty \) in the extended plane. The subgroup produced by taking all powers of \( g_\infty \) and its inverse \( g_\infty^{-1}(z) = z - 1 \) is denoted as \( M_\infty = \langle g_\infty \rangle = \{ \ldots, z - 2, z - 1, z, z + 1, z + 2, \ldots \} \).

Every point in the plane shifts one unit to the right under the action of \( g_\infty \). The infinite half-strips in the following Manipulate are images of each other under powers of \( g_\infty \). For contrast, we also provide images of these infinite half-strip regions under the map \( g_2(z) = -1/z \). These images are bunched in a flower-like arrangement attached to the real axis at the origin. As the blue infinite regions are pushed from left to right, their magenta images echo their motion in a counterclockwise direction. These two actions are not produced by a single transformation. The two transformations that cause these actions are closely related to each other as algebraic conjugates, but more on that in a later section.

```mathematica
Manipulate[
Graphics[
{
LightGray,
Line[Flatten[Join[
{Table[ReIm[n - \[LeftTriangle] 1 \[RightTriangle]], {n, -2, 3}]},
Table[ReIm[m - \[LeftTriangle] 1 \[RightTriangle]], {n, -2, 2}]}
], 1]],
(* background figures *)
Lighter@Blue,
Line@
ReIm[8 (t7 - 1/2) + Table[ReIm[n - \[LeftTriangle] 1 \[RightTriangle]], {n, -1, 1}]],
(* lines on vertical fundamental region *)
EdgeForm[Black],
Polygon@
ReIm[8 (t7 - 1/2) + Table[ReIm[n - \[LeftTriangle] 1 \[RightTriangle]], {n, -1, 1}]],
(* polygons on vertical fundamental region *)
Lighter@Magenta,
```
\[
\text{Line@ReIm}\left[\frac{-1}{8 \left(t7 - 1/2\right)} - \frac{1}{\text{Table}[\left(-1, (n,-1,1)\right)]}\right],
\]

\[
\text{Polygon@ReIm}\left[\left(-1\right)\frac{1}{8 \left(t7 - 1/2\right)} - \frac{1}{\text{Table}[\left(-1, (n,-1,1)\right)]}\right]
\]

\[
(* \text{polygons on fundamental region at origin} *)
\]

\[
\text{PlotRange} \rightarrow \{\{-2.5, 2.5\}, \{-1, 2.3\}\}, \text{Axes} \rightarrow \text{True},
\]

\[
\text{Ticks} \rightarrow \text{None}, \text{ImageSize} \rightarrow 1.2 \{400, 200\}
\]

\[
\{\{t7, .5, "\text{translation }\}, 0, 1, \text{Appearance} \rightarrow \"Labeled\"\},
\]

\[
\text{SaveDefinitions} \rightarrow \text{True}
\]

\[\]

**Figure 7.** This animation shows copies of fundamental regions moving back and forth, with corresponding regions anchored at the origin.
The hyperbolic isometry $g_\infty$ is notable among the elements of the modular group because it is also a Euclidean isometry. Under the hyperbolic metric, the magenta regions are each congruent to the half-strip regions in blue.

$\langle g_3 \rangle$

The third cyclic subgroup of $M$ that we consider is generated by the composition of the first two functions $g_\infty$ and $g_2$; define $g_3(x) = g_\infty \circ g_2(z) = \frac{z-1}{z}$. The subgroup generated by this element is denoted $M_3 = \langle g_3 \rangle = \{z, \frac{z-1}{z}, \frac{-1}{z-1}\}$, a subgroup of order 3.

$$g_3[z_] := \frac{z-1}{z}$$

Here is the fixed point of $g_3$.

$$\tau = \frac{1}{2} \left(1 + i \sqrt{3}\right);$$

$$g_3[\tau] = \tau \quad \text{// Simplify}$$

True

Define $\mu$ and its inverse $\nu$.

$$\mu[z_] := \frac{z - \tau}{z - \text{Conjugate}[\tau]}$$

$$\nu[z_] := \frac{-\text{Conjugate}[\tau] z + \tau}{-z + 1}$$

In Figure 8, the function $\mu$ moves the fixed point $\tau$ to the origin and $\nu$ moves it back.

$$\mu[\tau] \quad \text{// Simplify}$$

0

$$\nu[0]$$

$$\frac{1}{2} \left(1 + i \sqrt{3}\right)$$

The function $R3$ is an order-three hyperbolic rotation made continuous; it is used in Figures 8 and 11.

$$R3[F\_, t\_] :=$$

Table[
  ReIm@Table[\nu[\mu[[j]]] Exp[2/3 Pi I t]], {j, Length[F]}][[i]], {i, 3}]
Manipulate[
  Graphics[
    
    EdgeForm[Black],
    LightGray,
    Line@Flatten[Join[
      Table[ReIm[n - \frac{1}{\text{RightTriangle}}], {n, -1, 2}],
      Table[ReIm[m - \frac{1}{n - 1/\text{RightTriangle}}], {n, -1, 2}],
      {m, -1, 2}]]], 1],
    (* background figures *)
    Table[
      {{Lighter@Blue, Darker@Green, Lighter@Red}@[m],
        Line@R3[{LeftTriangle, 1 - \frac{1}{\text{LeftTriangle}}},
         1 - \frac{1}{\text{LeftTriangle}}, \text{t8}][[m]],
        Polygon@R3[{RightTriangle, 1 - \frac{1}{\text{RightTriangle}}},
         1 - \frac{1}{\text{RightTriangle}}, \text{t8}][[m]],
        {m, 1, 3}]
    ],
    PlotRange -> {{-1, 2}, {-1, 3}}, Axes -> True, Ticks -> None,
    ImageSize -> 400,
    
    {{t8, 0, "rotation ", 0, 3, Appearance -> "Labeled"},
     SaveDefinitions -> True}]
]
\[ g_3(z) = \frac{z-1}{z}. \]

\textbf{Figure 8.} Action of the order-three element \( g_3(z) = \frac{z-1}{z}. \)

A red fundamental region and a green fundamental region are shown associated with the blue fundamental region attached to the origin in the animation’s first frame. We include these in order to provide a better orientation for the scene. Of special interest is how the point of the blue region on the \( x \) axis moves as the rotation takes place. The point begins at the origin and slides toward the right along the positive \( x \) axis. The blue lines of the cluster become vertical precisely when that point arrives at the point at infinity! The point continues by sliding along the negative \( x \) axis to arrive back at the origin. It is fair to say that the motion of a point as it passes through the origin is a “mirror image” of the motion of the point as it passes through the point at infinity. The function used here is a composition of Möbius transformations that is described and demonstrated in Figure 12.
The rotations we saw in the action of \( g_2 \) and \( g_3 \) are of orders two and three, respectively. That is to say, after the rotation is repeated a number of times, all points are back to their original positions. In contrast, the function \( g_\infty \) generates an infinite subgroup. When we iterate \( g_\infty \), the right shifts accumulate at the point at infinity. Points in the left half-plane get repelled by infinity, while points in the right half-plane get attracted to infinity. Of course, since all points in the left half-plane eventually map to points in the right half-plane, all points are, in some sense, simultaneously attracted to and repelled by infinity under the action of \( g_\infty \). Indeed, the point at infinity is the single fixed point for the action of \( g_\infty \).

The transformation \( g_h(z) = \frac{2z+1}{z+1} \) in the modular group generates an infinite subgroup \( M_h \) that differs from \( M_\infty \) in the sense that \( g_h \) has two distinct fixed points, an attractor and a repeller.

Define the well-known golden ratio \( \phi \); its reciprocal is \( 1 / \phi = 1 - \phi \).

\[
\phi = 1 / 2 \left(1 + \sqrt{5}\right); \]

\[
1 / \phi = \phi - 1
\]

True

This defines \( g_h \) with fixed points \( \phi \) and \( 1 - \phi \).

\[
gh[z_] := \frac{2z+1}{z+1}
\]

\[
gh @@ \{\phi, 1 - \phi\} // \text{Simplify}
\]

\[
\left\{\frac{1}{2} \left(1 + \sqrt{5}\right), \frac{1}{2} \left(1 - \sqrt{5}\right)\right\}
\]

Define \( \alpha \) to move the fixed points \( \phi \) and \( 1 - \phi \) to infinity and zero, respectively.

\[
\alpha[z_] := \frac{z - (1 - \phi)}{z - \phi}
\]

\( \alpha[\phi] // \text{Simplify} // \text{Quiet} \)

\( \text{ComplexInfinity} \)

\( \alpha[1 - \phi] // \text{Simplify} \)

0
Define $\beta$ to be the inverse of $\alpha$; $\beta$ sends infinity back to $\phi$ and zero back to $1 - \phi$.

$$\beta[z_] := \frac{-\phi z + (1 - \phi)}{-z + 1}$$

Limit[$\beta[z], z \to \infty$]

$$\frac{1}{2} \left(1 + \sqrt{5}\right)$$

$\beta[0] // Simplify$

$$\frac{1}{2} \left(1 - \sqrt{5}\right)$$

For Figure 9, $\gamma$ makes the hyperbolic translation continuous. The ratio $\frac{\alpha[1]}{\alpha[0]}$ is the length of the hyperbolic translation.

$$\gamma[z_, e_] := \beta \left(\frac{\alpha[1]}{\alpha[0]}\right)^e \alpha[z]$$

Manipulate[

Module[{topOfUnitCircle, MovingRegion},

topOfUnitCircle = Table[Cos\left[\frac{2 \pi k}{114}\right] + I Sin\left[\frac{2 \pi k}{114}\right],

\{k, 0, 57\}];

MovingRegion = Join[

Table[n + VerticalStrip, \{n, -4, 4\}],

\frac{1}{1 - \frac{1}{\text{Table}[n + \text{VerticalStrip}, \{n, 0, 4\}]}}

\frac{-1}{\text{Table}[\text{VerticalStrip} + n, \{n, -4, 0\}]}\right]

\frac{-2 - \frac{1}{\text{Table}[n + \text{VerticalStrip}, \{n, 0, 4\}]}}{\text{Table}[n + \text{VerticalStrip}, \{n, 0, 4\}]}

];

Graphics[

\{Lighter@Gray,

Line@ReIm@Join[

Table[n + VerticalStrip, \{n, -4, 4\}],

\}

\}
\begin{verbatim}
-1
  Table[n + VerticalStrip, {n, -2, 2}]
]

(* background figures *)

InfiniteLine[{{-1, 0}, {1, 0}}],
Blue, Line@Re[Im[y[MovingRegion, t9]]],
Red, Circle[{-1, 0}, 1, {0, Pi}],
Darker[Green], Circle[{{2, 0}, 1, {0, Pi}}],
Cyan, Line@Re[Im[gh[topOfUnitCircle + 2]],
(* boundary circles *)

Cyan, Line@Re[Im[y[topOfUnitCircle + 2, t9]],
Darker[Green], Line@Re[Im[y[topOfUnitCircle - 1, t9]],
(* boundary of moving region *)

PointSize[.01],
Red, Point[{-1 - \phi, 0}],
Darker[Green], Point[{\phi, 0}]
(* fixed points *)

],
AspectRatio -> Automatic, PlotRange -> {{-4, 4}, {-2, 4}},
ImageSize -> 500]

{{t9, 0, "translate"}, 0, 1, Appearance -> "Labeled"},
SaveDefinitions -> True
\end{verbatim}
\[ g_h(z) = \frac{2z+1}{z+1}. \]

The \textit{manipulate} depicts the action of \( g_h \) on fundamental regions in the plane. All points exterior to the red circle on the left are mapped to the interior of the green circle on the right. The animation begins with regions that lie exterior to the red and green circles. These regions are all mapped to the area between the green circle and smaller cyan circle. If the action of \( g_h \) were to be repeated, the regions would be mapped into the interior of increasingly small circles inside the smallest (cyan) circle shown. The attracting fixed point for \( g_h \) lies within these shrinking, nested circles.

The rotations and translations we have seen as examples are intimately related to Euclidean rotations and translations, as discussed in Section 6. The transformation \( g_h \) is related in a similar way to a Euclidean dilation, which turns a figure into a similar but not congruent image figure. A curious characteristic of hyperbolic space is that the distinction between similarity and congruence disappears. In the hyperbolic plane, it is enough for two figures to have the same angles to guarantee congruence. In marked contrast to Euclidean space, equal angles guarantee that corresponding side lengths are equal in the hyperbolic metric.

\section*{Generators of \textit{M}}

The entire group \( M \) can be generated by the two functions \( g_2 \) and \( g_\infty \). In symbols, \( M = \langle g_2, g_\infty \rangle \). Establishing this fact requires tools from linear algebra about which we make only a few brief comments. The group of \( 2 \times 2 \) matrices with real number entries and with nonzero determinants is denoted by \( \text{GL}[2, \mathbb{R}] \). This group has been studied extensively and much is known about it. Thus, there are great advantages for any group that can be represented as a subgroup of \( \text{GL}[2, \mathbb{R}] \). While the modular group cannot be represented in
exactly this way, it almost can be. For instance, the elements \(
\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}
\) and \(
\begin{pmatrix} -1 & -2 \\ -1 & -3 \end{pmatrix}
\) are considered distinct in \(\text{GL}[2, \mathbb{R}]\). On the other hand, the actions of the two associated Möbius transformations are identical, since \(\frac{z-2}{z-3} = \frac{-cz+2}{-cz+3} = \frac{z+2}{z+3}\). In general, for every element in the modular group, there are two associated elements in \(\text{GL}[2, \mathbb{R}]\).

A remarkable feature of Möbius transformations is that the group operation of compositions produces coefficients that are identical to the results of \(2 \times 2\) matrix multiplication. To see this, consider the two Möbius transformations \(\frac{az+b}{cz+d}\) and \(\frac{ez+f}{gz+h}\). First, multiply the associated \(2 \times 2\) matrices.

\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}.
\]

Second, carry out the composition of the two functions.

\[
\frac{a\frac{ez+f}{gz+h}+b}{c\frac{ez+f}{gz+h}+d} = \frac{a(ez+f)+b(gz+h)}{c(ez+f)+d(gz+h)} = \frac{(ae+bg)z+(af+bh)}{(ce+dg)z+(cf+dh)}.
\]

The coefficients and the four matrix entries are the same!

In this way, the group operation of composition of functions in the modular group can be replaced with the group operation of matrix multiplication in \(\text{GL}[2, \mathbb{R}]\). It is down this path we would travel if we were to present a complete proof of the claim that the modular group is generated by the two elements \(g_2\) and \(g_∞\).

A major part of this claim is that any element of \(M\) can be written as a composition of \(g_2\) and \(g_∞\). Consider the following examples of compositions.

\[
g_∞ \circ g_2(z) = \frac{z-1}{z}.
\]

\[
g_∞ \circ g_∞ \circ g_2(z) = \frac{2z-1}{z}.
\]

\[
g_∞ \circ g_∞ \circ g_2 \circ g_∞ \circ g_∞ \circ g_2(z) = \frac{3z-2}{2z-1}.
\]

Each of these functions has a coefficient matrix with determinant equal to one. A worthy exercise for undergraduate mathematics students is to verify by direct computations that each equality holds for the indicated compositions.
5. Conjugate Subgroups

The four examples of cyclic subgroups outlined in Section 4 give a complete description of the four types of subgroups possible in the modular group. Any subgroup of the modular group is *conjugate* to a subgroup generated by \( g_2, g_3 \), an iterate of \( g_\infty \) or an iterate of an element of the same type as \( g_h \).

Recall that in a group \( G \), a subgroup \( H \) is *conjugate* to a subgroup \( K \) if there exists an element \( g \in G \) such that the entire subgroup \( H \) can be generated by computing \( g k g^{-1} \) for every \( k \in K \). More compactly, we write \( H = \{ g k g^{-1} | k \in K \} \), and even more compactly, \( H = g K g^{-1} \).

Here we define the hyperbolic translation \( \eta \) that relates two order-three elements: \( g_3 \) and \( f \).

\[
\eta[z_\_] := \frac{z}{z + 1}
\]

Consider a function \( f \) in the modular group that generates an order-three subgroup (Figure 10).

\[
f[F_, t_] :=
\text{Table}[
\text{ReIm@Table}[\eta[\nu[F[[j]]] \text{Exp}[2/3 \text{Pi I t}]],
{\nu, \text{Length}[F]][[i]], (i, 3)]
\]

\[
\text{RotatingRegions} = \{
\text{Table}[n + \text{VerticalStrip}, \{n, 0, 2\}],
\text{Table}[-1/\text{VerticalStrip} + n + 1, \{n, 0, 2\}],
\text{Table}[-1/\text{VerticalStrip} + n - 1, \{n, 0, 2\}]
\};
\]

\[
\text{RotatingRegionsRight} = \{
\text{Table}[\text{RightVerticalStrip} + n, \{n, 0, 2\}],
\text{Table}[-1/(\text{RightVerticalStrip} + n) + 1, \{n, 0, 2\}],
\text{Table}[-1/(\text{RightVerticalStrip} + n - 1), \{n, 0, 2\}]
\};
Manipulate[
  Graphics[
    
    
    
    
    
  ]
]
Figure 10. The action of the order-three function \( f \) on a selection of fundamental regions.

We view side by side the actions of \( f \) on the right and \( \eta^{-1} g_3 \eta \) on the left. For the figure on the left, first the function \( \eta \) moves the fixed point of \( f \) onto the fixed point of \( g_3 \). Then the function \( g_3 \) rotates the attached fundamental regions, as we have seen it do before, while at the same time the function \( f \) acts on the right-hand figure. Finally, the inverse of \( \eta \) returns the fixed point and associated regions to the original position, except that the fundamental regions have been rotated in the same way as those on the right. Thus, the final results are the same in both cases.

The function \( \zeta \) is used for the continuous motion in Figure 11.

\[
\zeta[z_\_\_, t\_] := \frac{z + t - 1}{z + t}
\]
graphics11[t_, table1_, table2_] := Framed@Graphics[
{EdgeForm[Black],
LightGray,
Line[Flatten[Join[
    {Table[ReIm/@ (-1/RightTriangle + n), {n, -1, 2}]},
    Table[
        ReIm/@ (-1/Table[-1/RightTriangle + n, {n, -1, 2}] +
         m), {m, -1, 2}]], 1]],
    table1, table2], PlotRange -> {{-.6, 1.5}, {-1, 2}},
    ImageSize -> {250, 250}
}]

Manipulate[
    Grid[{
        Which[
            t11 <= 1/3,
            {graphics11[1 - 3 t11, 
                Table[{{Lighter@Red, Lighter@Blue, Darker@Green}[[n]], Polygon[
                    ReIm/@ζ[RotatingRegions, 1 - 3 t11][[n, 2]]],
                    {n, 3}],
                Table[{{Lighter@Red, Lighter@Blue, Darker@Green}[[n]],
                    Line[ReIm/@ζ[RotatingRegionsRight, 1 - 3 t11][[n, 2]]],
                    {n, 3}]
                ],
                Spacer[10],
                graphics11[1,
                    Table[{{Lighter@Red, Lighter@Blue, Darker@Green}[[n]], Polygon[
                        ReIm/@ζ[RotatingRegions, 1][[n, 2]]}, {n, 3}],
                    Table[{{Lighter@Red, Lighter@Blue, Darker@Green}[[n]],
                        Line[ReIm/@ζ[RotatingRegionsRight, 1][[n, 2]]],
                        {n, 3}]
                    ]}
                },
            1/3 < t11 <= 2/3,
            {graphics11[3 (t11 - 1/3), 
                Table[{{Darker@Green, Lighter@Red, Lighter@Blue}[[n]], Polygon[
                    R3[RotatingRegions, 3 (t11 - 1/3 - del)][[n, 2]]],
                    {n, 3}],
                Table[{{Darker@Green, Lighter@Red, Lighter@Blue}[[n]]
}

The Modular Group
Figure 11. The action of \( \eta^{-1} g_3 \eta \) and \( f \).

This animation demonstrates what it means for two functions to be conjugate equivalent.

# 6. Coding Considerations

All functions in the modular group are abruptly discontinuous; that is, their actions move triangular regions onto other regions all in one jump. The facility to produce transformations that seem continuous is due to the following.

Every element of the modular group is conjugate equivalent to one of three Euclidean transformations, namely a rotation about the origin, a scaling from the origin or a rigid translation of the entire plane ([2], pp. 12–20).

These Euclidean transformations have very simple continuous forms:

- Rotation: \( \mathcal{R}_\theta(z) = e^{i\theta} z, \ \theta \in [0, 2\pi) \).
- Scaling: \( \mathcal{S}_k(z) = k z, \ k \in (0, \infty) \).
- Translation: \( \mathcal{T}_a(z) = z + a, \ a \in \mathbb{C} \).

The left-hand side of Figure 12 shows the fixed point of \( g_3 \) translated to the origin. Following this transformation, all circles that passed through the original fixed point become straight lines passing through the origin. A Euclidean rotation about the origin accomplishes the desired rearrangement of the regions. Finally, translating the fixed points back to their original positions maps the fundamental regions to their proper, final positions. We see that the final results are the same for the right and left animations.

Indeed, each frame in the right-hand animation was computed by composing the functions that are explicitly portrayed in the left-hand animation.
\begin{align*}
generalLFT[z_-, z_1, w_1, z_2, w_2, z_3] := \quad & N[-(-w_1 + w_2) \left( \frac{w_2}{w_1 - w_2} + \frac{(z - z_2)(z_1 - z_3)}{(z_1 - z_2)(z - z_3)} \right)] \\
(* z_3 -> \text{Infinity} *)
moveToOrigin[z_, t_] := 
\text{generalLFT}[z, .5 + \text{Sqrt}[3] / 2 (2 - t) I, 0, \quad & .5 - \text{Sqrt}[3] / 2 (1 - t) I, -(1 + 2.5 (1 - t)) I, \quad & .5 - (\text{Sqrt}[3] / 2 + 2 \text{Sqrt}[3] / 2 (1 - t)) I - \\
\text{generalLFT}[.5 + I \text{Sqrt}[3] / 2, \quad & .5 + (\text{Sqrt}[3] / 2 + \text{Sqrt}[3] / 2 (1 - t)) I, 0, \quad & .5 - \text{Sqrt}[3] / 2 (1 - t) I, -(1 + 2.5 (1 - t)) I, \quad & .5 - (\text{Sqrt}[3] / 2 + 2 \text{Sqrt}[3] / 2 (1 - t)) I + \\
(.5 + I \text{Sqrt}[3] / 2) (1 - t)]
\end{align*}

\text{graphics12[table_] := Framed[Graphics[
{LightGray, 
Circle[{0, 0}, 1], Line[{{-10, 0}, {10, 0}}] 
(* background figures *), 
EdgeForm[Black], 
table 
}], 
PlotRange -> {{-1.5, 1.5}, {-1, 2}}, 
ImageSize -> {250, 250}]]

\text{Manipulate[}
\text{Which[}
\text{t12 < 0,}
\text{Grid[}
{ 
\text{graphics12[}
\text{Table[(
{Lighter@Blue, Darker@Green, Lighter@Red}[[m]], 
Line[
ReIm/@
((LeftTriangle, -1/LeftTriangle + 1, 
1/(-1/LeftTriangle + 1) + 1)[[m]]), 
Polygon[
ReIm/@
((RightTriangle, -1/RightTriangle + 1, 
1/(-1/RightTriangle + 1) + 1)[[m]]),
{m, 1, 3}]
(* arms of pinwheel *)
], 
Spacer[10],
graphics12[}
Table[
  {Lighter@Blue, Darker@Green, Lighter@Red}[[m]],
  Line[
    ReIm/@
    ({{LeftTriangle, -1 / LeftTriangle + 1, 
       -1 / (-1 / LeftTriangle + 1) + 1}[[m]]},
     Polygon[
      ReIm/@
      ({{RightTriangle, -1 / RightTriangle + 1, 
         -1 / (-1 / RightTriangle + 1) + 1}[[m]]},
       {m, 1, 3}]
     ]
   ]
  ]},
0 ≤ t12 && t12 ≤ 1 / 3,
Grid[
  {graphics12[
   Table[
     {Lighter@Blue, Darker@Green, Lighter@Red}[[m]],
     Line[
       ReIm/@
       (moveToOrigin[{{LeftTriangle, -1 / LeftTriangle + 1, 
          -1 / (-1 / LeftTriangle + 1) + 1}, 3 t12}[[m]]],
        Polygon[
          ReIm/@
          (moveToOrigin[{{RightTriangle, 
             -1 / RightTriangle + 1, 
             -1 / (-1 / RightTriangle + 1) + 1}, 3 t12}[[
                 m]]},
                   {m, 1, 3}]
          ]
        ],
     ],
     Spacer[10],
     graphics12[
      Table[
       {Lighter@Blue, Darker@Green, Lighter@Red}[[m]],
       Line[
         ReIm/@
         ({{LeftTriangle, -1 / LeftTriangle + 1, 
            -1 / (-1 / LeftTriangle + 1) + 1}[[m]]},
          Polygon[
           ReIm/@
           ({{RightTriangle, -1 / RightTriangle + 1, 
               -1 / (-1 / RightTriangle + 1) + 1}[[m]]},
             {m, 1, 3}]
           ]
         ],
      ]
     ]},
t12 > 1 / 3 && t12 ≤ 2 / 3,
Grid[
{
  graphics12[
    Table[
      {Lighter@Blue, Darker@Green, Lighter@Red}[[m]],
      Line[
        ReIm/@(E^(I (t12 - 1/3) 2 Pi)
          moveToOrigin[{LeftTriangle, -1 / LeftTriangle + 1, -1 / (-1 / LeftTriangle + 1) + 1}, 1][[m]]),
        Polygon[
          ReIm/@(E^(I (t12 - 1/3) 2 Pi)
            moveToOrigin[{RightTriangle, -1 / RightTriangle + 1, -1 / (-1 / RightTriangle + 1) + 1}, 1][[m]]),
          {m, 1, 3}
        ]
      ],
      Spacer[10],
      graphics12[
        Table[
          {Lighter@Blue, Darker@Green, Lighter@Red}[[m]],
          Line[R3[{LeftTriangle, -1 / LeftTriangle + 1, -1 / (-1 / LeftTriangle + 1) + 1}, 3 (t12 - 1/3) ][[m]]],
          Polygon[
            R3[{RightTriangle, -1 / RightTriangle + 1, -1 / (-1 / RightTriangle + 1) + 1}, 3 (t12 - 1/3) ][[m]]],
          {m, 1, 3}
        ]
      ]
    ],
    t12 > 2 / 3 && t12 <= 1,
  Grid[
    {graphics12[
      Table[
        {Lighter@Blue, Darker@Green, Lighter@Red}[[m]],
        Line[
          ReIm/@
          (moveToOrigin[{-1 / (-1 / LeftTriangle + 1) + 1, LeftTriangle, -1 / LeftTriangle + 1},
            1 - 3 (t12 - 2 / 3)][[m]])],
          Polygon[
            ReIm/@
            (moveToOrigin[{-1 / (-1 / RightTriangle + 1) + 1, RightTriangle, -1 / RightTriangle + 1},
              1 - 3 (t12 - 2 / 3)][[m]])],
          {m, 1, 3}
        ]
    ]
  ]
}
In this way, the action of the hyperbolic motions can be animated as continuous because the Euclidean rotations, translations and dilations can all be coded as continuous functions.

\section*{References}


About the Authors

Paul R. McCreary  
The Evergreen State College-Tacoma  
Tacoma, WA  
mccrearp@evergreen.edu

Teri Jo Murphy  
Department of Mathematics & Statistics  
Northern Kentucky University  
Highland Heights, KY

Christian Carter  
Department of Mathematics  
Xavier University of Louisiana  
New Orleans, LA