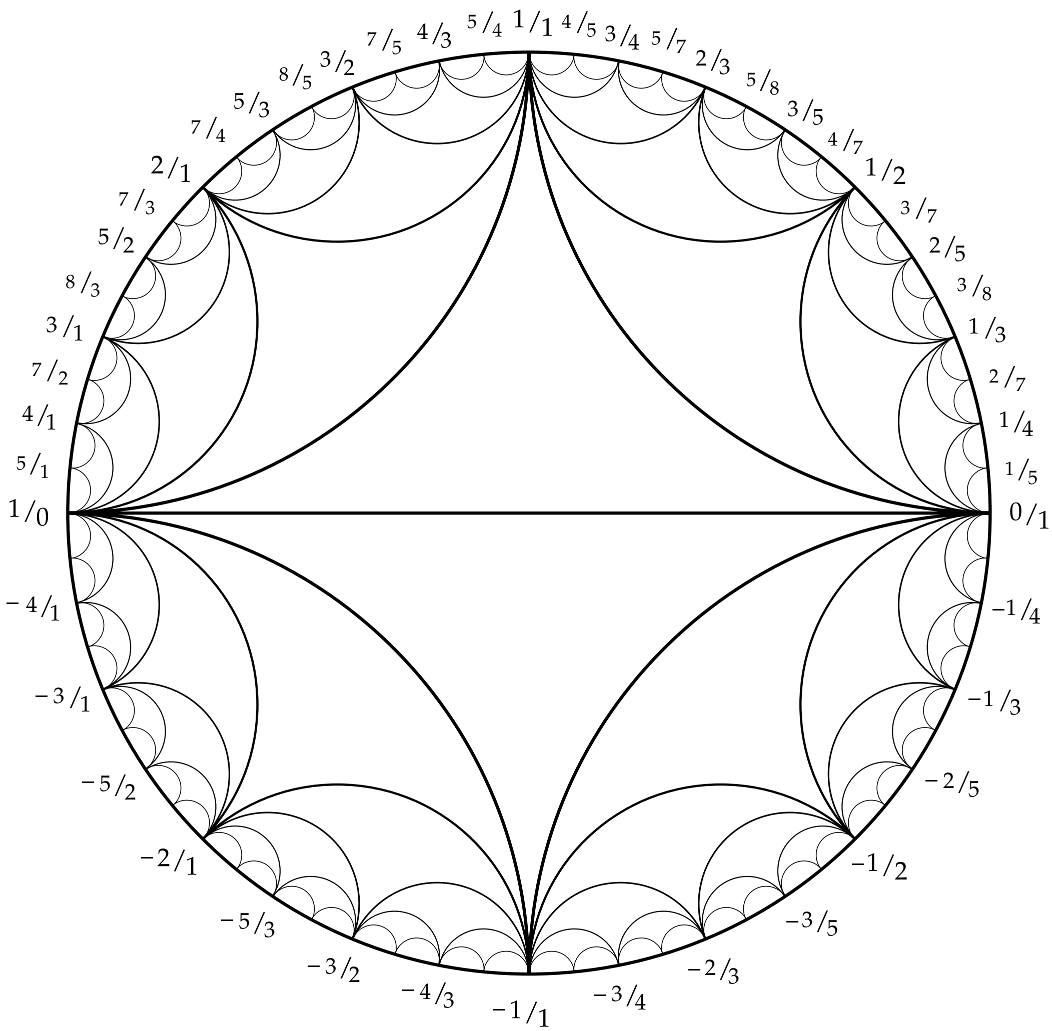
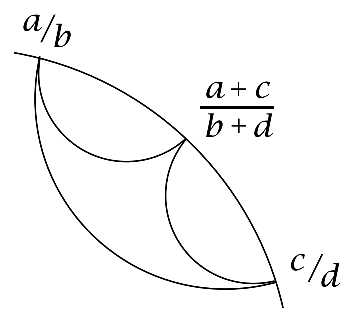
1.1 Constructing the Farey Diagram

Our goal is to use geometry to study numbers. Of the various kinds of numbers, the simplest are integers, along with their ratios, the rational numbers. The large figure below shows a very interesting diagram displaying rational numbers and certain relations between them that we will be exploring. This diagram, along with several variants of it that will be introduced later, is known as the *Farey diagram*. The origin of the name will be explained when we get to one of these variants.



What is shown here is not the whole diagram but only a finite part of it. The actual diagram has infinitely many curvilinear triangles, getting smaller and smaller out near the boundary circle. The diagram can be constructed by first inscribing the two big triangles in the circle, then adding the four triangles that share an edge with the two big triangles, then the eight triangles sharing an edge with these four, then sixteen more triangles, and so on forever. With a little practice one can draw the diagram without lifting one's pencil from the paper: First draw the outer circle starting at the left or right side, then the diameter, then make the two large triangles, then the four next-largest triangles, etc.

The vertices of all the triangles are labeled with fractions a/b, including the fraction $1/0$ for ∞, according to the following scheme. In the upper half of the diagram first label the vertices of the big triangles $0/1$, $1/1$, and $1/0$ as shown. Then by induction, if the labels at the two ends of the long edge of a triangle are a/b and c/d, the label on the third vertex of the triangle is $\frac{a+c}{b+d}$. This fraction is called the *mediant* of a/b and c/d.



The labels in the lower half of the diagram follow the same scheme, starting with the labels $0/1$, $-1/1$, and $-1/0$ on the large triangle. Using $-1/0$ instead of $1/0$ as the label of the vertex at the far left means that we are regarding $+\infty$ and $-\infty$ as the same. The labels in the lower half of the diagram are the negatives of those in the upper half, and the labels in the left half are the reciprocals of those in the right half.

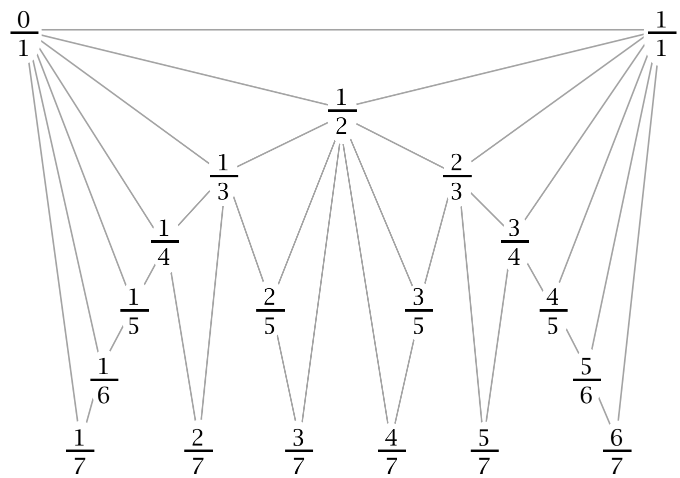
The labels occur in their proper order around the circle, increasing from $-\infty$ to $+\infty$ as one goes around the circle in the counterclockwise direction. To see why this is so, it suffices to look at the upper half of the diagram where all numbers are positive. What we want to show is that the mediant $\frac{a+c}{b+d}$ is always a number between $\frac{a}{b}$ and $\frac{c}{d}$ (hence the term "mediant"). Thus we want to see that if $\frac{a}{b} > \frac{c}{d}$ then $\frac{a}{b} > \frac{a+c}{b+d} > \frac{c}{d}$. Since we are dealing with positive numbers, the inequality $\frac{a}{b} > \frac{c}{d}$ is equivalent to $ad > bc$, and $\frac{a}{b} > \frac{a+c}{b+d}$ is equivalent to $ab + ad > ab + bc$ which follows from $ad > bc$. Similarly, $\frac{a+c}{b+d} > \frac{c}{d}$ is equivalent to $ad + cd > bc + cd$ which also follows from $ad > bc$.

We will show in the next section that the mediant rule for labeling vertices in the diagram automatically produces labels that are fractions in lowest terms. It is not immediately apparent why this should be so. For example, the mediant of $1/3$ and $2/3$ is $3/6$, which is not in lowest terms, and the mediant of $2/7$ and $3/8$ is $5/15$, again not in lowest terms. Somehow cases like this don't occur in the diagram.

Another non-obvious fact about the diagram is that all rational numbers occur eventually as labels of vertices. This will be shown in the next section as well.

Farey Series

We can build the set of rational numbers by starting with the integers and then inserting in succession all the halves, thirds, fourths, fifths, sixths, and so on. Let us look at what happens if we restrict to rational numbers between 0 and 1. Starting with 0 and 1 we first insert $1/2$, then $1/3$ and $2/3$, then $1/4$ and $3/4$, skipping $2/4$ which we already have, then inserting $1/5$, $2/5$, $3/5$, and $4/5$, then $1/6$ and $5/6$, etc. This process can be pictured as in the following diagram:



The interesting thing to notice is:

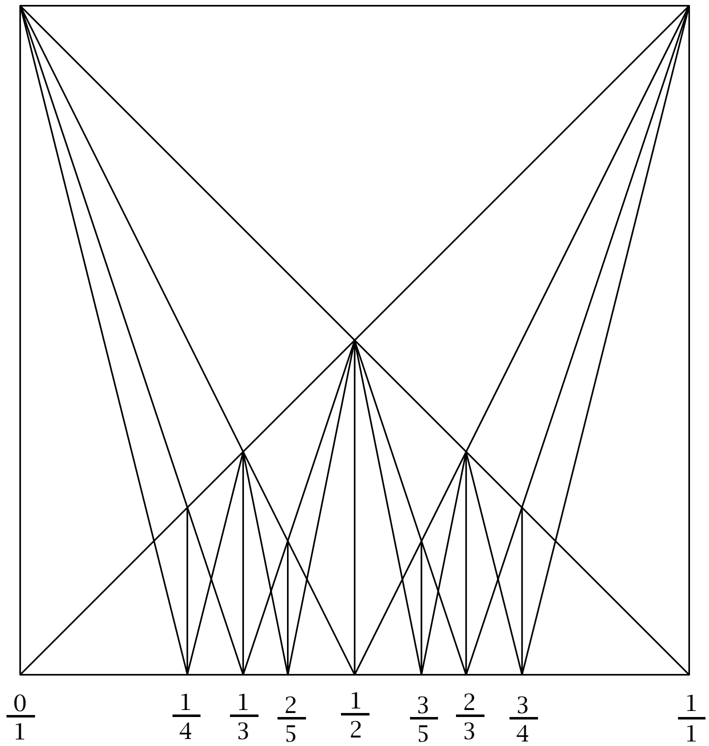
> *Each time a new number is inserted, it forms the third vertex of a triangle whose other two vertices are its two nearest neighbors among the numbers already listed, and if these two neighbors are a/b and c/d then the new vertex is exactly the mediant $\frac{a+c}{b+d}$.*

The discovery of this curious phenomenon in the early 1800s was initially attributed to a geologist and amateur mathematician named Farey, although it turned out that he was not the first person to have noticed it. In spite of this confusion, the sequence of fractions a/b between 0 and 1 with denominator less than or equal to a given number n is usually called the nth *Farey series* F_n. For example, here is F_7:

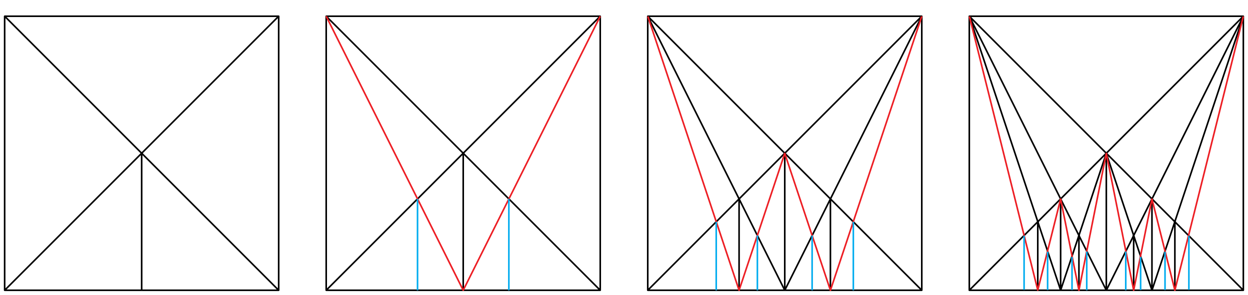
$$\frac{0}{1}\ \frac{1}{7}\ \frac{1}{6}\ \frac{1}{5}\ \frac{1}{4}\ \frac{2}{7}\ \frac{1}{3}\ \frac{2}{5}\ \frac{3}{7}\ \frac{1}{2}\ \frac{4}{7}\ \frac{3}{5}\ \frac{2}{3}\ \frac{5}{7}\ \frac{3}{4}\ \frac{4}{5}\ \frac{5}{6}\ \frac{6}{7}\ \frac{1}{1}$$

These numbers trace out the up-and-down path across the bottom of the figure above. For the next Farey series F_8 we would insert $1/8$ between $0/1$ and $1/7$, $3/8$ between $1/3$ and $2/5$, $5/8$ between $3/5$ and $2/3$, and finally $7/8$ between $6/7$ and $1/1$.

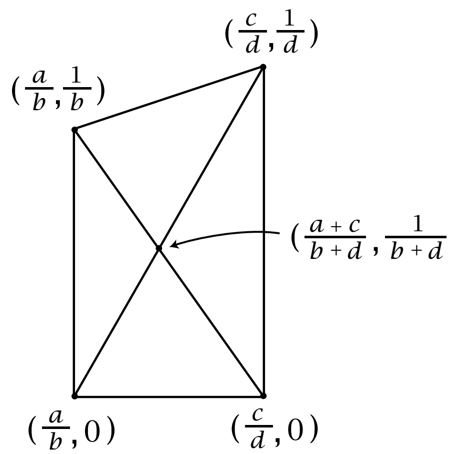
There is a cleaner way to draw the preceding diagram using straight lines in a square:



One can construct this diagram in stages, as indicated in the sequence of figures below. Start with a square together with its diagonals and a vertical line from their intersection point down to the bottom edge of the square. Next, connect the resulting midpoint of the lower edge of the square to the two upper corners of the square and drop vertical lines down from the two new intersection points this produces. Now add a W-shaped zigzag and drop verticals again. It should then be clear how to continue.



A nice feature of this construction is that if we start with a square whose sides have length 1 and place this square so that its bottom edge lies along the x-axis with the lower left corner of the square at the origin, then the construction assigns labels to the vertices along the bottom edge of the square that are exactly the x coordinates of these points. Thus the vertex labeled $1/2$ really is at the midpoint of the bottom edge of the square, and the vertices labeled $1/3$ and $2/3$ really are $1/3$ and $2/3$ of the way along this edge, and so forth. In order to verify this fact the key observation is the following: For a vertical line segment in the diagram whose lower endpoint is at the point $(\frac{a}{b}, 0)$ on the x-axis, the upper endpoint is at the point $(\frac{a}{b}, \frac{1}{b})$. This is obviously true at the first stage of the construction, and it continues to hold at each successive stage since for a quadrilateral whose four vertices have coordinates as shown in the figure at the right, the two diagonals intersect at the point $(\frac{a+c}{b+d}, \frac{1}{b+d})$. For example, to verify that $(\frac{a+c}{b+d}, \frac{1}{b+d})$ is on the line from $(\frac{a}{b}, 0)$ to $(\frac{c}{d}, \frac{1}{d})$ it suffices to show that the line segments from $(\frac{a}{b}, 0)$ to $(\frac{a+c}{b+d}, \frac{1}{b+d})$ and from $(\frac{a+c}{b+d}, \frac{1}{b+d})$ to $(\frac{c}{d}, \frac{1}{d})$ have the same slope. These slopes are



$$\frac{1/(b+d) - 0}{(a+c)/(b+d) - a/b} \cdot \frac{b(b+d)}{b(b+d)} = \frac{b}{b(a+c) - a(b+d)} = \frac{b}{bc - ad}$$

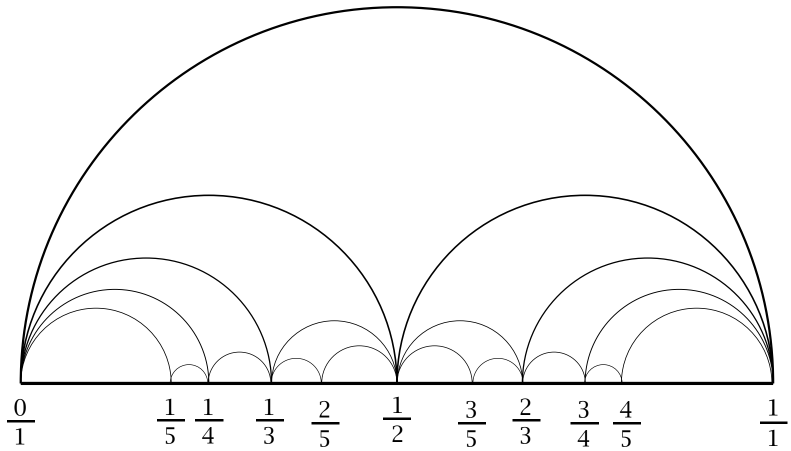
and

$$\frac{1/d - 1/(b+d)}{c/d - (a+c)/(b+d)} \cdot \frac{d(b+d)}{d(b+d)} = \frac{b+d-d}{c(b+d) - d(a+c)} = \frac{b}{bc - ad}$$

so they are equal. The same argument works for the other diagonal, just by interchanging $\frac{a}{b}$ and $\frac{c}{d}$.

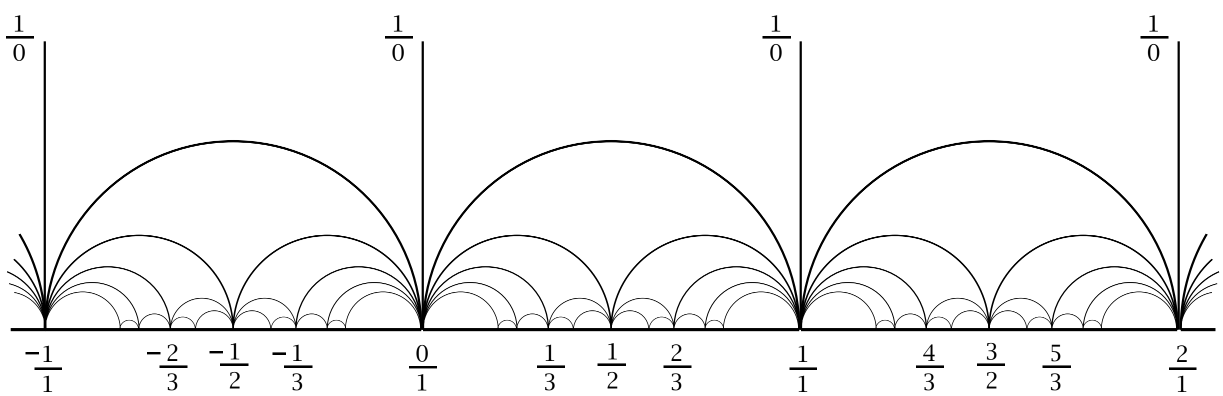
Going back to the square diagram, this fact that we have just shown implies that the successive Farey series can be obtained by taking the vertices that lie above the line $y = \frac{1}{2}$, then the vertices above $y = \frac{1}{3}$, then above $y = \frac{1}{4}$, and so on. Here we are assuming the two properties of the Farey diagram that will be shown in the next section, that all rational numbers occur eventually as labels on vertices, and that these labels are always fractions in lowest terms.

In the square diagram, the most important thing for our purposes is the triangles, not the vertical lines. We can get rid of all the vertical lines by shrinking each one to its lower endpoint, converting each triangle into a curvilinear triangle with semicircles as edges, as shown in the diagram below.



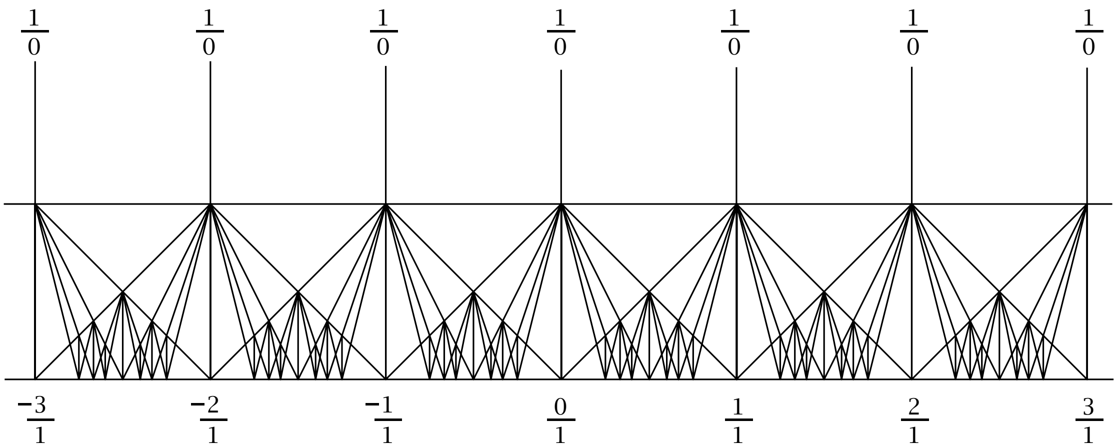
This looks more like a portion of the Farey diagram we started with at the beginning of the chapter, but with the outer boundary circle straightened into a line. The advantage of the new version is that the labels on the vertices are exactly in their correct places along the x-axis, so the vertex labeled $\frac{a}{b}$ is exactly at the point $\frac{a}{b}$ on the x-axis.

This diagram can be enlarged so as to include similar diagrams for fractions between all pairs of adjacent integers, not just 0 and 1, all along the x-axis:

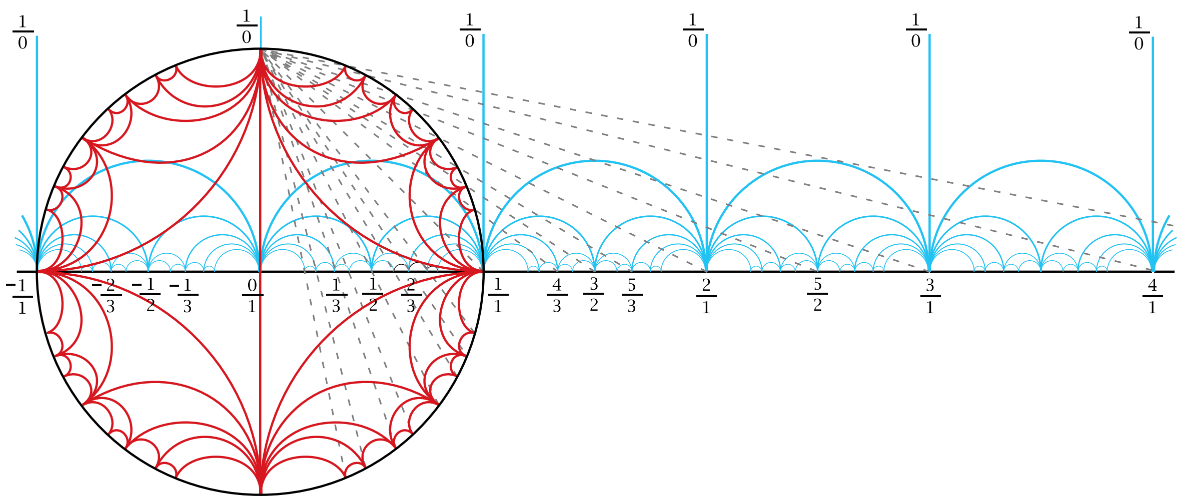


We can also put in vertical lines at the integer points, extending upward to infinity. These correspond to the edges having one endpoint at the vertex $1/0$ in the original Farey diagram.

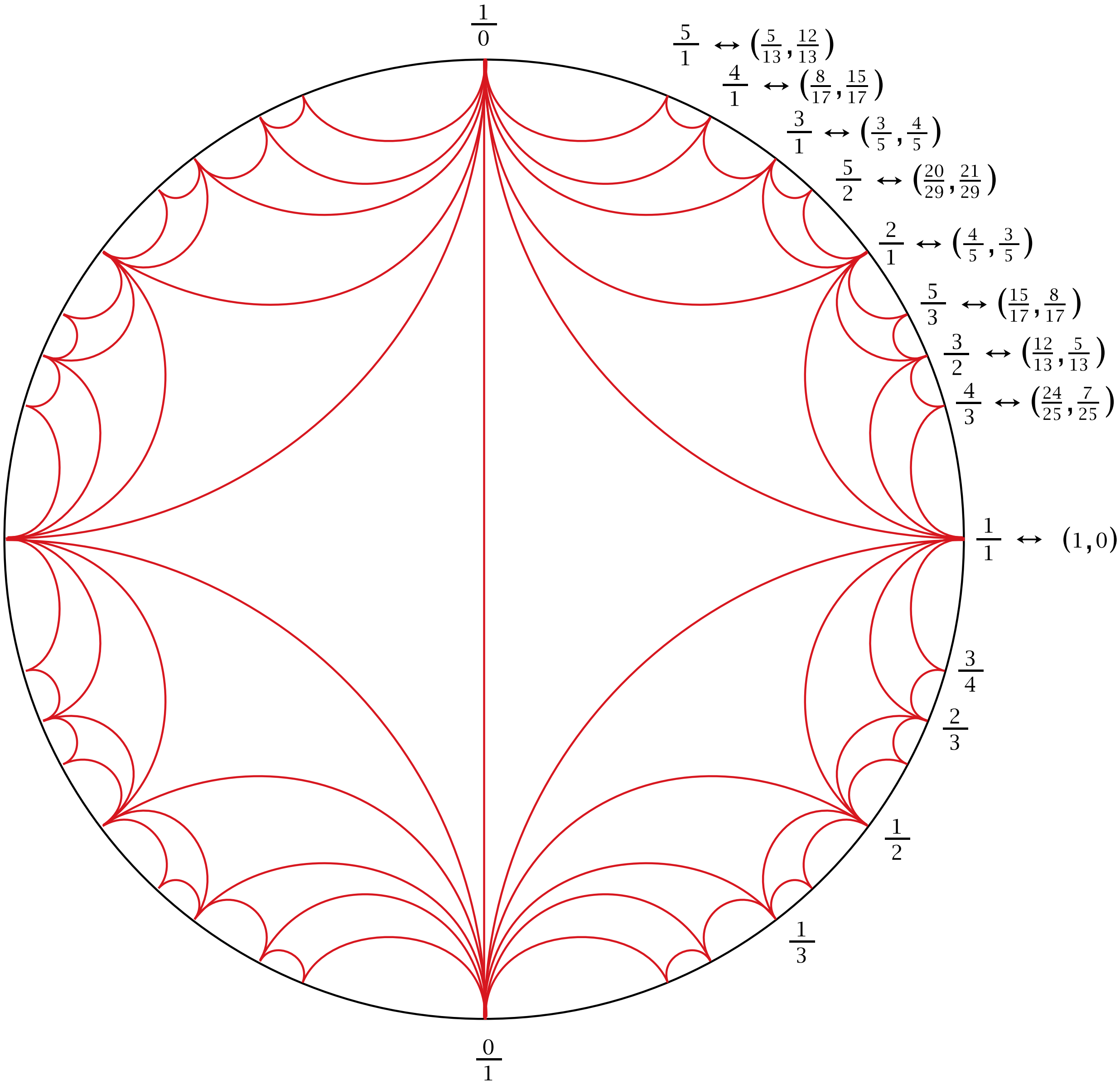
All these diagrams are variants of the Farey diagram we started with at the beginning of the chapter. Let us call the diagram we have just drawn the *standard Farey diagram* and the one at the beginning of the chapter the *circular Farey diagram*. We could also form a variant of the Farey diagram from copies of the square:



Next we describe a variant of the circular Farey diagram that is closely related to Pythagorean triples. Recall from Chapter 0 that rational points (x, y) on the unit circle correspond to rational points p/q on the x-axis by means of lines through the point $(0,1)$ on the circle. In formulas, $(x,y) = (\frac{2pq}{p^2+q^2}, \frac{p^2-q^2}{p^2+q^2})$. Using this correspondence, we can label the rational points on the circle by the corresponding rational points on the x-axis and then construct a new Farey diagram in the circle by filling in triangles by the mediant rule just as before.



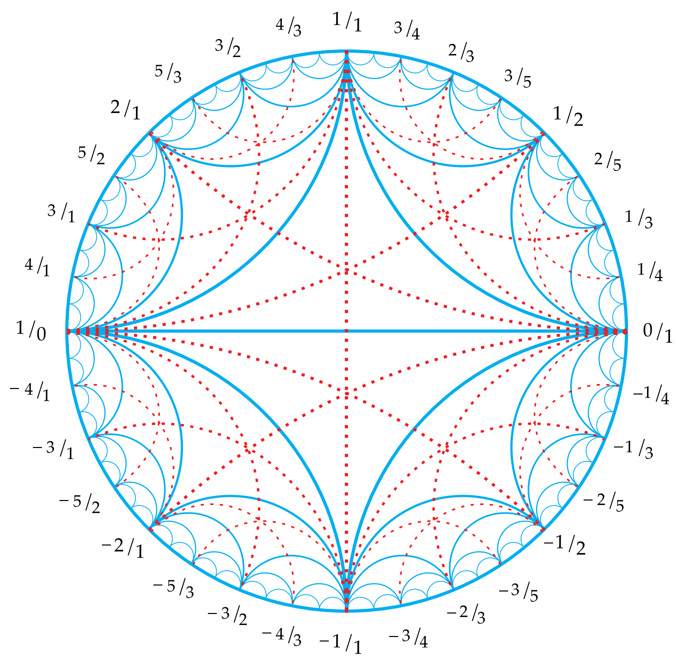
The result is a version of the circular Farey diagram that is rotated by 90 degrees to put $1/0$ at the top of the circle, and there are also some perturbations of the positions of the other vertices and the shapes of the triangles. The next figure shows an enlargement of the new part of the diagram, with the vertices labeled by both the fraction p/q and the coordinates (x, y) of the vertex:



The construction we have described for the Farey diagram involves an inductive process, where more and more triangles are added in succession. With a construction like this it is not easy to tell by a simple calculation whether or not two given rational numbers a/b and c/d are joined by an edge in the diagram. Fortunately there is such a criterion:

Two rational numbers a/b and c/d are joined by an edge in the Farey diagram exactly when the determinant $ad - bc$ of the matrix $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$ is ± 1. This applies also when one of a/b or c/d is $\pm 1/0$.

We will prove this in the next section. What it means in terms of the standard Farey diagram is that if one were to start with the upper half of the xy-plane and insert vertical lines through all the integer points on the x-axis, and then insert semicircles perpendicular to the x-axis joining each pair of rational points a/b and c/d such that $ad - bc = \pm 1$, then no two of these vertical lines or semicircles would cross, and they would divide the upper half of the plane into non-overlapping triangles. This is really quite remarkable when you think about it, and it does not happen for other values of the determinant besides ± 1. For example, for determinant ± 2 the edges would be the dotted lines in the figure below. Here there are three lines crossing in each triangle of the original Farey diagram, and these lines divide each triangle of the Farey diagram into six smaller triangles.



1.2 Continued Fractions

Here are two typical examples of continued fractions:

$$\frac{7}{16} = \cfrac{1}{2 + \cfrac{1}{3 + \cfrac{1}{2}}} \qquad \frac{67}{24} = 2 + \cfrac{1}{1 + \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{4}}}}$$

To compute the value of a continued fraction one starts in the lower right corner and works one's way upward. For example in the continued fraction for $\frac{7}{16}$ one starts with $3 + \frac{1}{2} = \frac{7}{2}$, then taking 1 over this gives $\frac{2}{7}$, and adding the 2 to this gives $\frac{16}{7}$, and finally 1 over this gives $\frac{7}{16}$.

Here is the general form of a continued fraction:

$$\frac{p}{q} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n}}}$$

To write this in more compact form on a single line one can write it as

$$\frac{p}{q} = a_0 + 1\big/a_1 + 1\big/a_2 + \cdots + 1\big/a_n$$

For example:

$$\frac{7}{16} = 1\big/2 + 1\big/3 + 1\big/2 \qquad \frac{67}{24} = 2 + 1\big/1 + 1\big/3 + 1\big/1 + 1\big/4$$

To compute the continued fraction for a given rational number one starts in the upper left corner and works one's way downward, as the following example shows:

$$\frac{67}{24} = 2 + \frac{19}{24} = 2 + \frac{1}{24/19} = 2 + \frac{1}{1 + 5/19} = 2 + \cfrac{1}{1 + \cfrac{1}{19/5}}$$

$$= 2 + \cfrac{1}{1 + \cfrac{1}{3 + 4/5}} = 2 + \cfrac{1}{1 + \cfrac{1}{3 + \cfrac{1}{5/4}}} = 2 + \cfrac{1}{1 + \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{4}}}}$$

If one is good at mental arithmetic and the numbers aren't too large, only the final form of the answer needs to be written down: $\frac{67}{24} = 2 + 1\big/1 + 1\big/3 + 1\big/1 + 1\big/4$.

This process is known as the *Euclidean Algorithm*. It consists of repeated division, at each stage dividing the previous remainder into the previous divisor. The procedure for $67/24$ is shown at the right. Note that the numbers in the shaded box are the numbers a_i in the continued fraction. These are the quotients of the successive divisions. They are sometimes called the *partial quotients* of the original fraction.

$$\begin{array}{rcl}
67 & = & 2 \cdot 24 + 19 \\
24 & = & 1 \cdot 19 + 5 \\
19 & = & 3 \cdot 5 + 4 \\
5 & = & 1 \cdot 4 + 1 \\
4 & = & 4 \cdot 1 + 0
\end{array}$$

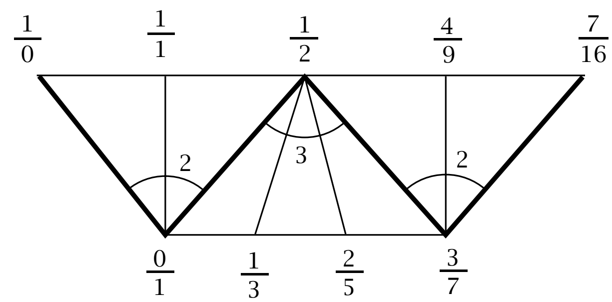
One of the classical uses for the Euclidean algorithm is to find the greatest common divisor of two given numbers. If one applies the algorithm to two numbers p and q, dividing the smaller into the larger, then the remainder into the first divisor, and so on, then the greatest common divisor of p and q turns out to be the last nonzero remainder. For example, starting with $p = 72$ and $q = 201$ the calculation is shown at the right, and the last nonzero remainder is 3, which is the greatest common divisor of 72 and 201. (In fact the fraction $201/72$ equals $67/24$, which explains why the successive quotients for this example are the same as in the preceding example.) It is easy to see from the displayed equations why 3 has to be the greatest common divisor of 72 and 201, since from the first equation it follows that any divisor of 72 and 201 must also divide 57, then the second equation shows it must divide 15, the third equation then shows it must divide 12, and the fourth equation shows it must divide 3, the last nonzero remainder. Conversely, if a number divides the last nonzero remainder 3, then the last equation shows it must also divide the 12, and the next-to-last equation then shows it must divide 15, and so on until we conclude that it divides all the numbers not in the shaded rectangle, including the original two numbers 72 and 201. The same reasoning applies in general.

$$\begin{array}{rcl}
201 & = & 2 \cdot 72 + 57 \\
72 & = & 1 \cdot 57 + 15 \\
57 & = & 3 \cdot 15 + 12 \\
15 & = & 1 \cdot 12 + 3 \\
12 & = & 4 \cdot 3 + 0
\end{array}$$

A more obvious way to try to compute the greatest common divisor of two numbers would be to factor each of them into a product of primes, then look to see which primes occurred as factors of both, and to what power. But to factor a large number into its prime factors is a very laborious and time-consuming process. For example, even a large computer would have a hard time factoring a number of a hundred digits into primes, so it would not be feasible to find the greatest common divisor of a pair of hundred-digit numbers this way. However, the computer would have no trouble at all applying the Euclidean algorithm to find their greatest common divisor.

Having seen what continued fractions are, let us now see what they have to do with the Farey diagram. Some examples will illustrate this best, so let us first look at the continued fraction for $7/16$ again. This has $2, 3, 2$ as its sequence of partial quotients. We use these three numbers to build a strip of three large triangles subdivided into 2, 3, and 2 smaller triangles, from left to right:

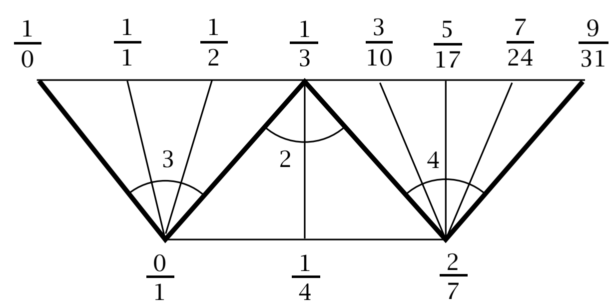
$$\frac{7}{16} = \cfrac{1}{2 + \cfrac{1}{3 + \cfrac{1}{2}}}$$



We can think of the diagram as being formed from three “fans", where the first fan is made from the first 2 small triangles, the second fan from the next 3 small triangles, and the third fan from the last 2 small triangles. Now we begin labeling the vertices of this strip. On the left edge we start with the labels $1/0$ and $0/1$. Then we use the mediant rule for computing the third label of each triangle in succession as we move from left to right in the strip. Thus we insert, in order, the labels $1/1$, $1/2$, $1/3$, $2/5$, $3/7$, $4/9$, and finally $7/16$.

Was it just an accident that the final label was the fraction $7/16$ that we started with, or does this always happen? Doing more examples should help us decide. Here is a second example:

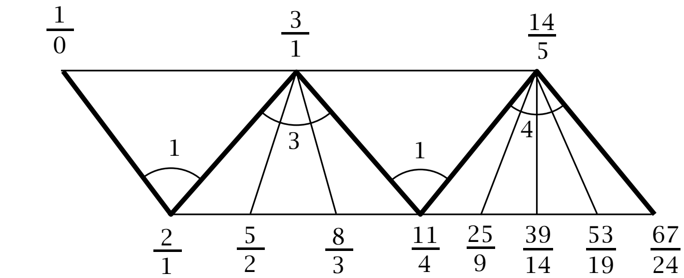
$$\frac{9}{31} = \cfrac{1}{3 + \cfrac{1}{2 + \cfrac{1}{4}}}$$



Again the final vertex on the right has the same label as the fraction we started with. The reader is encouraged to try more examples to make sure we are not rigging things to get a favorable outcome by only choosing examples that work.

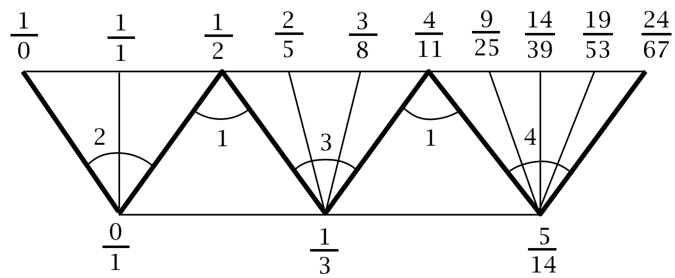
In fact this always works for fractions p/q between 0 and 1. For fractions larger than 1 the procedure works if we modify it by replacing the label $0/1$ with the initial integer $a_0/1$ in the continued fraction $a_0 + {}^1\!/_{a_1} + {}^1\!/_{a_2} + \cdots + {}^1\!/_{a_n}$. This is illustrated by the $67/24$ example:

$$\frac{67}{24} = 2 + \cfrac{1}{1 + \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{4}}}}$$



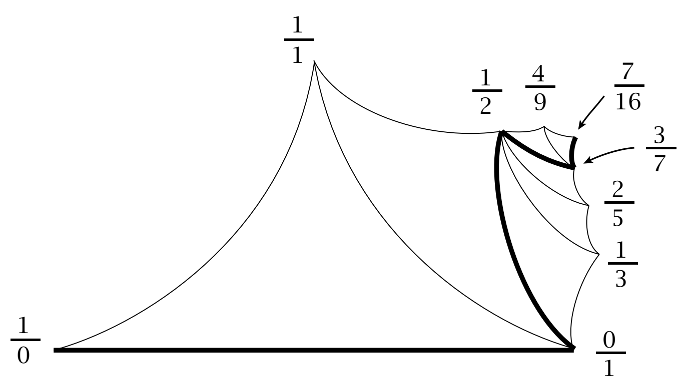
For comparison, here is the corresponding strip for the reciprocal, $24/67$:

$$\frac{24}{67} = \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{4}}}}}$$

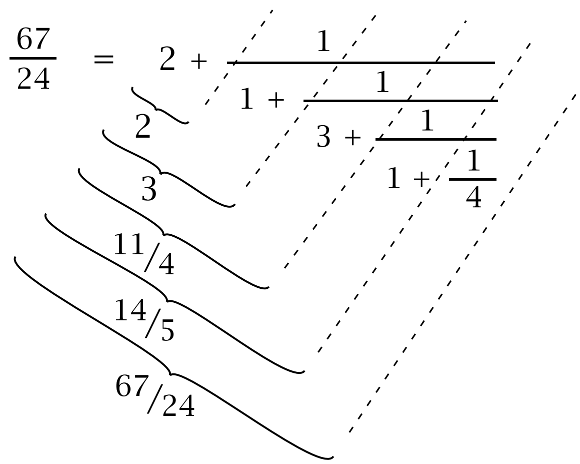


Now let us see how all this relates to the Farey diagram. Since the rule for labeling vertices in the triangles along the horizontal strip for a fraction p/q is the mediant rule, each of the triangles in the strip is a triangle in the Farey diagram, somewhat distorted in shape, and the strip of triangles can be regarded as a sequence of adjacent triangles in the diagram. Here is what this looks like for the fraction $7/16$ in the circular Farey diagram, slightly distorted for the sake of visual clarity:

$$\frac{7}{16} = \cfrac{1}{2 + \cfrac{1}{3 + \cfrac{1}{2}}}$$



In the strip of triangles for a fraction p/q there is a zigzag path from $1/0$ to p/q that we have indicated by the heavily shaded edges. The vertices that this zigzag path passes through have a special significance. They are the fractions that occur as the values of successively larger initial portions of the continued fraction, as illustrated in the following example:

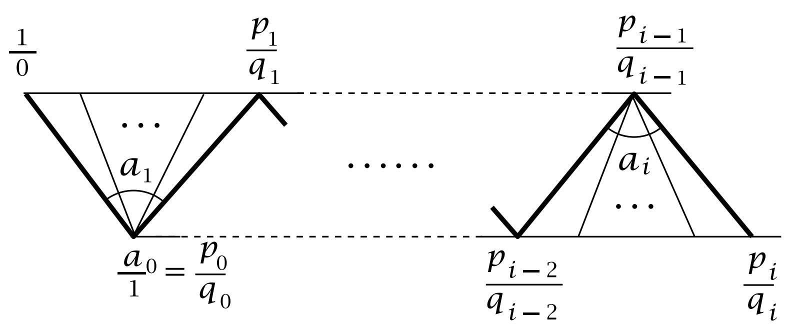


These fractions are called the *convergents* for the given fraction. Thus the convergents for $67/24$ are 2, 3, $11/4$, $14/5$, and $67/24$ itself.

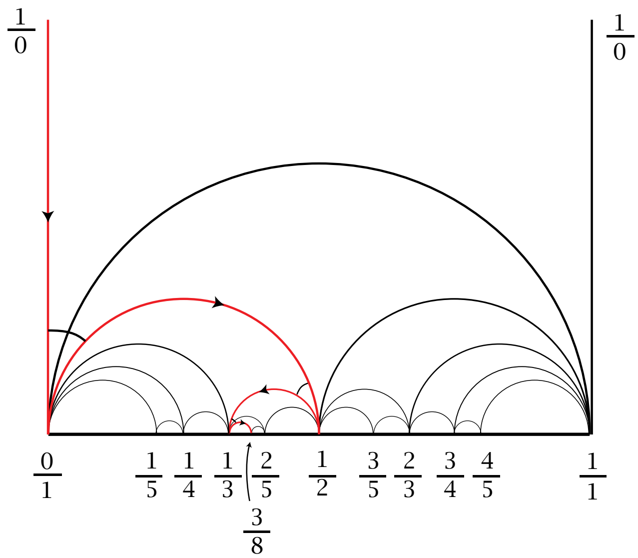
From the preceding examples one can see that each successive vertex label p_i/q_i along the zigzag path for a continued fraction $\frac{p}{q} = a_0 + 1/a_1 + 1/a_2 + \cdots + 1/a_n$ is computed in terms of the two preceding vertex labels according to the rule

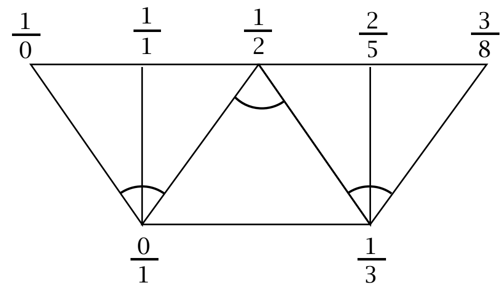
$$\frac{p_i}{q_i} = \frac{a_i p_{i-1} + p_{i-2}}{a_i q_{i-1} + q_{i-2}}$$

This is because the mediant rule is being applied a_i times, 'adding' p_{i-1}/q_{i-1} to the previously obtained fraction each time until the next label p_i/q_i is obtained.



It is interesting to see what the zigzag paths corresponding to continued fractions look like in the standard Farey diagram. The next figure shows the simple example of the continued fraction for $3/8$. We can see here that the five triangles of the strip correspond to the four curvilinear triangles lying directly above $3/8$ in the Farey diagram, plus the fifth 'triangle' extending upward to infinity, bounded on the left and right by the vertical lines above $0/1$ and $1/1$, and bounded below by the semicircle from $0/1$ to $1/1$.





This example is typical of the general case, where the zigzag path for a continued fraction $\frac{p}{q} = a_0 + 1/a_1 + 1/a_2 + \cdots + 1/a_n$ becomes a 'pinball path' in the standard Farey diagam, starting down the vertical line from $1/0$ to $a_0/1$, then turning left across a_1 triangles, then right across a_2 triangles, then left across a_3 triangles, continuing to alternate left and right turns until reaching the final vertex p/q. Two consequences of this are:

(1) The convergents are alternately smaller than and greater than p/q.

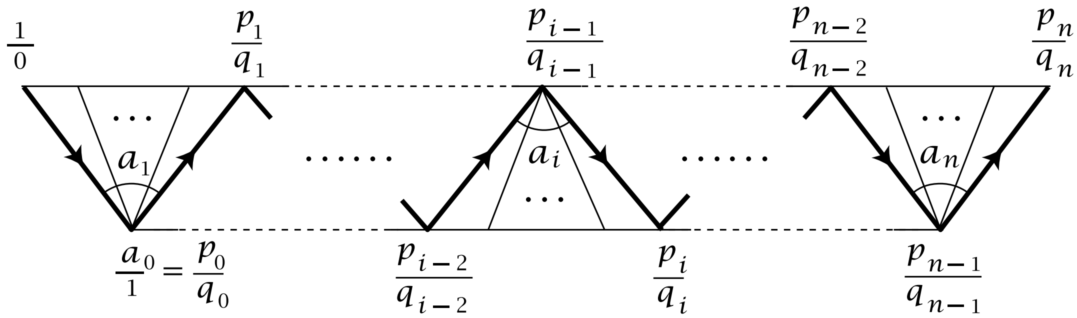
(2) The triangles that form the strip of triangles for p/q are exactly the triangles in the Farey diagram that lie directly above the point p/q on the x-axis.

Here is a general statement describing the relationship between continued fractions and the Farey diagram that we have observed in all our examples so far:

Theorem. *The convergents for the continued fraction* $\frac{p}{q} = a_0 + 1/a_1 + 1/a_2 + \cdots + 1/a_n$ *are the vertices along a zigzag path consisting of a finite sequence of edges in the Farey diagram, starting at* $1/0$ *and ending at* p/q*. The path starts along the edge from* $1/0$ *to* $a_0/1$*, then turns left across a fan of* a_1 *triangles, then right across a fan of* a_2 *triangles, etc., finally ending at* p/q*.*

In particular, since every positive rational number has a continued fraction expansion, we see that every positive rational number occurs eventually as the label of some vertex in the positive half of the diagram. All negative rational numbers then occur as labels in the negative half.

Proof of the Theorem: The continued fraction $\frac{p}{q} = a_0 + 1/a_1 + 1/a_2 + \cdots + 1/a_n$ determines a strip of triangles:



We will show that the label p_n/q_n on the final vertex in this strip is equal to p/q, the value of the continued fraction. Replacing n by i, we conclude that this holds also for each initial seqment $a_0 + 1/a_1 + 1/a_2 + \cdots + 1/a_i$ of the continued fraction. This is just saying that the vertices p_i/q_i along the strip are the convergents to p/q, which is what the theorem claims.

To prove that $p_n/q_n = p/q$ we will use 2×2 matrices. Consider the product

$$P = \begin{pmatrix} 1 & a_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix}$$

We can multiply this product out starting either from the left or from the right. Suppose first that we multiply starting at the left. The initial matrix is $\begin{pmatrix} 1 & a_0 \\ 0 & 1 \end{pmatrix}$ and we can view the two columns of this matrix as the two fractions $1/0$ and $a_0/1$ labeling the left edge of the strip of triangles. When we multiply this matrix by the next matrix we get

$$\begin{pmatrix} 1 & a_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} = \begin{pmatrix} a_0 & 1 + a_0 a_1 \\ 1 & a_1 \end{pmatrix} = \begin{pmatrix} p_0 & p_1 \\ q_0 & q_1 \end{pmatrix}$$

The two columns here give the fractions at the ends of the second edge of the zigzag path. The same thing happens for subsequent matrix multiplications, as multiplying by the next matrix in the product takes the matrix corresponding to one edge of the zigzag path to the matrix corresponding to the next edge:

$$\begin{pmatrix} p_{i-2} & p_{i-1} \\ q_{i-2} & q_{i-1} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_i \end{pmatrix} = \begin{pmatrix} p_{i-1} & p_{i-2} + a_i p_{i-1} \\ q_{i-1} & q_{i-2} + a_i q_{i-1} \end{pmatrix} = \begin{pmatrix} p_{i-1} & p_i \\ q_{i-1} & q_i \end{pmatrix}$$

In the end, when all the matrices have been multiplied, we obtain the matrix corresponding to the last edge in the strip from p_{n-1}/q_{n-1} to p_n/q_n. Thus the second column of the product P is p_n/q_n, and what remains is to show that this equals the value p/q of the continued fraction $a_0 + 1/a_1 + 1/a_2 + \cdots + 1/a_n$.

The value of the continued fraction $a_0 + 1/\!\!\nearrow_{a_1} + 1/\!\!\nearrow_{a_2} + \cdots + 1/\!\!\nearrow_{a_n}$ is computed by working from right to left. If we let r_i/s_i be the value of the tail $1/\!\!\nearrow_{a_i} + 1/\!\!\nearrow_{a_{i+1}} + \cdots + 1/\!\!\nearrow_{a_n}$ of the continued fraction, then $r_n/s_n = 1/a_n$ and we have

$$\frac{r_i}{s_i} = \frac{1}{a_i + \dfrac{r_{i+1}}{s_{i+1}}} = \frac{s_{i+1}}{a_i s_{i+1} + r_{i+1}} \qquad \text{and finally} \qquad \frac{p}{q} = a_0 + \frac{r_1}{s_1} = \frac{a_0 s_1 + r_1}{s_1}$$

In terms of matrices this implies that we have

$$\begin{pmatrix} r_n \\ s_n \end{pmatrix} = \begin{pmatrix} 1 \\ a_n \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 \\ 1 & a_i \end{pmatrix}\begin{pmatrix} r_{i+1} \\ s_{i+1} \end{pmatrix} = \begin{pmatrix} s_{i+1} \\ r_{i+1} + a_i s_{i+1} \end{pmatrix} = \begin{pmatrix} r_i \\ s_i \end{pmatrix}$$

$$\text{and} \quad \begin{pmatrix} 1 & a_0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} r_1 \\ s_1 \end{pmatrix} = \begin{pmatrix} r_1 + a_0 s_1 \\ s_1 \end{pmatrix} = \begin{pmatrix} p \\ q \end{pmatrix}$$

This means that when we multiply out the product P starting from the right, then the second columns will be successively $\begin{pmatrix} r_n \\ s_n \end{pmatrix}, \begin{pmatrix} r_{n-1} \\ s_{n-1} \end{pmatrix}, \cdots, \begin{pmatrix} r_1 \\ s_1 \end{pmatrix}$ and finally $\begin{pmatrix} p \\ q \end{pmatrix}$. We already showed this second column is $\begin{pmatrix} p_n \\ q_n \end{pmatrix}$, so $p/q = p_n/q_n$ and the proof is complete. ◻

An interesting fact that can be deduced from the preceding proof is that for a continued fraction $1/\!\!\nearrow_{a_1} + 1/\!\!\nearrow_{a_2} + \cdots + 1/\!\!\nearrow_{a_n}$ with no initial integer a_0, if we reverse the order of the numbers a_i, this leaves the denominator unchanged. For example

$$1/\!\!\nearrow_2 + 1/\!\!\nearrow_3 + 1/\!\!\nearrow_4 = \frac{13}{30} \qquad \text{and} \qquad 1/\!\!\nearrow_4 + 1/\!\!\nearrow_3 + 1/\!\!\nearrow_2 = \frac{7}{30}$$

To see why this must always be true we use the operation of transposing a matrix to interchange its rows and columns. For a 2×2 matrix this just amounts to interchanging the upper-right and lower-left entries:

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}^T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Transposing a product of matrices reverses the order of the factors: $(AB)^T = B^T A^T$, as can be checked by direct calculation. In the product

$$\begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & a_2 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix} = \begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix}$$

the individual matrices on the left side of the equation are symmetric with respect to transposition, so the transpose of the product is obtained by just reversing the order of the factors:

$$\begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & a_{n-1} \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ 1 & a_1 \end{pmatrix} = \begin{pmatrix} p_{n-1} & q_{n-1} \\ p_n & q_n \end{pmatrix}$$

Thus the denominator q_n is unchanged, as claimed.

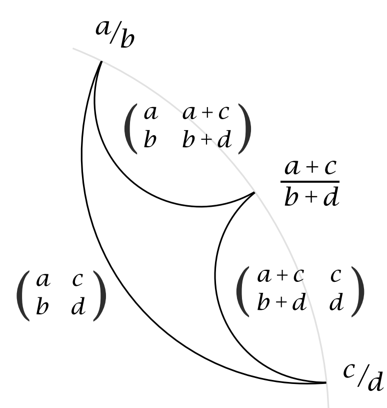
There is also a fairly simple relationship between the numerators. In the example of $13/30$ and $7/30$ we see that the product of the numerators, 91, is congruent to 1 modulo the denominator. In the general case the product of the numerators is $p_n q_{n-1}$ and this is congruent to $(-1)^{n+1}$ modulo the denominator q_n. To verify this, we note that the determinant of each factor $\begin{pmatrix} 0 & 1 \\ 1 & a_i \end{pmatrix}$ is -1 so since the determinant of a product is the product of the determinants, we have $p_{n-1}q_n - p_n q_{n-1} = (-1)^n$, which says that $p_n q_{n-1}$ is congruent to $(-1)^{n+1}$ modulo q_n.

Determinants Determine Edges

We constructed the Farey diagram by an inductive procedure, inserting successive edges according to the mediant rule, but there is another rule that can be used to characterize the edges in the diagram:

Theorem. *In the Farey diagram, two vertices labeled a/b and c/d are joined by an edge if and only if the determinant $ad-bc$ of the matrix $\left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right)$ is equal to ± 1.*

Proof: First we show that for an arbitrary edge in the diagram joining a/b to c/d, the associated matrix $\left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right)$ has determinant ± 1. This is obviously true for the edges in the two largest triangles in the circular version of the diagram. For the smaller triangles we proceed by induction. The figure at the right shows the three matrices corresponding to the edges of one of these smaller triangles. By induction we assume we know that $ad-bc = \pm 1$ for the long edge of the triangle. Then the determinant condition holds also for the two shorter edges of the triangle since $a(b+d) - b(a+c) = ad-bc$ and $(a+c)d - (b+d)c = ad-bc$.



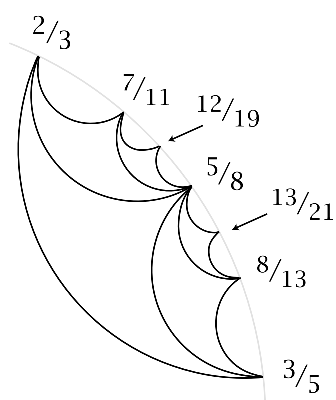
Before proving the converse let us pause to apply what we have shown so far to deduce a basic fact about the Farey diagram that was mentioned but not proved when we first constructed the diagram:

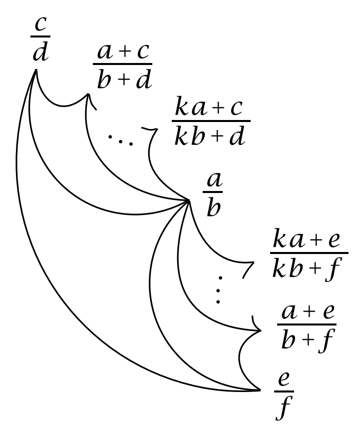
Corollary. *The mediant rule for labeling the vertices in the Farey diagram always produces labels a/b that are fractions in lowest terms.*

Proof: Consider an edge joining a vertex labeled a/b to some other vertex labeled c/d. By the preceding proposition we know that $ad-bc = \pm 1$. This equation implies that a and b can have no common divisor greater than 1 since any common divisor of

a and b must divide the products ad and bc, hence also the difference $ad-bc = \pm 1$, but the only divisors of ± 1 are ± 1. $\square$

Now we return to proving the converse half of the theorem, which says that there is an edge joining a/b to c/d whenever $ad-bc = \pm 1$. To do this we will examine how all the edges emanating from a fixed vertex a/b are related. To begin, if $a/b = 0/1$ then the matrices $\begin{pmatrix} 0 & c \\ 1 & d \end{pmatrix}$ with determinant ± 1 are the matrices $\begin{pmatrix} 0 & \pm 1 \\ 1 & d \end{pmatrix}$, and these correspond exactly to the edges in the diagram from $0/1$ to $\pm 1/d$. There is a similar exact correspondence for the edges from $1/0$. For the other vertices a/b, the example $a/b = 5/8$ is shown in the left half of the figure below. The first edges drawn to this vertex come from $2/3$ and $3/5$, and after this all the other edges from $5/8$ are drawn in turn. As one can see, they are all obtained by adding $(5,8)$ to $(2,3)$ or $(3,5)$ repeatedly. If we choose any one of these edges from $5/8$, say the edge to $2/3$ for example, then the edges from $5/8$ have their other endpoints at the fractions $(2+5k)/(3+8k)$ as k ranges over all integers, with positive values of k giving the edges on the upper side of the edge to $2/3$ and negative values of k giving the edges on the lower side of the edge to $2/3$.





The same thing happens for an arbitrary value of a/b as shown in the right half of the figure, where a/b initially arises as the mediant of c/d and e/f. In this case if we choose the edge to c/d as the starting edge, then the other edges go from a/b to $(c+ka)/(d+kb)$. In particular, when $k=-1$ we get the edge to $(c-a)/(d-b) = (a-c)/(b-d) = e/f$.

To finish the argument we need to know how the various matrices $\left(\begin{smallmatrix} a & x \\ b & y \end{smallmatrix}\right)$ of determinant $ay-bx = \pm 1$ having the same first column are related. This can be deduced from the following result about integer solutions of linear equations with integer coefficients:

Lemma. *Suppose a and b are integers with no common divisor. If one solution of $ay - bx = n$ is $(x, y) = (c, d)$, then the general solution is $(x, y) = (c + ka, d + kb)$ for k an arbitrary integer.*

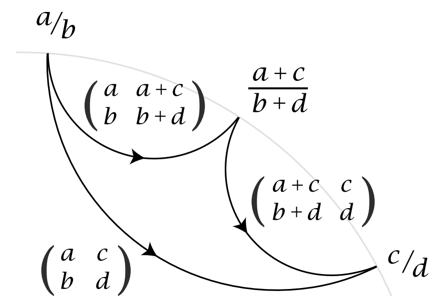
The proof will use the same basic argument as is used in linear algebra to show that the general solution of a system of nonhomogeneous linear equations is obtained from any particular solution by adding the general solution of the associated system of homogeneous equations.

Proof: One solution $(x, y) = (c, d)$ of $ay - bx = n$ is given. For an arbitrary solution (x, y) we look at the difference $(x_0, y_0) = (x - c, y - d)$. This satisfies $ay_0 - bx_0 = 0$, or in other words, $ay_0 = bx_0$. Since a and b have no common divisors, the equation $ay_0 = bx_0$ implies that x_0 must be a multiple of a and y_0 must be a multiple of b, in fact the same multiple in both cases so that the equation becomes $a(kb) = b(ka)$. Thus we have $(x_0, y_0) = (ka, kb)$ for some integer k. Thus every solution of $ay - bx = n$ has the form $(x, y) = (c + x_0, d + y_0) = (c + ka, d + kb)$, and it is clear that these formulas for x and y give solutions for all values of k. $\square$

Now we can easily finish the proof of the theorem. The lemma in the cases $n = \pm 1$ implies that the edges in the Farey diagram with a/b at one endpoint account for all matrices $\begin{pmatrix} a & x \\ b & y \end{pmatrix}$ of determinant $ay - bx = \pm 1$. $\square$

There is some ambiguity in the correspondence between edges of the Farey diagram and matrices $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$ of determinant ± 1. For one thing, either column of the matrix can be multiplied by -1, changing the sign of the determinant without changing the value of the fractions a/b and c/d. This ambiguity can be eliminated by choosing all of a, b, c, and d to be positive for edges in the upper half of the circular Farey diagram, and choosing just the numerators a and c to be negative for edges in the lower half of the diagram. The only other ambiguity is that both $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$ and $\begin{pmatrix} c & a \\ d & b \end{pmatrix}$ correspond to the same edge. This ambiguity can be eliminated by orienting the edges by placing an arrowhead on each edge pointing from the vertex corresponding to the first column of the matrix to the vertex corresponding to the second column. Changing the orientation of an edge switches the two columns of the matrix, which changes the sign of the determinant.

The identity matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ has determinant $+1$ and corresponds to the edge from $1/0$ to $0/1$ oriented from left to right in the circular diagram. We can use this orientation to give orientations to all other edges when we build the diagram using the mediant rule. In the upper half of the diagram this



makes all edges be oriented toward the right, or in other words from a/b to c/d with $a/b > c/d$. With this orientation, all the corresponding matrices have determinant $+1$ since $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ has determinant $+1$ and we have seen that the determinant doesn't change when we add new edges by the mediant rule. When we use the mediant rule to construct the lower half of the diagram we have to start with $-1/0$ instead of $1/0$. This means that we are starting with the matrix $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ instead of $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Since the determinant of $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ is -1, this means that the edges in the lower half of the diagram, when oriented toward the right as in the upper half, correspond to matrices of determinant -1.

The Diophantine Equation ax+by=n

The Euclidean algorithm and continued fractions can be used to compute all the integer solutions of a linear equation $ax + by = n$ where a, b, and n are given integers. We can assume neither a nor b is zero, otherwise the equation is rather trivial. Changing the signs of x or y if necessary, we can rewrite the equation in the form $ax - by = n$ where a and b are both positive.

If a and b have greatest common divisor $d > 1$, then since d divides a and b it must divide $ax - by$, so d must divide n if the equation is to have any solutions at all. If d does divide n we can divide both sides of the equation by d to get a new equation of the same type as the original one and having the same solutions, but with the new coefficients a and b having no common divisors. For example, the equation $6x - 15y = 21$ reduces in this way to the equation $2x - 5y = 7$. Thus we can assume from now on that a and b have no common divisors.

The Lemma from a page or two back shows how to find the general solution of $ax - by = n$ once we have found one particular solution. To find a particular solution it suffices to do the case $n = 1$ since if we have a solution of $ax - by = 1$, we can multiply x and y by n to get a solution of $ax - by = n$. For small values of a and b a solution of $ax - by = 1$ can be found more or less by inspection since the equation $ax - by = 1$ says that we have multiples of a and b that differ by 1. For example, for the equation $2x - 5y = 1$ the smallest multiples of 2 and 5 that differ by 1 are $2 \cdot 3$ and $5 \cdot 1$, so a solution of $2x - 5y = 1$ is $(x, y) = (3, 1)$. A solution of $2x - 5y = 7$ is then $(x, y) = (21, 7)$. By the earlier Lemma, the general solution of $2x - 5y = 7$ is $(x, y) = (21 + 5k, 7 + 2k)$ for arbitrary integers k. The smallest positive solution is $(6, 1)$, obtained by setting $k = -3$. This means we could also write the general solution as $(6+5k, 1+2k)$.

Solutions of $ax - by = 1$ always exist when a and b have no common divisors, and a way to find one is to find an edge in the Farey diagram with a/b at one end of the edge. This can be done by using the Euclidean algorithm to compute the strip of triangles from $1/0$ to a/b. As an example, let us solve $67x - 24y = 1$. We already computed the strip of triangles for $67/24$ earlier in this section. The vertex preceding $67/24$ in the zigzag path is $14/5$ and this vertex lies above $67/24$ so we have $14/5 > 67/24$ and hence the matrix $\begin{pmatrix} 14 & 67 \\ 5 & 24 \end{pmatrix}$ has determinant $+1$. Thus one solution of $67x - 24y = 1$ is $(x,y) = (-5,-14)$ and the general solution is $(x,y) = (-5+24k, -14+67k)$. We could also use the edge from $53/19$ to $67/24$, so $\begin{pmatrix} 67 & 53 \\ 24 & 19 \end{pmatrix}$ has determinant $+1$, yielding another formula for the general solution $(19+24k, 53+67k)$.

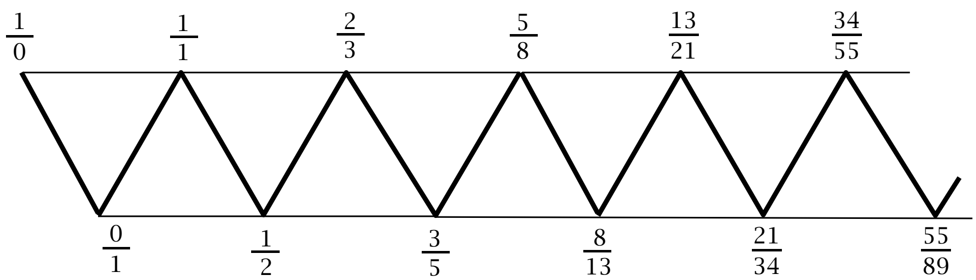
From a geometric point of view, finding the integer solutions of $ax + by = n$ is finding the points on the line $ax + by = n$ in the xy-plane having both coordinates integers. The points in the plane having both coordinates integers form a square grid called the *integer lattice.* Thus we wish to see which points in the integer lattice lie on the line $ax + by = n$. This equation can be written in the form $y = mx + b$ where slope m and y-intercept b are both rational. Conversely, an equation $y = mx + b$ with m and b rational can be written as an equation $ax + by = n$ with a, b, and n integers by multiplying through by a common denominator of m and b. Sometimes the equation $ax + by = n$ has no integer solutions, as we have seen, namely when n is not a multiple of the greatest common divisor of a and b, for example the equation $2x + 2y = 1$. In these cases the line $ax + by = n$ passes through no integer lattice points. In the opposite case that there does exist an integer solution, there are infinitely many, and they correspond to integer lattice points spaced at equal intervals along the line.

Infinite Continued Fractions

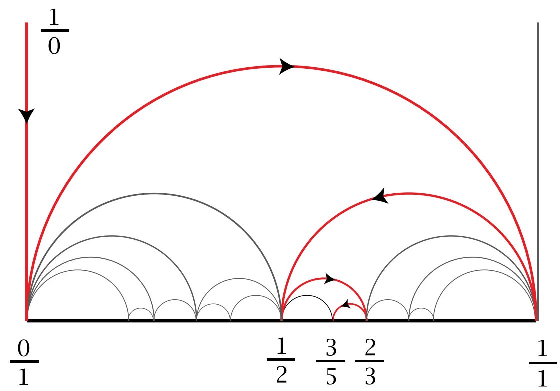
We have seen that all rational numbers can be represented as continued fractions $a_0 + {}^1\!\diagup_{a_1} + {}^1\!\diagup_{a_2} + \cdots + {}^1\!\diagup_{a_n}$, but what about irrational numbers? It turns out that these can be represented as *infinite* continued fractions $a_0 + {}^1\!\diagup_{a_1} + {}^1\!\diagup_{a_2} + {}^1\!\diagup_{a_3} + \cdots$. A simple example is ${}^1\!\diagup_{1} + {}^1\!\diagup_{1} + {}^1\!\diagup_{1} + \cdots$, or in its expanded form:

$$\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\ddots}}}}$$

The corresponding strip of triangles is infinite:



Notice that these fractions after $1/0$ are the successive ratios of the famous Fibonacci sequence $0, 1, 1, 2, 3, 5, 8, 13, 21, \cdots$ where each number is the sum of its two predecessors. The sequence of convergents is thus $0/1, 1/1, 1/2, 2/3, 3/5, 5/8, 8/13, \cdots$, the vertices along the zigzag path. The way this zigzag path looks in the standard Farey diagram is shown in the figure at the right. What happens when we follow this path farther and farther? The path consists of an infinite sequence of semicircles, each one shorter than the preceding one and sharing a common endpoint. The left endpoints of the semicircles form an increasing sequence of numbers which have to be approaching a certain limiting value x. We know x has to be finite since it is certainly less than each of the right-hand endpoints of the semicircles, the convergents $1/1, 2/3, 5/8, \cdots$. Similarly the right endpoints of the semicircles form a decreasing sequence of numbers approaching a limiting value y greater than each of the left-hand endpoints $0/1, 1/2, 3/5, \cdots$. Obviously $x \le y$. Is it pos-



sible that x is not equal to y? If this happened, the infinite sequence of semicircles would be approaching the semicircle from x to y. Above this semicircle there would then be an infinite number of semicircles, all the semicircles in the infinite sequence. Between x and y there would have to be a rational numbers p/q (between any two real numbers there is always a rational number), so above this rational number there would be an infinite number of semicircles, hence an infinite number of triangles in the Farey diagram. But we know that there are only finitely many triangles above any rational number p/q, namely the triangles that appear in the strip for the continued fraction for p/q. This contradiction shows that x has to be equal to y. Thus the sequence of convergents along the edges of the infinite strip of triangles converges to a unique real number x. (This is why the convergents are called convergents.)

This argument works for arbitrary infinite continued fractions, so we have shown the following general result:

Proposition. *For every infinite continued fraction* $a_0 + {}^1\!\!\nearrow\!_{a_1} + {}^1\!\!\nearrow\!_{a_2} + {}^1\!\!\nearrow\!_{a_3} + \cdots$ *the convergents converge to a unique limit.*

This limit is by definition the value of the infinite continued fraction. There is a simple method for computing the value in the example involving Fibonacci numbers. We begin by setting

$$x = {}^1\!\!\nearrow\!_1 + {}^1\!\!\nearrow\!_1 + {}^1\!\!\nearrow\!_1 + \cdots$$

Then if we take the reciprocals of both sides of this equation we get

$$\frac{1}{x} = 1 + {}^1\!\!\nearrow\!_1 + {}^1\!\!\nearrow\!_1 + {}^1\!\!\nearrow\!_1 + \cdots$$

The right side of this equation is just $1 + x$, so we can easily solve for x:

$$\begin{aligned} \frac{1}{x} &= 1 + x \\ 1 &= x + x^2 \\ x^2 + x - 1 &= 0 \\ x &= \frac{-1 \pm \sqrt{5}}{2} \end{aligned}$$

We know x is positive, so this rules out the negative root and we are left with the final value $x = (-1 + \sqrt{5})/2$. This number, approximately $.618$, goes by the name of the golden ratio. It has many interesting properties.

Proposition. *Every irrational number has an expression as an infinite continued fraction, and this continued fraction is unique.*

Proof: In the Farey diagram consider the vertical line L going upward from a given irrational number x on the x-axis. The lower endpoint of L is not a vertex of the Farey diagram since x is irrational. Thus as we move downward along L we cross a sequence of triangles, entering each triangle by crossing its upper edge and exiting the triangle by crossing one of its two lower edges. When we exit one triangle we are entering another, the one just below it, so the sequence of triangles and edges we cross must be infinite. The left and right endpoints of the edges in the sequence must be approaching the single point x by the argument we gave in the preceding proposition, so the edges themselves are approaching x. Thus the triangles in the sequence form a single infinite strip consisting of an infinite sequence of fans with their pivot vertices on alternate sides of the strip. The zigzag path along this strip gives a continued fraction for x.

For the uniqueness, we have seen that an infinite continued fraction for x corresponds to a zigzag path in the infinite strip of triangles lying above x. This set of triangles is unique so the strip is unique, and there is only one path in this strip that starts at $1/0$ and then does left and right turns alternately, starting with a left turn. The initial turn must be to the left because the first two convergents are a_0 and $a_0 + \frac{1}{a_1}$, with $a_0 + \frac{1}{a_1} > a_0$ since $a_1 > 0$. After the path traverses the first edge, no subsequent edge of the path can go along the border of the strip since this would entail two successive left turns or two successive right turns. $\square$

The arguments we have just given can be used to prove a fact about the standard Farey diagram that we have been taking more or less for granted. This is the fact that the triangles in the diagram completely cover the upper halfplane. In other words, every point (x, y) with $y > 0$ lies either in the interior of some triangle or on the common edge between two triangles. To see why, consider the vertical line L in the upper halfplane through the given point (x, y). If x is an integer then (x, y) is on one of the vertical edges of the diagram. Thus we can assume x is not an integer and hence L is not one of the vertical edges of the diagram. The line L will then be contained in the strip of triangles corresponding to the continued fraction for x. This is a finite strip if x is rational and an infinite strip if x is irrational. In either case the point (x, y), being in L, will be in one of the triangles of the strip or on an edge separating two triangles in the strip. This proves what we wanted to prove.

To compute the infinite continued fraction $a_0 + {}^1\!/\!_{a_1} + {}^1\!/\!_{a_2} + {}^1\!/\!_{a_3} + \cdots$ for a given irrational number x we can follow the same procedure as for rational numbers, but it doesn't terminate after a finite number of steps. Recall the original example that we did:

$$\frac{67}{24} = 2 + \frac{19}{24} = 2 + \cfrac{1}{24/19} = 2 + \cfrac{1}{1 + 5/19} = 2 + \cfrac{1}{1 + \cfrac{1}{19/5}}$$

$$= 2 + \cfrac{1}{1 + \cfrac{1}{3 + 4/5}} = 2 + \cfrac{1}{1 + \cfrac{1}{3 + \cfrac{1}{5/4}}} = 2 + \cfrac{1}{1 + \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{4}}}}$$

The sequence of steps is the following:

(1) Write $x = a_0 + r_1$ where a_0 is an integer and $0 \le r_1 < 1$

(2) Write $1/r_1 = a_1 + r_2$ where a_1 is an integer and $0 \le r_2 < 1$

(3) Write $1/r_2 = a_2 + r_3$ where a_2 is an integer and $0 \le r_3 < 1$

and so on, repeatedly. Thus one first finds the largest integer $a_0 \le x$, with r_1 the 'remainder', then one inverts r_1 and finds the greatest integer $a_1 \le 1/r_1$, with r_2 the remainder, etc.

Here is how this works for $x = \sqrt{2}$:

(1) $\sqrt{2} = 1 + (\sqrt{2} - 1)$ where $a_0 = 1$ since $\sqrt{2}$ is between 1 and 2. Before going on to step (2) we have to compute $\frac{1}{r_1} = \frac{1}{\sqrt{2}-1}$. Multiplying numerator and denominator by $\sqrt{2} + 1$ gives $\frac{1}{\sqrt{2}-1} = \frac{1}{\sqrt{2}-1} \cdot \frac{\sqrt{2}+1}{\sqrt{2}+1} = \sqrt{2} + 1$. This is the number we use in the next step.

(2) $\sqrt{2} + 1 = 2 + (\sqrt{2} - 1)$ since $\sqrt{2} + 1$ is between 2 and 3.

Notice that something unexpected has happened: The remainder $r_2 = \sqrt{2} - 1$ is exactly the same as the previous remainder r_1. There is then no need to do the calculation of $\frac{1}{r_2} = \frac{1}{\sqrt{2}-1}$ since we know it will have to be $\sqrt{2} + 1$. This means that the next step (3) will be exactly the same as step (2), and the same will be true for all subsequent steps. Hence we get the continued fraction

$$\sqrt{2} = 1 + {}^{1}\!\nearrow\!{}_{2} + {}^{1}\!\nearrow\!{}_{2} + {}^{1}\!\nearrow\!{}_{2} + \cdots$$

We can check this calculation by finding the value of the continued fraction in the same way that we did earlier for ${}^{1}\!\nearrow\!{}_{1} + {}^{1}\!\nearrow\!{}_{1} + {}^{1}\!\nearrow\!{}_{1} + \cdots$. First we set $x = {}^{1}\!\nearrow\!{}_{2} + {}^{1}\!\nearrow\!{}_{2} + {}^{1}\!\nearrow\!{}_{2} + \cdots$. Taking reciprocals gives $1/x = 2 + {}^{1}\!\nearrow\!{}_{2} + {}^{1}\!\nearrow\!{}_{2} + {}^{1}\!\nearrow\!{}_{2} + \cdots = 2 + x$. This leads to the quadratic equation $x^2 + 2x - 1 = 0$, which has roots $x = -1 \pm \sqrt{2}$. Since x is positive we can discard the negative root. Thus we have $-1 + \sqrt{2} = {}^{1}\!\nearrow\!{}_{2} + {}^{1}\!\nearrow\!{}_{2} + {}^{1}\!\nearrow\!{}_{2} + \cdots$. Adding 1 to both sides of this equation gives the formula for $\sqrt{2}$ as a continued fraction.

We can get good rational approximations to $\sqrt{2}$ by computing the convergents in its continued fraction $1 + {}^1\!/\!_2 + {}^1\!/\!_2 + {}^1\!/\!_2 + \cdots$. It's a little easier to compute the convergents in $2 + {}^1\!/\!_2 + {}^1\!/\!_2 + {}^1\!/\!_2 + \cdots = 1 + \sqrt{2}$ and then subtract 1 from each of these. For $2 + {}^1\!/\!_2 + {}^1\!/\!_2 + {}^1\!/\!_2 + \cdots$ there is a nice pattern to the convergents:

$$\frac{2}{1}, \frac{5}{2}, \frac{12}{5}, \frac{29}{12}, \frac{70}{29}, \frac{169}{70}, \frac{408}{169}, \frac{985}{408}, \cdots$$

Notice that the sequence of numbers $1, 2, 5, 12, 29, 70, 169, \cdots$ is constructed in a way somewhat analogous to the Fibonacci sequence, except that each number is *twice* the preceding number plus the number before that. (It's easy to see why this has to be true, because each convergent is constructed from the previous one by inverting the fraction and adding 2.) After subtracting 1 from each of these fractions we get the convergents to $\sqrt{2}$:

$$\begin{aligned}
\sqrt{2} &= 1.41421356\cdots \\
1/1 &= 1.00000000\cdots \\
3/2 &= 1.50000000\cdots \\
7/5 &= 1.40000000\cdots \\
17/12 &= 1.41666666\cdots \\
41/29 &= 1.41379310\cdots \\
99/70 &= 1.41428571\cdots \\
239/169 &= 1.41420118\cdots \\
577/408 &= 1.41421568\cdots
\end{aligned}$$

We can compute the continued fraction for $\sqrt{3}$ by the same method as for $\sqrt{2}$, but something slightly different happens:

(1) $\sqrt{3} = 1 + (\sqrt{3} - 1)$ since $\sqrt{3}$ is between 1 and 2. Computing $\frac{1}{\sqrt{3}-1}$, we have $\frac{1}{\sqrt{3}-1} = \frac{1}{\sqrt{3}-1} \cdot \frac{\sqrt{3}+1}{\sqrt{3}+1} = \frac{\sqrt{3}+1}{2}$.

(2) $\frac{\sqrt{3}+1}{2} = 1 + (\frac{\sqrt{3}-1}{2})$ since the numerator $\sqrt{3}+1$ of $\frac{\sqrt{3}+1}{2}$ is between 2 and 3. Now we have a remainder $r_2 = \frac{\sqrt{3}-1}{2}$ which is different from the previous remainder $r_1 = \sqrt{3} - 1$, so we have to compute $\frac{1}{r_2} = \frac{2}{\sqrt{3}-1}$, namely $\frac{2}{\sqrt{3}-1} = \frac{2}{\sqrt{3}-1} \cdot \frac{\sqrt{3}+1}{\sqrt{3}+1} = \sqrt{3} + 1$.

(3) $\sqrt{3} + 1 = 2 + (\sqrt{3} - 1)$ since $\sqrt{3} + 1$ is between 2 and 3.

Now this remainder $r_3 = \sqrt{3} - 1$ is the same as r_1, so instead of the same step being repeated infinitely often, as happened for $\sqrt{2}$, the same two steps will repeat infinitely often. This means we get the continued fraction

$$\sqrt{3} = 1 + {}^1\!/\!_1 + {}^1\!/\!_2 + {}^1\!/\!_1 + {}^1\!/\!_2 + {}^1\!/\!_1 + {}^1\!/\!_2 + \cdots$$

Checking this takes a little more work than before. We begin by isolating the part of the continued fraction that repeats periodically, so we set

$$x = {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + \cdots$$

Taking reciprocals, we get

$$\frac{1}{x} = 1 + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + \cdots$$

Subtracting 1 from both sides gives

$$\frac{1}{x} - 1 = {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + \cdots$$

The next step will be to take reciprocals of both sides, so before doing this we rewrite the left side as $\frac{1-x}{x}$. Then taking reciprocals gives

$$\frac{x}{1-x} = 2 + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + \cdots$$

Hence

$$\frac{x}{1-x} - 2 = {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + \cdots = x$$

Now we have the equation $\frac{x}{1-x} - 2 = x$ which can be simplified to the quadratic equation $x^2 + 2x - 2 = 0$, with roots $x = -1 \pm \sqrt{3}$. Again the negative root is discarded, and we get $x = -1 + \sqrt{3}$. Thus $\sqrt{3} = 1 + x = 1 + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + \cdots$.

To simplify the notation we will write a bar over a block of terms in a continued fraction that repeat infinitely often, for example

$$\sqrt{2} = 1 + \overline{{}^{1}\!\diagup_{2}} \qquad \text{and} \qquad \sqrt{3} = 1 + \overline{{}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2}}$$

It is true in general that for every positive integer n that is not a square, the continued fraction for $\sqrt{n}$ has the form $a_0 + \overline{{}^{1}\!\diagup_{a_1} + {}^{1}\!\diagup_{a_2} + \cdots + {}^{1}\!\diagup_{a_k}}$. The length of the period can be large, for example

$$\sqrt{46} = 6 + \overline{{}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{3} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{6} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{3} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{12}}$$

This example illustrates two other curious facts about the continued fraction for an irrational number $\sqrt{n}$:

(i) The last term of the period (12 in the example) is always twice the integer a_0 (the initial 6).

(ii) If the last term of the period is omitted, the preceding terms in the period form a palindrome, reading the same backwards as forwards.

We will see in the next chapter why these two properties have to be true.

It is natural to ask exactly which irrational numbers have continued fractions that are periodic, or at least *eventually* periodic, like for example

$$1/2+1/4+\overline{1/3+1/5+1/7} = 1/2+1/4+1/3+1/5+1/7+1/3+1/5+1/7+1/3+1/5+1/7+\cdots$$

The answer is given by a theorem of Lagrange from around 1766:

Lagrange's Theorem. *The numbers whose continued fractions are eventually periodic are exactly the numbers of the form $a+b\sqrt{n}$ where a and b are rational numbers and n is a positive integer that is not a square.*

These numbers $a+b\sqrt{n}$ are called *quadratic irrationals* because they are roots of quadratic equations with integer coefficients. The easier half of the theorem is the statement that the value of an eventually periodic infinite continued fraction is always a quadratic irrational. This can be proved by showing that the method we used for finding a quadratic equation satisfied by an eventually periodic continued fraction works in general. Rather than following this purely algebraic approach, however, we will develop a more geometric version of the procedure in the next section, so we will wait until then to give the argument that proves this half of Lagrange's Theorem. The more difficult half of the theorem is the assertion that the continued fraction expansion of every quadratic irrational is eventually periodic. It is not at all apparent from the examples of $\sqrt{2}$ and $\sqrt{3}$ why this should be true in general, but in the next chapter we will develop some theory that will make it clear.

What can be said about the continued fraction expansions of irrational numbers that are not quadratic, such as $\sqrt[3]{2}$, π, or e, the base for natural logarithms? It happens that e has a continued fraction whose terms have a very nice pattern, even though they are not periodic or eventually periodic:

$$e = 2+\underbrace{1/1+1/2+1/1}+\underbrace{1/1+1/4+1/1}+\underbrace{1/1+1/6+1/1}+\cdots$$

where the terms are grouped by threes with successive even numbers as middle denominators. Even simpler are the continued fractions for certain numbers built from e that have arithmetic progressions for their denominators:

$$\frac{e-1}{e+1} = 1/2+1/6+1/10+1/14+\cdots$$

$$\frac{e^2-1}{e^2+1} = 1/1+1/3+1/5+1/7+\cdots$$

The last two formulas were found by Lambert in 1770, while the expression for e itself was found by Euler in 1737.

For $\sqrt[3]{2}$ and π, however, the continued fractions have no known pattern. For π the continued fraction begins

$$\pi = 3 + {}^{1}\!\nearrow\!{}_{7} + {}^{1}\!\nearrow\!{}_{15} + {}^{1}\!\nearrow\!{}_{1} + {}^{1}\!\nearrow\!{}_{292} + \cdots$$

Here the first four convergents are 3, $22/7$, $333/106$, and $355/113$. We recognize $22/7$ as the familiar approximation $3\frac{1}{7}$ to π. The convergent $355/113$ is a particularly good approximation to π since its decimal expansion begins 3.14159282 whereas $\pi = 3.1415926535\cdots$. It is no accident that the convergent $355/113$ obtained by truncating the continued fraction just before the 292 term gives a good approximation to π since it is a general fact that a convergent immediately preceding a large term in the continued fraction always gives an especially good approximation, because the next jump in the zigzag path in the Farey diagram will be rather small, and all succeeding jumps will of course be smaller still.

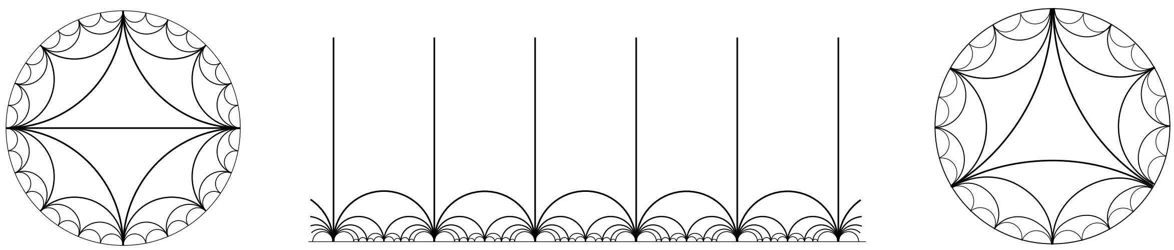
There are nice continued fractions for π if one allows numerators larger than 1, as in the following formula discovered by Euler:

$$\pi = 3 + {}^{1^2}\!\nearrow\!{}_{6} + {}^{3^2}\!\nearrow\!{}_{6} + {}^{5^2}\!\nearrow\!{}_{6} + {}^{7^2}\!\nearrow\!{}_{6} + \cdots$$

However, it is the continued fractions with numerator 1 that have the nicest properties, so we will not consider the more general sort in this book.

1.3 Linear Fractional Transformations

One thing one notices about the various versions the Farey diagram is their symmetry. For the circular Farey diagram the symmetries are the reflections across the horizontal and vertical axes and the 180 degree rotation about the center. For the standard Farey diagram in the upper halfplane there are symmetries that translate the diagram by any integer distance to the left or the right, as well as reflections across certain vertical lines, the vertical lines through an integer or half-integer point on the x-axis. The Farey diagram could also be drawn to have 120 degree rotational symmetry and three reflectional symmetries.



Our purpose in this section is to study all possible symmetries of the Farey diagram, where we interpret the word “symmetry" in a broader sense than the familiar meaning from Euclidean geometry. For our purposes, symmetries will be invertible transformations that take vertices to vertices and edges to edges. (It follows that triangles are sent to triangles.) There are simple algebraic formulas for these more general symmetries, and these formulas lead to effective means of calculation. One of the applications will be to computing the values of periodic or eventually periodic continued fractions.

From linear algebra one is familiar with the way in which 2×2 matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ correspond to linear transformations of the plane $\mathbb{R}^2$, transformations of the form

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix}$$

In our situation we are going to restrict a, b, c, d, x, y to be integers. Then by associating to a pair (x, y) the fraction x/y one obtains a closely related transformation

$$T\left(\frac{x}{y}\right) = \frac{ax+by}{cx+dy} = \frac{a\left(\frac{x}{y}\right)+b}{c\left(\frac{x}{y}\right)+d}$$

If we set $z = x/y$ then T can also be written in the form

$$T(z) = \frac{az+b}{cz+d}$$

Such a transformation is called a *linear fractional transformation* since it is defined by a fraction whose numerator and denominator are linear functions.

In the formula $T(x/y) = (ax+by)/(cx+dy)$ there is no problem with allowing $x/y = 1/0$ just by setting $(x,y) = (1,0)$, and the result is that $T(1/0) = a/c$. The value $T(x/y) = (ax+by)/(cx+dy)$ can also sometimes be $1/0$. This means that T defines a function from vertices of the Farey diagram to vertices of the Farey diagram. We would like T to take edges of the diagram to edges of the diagram, and the following result gives a condition for this to happen.

Proposition. *If the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ has determinant ± 1 then the associated linear fractional transformation T takes each pair of vertices in the Farey diagram that lie at the ends of an edge of the diagram to another such pair of vertices.*

Proof: We showed in the previous section that two vertices labeled p/q and r/s are joined by an edge in the diagram exactly when $ps - qr = \pm 1$, or in other words when the matrix $\left(\begin{smallmatrix} p & r \\ q & s \end{smallmatrix}\right)$ has determinant ± 1. The two columns of the product matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} p & r \\ q & s \end{smallmatrix}\right)$ correspond to the two vertices $T(p/q)$ and $T(r/s)$, by the definition of matrix multiplication:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} p & r \\ q & s \end{pmatrix} = \begin{pmatrix} ap+bq & ar+bs \\ cp+dq & cr+ds \end{pmatrix}$$

The proposition can then be restated as saying that if $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} p & r \\ q & s \end{smallmatrix}\right)$ each have determinant ± 1 then so does their product $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} p & r \\ q & s \end{smallmatrix}\right)$. But it is a general fact about determinants that the determinant of a product is the product of the determinants. (This is easy to prove by a direct calculation in the case of 2×2 matrices.) So the product of two matrices of determinant ± 1 has determinant ± 1. $\square$

As notation, we will use $LF(\mathbb{Z})$ to denote the set of all linear fractional transformations $T(x/y) = (ax+by)/(cx+dy)$ with coefficients a,b,c,d in $\mathbb{Z}$ such that the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ has determinant ± 1.

Changing the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ to its negative $\left(\begin{smallmatrix} -a & -b \\ -c & -d \end{smallmatrix}\right)$ produces the same linear fractional transformation since $(-ax-by)/(-cx-dy) = (ax+by)/(cx+dy)$. This is in fact the only way that different matrices can give the same linear fractional transformation T, as we will see later in this section. Note that changing $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ to its negative $\left(\begin{smallmatrix} -a & -b \\ -c & -d \end{smallmatrix}\right)$ does not change the determinant. Thus each linear fractional transformation in $LF(\mathbb{Z})$ has a well-defined determinant, either $+1$ or -1. Later in this section we will also see how the distinction between determinant $+1$ and determinant -1 has a geometric interpretation in terms of orientations.

A useful fact about $LF(\mathbb{Z})$ is that each transformation T in $LF(\mathbb{Z})$ has an inverse T^{-1} in $LF(\mathbb{Z})$ because the inverse of a 2×2 matrix is given by the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

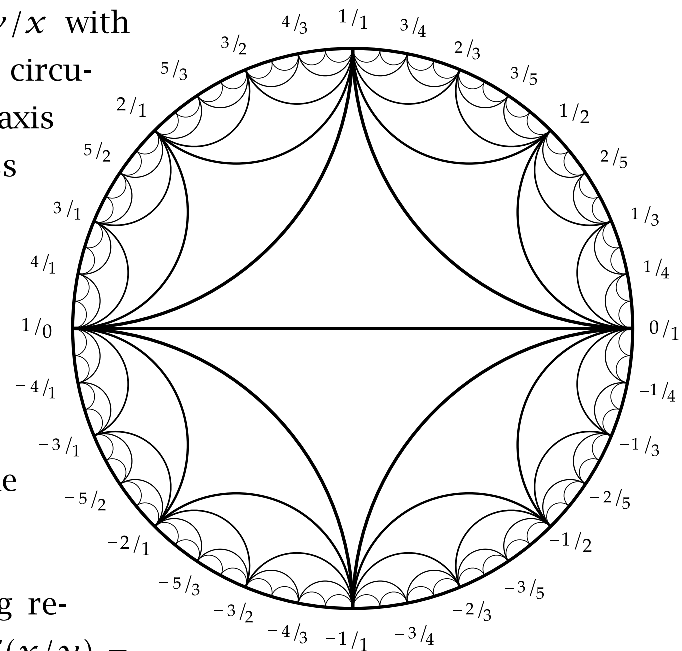
Thus if a,b,c,d are integers with $ad-bc=\pm 1$ then the inverse matrix also has integer entries and determinant ± 1. The factor $\frac{1}{ad-bc}$ is ± 1 so it can be ignored since the matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $-\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ determine the same linear fractional transformation, as we observed in the preceding paragraph.

The preceding proposition says that each linear fractional transformation T in $LF(\mathbb{Z})$ not only sends vertices of the Farey diagram to vertices, but also edges to edges. It follows that T must take triangles in the diagram to triangles in the diagram, since triangles correspond to sets of three vertices, each pair of which forms the endpoints of an edge. Since each transformation T in $LF(\mathbb{Z})$ has an inverse in $LF(\mathbb{Z})$, this implies that T gives a one-to-one (injective) and onto (surjective) transformation of vertices, and also of edges and triangles. For example, if two edges e_1 and e_2 have the same image $T(e_1)=T(e_2)$ then we must have $T^{-1}(T(e_1))=T^{-1}(T(e_2))$ or in other words $e_1=e_2$, so T cannot send two different edges to the same edge, which means it is one-to-one on edges. Also, every edge e_1 is the image $T(e_2)$ of some edge e_2 since we can write $e_1=T(T^{-1}(e_1))$ and let $e_2=T^{-1}(e_1)$. The same reasoning works with vertices and triangles as well as edges.

A useful property of linear fractional transformations that we will use repeatedly is that the way an element of $LF(\mathbb{Z})$ acts on the Farey diagram is uniquely determined by where a single triangle is sent. This is because once one knows where one triangle goes, this uniquely determines where the three adjacent triangles go, and this in turn determines where the six new triangles adjacent to these three go, and so on.

Seven Types of Transformations

We will now give examples illustrating seven different ways that elements of $LF(\mathbb{Z})$ can act on the Farey diagram.



(1) The transformation $T(x/y) = y/x$ with matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ gives a reflection of the circular Farey diagram across its vertical axis of symmetry. This is a reflection across a line perpendicular to an edge of the diagram.

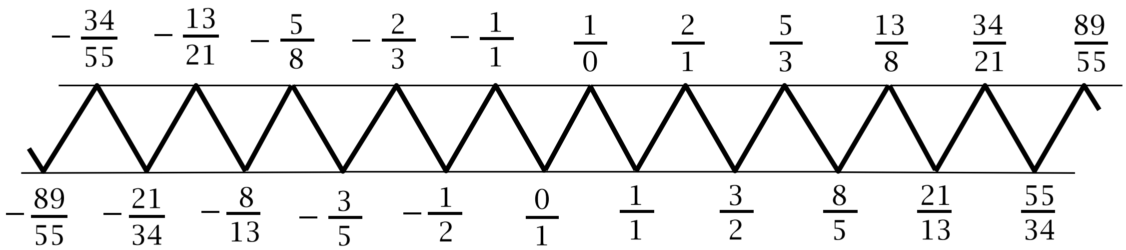
(2) The reflection across the horizontal axis of symmetry is the element $T(x/y) = -x/y$ with matrix $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$. This is a reflection across an edge of the diagram.

(3) If we compose the two preceding reflections we get the transformation $T(x/y) = -y/x$ with matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. This rotates the Farey diagram 180 degrees about its center, interchanging $1/0$ and $0/1$ and also interchanging $1/1$ and $-1/1$. Thus it rotates the diagram 180 degrees about the centerpoint of an edge.

(4) Consider $T(x/y) = y/(y-x)$ corresponding to the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}$. This has the effect of "rotating" the triangle $\langle 1/0, 0/1, 1/1 \rangle$ about its centerpoint, taking $1/0$ to $0/1$, $0/1$ to $1/1$ and $1/1$ back to $1/0$. The whole Farey diagram is then "rotated" about the same point.

(5) Next let $T(x/y) = x/(x+y)$, corresponding to the matrix $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. In particular $T(0/1) = 0/1$, so $0/1$ is a fixed point of T, a point satisfying $T(z) = z$. Also we have $T(1/0) = 1/1$ and more generally $T(1/n) = 1/(n+1)$. Thus the triangle $\langle 0/1, 1/0, 1/1 \rangle$ is taken to the triangle $\langle 0/1, 1/1, 1/2 \rangle$. This implies that T is a "rotation" of the Farey diagram about the vertex $0/1$, taking each triangle with $0/1$ as a vertex to the next triangle in the clockwise direction about this vertex.

(6) A different sort of behavior is exhibited by $T(x/y) = (2x+y)/(x+y)$ corresponding to $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. To visualize T as a transformation of the Farey diagram let us look at the infinite strip



We claim that T translates the whole strip one unit to the right. To see this, notice first that since T takes $1/0$ to $2/1$, $0/1$ to $1/1$, and $1/1$ to $3/2$, it takes the triangle $\langle 1/0, 0/1, 1/1\rangle$ to the triangle $\langle 2/1, 1/1, 3/2\rangle$. This implies that T takes the triangle just to the right of $\langle 1/0, 0/1, 1/1\rangle$ to the triangle just to the right of $\langle 2/1, 1/1, 3/2\rangle$, and similarly each successive triangle is translated one unit to the right. The same argument shows that each successive triangle to the left of the original one is also translated one unit to the right. Thus the whole strip is translated one unit to the right.

(7) Using the same figure as in the preceding example, consider the transformation $T(x/y) = (x+y)/x$ corresponding to the matrix $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. This sends the triangle $\langle 1/0, 0/1, 1/1\rangle$ to $\langle 1/1, 1/0, 2/1\rangle$ which is the next triangle to the right in the infinite strip. Geometrically, T translates the first triangle half a unit to the right and reflects it across the horizontal axis of the strip. It follows that the whole strip is translated half a unit to the right and reflected across the horizontal axis. Such a motion is sometimes referred to as a glide-reflection. Notice that performing this motion twice in succession yields a translation of the strip one unit to the right, the transformation in the preceding example.

Thus we have seven types of symmetries of the Farey diagram: reflections across an edge or a line perpendicular to an edge; rotations about the centerpoint of an edge or a triangle, or about a vertex; and translations and glide-reflections of periodic infinite strips. (Not all periodic strips have glide-reflection symmetries.) It is a true fact, though we won't prove it here, that every element of $LF(\mathbb{Z})$ acts on the Farey diagram in one of these seven ways, except for the identity transformation $T(x/y) = x/y$ of course.

Specifying Where a Triangle Goes

As we observed earlier, the action of an element of $LF(\mathbb{Z})$ on the Farey diagram is completely determined by where it sends a single triangle. Now we will see that there always exists an element of $LF(\mathbb{Z})$ sending any triangle to any other triangle, and in fact, one can do this specifying where each individual vertex of the triangle goes.

As an example, suppose we wish to find an element T of $LF(\mathbb{Z})$ that takes the triangle $\langle 2/5, 1/3, 3/8\rangle$ to the triangle $\langle 5/8, 7/11, 2/3\rangle$, preserving the indicated ordering of the vertices, so $T(2/5) = 5/8$, $T(1/3) = 7/11$, and $T(3/8) = 2/3$. For this problem to even make sense we might want to check first that these really are triangles in the Farey diagram. In the first case, $\langle 2/5, 1/3\rangle$ is an edge since the matrix $\left(\begin{smallmatrix} 2 & 1 \\ 5 & 3 \end{smallmatrix}\right)$ has determinant 1, and there is a triangle joining this edge to $3/8$ since $3/8$ is the mediant of $2/5$ and $1/3$. For the other triangle, the determinant of $\left(\begin{smallmatrix} 5 & 2 \\ 8 & 3 \end{smallmatrix}\right)$ is -1 and the mediant of $5/8$ and $2/3$ is $7/11$.

As a first step toward constructing the desired transformation T we will do something slightly weaker: We construct a transformation T taking the edge $\langle 2/5, 1/3\rangle$ to the edge $\langle 5/8, 7/11\rangle$. This is rather easy if we first notice the general fact that the transformation $T(x/y) = (ax+by)/(cx+dy)$ with matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ takes $1/0$ to a/c and $0/1$ to b/d. Thus the transformation T_1 with matrix $\left(\begin{smallmatrix} 2 & 1 \\ 5 & 3 \end{smallmatrix}\right)$ takes $\langle 1/0, 0/1\rangle$ to $\langle 2/5, 1/3\rangle$, and the transformation T_2 with matrix $\left(\begin{smallmatrix} 5 & 7 \\ 8 & 11 \end{smallmatrix}\right)$ takes $\langle 1/0, 0/1\rangle$ to $\langle 5/8, 7/11\rangle$. Then the product

$$T_2T_1^{-1} = \begin{pmatrix} 5 & 7 \\ 8 & 11 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}^{-1}$$

takes $\langle 2/5, 1/3\rangle$ first to $\langle 1/0, 0/1\rangle$ and then to $\langle 5/8, 7/11\rangle$. Doing the calculation, we get

$$\begin{pmatrix} 5 & 7 \\ 8 & 11 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} 5 & 7 \\ 8 & 11 \end{pmatrix}\begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix} = \begin{pmatrix} -20 & 9 \\ -31 & 14 \end{pmatrix}$$

This takes the edge $\langle 2/5, 1/3\rangle$ to the edge $\langle 5/8, 7/11\rangle$, but does it do the right thing on the third vertex of the triangle $\langle 2/5, 1/3, 3/8\rangle$, taking it to the third vertex of $\langle 5/8, 7/11, 2/3\rangle$? This is not automatic since there are always two triangles containing a given edge, and in this case the other triangle having $\langle 5/8, 7/11\rangle$ as an edge is $\langle 5/8, 7/11, 12/19\rangle$ since $12/19$ is the mediant of $5/8$ and $7/11$. In fact, if we compute what our T does to $3/8$ we get

$$\begin{pmatrix} -20 & 9 \\ -31 & 14 \end{pmatrix}\begin{pmatrix} 3 \\ 8 \end{pmatrix} = \begin{pmatrix} 12 \\ 19 \end{pmatrix}$$

so we don't have the right T yet. To fix the problem, notice that we have a little flexibility in the choice of a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ taking $1/0$ to a/c and $0/1$ to b/d since we can multiply either column by -1 without affecting the fractions a/b and c/d. It doesn't matter which column we multiply by -1 since multiplying both columns by

-1 multiplies the whole matrix by -1 which doesn't change the associated element of $LF(\mathbb{Z})$, as noted earlier. In the case at hand, suppose we change the sign of the first column of $\begin{pmatrix} 5 & 7 \\ 8 & 11 \end{pmatrix}$. Then we get

$$\begin{pmatrix} -5 & 7 \\ -8 & 11 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} -5 & 7 \\ -8 & 11 \end{pmatrix}\begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix} = \begin{pmatrix} -50 & 19 \\ -79 & 30 \end{pmatrix}$$

This fixes the problem since

$$\begin{pmatrix} -50 & 19 \\ -79 & 30 \end{pmatrix}\begin{pmatrix} 3 \\ 8 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Here is a general statement summarizing what we saw in this one example:

Proposition. *(a) For any two triangles $\langle p/q, r/s, t/u\rangle$ and $\langle p'/q', r'/s', t'/u'\rangle$ in the Farey diagram there is a unique element T in $LF(\mathbb{Z})$ taking the first triangle to the second triangle preserving the ordering of the vertices, so $T(p/q) = p'/q'$, $T(r/s) = r'/s'$, and $T(t/u) = t'/u'$.*
(b) The matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ representing a given transformation T in $LF(\mathbb{Z})$ is unique except for replacing it by $\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$.

Proof: As we saw in the example above, there is a composition $T_2T_1^{-1}$ taking the edge $\langle p/q, r/s\rangle$ to $\langle p'/q', r'/s'\rangle$, where T_1 has matrix $\begin{pmatrix} p & r \\ q & s \end{pmatrix}$ and T_2 has matrix $\begin{pmatrix} p' & r' \\ q' & s' \end{pmatrix}$. If this composition $T_2T_1^{-1}$ does not take t/u to t'/u' we modify T_2 by changing the sign of one of its columns, say the first column. Thus we change $\begin{pmatrix} p' & r' \\ q' & s' \end{pmatrix}$ to $\begin{pmatrix} -p' & r' \\ -q' & s' \end{pmatrix}$, which equals the product $\begin{pmatrix} p' & r' \\ q' & s' \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$. The matrix $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ corresponds to the transformation $R(x/y) = -x/y$ reflecting the Farey diagram across the edge $\langle 1/0, 0/1\rangle$. Thus we are replacing $T_2T_1^{-1}$ by $T_2RT_1^{-1}$, inserting a reflection that interchanges the two triangles containing the edge $\langle 1/0, 0/1\rangle$. By inserting R we change where the composition $T_2T_1^{-1}$ sends the third vertex t/u of the triangle $\langle p/q, r/s, t/u\rangle$, so we can guarantee that t/u is taken to t'/u'. This proves part (a).

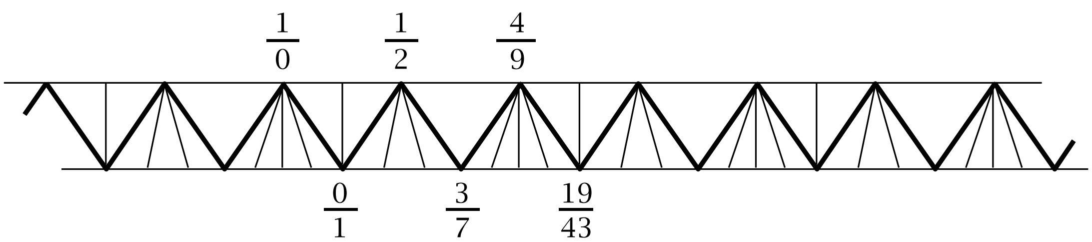
For part (b), note first that the transformation T determines the values $T(1/0) = a/c$ and $T(0/1) = b/d$. The fractions a/c and b/d are in lowest terms (because $ad - bc = \pm 1$) so this means that we know the two columns of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ up to multiplying either or both columns by -1. We need to check that changing the sign of one column without changing the sign of the other column gives a different transformation. It doesn't matter which column we change since $\begin{pmatrix} -a & b \\ -c & d \end{pmatrix} = -\begin{pmatrix} a & -b \\ c & -d \end{pmatrix}$. As we saw in part (a), changing the sign in the first column amounts to replacing T by the composition TR, but this is a different transformation from T since it has a different effect on the triangles containing the edge $\langle 1/0, 0/1\rangle$. $\square$

Continued Fractions Again

Linear fractional transformations can be used to compute the values of periodic or eventually periodic continued fractions, and to see that these values are always quadratic irrational numbers. To illustrate this, consider the periodic continued fraction

$$\overline{{}^1\!\diagup_{2} + {}^1\!\diagup_{3} + {}^1\!\diagup_{1} + {}^1\!\diagup_{4}}$$

The associated periodic strip in the Farey diagram is the following:



We would like to compute the element T of $LF(\mathbb{Z})$ that gives the rightward translation of this strip that exhibits the periodicity. A first guess is the T with matrix $\begin{pmatrix} 4 & 19 \\ 9 & 43 \end{pmatrix}$ since this sends $\langle 1/0, 0/1\rangle$ to $\langle 4/9, 19/43\rangle$. This is actually the correct T since it sends the vertex $1/1$ just to the right of $1/0$, which is the mediant of $1/0$ and $0/1$, to the vertex $(4+19)/(9+43)$ just to the right of $4/9$, which is the mediant of $4/9$ and $19/43$. This is a general fact since $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} a+b \\ c+d \end{pmatrix}$.

The sequence of fractions labeling the vertices along the zigzag path in the strip moving toward the right are the convergents to $\overline{{}^1\!\diagup_{2} + {}^1\!\diagup_{3} + {}^1\!\diagup_{1} + {}^1\!\diagup_{4}}$. Call these convergents $z_1, z_2, \cdots$ and their limit z. When we apply the translation T we are taking each convergent to a later convergent in the sequence, so both the sequence $\{z_n\}$ and the sequence $\{T(z_n)\}$ converge to z. Thus we have

$$T(z) = T(\lim z_n) = \lim T(z_n) = z$$

where the middle equality uses the fact that T is continuous. (Note that a linear fractional transformation $T(z) = \frac{az+b}{cz+d}$ is defined for real values of z, not just rational values $z = x/y$, when $T(x/y) = (ax+by)/(cx+dy) = (a\frac{x}{y}+b)/(c\frac{x}{y}+d)$.)

In summary, what we have just argued is that the value z of the periodic continued fraction satisfies the equation $T(z) = z$, or in other words, $\frac{4z+19}{9z+43} = z$. This can be rewritten as $4z + 19 = 9z^2 + 43z$, which simplifies to $9z^2 + 39z - 19 = 0$. Computing the roots of this quadratic equation, we get

$$z = \frac{-39 \pm \sqrt{39^2 + 4\cdot 9\cdot 19}}{18} = \frac{-39 \pm 3\sqrt{13^2 + 4\cdot 19}}{18} = \frac{-13 \pm \sqrt{245}}{6} = \frac{-13 \pm 7\sqrt{5}}{6}$$

The positive root is the one that the right half of the infinite strip converges to, so we have

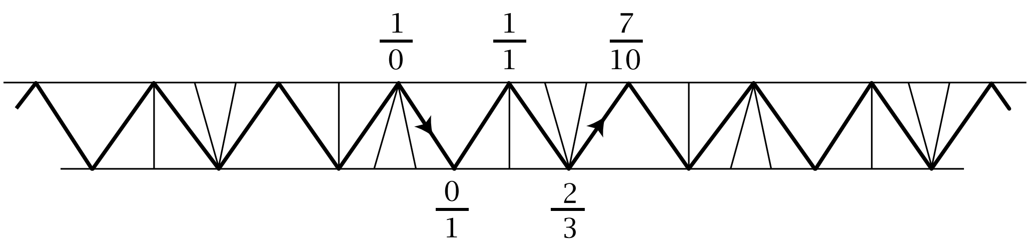
$$\frac{-13+7\sqrt{5}}{6} = \overline{{}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{3} + {}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{4}}$$

Incidentally, the other root $(-13-7\sqrt{5})/6$ has an interpretation in terms of the diagram as well: It is the limit of the numbers labeling the vertices of the zigzag path moving off to the left rather than to the right. This follows by the same sort of argument as above.

If a periodic continued fraction has period of odd length, the transformation giving the periodicity is a glide-reflection of the periodic strip rather than a translation. As an example, consider

$$\overline{{}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{3}}$$

Here the periodic strip is



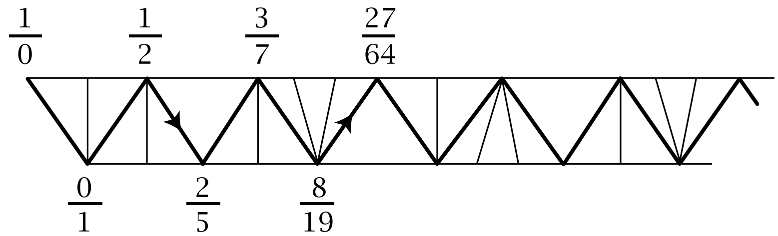
The transformation T with matrix $\begin{pmatrix} 2 & 7 \\ 3 & 10 \end{pmatrix}$ takes $\langle 1/0, 0/1 \rangle$ to $\langle 2/3, 7/10 \rangle$ and the mediant $1/1$ of $1/0$ and $0/1$ to the mediant $9/13$ of $2/3$ and $7/10$ so this transformation is a glide-reflection of the strip. The equation $T(z) = z$ becomes $\frac{2z+7}{3z+10} = z$, which simplifies to $2z+7 = 3z^2+10z$ and then $3z^2+8z-7=0$, with roots $(-4 \pm \sqrt{37})/3$. The positive root gives

$$\frac{-4+\sqrt{37}}{3} = \overline{{}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{3}}$$

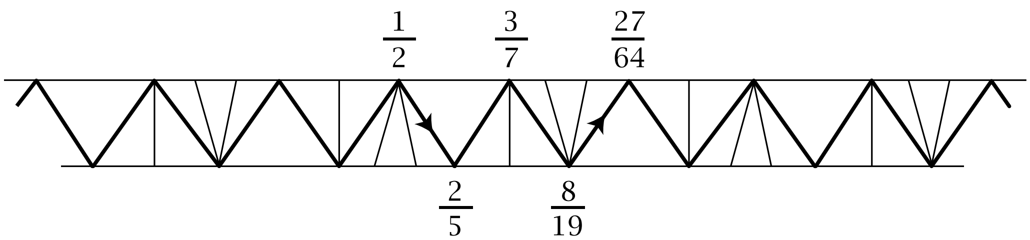
Continued fractions that are only eventually periodic can be treated in a similar fashion. For example, consider

$${}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{2} + \overline{{}^{1}\!\diagup_{1} + {}^{1}\!\diagup_{2} + {}^{1}\!\diagup_{3}}$$

The corresponding infinite strip is



In this case if we discard the triangles corresponding to the initial nonperiodic part of the continued fraction, ${}^{1}\!\!\nearrow_{2} + {}^{1}\!\!\nearrow_{2}$, and then extend the remaining periodic part in both directions, we obtain a periodic strip that is carried to itself by the glide-reflection T taking $\langle 1/2, 2/5\rangle$ to $\langle 8/19, 27/64\rangle$:



We can compute T as the composition $\langle 1/2, 2/5\rangle \to \langle 1/0, 0/1\rangle \to \langle 8/19, 27/64\rangle$ corresponding to the product

$$\begin{pmatrix} 8 & 27 \\ 19 & 64 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}^{-1} = \begin{pmatrix} 8 & 27 \\ 19 & 64 \end{pmatrix}\begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -14 & 11 \\ -33 & 26 \end{pmatrix}$$

Since this transformation takes $3/7$ to the mediant $(8+27)/(19+64)$, it is the glide-reflection we want. Now we solve $T(z) = z$. This means $\frac{-14z+11}{-33z+26} = z$, which reduces to the equation $33z^2 - 40z + 11 = 0$ with roots $z = (20 \pm \sqrt{37})/33$. Both roots are positive, and we want the smaller one, $(20-\sqrt{37})/33$, because along the top edge of the strip the numbers decrease as we move to the right, approaching the smaller root, and they increase as we move to the left, approaching the larger root. Thus we have

$$(20-\sqrt{37})/33 = {}^{1}\!\!\nearrow_{2} + {}^{1}\!\!\nearrow_{2} + \overline{{}^{1}\!\!\nearrow_{1} + {}^{1}\!\!\nearrow_{2} + {}^{1}\!\!\nearrow_{3}}$$

Notice that $\sqrt{37}$ occurs in both this example and the preceding one where we computed the value of $\overline{{}^{1}\!\!\nearrow_{1} + {}^{1}\!\!\nearrow_{2} + {}^{1}\!\!\nearrow_{3}}$. This is not just an accident. It had to happen because to get from $\overline{{}^{1}\!\!\nearrow_{1} + {}^{1}\!\!\nearrow_{2} + {}^{1}\!\!\nearrow_{3}}$ to ${}^{1}\!\!\nearrow_{2} + {}^{1}\!\!\nearrow_{2} + \overline{{}^{1}\!\!\nearrow_{1} + {}^{1}\!\!\nearrow_{2} + {}^{1}\!\!\nearrow_{3}}$ one adds 2 and inverts, then adds 2 and inverts again, and each of these operations of adding an integer or taking the reciprocal takes place within the field $\mathbb{Q}(\sqrt{37})$ consisting of numbers of the form $a + b\sqrt{37}$ with a and b rational. More generally, this argument shows that any eventually periodic continued fraction whose periodic part is $\overline{{}^{1}\!\!\nearrow_{1} + {}^{1}\!\!\nearrow_{2} + {}^{1}\!\!\nearrow_{3}}$ has as its value some number in the field $\mathbb{Q}(\sqrt{37})$. However, not all irrational numbers in this field have eventually periodic continued fractions with periodic part $\overline{{}^1\!/\!_1 + {}^1\!/\!_2 + {}^1\!/\!_3}$. For example, the continued fraction for $\sqrt{37}$ itself is $6 + \overline{{}^1\!/\!_{12}}$, with a different periodic part. (Check this by computing the value of this continued fraction.)

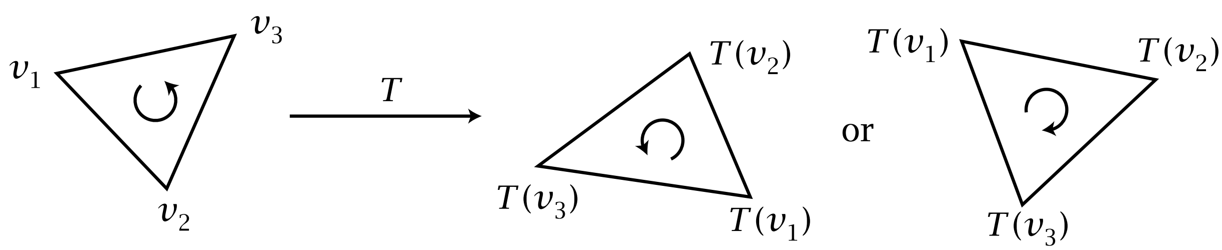
One Half of Langrange's Theorem

The procedure we have used in these examples works in general for any irrational number z whose continued fraction is eventually periodic. From the periodic part of the continued fraction one constructs a periodic infinite strip in the Farey diagram, where the periodicity is given by a linear fractional transformation $T(z) = \frac{az+b}{cz+d}$ with integer coefficients, with T either a translation or a glide-reflection of the strip. As we argued in the first example, the number z satisfies the equation $T(z) = z$. This becomes the quadratic equation $az + b = cz^2 + dz$ with integer coefficients, or in simpler form, $cz^2 + (d - a)z - b = 0$. By the quadratic formula, the roots of this equation have the form $A + B\sqrt{n}$ for some rational numbers A and B and some integer n. We know that the real number z is a root of the equation so n can't be negative, and it can't be a square since z is irrational.

Thus we have an argument that proves one half of Lagrange's Theorem, the statement that a number whose continued fraction is periodic or eventually periodic is a quadratic irrational. There is one technical point that should be addressed, however. Could the leading coefficient c in the quadratic equation $cz^2 + (d - a)z - b = 0$ be zero? If this were the case then we couldn't apply the quadratic formula to solve for z, so we need to show that c cannot be zero. We do this in the following way. If c were zero the equation would become the linear equation $(d - a)z - b = 0$. If the coefficient of z in this equation is nonzero, we have only one root, $z = b/(d - a)$, a rational number contrary to the fact that z is irrational since its continued fraction is infinite. Thus we are left with the possibility that $c = 0$ and $a = d$, so the equation for z reduces to the equation $b = 0$. Then the transformation T would have the form $T(z) = \frac{az}{a} = z$ so it would be the identity transformation. However we know it is a genuine translation or a glide-reflection, so it is not the identity. We conclude from all this that c cannot be zero, and the technical point is taken care of.

Orientations

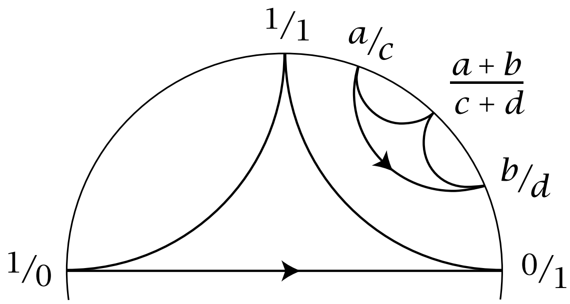
Elements of $LF(\mathbb{Z})$ are represented by integer matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ of determinant ± 1. The distinction between determinant $+1$ and -1 has a very nice geometric interpretation in terms of orientations, which can be described in terms of triangles. A triangle in the Farey diagram can be oriented by choosing either the clockwise or counterclockwise ordering of its three vertices. An element T of $LF(\mathbb{Z})$ takes each triangle to another triangle in a way that either preserves the two possible orientations or reverses them.



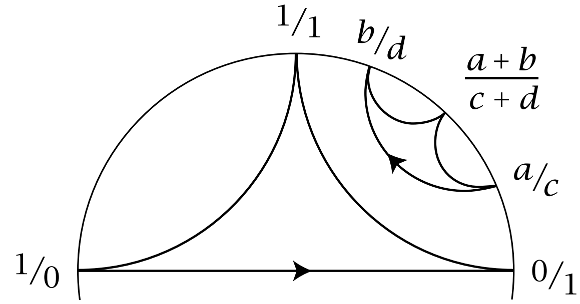
For example, among the seven types of transformations we looked at earlier, only reflections and glide-reflections reverse the orientations of triangles. Note that if a transformation T preserves the orientation of one triangle, it has to preserve the orientation of the three adjacent triangles, and then of the triangles adjacent to these, and so on for all the triangles. Similarly, if the orientation of one triangle is reversed by T, then the orientations of all triangles are reversed.

Proposition. *A transformation $T(x/y) = (ax + by)/(cx + dy)$ in $LF(\mathbb{Z})$ preserves orientations of triangles in the Farey diagram when the determinant $ad - bc$ is $+1$ and reverses the orientations when the determinant is -1.*

Proof: We will first prove a special case and then deduce the general case from the special case. The special case is that a, b, c, d are all positive or zero. The transformation T with matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ takes the edge $\langle 1/0, 0/1 \rangle$ in the circular Farey diagram to the edge $\langle a/c, b/d \rangle$, and if a, b, c, d are all positive or zero, this edge lies in the upper half of the diagram. Since $T(1/1) = (a+b)/(c+d)$, the triangle $\langle 1/0, 0/1, 1/1 \rangle$ is taken to the triangle $\langle a/c, b/d, (a+b)/(c+d) \rangle$ whose third vertex $(a+b)/(c+d)$ lies above the edge $\langle a/c, b/d \rangle$, by the way the Farey diagram was constructed using mediants, since we assume a, b, c, d are positive or zero. We know that the edge $\langle a/c, b/d \rangle$ is oriented to the right if $ad - bc = +1$ and to the left if $ad - bc = -1$. This means that T preserves the orientation of the triangle $\langle 1/0, 0/1, 1/1 \rangle$ if the determinant is $+1$ and reverses the orientation if the determinant is -1.



$ad - bc = +1$



$ad - bc = -1$

This proves the special case.

The general case can be broken into two subcases, according to whether the edge $\langle a/c, b/d\rangle$ lies in the upper or the lower half of the diagram. If $\langle a/c, b/d\rangle$ lies in the upper half of the diagram, then after multiplying one or both columns of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ by -1 if necessary, we will be in the special case already considered. Multiplying both columns by -1 doesn't affect T. Multiplying one column by -1 corresponds to first reflecting across the edge $\langle 1/0, 0/1\rangle$, as we have seen earlier. Modifying T in this way changes the sign of the determinant and it also changes whether T preserves or reverses orientation, so the special case already proved implies the case that T takes $\langle 1/0, 0/1\rangle$ to an edge in the upper half of the diagram.

The remaining possibility is that T takes the edge $\langle 1/0, 0/1\rangle$ to an edge in the lower half of the diagram. In this case if we follow T by reflection across the edge $\langle 1/0, 0/1\rangle$ we get a new transformation taking $\langle 1/0, 0/1\rangle$ to an edge in the upper half of the diagram. As before, composing with this reflection changes T from orientation-preserving to orientation-reversing and vice versa, and it also changes the sign of the determinant since the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is changed to $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} -a & -b \\ c & d \end{pmatrix}$, so this case follows from the previous case. $\square$

Computational note: To determine whether a matrix representing an element of $LF(\mathbb{Z})$ has determinant $+1$ or -1 it suffices to compute just the last digit of the determinant, and this can be done using just the last digit of the entries in the matrix. This is easy to do in one's head even if the entries in the matrix have many digits.

We will let $LF^+(\mathbb{Z})$ denote the elements of $LF(\mathbb{Z})$ corresponding to matrices of determinant $+1$.

Proposition. *For any two edges $\langle p/q, r/s\rangle$ and $\langle p'/q', r'/s'\rangle$ of the Farey diagram there exists a unique element $T \in LF^+(\mathbb{Z})$ taking the first edge to the second edge* *preserving the ordering of the vertices, so* $T(p/q) = p'/q'$ *and* $T(r/s) = r'/s'$.

Proof: We already know that there exists an element T in $LF(\mathbb{Z})$ with $T(p/q) = p'/q'$ and $T(r/s) = r'/s'$, and in fact there are exactly two choices for T which are distinguished by which of the two triangles containing $\langle p'/q', r'/s' \rangle$ a triangle containing $\langle p/q, r/s \rangle$ is sent to. One of these choices will make T preserve orientation and the other will make T reverse orientation. So there is only one choice where the determinant is $+1$. $\square$