Quadratic Forms: A Geometric Approach

Featuring a dizzying number of circles

Keith Sheppard

260957189

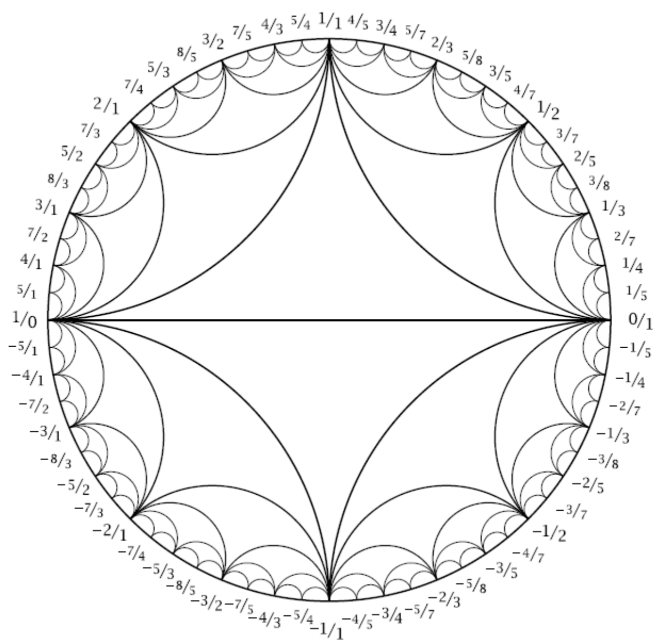


Figure 1: The Farey diagram (from *Topology of Numbers* by Hatcher).

A Directed Reading Program (DRP) project

Department of Mathematics and Statistics
McGill University

July 13, 2023

Introduction

While the theory of quadratic forms as pioneered by Gauss is relatively recent (by mathematical standards), their history is extensive. One of the most ubiquitous examples of quadratic forms in practice is in the **Pythagorean theorem**: in any right triangle with sides a, b and hypotenuse c,

$$a^2 + b^2 = c^2. \tag{1}$$

In school, we often come across simple integer solutions to equation (1), which we write as **Pythagorean triples**. Some examples include $(a, b, c) = (3, 4, 5)$, $(5, 12, 13)$, and so on. However, to the mathematician, these few solutions provide a very incomplete picture. It is natural to take mathematical problems such as these and think about them more generally. For example, we might ask:

- How many Pythagorean triples are there?
- How many Pythagorean triples exist with $c = 7$, or with $b = 15$, or with $a = 29$?
- Do the equations $a^3 + b^3 = c^3$ and $a^4 + b^4 = c^4$ also have integer solutions?

The answer to the third question (which is, "no") comes from Fermat's Last Theorem, which took hundreds of years of work by the world's greatest mathematical minds to answer. As such, its proof is beyond the scope of this article. However, even without understanding every step of the proof, it quickly becomes clear that problems like these, while seemingly within the realm of simple trigonometry, draw theory and techniques from advanced mathematics. And at the heart of extensive theory in this area is the *quadratic form*.

While the name might sound intimidating, a **quadratic form** is simply a homogeneous degree-two polynomial: that is, a polynomial where every term has degree two. These include $5xy + 6y^2$, $2xz + z^2 + 3xy$, and of course $x^2 + y^2$. For simplicity, we will restrict our study to **binary** quadratic forms which have only two variables, x and y.

Notice, then, that the Pythagorean theorem simply describes some **representations** (i.e. integer solutions) of a special case of the most general quadratic form,

$$ax^2 + bxy + cy^2,$$

where $a = c = 1$ and $b = 0$. So perhaps we can explore more about quadratic forms as a whole, especially with regard to the Pythagorean theorem, by considering the question:

When is a number n represented by the quadratic form $ax^2 + bxy + cy^2$, where a, b, and c are integers?

And it is that question that we will address in this article.

Pythagorean Triples and the Unit Circle

We first go back to the Pythagorean theorem to begin the discussion of quadratic forms. One of the most common ways to generate Pythagorean triples is using (a variant of) **Euclid's formula**:

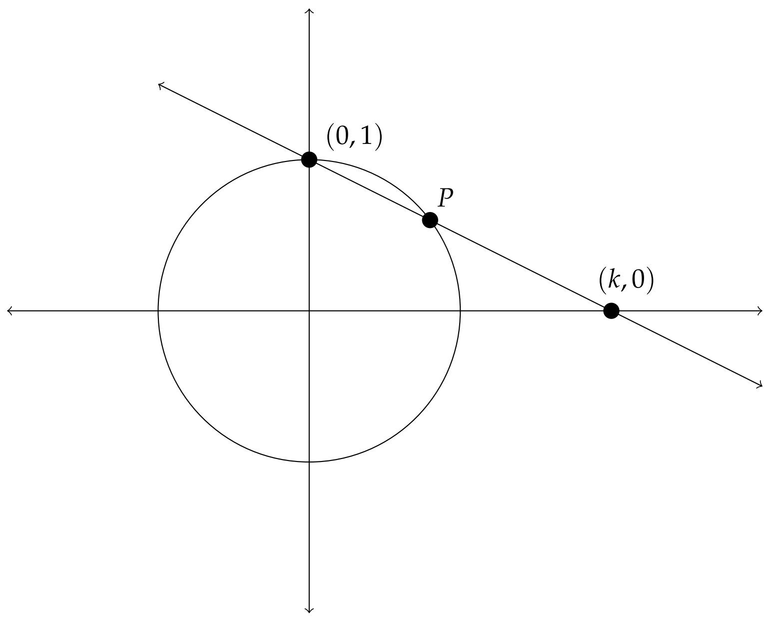
$$p, q \in \mathbb{Z} \implies (2pq, p^2 - q^2, p^2 + q^2) \text{ is a Pythagorean triple}$$

There are several derivations of this result. One common way is to observe that

$$a^2 + b^2 = c^2 \implies b^2 = c^2 - a^2 = (c + a)(c - a),$$

where if a and c are both odd, $c + a \equiv 2m$ and $c - a \equiv 2n$ are both even. From there $c = m + n$, $a = m - n$, and b being forced to be even results in $m = p^2$ and $n = q^2$ being square numbers when m and n share no common factors (a corollary of the fundamental theorem of arithmetic).

Recall that we first asked how many Pythagorean triples exist. We usually answer this question mathematically by finding a bijection between two sets, one of which whose size we know. We can do this here as follows. Take any point on the **unit circle** $x^2 + y^2 = 1$; it is easiest to use $(0, 1)$. Now draw any line with nonzero slope through this point. It will intersect the unit circle at some point P:



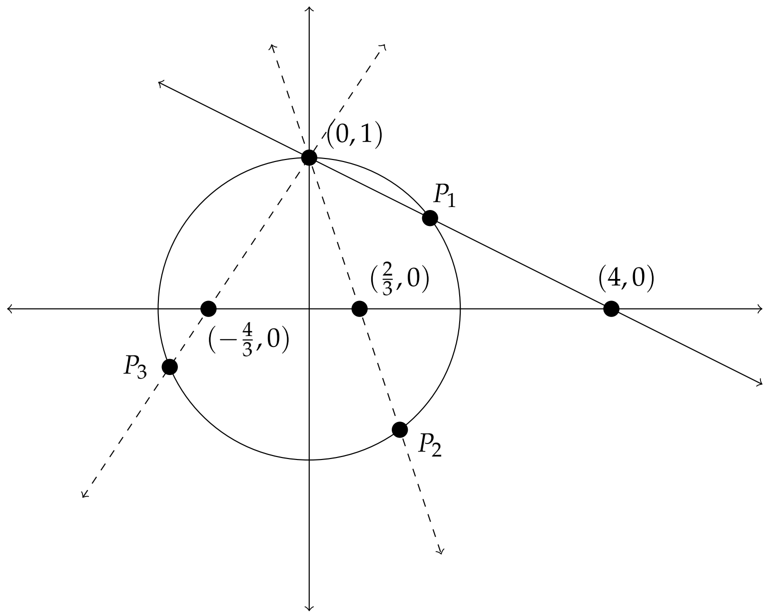
Notice this line must have equation $y = -kx + 1$ (to pass through the x and y axes at the desired points). This means that we can solve for the x coordinate at P:

$$\begin{aligned} y = 1 - kx &\implies y^2 = 1 - 2kx + k^2x^2 \\ &\implies 1 - x^2 = 1 - 2kx + k^2x^2 \\ &\implies (k^2 + 1)x^2 - 2kx = 0 \\ &\implies x[(k^2 + 1)x - 2k] = 0 \\ &\implies x = \frac{2k}{k^2 + 1} \end{aligned}$$

whereby $y = 1 - \frac{2k^2}{k^2+1} = \frac{1-k^2}{1+k^2}$. So if k is rational, the point on the circle must be rational; in other words, when $k = \frac{p}{q}$,

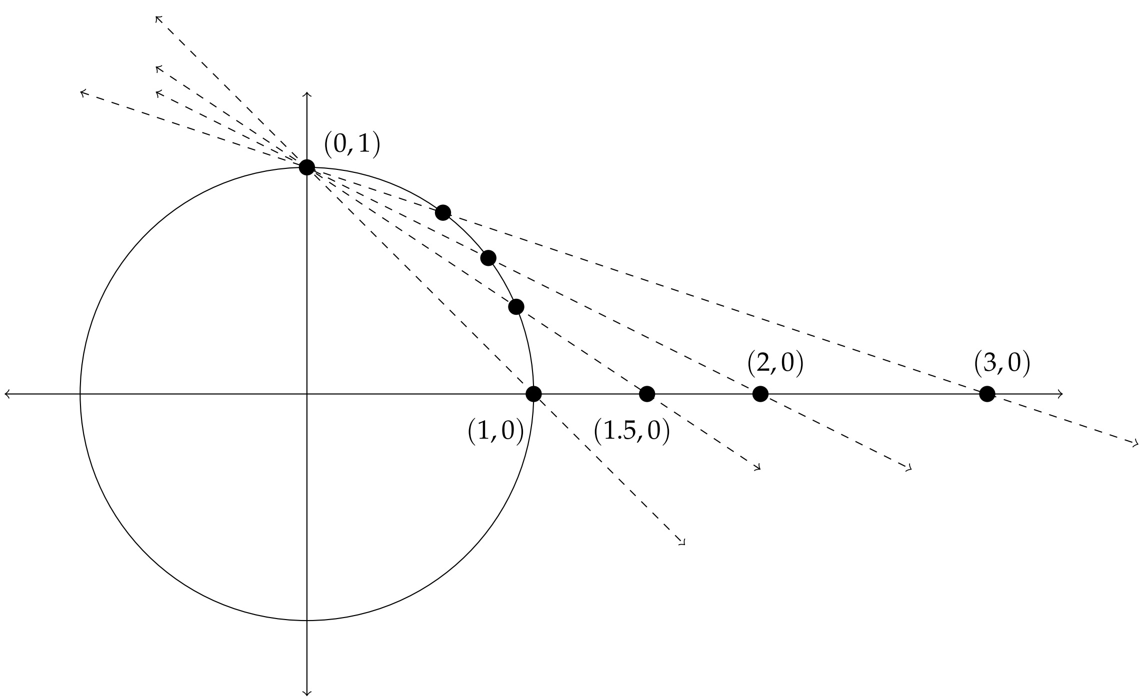
$$x = \frac{2k}{k^2 + 1} = \frac{2p}{q(\frac{p^2}{q^2} + 1)} = \frac{2pq}{p^2 + q^2} \qquad\qquad y = \frac{1 - k^2}{1 + k^2} = \frac{1 - \frac{p^2}{q^2}}{1 + \frac{p^2}{q^2}} = \frac{q^2 - p^2}{q^2 + p^2}$$

which leads us to Euclid's original formula. This shows how the unit circle and Pythagorean triples are related. In particular, it highlights that every rational point on the unit circle corresponds to a Pythagorean triple. Further, we also conclude that there must be infinitely many Pythagorean triples given there are infinitely many rational points on the x axis.



The Farey Diagram

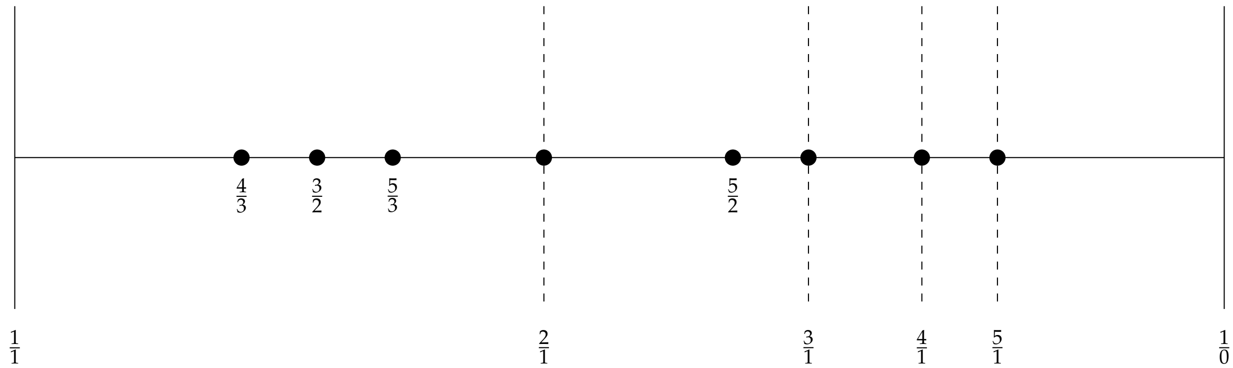
Let us go through the plotting process mentioned previously using "nice" (i.e. easy to work with) points on the x axis. For example:



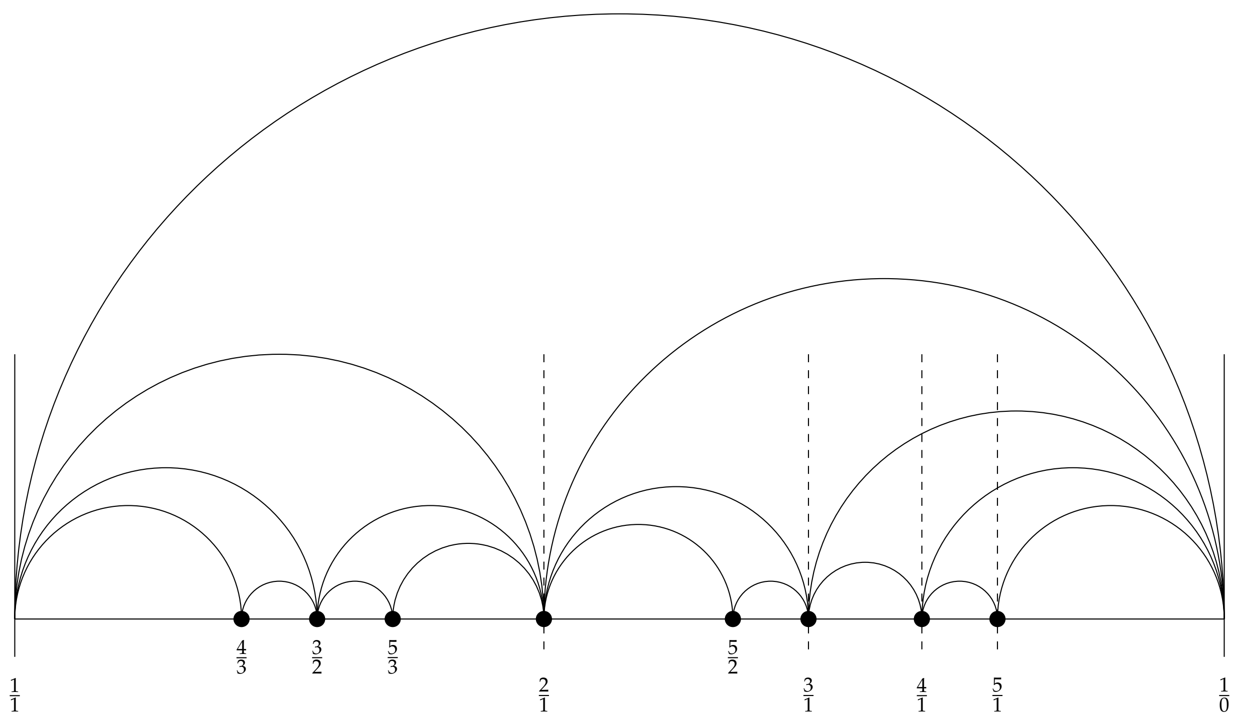
To get a better sense of how these points are distributed, we can change coordinates so that the quarter circle on which we were plotting points becomes a horizontal line. To do this, use the transformation

$$f(x, y) = \left(\frac{1 + \sqrt{1 - x^2}}{x}, 0 \right).$$

Notice that this transformation simply extracts the value of k (by solving $x = \frac{2k}{k^2+1}$ for k) and makes it the x coordinate. Then we reach the graph



where we could continue plotting points *ad infinitum*. Now comes the strange part: we now connect two fractions $\frac{a}{b}$ and $\frac{c}{d}$ with a concave semicircle if the matrix $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$ has determinant ± 1.[1] This gives a new graph



[1] Where does this condition come from? A brief explanation is the following. There are many examples of binary quadratic forms, and some share similar properties. This is described in a property called *equivalence*, which happens when a linear transformation can take one quadratic form and turn it into another. However, these transformations must use coefficients that form a matrix with determinant ± 1. This is because the group these matrices are a part of – called $SL_2(\mathbb{Z})$ – has an important feature which is that its transformations preserve structures like area and orientation, something that comes up when examining symmetries in the Farey diagram.

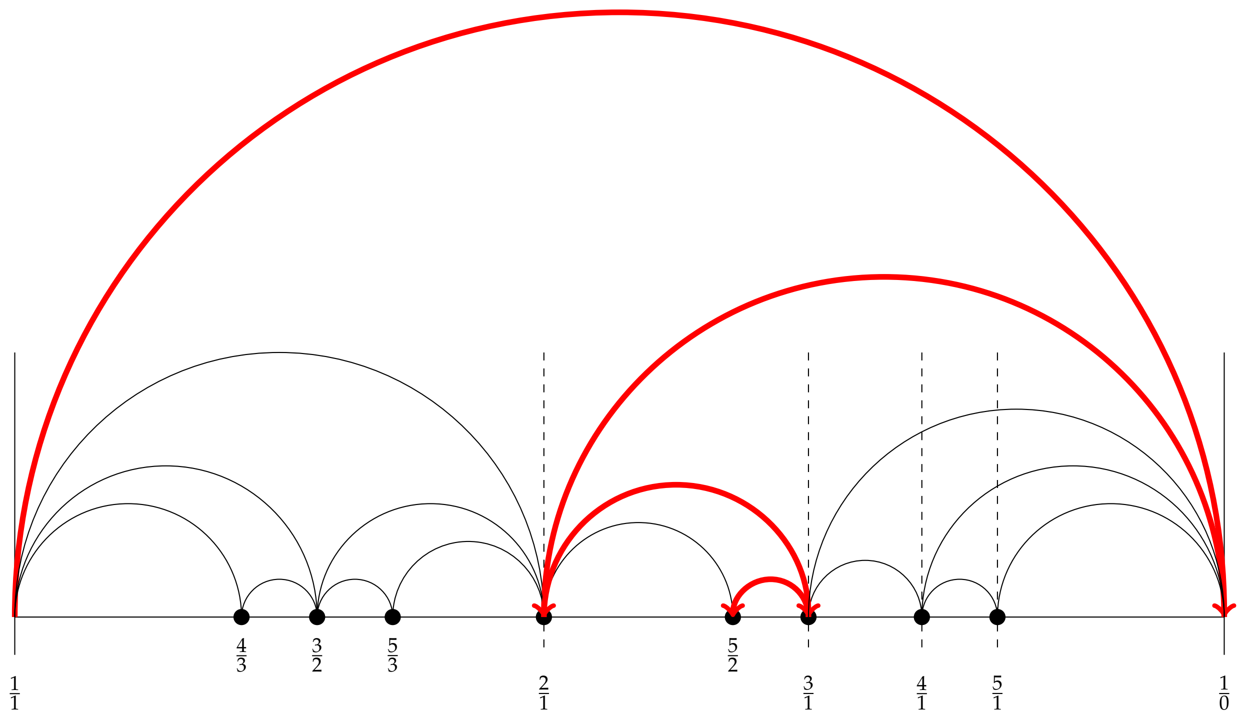
We can use our seemingly obscure rule we describe about creating a matrix with determinant ± 1 to come up with another, simpler way to determine when two points should be connected by a semicircle: start with $\frac{1}{1}$ and $\frac{1}{0}$. We know that the matrix $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ has determinant -1, but the matrices

$$\begin{pmatrix} 1 & 1+1 \\ 1 & 1+0 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 1+1 & 1 \\ 1+0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$$

also have determinant -1. So we draw a semicircle between $\frac{1}{1}$ and $\frac{2}{1}$, and between $\frac{2}{1}$ and $\frac{1}{0}$. In mathematical parlance, taking two fractions $\frac{a}{b}$ and $\frac{c}{d}$ and producing the new fraction

$$\frac{a+c}{b+d}$$

results in what we call the **mediant** of the two fractions. So we can draw semicircles between these points by simply starting at $\frac{1}{1}$ and $\frac{0}{1}$, and finding successive mediants. As an example:



gives us the path to $\frac{5}{2}$. Soon we will see that this technique, when applied infinitely, gives converging subsequences to various irrational numbers. With this in mind, can you determine what our sequence of mediants

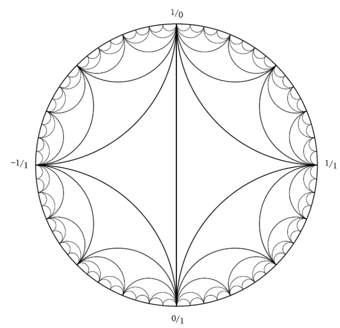
$$\frac{1}{1}, \frac{1}{0}, \frac{2}{1}, \frac{3}{1}, \frac{5}{2}, \frac{8}{3}, \frac{13}{5}, \ldots$$

converges to?[2]

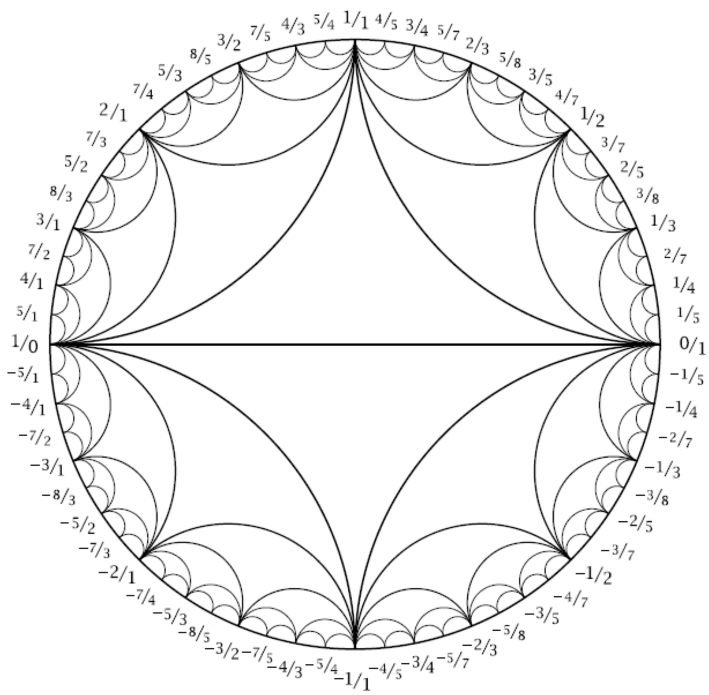
Notice that we only worked with a quarter of the unit circle; you might be confused, because earlier we showed that the process of finding rational points applies to the entire unit circle, not just the part in the first quadrant. However, the process of transforming the remaining quarter-arcs and drawing semicircles

[2]Perhaps you are familiar with the fact that the ratio of two successive Fibonacci numbers has special mathematical significance: this will also apply here, but a slight correction needs to be made to get the right answer. You can see this after calculating further terms.

connecting mediants is the same. When we transform our new diagram back into Cartesian coordinates (i.e. apply f^{-1}), we come across a circle with various semicircles on the inside. These circles begin and end at various points on the circumference of a circle, which seem distributed haphazardly (other than the fact that a mediant of two fractions is located somewhere in between those fractions on the line). It would be nice if the mediant of two given fractions on the circle was also the midpoint between them along its perimeter, like so:



This is one version what is called the **Farey diagram**, different from f^{-1} in that its points are irrational, as opposed to the rational points of f^{-1}[3]. There are several others: one of them we achieve by rotating the image counter-clockwise to get Hatcher's version of the Farey diagram on the cover of this article:



There is an interesting application of the Farey diagram called the **Farey sequence**. A Farey sequence is a sequence of reduced integer fractions whose denominators are not greater than a given integer. For

[3] This should make sense if you consider how the unit circle is used in trigonometry: for the quarter arc of the unit circle in the first quadrant, the point that divides it in two is naturally 45° in polar coordinates, or $(\cos(45^\circ), \sin(45^\circ)) = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ in Cartesian coordinates. $\sqrt{2}$, of course, is irrational, and so are the proceeding midpoints

$$\left(\cos\left(\frac{\theta}{2}\right), \sin\left(\frac{\theta}{2}\right)\right) = \left(\sqrt{\frac{1+\cos(\theta)}{2}}, \sqrt{\frac{1-\cos(\theta)}{2}}\right) = \left[\frac{1}{2}(e^{\frac{i\theta}{2}} + e^{-\frac{i\theta}{2}}), \frac{i}{2}(e^{-\frac{i\theta}{2}} - e^{\frac{i\theta}{2}})\right]$$

example, the fourth Farey sequence is given by

$$F_4 = \left\{0,\ \frac{1}{4},\ \frac{1}{3},\ \frac{1}{2},\ \frac{2}{3},\ \frac{3}{4},1\right\},$$

which you can achieve by tracing with your pen along equal length semicircles on the Farey diagram going from $\frac{0}{1}$ to $\frac{1}{1}$. And what makes these sequences useful is that they describe *all* nonnegative fractions with a denominator less than your choosing. This could come up, for example, if you have the time (and your friends have the patience) to come up with the optimal distribution of pizza at a party.

Many other interesting features of the Farey diagram, as alluded to in the first footnote, are best explained with higher level math concepts like hyperbolic geometry, the upper half plane $\mathbb{H}$, and group actions. These include the fact that none of the semicircles on the Farey diagram intersect and that the Farey diagram is both vertically and horizontally symmetric.

The Topograph

Whether through our exercise of plotting rational points or through seeing the Farey diagram itself, we should be convinced of the fact that every rational number is found on the Farey diagram. And this is helpful because it gives us endless pairs of integers that we can input into any quadratic form. Take, for example, the perfect-square binomial $x^2 + 2xy + y^2$. Starting with the first four points of our Farey diagram, we find that

$$(x, y) = (1, -1.5) \implies x^2 + 2xy + y^2 = 1 \qquad (x, y) = (0, 1) \implies x^2 + 2xy + y^2 = 1$$

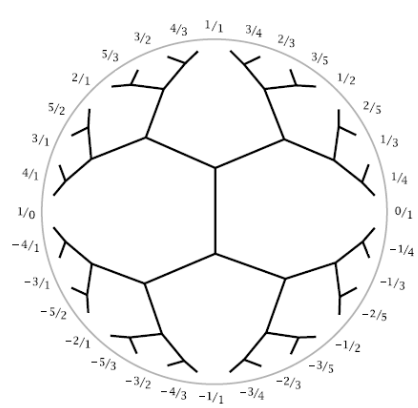
$$(x, y) = (1, 1) \implies x^2 + 2xy + y^2 = 4 \qquad (x, y) = (-1, 1) \implies x^2 + 2xy + y^2 = 0$$

The polynomial $x^2 + 2xy + y^2$ is particularly nice in that

$$(x, y) = (a, b) \implies x^2 + 2xy + y^2 = a^2 + 2ab + b^2 \impliedby (x, y) = (b, a)$$

So there is a certain "symmetry" in the outputs of this quadratic form.

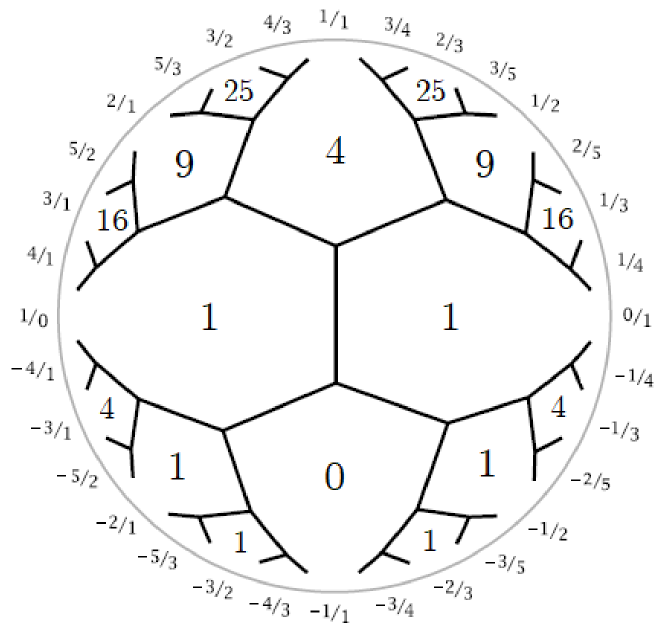
To examine this more closely, instead of putting some integer pairs from the Farey diagram into our quadratic form, let us treat the Farey diagram as a function. Our input will be the quadratic form, and the output will be its **topograph**. A topograph is constructed as follows. Every region in the Farey diagram is a (warped) triangle: take this triangle and put a point at its geometric center. Then, link points in adjacent triangles together. Without the various semicircles, this gives



where, again, we can continue drawing those "branches" at the edges as long as we have patience for.

Into each region, let us input the value of a given quadratic form using the fraction $\frac{p}{q}$ on the Farey diagram that it is closest to. So, using our analysis at the start of this section, the four largest regions would contain the numbers 4, 1, 0, and 1 if we start at the top and write write clockwise.

Similarly, in the next four largest regions we can place the values $2^2+2(2)(1)+1^2=9$, $1^2+2(1)(2)+2^2=9$, $(-1)^2+2(-1)(2)+2^2=1$ and $(-2)^2+2(-2)(1)+1^2=1$. Continuing the process gives a taste of the output of our function:



One might ask whether there is a shortcut to create a topograph given a quadratic form: after all, it would be pretty tedious to have to do all the calculations by hand. It turns out there is, and we will call it the *arithmetic progression rule* for a reason we will prove now.

Take any quadratic form $ax^2+bxy+cy^2$, and two fractions $\frac{m_1}{n_1}, \frac{m_2}{n_2}$, and their mediant $\frac{m_1+m_2}{n_1+n_2}$. If we plug these fractions into the given quadratic form, we get

$$(x,y)=(m_1,n_1) \implies am_1^2+bm_1n_1+cn_1^2$$

$$(x,y)=(m_2,n_2) \implies am_2^2+bm_2n_2+cn_2^2$$

$$(x,y)=(m_1+m_2,n_1+n_2) \implies am_1^2+bm_1n_1+cn_1^2+am_2^2+bm_2n_2+cn_2^2+b(m_2n_1+m_1n_2)+2(am_1m_2+cn_1n_2)$$

where the output of (m_1+m_2, n_1+n_2) is just the sum of the outputs of (m_1,n_1) and (m_2+n_2) along with some extraneous terms.

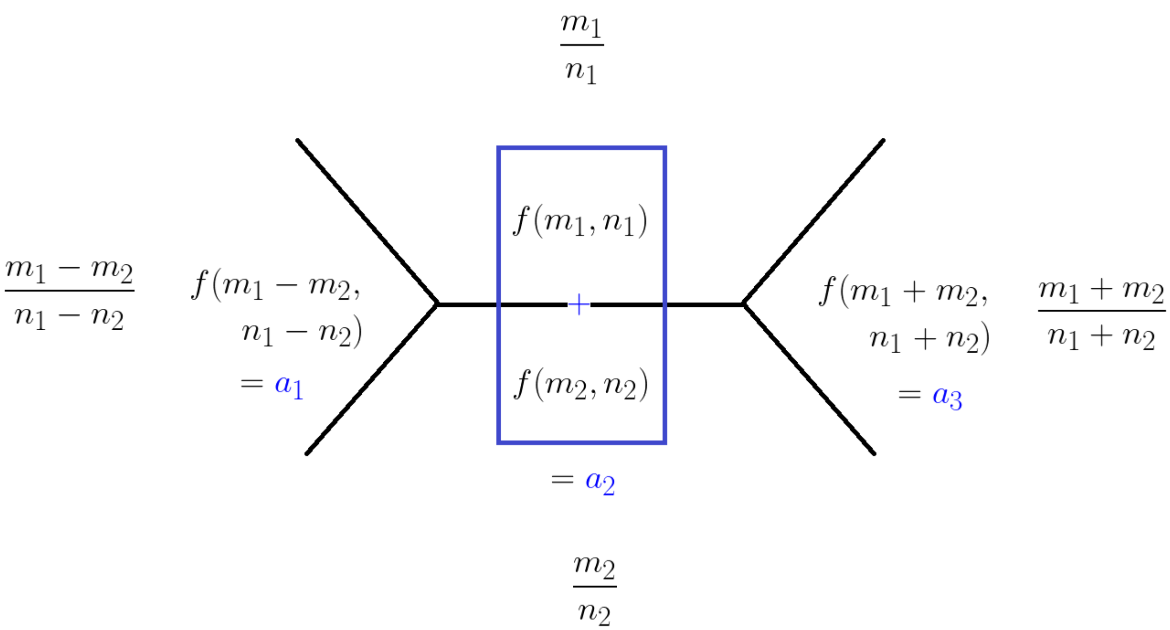
Now suppose we wanted to find a fraction that, along with $\frac{m_1+m_2}{n_1+n_2}$, would give a mediant of $\frac{m_1}{n_1}$ (this fraction is arbitrarily chosen of the two). Some backwards algebra gives the fraction $\frac{m_1-m_2}{n_1-n_2}$. And when we input this fraction into our quadratic form we get

$$(x,y)=(m_1-m_2,n_1-n_2) \implies am_1^2+bm_1n_1+cn_1^2+am_2^2+bm_2n_2+cn_2^2-b(m_2n_1+m_1n_2)-2(am_1m_2+cn_1n_2)$$

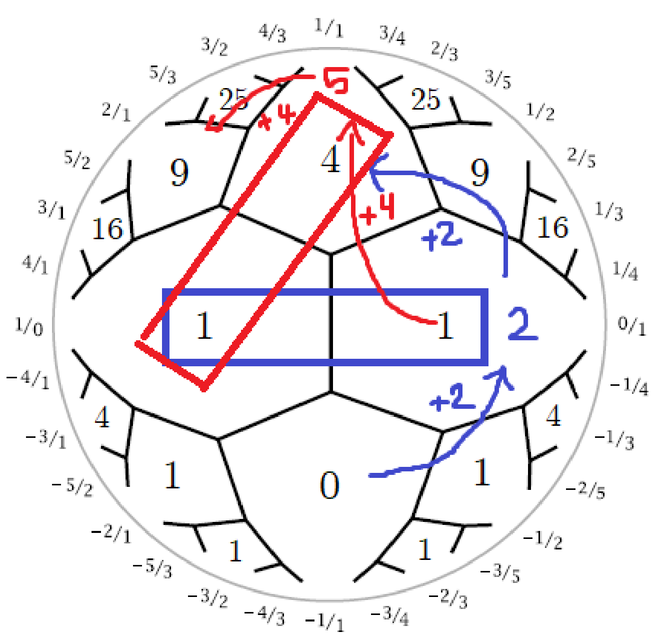
where the extraneous terms are now subtracted instead of added. If we set $f(m,n) = am^2 + bmn + cn^2$, this means that the sequence

$$f(m_1 - m_2, n_1 - n_2),\ f(m_1, n_1) + f(m_2, n_2),\ f(m_1 + m_2, n_1 + n_2)$$

is an arithmetic progression with difference $b(m_2 n_1 + m_1 n_2) + 2(am_1 m_2 + cn_1 n_2)$. As an image,

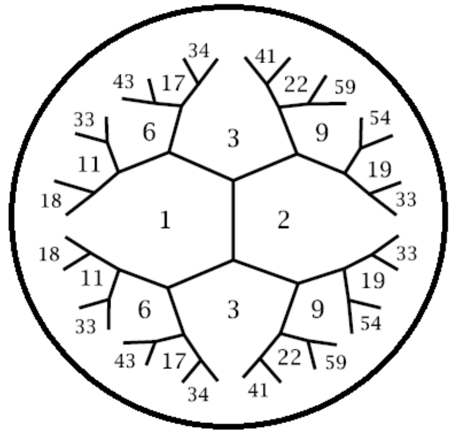


We can reinforce this trend by examining the topograph we previously constructed:



where there are many more junctions where we find that this "arithmetic progression rule" holds. What this implies is that for any quadratic form, given its outputs of $(1,0)$,$(0,1)$, and $(1,1)$ (or, $(-1,1)$), one can fully construct its topograph. As another example,

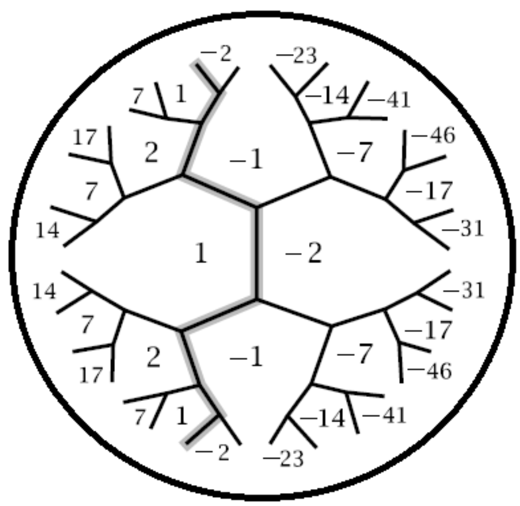
$$x^2 + 2y^2 \quad \Longrightarrow$$



We can see that the numbers in the topograph progressively get larger as one approaches the perimeter of the circle. From this we can conclude, for example, that there are no integer solutions $x^2 + 2y^2 = 7$ and $x^2 + 2y^2 = 10$.

Among every topograph, there is the property that outputs of the quadratic form get larger as we move to the edge of the circle. This can be called the *monotonicity property*. Some topographs also exhibit various symmetries depending on the nature of the quadratic form, as we saw earlier. However, when we introduce negative coefficients (and thus the possibility of negative outputs), more complex patterns start to emerge. For example,

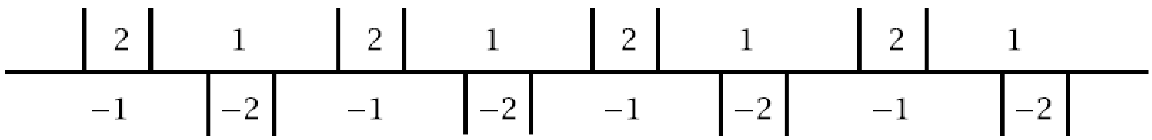
$$x^2 - 2y^2 \quad \Longrightarrow$$



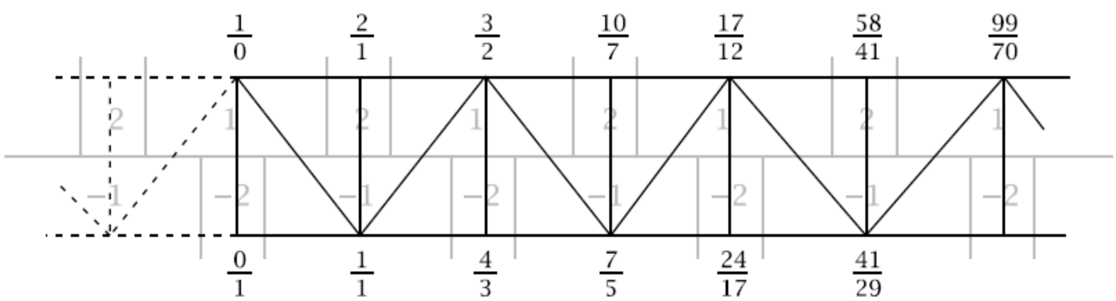
where the topograph has a distinct partition between positive and negative outputs that was named the **river** by Conway. Ignoring changes in sign, we can see in the topograph of $x^2 - 2y^2$ that on both sides of the river seem to be the values 1 and 2. Rivers like this one always have a pattern of numbers that repeats itself, and it is not these numbers themselves but their corresponding inputs that serve as motivation for the next topic.

Continued Fractions

If we extract the river of the topograph of $x^2 - 2y^2$, we get



where the image continues infinitely in both directions. Notice the pattern that was described earlier. If we superimpose the relevant Farey diagram fractions and semicircles, this image becomes



where the sequence of fractions

$$\left\{\frac{1}{0}, \frac{0}{1}, \frac{1}{1}, \frac{2}{1}, \frac{3}{2}...\right\} = \{\infty, 0, 1, 2, 1.5, 1.333, 1.4, 1.42857..., 1.4166..., \}$$

seems to converge to $\sqrt{2}$. Why does this happen? Here, it is worth looking at the fractions along the diagonal lines, namely $\frac{1}{0}, \frac{1}{1}, \frac{3}{2}$, and so on (this is because they form an increasing, bounded subsequence which hence converges). These fractions are really

$$\left\{1, 1+\frac{1}{2}, 1+\frac{2}{5}, 1+\frac{5}{12}, ...\right\} = \left\{1, 1+\frac{1}{2}, 1+\frac{1}{\frac{5}{2}}, 1+\frac{1}{\frac{12}{5}}, ...\right\}$$

$$= \left\{1, 1+\frac{1}{2}, 1+\frac{1}{1+(1+\frac{1}{2})}, 1+\frac{1}{1+(1+\frac{2}{5})}, ...\right\}$$

In other words, we have the recursion

$$a_n = 1 + \frac{1}{1+a_{n-1}}.$$

This can be solved by taking limits:

$$\lim_{n\to\infty} a_n = 1 + \frac{1}{1+\lim\limits_{n\to\infty} a_{n-1}} = 1 + \frac{1}{1+\lim\limits_{n\to\infty} a_n}$$

whereby

$$\lim_{n\to\infty} a_n = \sqrt{2}.$$

So our suspicion was indeed correct. Is this just a coincidence, or is there something deeper here? Let us take a closer look at the previous sequence of fractions that we got our recursion from. There isn't anything stopping us from performing the same Euclidean-algorithm style division on the fractional denominators:

$$\begin{aligned}\left\{1, 1+\frac{1}{2}, 1+\frac{2}{5}, 1+\frac{5}{12}, \ldots\right\} &= \left\{1, 1+\frac{1}{2}, 1+\frac{1}{\frac{5}{2}}, 1+\frac{1}{\frac{12}{5}}, \ldots\right\} \\ &= \left\{1, 1+\frac{1}{2}, 1+\frac{1}{2+\frac{1}{2}}, 1+\frac{1}{2+\frac{2}{5}}, \ldots\right\} \\ &= \left\{1, 1+\frac{1}{2}, 1+\frac{1}{2+\frac{1}{2}}, 1+\frac{1}{2+\frac{1}{\frac{5}{2}}}, \ldots\right\} \\ &= \left\{1, 1+\frac{1}{2}, 1+\frac{1}{2+\frac{1}{2}}, 1+\frac{1}{2+\frac{1}{1+\frac{1}{2}}}, \ldots\right\}\end{aligned}$$

and so on. Knowing what we know about the limit of this sequence, we should be easily convinced that

$$\sqrt{2} = 1 + \cfrac{1}{2+\cfrac{1}{2+\cfrac{1}{\ddots}}}$$

This is an example of a **continued fraction**. It is quite miraculous when you think about it: given that $\sqrt{2}$ is a classic example of an irrational number, it seems unexpected that it is the result of an infinite fraction. A similar phenomenon happens when dealing with the topograph of $x^2 - 3y^2$: we get a sequence of fractions

$$\begin{aligned}\left\{1, \frac{5}{3}, \frac{19}{11}, \frac{71}{41}, \ldots\right\} &= \left\{1, 1+\frac{2}{3}, 1+\frac{8}{11}, 1+\frac{30}{41} \ldots\right\} \\ &= \left\{1, 1+\frac{1}{1+\frac{1}{2}}, 1+\frac{1}{1+\frac{1}{1+(1+\frac{2}{3})}}, \ldots\right\}\end{aligned}$$

whereby the recursion

$$a_n = 1 + \frac{1}{1+\frac{1}{1+a_{n-1}}}$$

gives

$$\lim_{n\to\infty} a_n = \sqrt{3}.$$

Continuing this process for non-squares seems to give a relation between $\sqrt{k}$ and the quadratic form $x^2 - ky^2$...

One simple way to construct these continued fractions given a river is to simply count the number of triangles of which each fraction on the diagonal zig-zag is a vertex. For example, using the topograph of $x^2 - 2y^2$ that we have seen gives a pattern of 1 triangle, then 2, then 2, then 2, ad infinitum. Thus

$$\sqrt{2} = {\color{red}1} + \cfrac{1}{{\color{red}2}+\cfrac{1}{{\color{red}2}+\ddots}}$$

Feel free to verify this with the topograph of $x^2 - 3y^2$.

Conclusion

There are several ways that people approach the study of quadratic forms. The more traditional approach involves using abstract algebra techniques to group certain types of quadratic forms together. Instead, we have examined quadratic forms using a geometric analysis involving the Farey diagram and the topograph, highlighting connections to the Pythagorean theorem and continued fraction representations of rational square roots.

There are many more elements of quadratic forms that are worth studying. One is the equivalence of quadratic forms, as alluded to earlier. This is helpful to mathematicians because it allows them to describe a whole host of equations by a simple set of properties. There is also the link between quadratic forms and linear transformations (with matrices) that also merits a much longer discussion. One specific quadratic form mentioned previously, $x^2 - ny^2$, is called **Pell's equation** as has its own theory and connections. The possibilities are endless!

Acknowledgements

Many thanks are in order for Hazem Hassan, who was my mentor for this semester of the Directed Reading Program. Our sessions were always very stimulating, and I hope this write-up lives up to all the content we learned during the semester. Also: your effortless drawings of the Farey diagram were always impressive, though I think the three of us got close to your level in the end! Thanks also to Adam Benzaari and Bronwyn Walsh for being great peers in our DRP group, always willing to entertain my quirky suggestions and silly jokes.

The book used during this DRP was *Topology of Numbers* by Allen Hatcher, from which many pictures in this article were taken. I still wonder how he created them; if it was using LaTeX, I would be seriously impressed (I can barely get LaTeX to output text on good days). Hatcher's book was also the source of this more geometric approach to quadratic forms, originally described by the late mathematician John Conway of "Conway's Game of Life" fame.