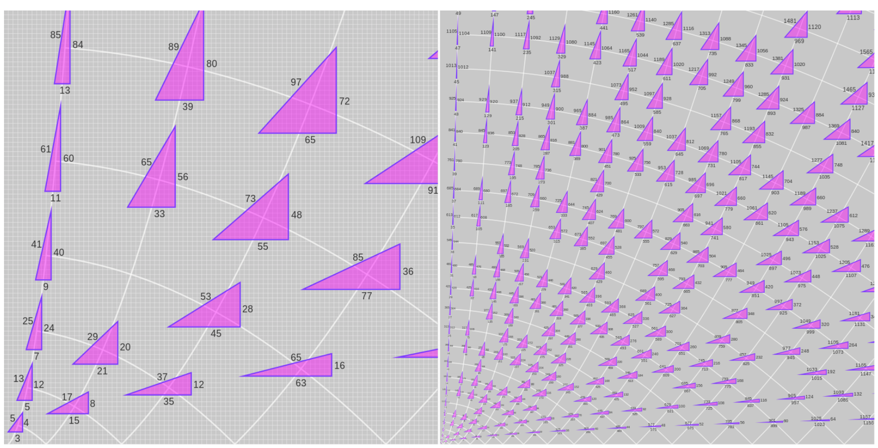
Pythagorean triples: Rational points on the circle

This chapter applies the quadratic formula in a breathtakingly beautiful way to explain a lovely formula for Pythagorean triples.

Pythagorean triples

A *Pythagorean triple* is a triple of whole numbers (a, b, c) such that $a^2 + b^2 = c^2$. It is easy to make a right triangle where the two legs are whole numbers, but it is relatively rare that the hypotenuse is also a whole number. For example, if the legs have length 1, then the hypotenuse has length $\sqrt{2}$. Because this is an irrational length, there is no whole-number scaling of this triangle whose hypotenuse would be a whole number. On the other hand, the right triangle with legs $a = 3$ and $b = 4$ has hypotenuse $c = 5$ because $9 + 16 = 25$.

There are other triples that make right triangles, like $(5, 12, 13)$ or $(8, 15, 17)$. In fact, there are *lots* more of these. Here is a picture of some.



Picture by Adam Cunningham and John Ringland, 2012
`https://en.wikipedia.org/wiki/File:PrimitivePythagoreanTriplesRev08.svg`

Two questions you probably have at this stage: Why are there gaps in the picture? What pattern do these numbers form? We will answer both of these with a beautiful application of our trusty friend the quadratic formula!!

Connection to rational points on the circle

To do understand this, first we need to understand the relationship between Pythagorean triples and points on the unit circle. This comes from the observation that every right triangle can be scaled to another right triangle with hypotenuse 1, and therefore a Pythagorean triple (a, b, c) is equivalent to the triple $(\frac{a}{c}, \frac{b}{c}, 1)$ such that $\frac{a^2}{c^2} + \frac{b^2}{c^2} = 1$. And this condition is equivalent to the point $(\frac{a}{c}, \frac{b}{c})$ lying on the unit circle $x^2 + y^2 = 1$. This is an excellent mathematical observation because it is simple but, as we shall see, allows us to truly understand Pythagorean triples in a new way.

Formula for rational points on the circle

It turns out there is a formula for rational points on the circle, and it's not a very complicated one. The formula is to choose a rational number t and take the coordinates

$$\left(\frac{t^2 - 1}{1 + t^2}, \frac{2t}{1 + t^2}\right).$$

For example, when $t = 2$, we get the point $(\frac{3}{5}, \frac{4}{5})$, and when $t = 4$ we get the point $(\frac{15}{17}, \frac{8}{17})$.

Once you've seen this formula, it's easy to check that it will always give you a point on the circle, because you can just expand the expression

$$\left(\frac{t^2 - 1}{1 + t^2}\right)^2 + \left(\frac{2t}{1 + t^2}\right)^2$$

and see that it always equals 1, for any value of t. Therefore if you plug in a rational number for t, you will get a rational point on the circle.

But this doesn't tell you at all how someone might have come up with that formula in the first place. Explaining that will use a truly beautiful idea.

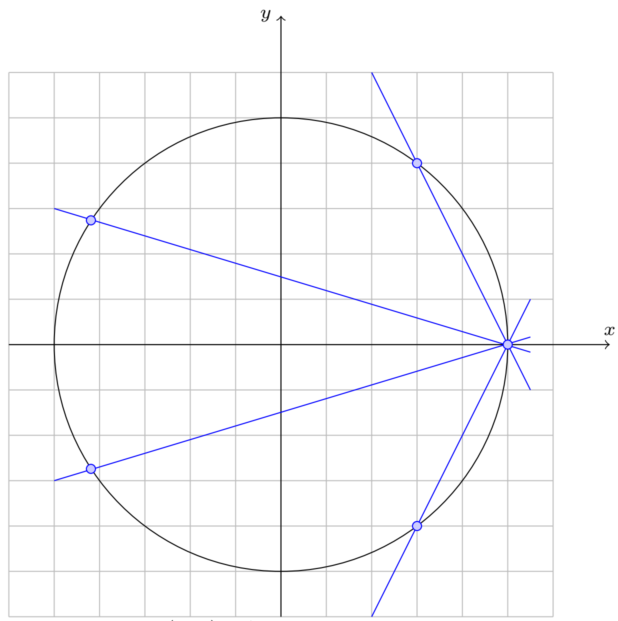
Discovering the formula: lines intersecting circles

Here is the idea: Every point (x, y) on the circle can be identified as the point of intersection between the circle and a line passing through $(1, 0)$ and (x, y). A line of slope t passing through $(1, 0)$ is given by the equation

$$y = t(x - 1)$$

and therefore the points where it intersects the unit circle are given by

$$x^2 + (tx - t)^2 = 1.$$



Line of slope t through $(1,0)$. As t varies, the line intersects the circle in different places.

Expanding, this means that

$$x^2 + t^2x^2 - 2t^2x + t^2 = 1,$$

which means

$$(1 + t^2)x^2 - (2t^2)x + (t^2 - 1) = 0.$$

Now we know how to express x in terms of these coefficients, because we know the quadratic formula! Before we do that though, let's just observe that these roots are going to be rational expressions, because we already know $x = -1$ has got to be one of them! The discriminant is

$$(2t^2)^2 - 4(1 + t^2)(t^2 - 1) = 4t^4 + 4(1 - t^4) = 4$$

and thus we have

$$x = \frac{2t^2 \pm \sqrt{(2t^2)^2 - 4(1+t^2)(t^2-1)}}{2(1+t^2)}$$

$$= \frac{2t^2 \pm 2}{2(1+t^2)}$$

$$= \frac{t^2 \pm 1}{1+t^2}$$

$$= -1 \text{ or } \frac{t^2-1}{1+t^2}.$$

The first of these corresponds to the intersection point we already knew, $(-1, 0)$, and the other one has coordinates

$$x = \frac{t^2-1}{1+t^2}$$

$$y = t(x-1) = t\left(\frac{t^2-1}{t^2+1} - \frac{t^2+1}{t^2+1}\right)$$

$$= \frac{-2t}{t^2+1}$$

If we had used the line of slope $-t$ instead of t, all of our algebra would be reflected across the x-axis, keeping the x-coordinate the same and changing the sign of the y-coordinate to the one given in the formula above!