MAT 302: LECTURE SUMMARY

The goal of today's lecture is to motivate elliptic curves. We start with a problem which was intensively studied two millennia ago by Greek mathematicians, and quite possibly even earlier by the Babylonians:

Problem. *Determine all integer solutions to* $x^2 + y^2 = z^2$.

There are some trivial solutions to this, for example $(1, 0, 1)$. More generally, we'll call any solution (x, y, z) *trivial* if $xyz = 0$. So the question becomes: what are the nontrivial solutions? A famous example is $(3, 4, 5)$; this immediately gives rise to infinitely many others, such as $(6, 8, 10)$, $(9, 12, 15)$, etc. We shall say that $(3, 4, 5)$ is a *primitive* solution, while the solutions which are multiples of it are *imprimitive*. Formally, a solution (x, y, z) is primitive iff the three numbers are relatively prime to each other. It is easy to see that any imprimitive solution is the multiple of a primitive one, so to find all integer solutions it suffices to find all the primitive ones.

Finding a nontrivial solution to $x^2 + y^2 = z^2$ is equivalent to solving

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

Thus, to solve our problem it suffices to find all *rational* solutions (X, Y) to

$$X^2 + Y^2 = 1. \tag{*}$$

Of course, it's not clear that this task is any easier! On the other hand, there are now only two variables, so that simplifies matters. Equation (*) should be very familiar: it's the equation of the unit circle. This makes it natural to try to attack the problem geometrically.

Our goal is to determine all rational points on the circle. (A point is called *rational* if both of its coordinates are rational.) There are a few obvious such points, for example, the point $(-1, 0)$; call this point $\mathcal{O}$. Pick any point $\mathcal{P} \neq \mathcal{O}$ on the circle, and consider the line passing through $\mathcal{P}$ and $\mathcal{O}$. This line must intersect the y-axis somewhere; this point is called the *projection* of $\mathcal{P}$ from $\mathcal{O}$ onto the y-axis. (Think of a light source at $\mathcal{O}$, and the projection as the location of $\mathcal{P}$'s shadow. Of course, this is more strictly accurate for points in the second quadrant than in the first, but it's still a helpful metaphor.) Note that projection from $\mathcal{O}$ is a bijection between the y-axis and the points of the circle other than $\mathcal{O}$. (Why is this?)

One advantage of projecting the circle onto the y-axis is that lines are easier to understand than circles. Moreover, a point on the circle is rational if and only if its projection is. (I'll give a proof of this below, but you should try to prove it on your own now.) Since the rational points on the y-axis are in bijection with $\mathbb{Q}$, we see that projection from $\mathcal{O}$ gives a bijection between $\mathbb{Q}$ and the rational points on the unit circle (other than $\mathcal{O}$ itself). If we can determine this bijection explicitly, we will be able to generate all rational points on the unit circle, and thus solve our initial problem.

Date: March 29th, 2011.

With this strategy in mind, we prove

Theorem 1. *A point $\mathcal{P}$ on the unit circle is rational if and only if its projection (from $\mathcal{O} = (-1, 0)$ onto the y-axis) is rational.*

Proof. Suppose $(0, t)$ is the projection of (x, y) from $\mathcal{O}$. In other words, t is the y-intercept of the line through $\mathcal{O}$ and (x, y). The line passing through $\mathcal{O}$ and (x, y) is the set of points

$$\Big\{\Big((x+1)\alpha - 1,\ y\alpha\Big) : \alpha \in \mathbb{R}\Big\}.$$

The y-intercept occurs when the x-coordinate is 0, i.e. when $\alpha = \frac{1}{x+1}$. Thus, for $(0, t)$ to be in the set we must have

(♣) $$t = \frac{y}{x+1}$$

from which we immediately deduce that if (x, y) is a rational point (other than $\mathcal{O}$) on the circle, its projection is also rational. Now we prove the converse of this statement, by finding explicit formulae for x and y in terms of t.

Since (x, y) is a point on the unit circle, $x^2 + y^2 = 1$. From (♣) we know that $y = (x+1)t$. Substituting and expanding gives

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

$x = -1$ is a root of this equation (can you see why this is *without* plugging -1 in?). Factoring the left hand side gives

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

We therefore find that

(†) $$x = \frac{1-t^2}{1+t^2}$$

which by (♣) implies

(‡) $$y = \frac{2t}{1+t^2}.$$

It follows that if $t \in \mathbb{Q}$, then x and y are rational. □

The equations (†) and (‡) give an explicit bijection from $\mathbb{Q}$ to the rational points on the circle. Writing $t = a/b$ with a and b integers, we conclude that *every* rational point on the unit circle can be written in the form

$$\left(\frac{b^2 - a^2}{b^2 + a^2},\ \frac{2ab}{b^2 + a^2}\right).$$

Thus, any nontrivial integer solution (x, y, z) to

$$x^2 + y^2 = z^2$$

can be written in the form $(b^2 - a^2, 2ab, b^2 + a^2)$ for some integers a and b. Conversely, any triple of this form gives a solution to the equation. We have therefore found a complete set of nontrivial

solutions to the equation. Not all of these are primitive, however. It is a good exercise to prove that the set of all nontrivial primitive solutions (up to permuting x and y) is the set

$$\Big\{\big(b^2-a^2, 2ab, b^2+a^2\big) : (a,b)=1\Big\}.$$

Actually, there was one statement I made above which was not entirely correct: (†) and (‡) don't quite give a bijection between $\mathbb{Q}$ and the rational points on the unit circle. Why not? Because $\mathcal{O}$ is missing! Although this doesn't affect our conclusions (the point $(-1,0)$ corresponds to a trivial solution of $x^2+y^2=z^2$) it's still a bit curious. Note that taking $t\to\infty$ in (†) and (‡) does yield the special point $\mathcal{O}$. Thus, one might be tempted to formally adjoin ∞ to $\mathbb{Q}$, and say there is a bijection between $\mathbb{Q}\cup\{\infty\}$ and the rational points on the unit circle.

There's a more geometric way to view this new infinite element. Recall the construction of the 'projection from $\mathcal{O}$' map: given any point $\mathcal{P}$ on the unit circle, draw the line going through both $\mathcal{P}$ and $\mathcal{O}$ and find its point of intersection with the y-axis: this point is the projection of $\mathcal{P}$. It is not clear how to carry out this procedure when $\mathcal{P}=\mathcal{O}$. Inspired by calculus, we take a sequence of points $\mathcal{P}$ on the circle which get closer and closer to $\mathcal{O}$. As this happens, the line containing both $\mathcal{P}$ and $\mathcal{O}$ becomes more and more vertical. It is therefore natural to say the line connecting $\mathcal{O}$ to itself is the vertical line through $\mathcal{O}$, i.e. the line tangent to the circle at $\mathcal{O}$. Of course, this line doesn't intersect the y-axis anywhere. However, we can imagine a 'point at infinity' where these two parallel lines *do* intersect. This is not as foreign as it seems at first. Think of longitudinal lines on the surface of the earth – they all run directly north-south, so they're all parallel, and yet they all intersect at the North and South poles. For a beautiful illustration of a closely related idea, look up the video 'Moebius transformations revealed' on youtube.

Now that we've solved the initial problem, we can look at a harder one: find all integer solutions (x,y,z) to $x^3+y^3=z^3$. As before, there are many trivial solutions in which one of the variables is 0. What are the nontrivial solutions? The same game as above (projecting points of the curve onto the y-axis) no longer works, because rational points on the curve don't necessarily project onto rational points on the y-axis. After a bit of playing around, you might find that making the substitutions $r=\frac{12}{x+y}$ and $s=36\left(\frac{x-y}{x+y}\right)$ transforms the original equation into

$$s^2=r^3-432.$$

Conversely, letting $x=\frac{36+s}{6r}$ and $y=\frac{36-s}{6r}$ gets us back to the original equation $x^3+y^3=1$. Thus, finding rational points on $x^3+y^3=1$ is equivalent to finding rational points on $s^2=r^3-432$. It's not obvious how to tackle this problem either, but it turns out that quite a lot can be said. Equations of the form $y^2=x^3+ax+b$ are called *elliptic curves*, and it is an important open problem to understand the set of rational points on such curves. We will discuss this topic, as well as the implications for cryptography, in the next lecture.