MATH 115, SUMMER 2012
LECTURE 26

JAMES MCIVOR

In the next few lectures we will try to use geometric ideas to get number theory results. The motivating example is our first proof of the Pythagorean triples theorem.

We took the (homogeneous) equation $x^2 + y^2 = z^2$ and divided through by z^2 to get the (inhomogeneous) equation $X^2 + Y^2 = 1$. The set of pairs (x, y) of real numbers satisfying this equation describes a curve (a circle) in the real plane $\mathbb{R}^2$.

We were interested in rational values of x, y satisfying this equation, since those give us back the integer Pythagorean triples. To find all rational points (x, y) on the circle, we first picked one, namely $(-1, 0)$, and looked at all lines with rational slope through this point. Each of them intersects the circle in one other point, and we obtain all rational points this way.

What are the key issues in this argument?

(1) (Essential) There is at least one rational point on the curve $X^2 + Y^2 = 1$.
(2) (Very important) For each line with rational slope, it passes through the circle in exactly one other point.
(3) (Not so important) The slope at this point is vertical (so we didn't double count any rational points)

We want to try to generalize this type of argument, and need to make sure we understand when these conditions are satisfied.

1. Plane Curve

Definition 1. A **plane curve of degree** d is a set of points (x, y) in $\mathbb{R}^2$ satisfying some polynomial[1] equation $f(x, y) = 0$.

Examples 1.1. (1) The circle above.
(2) The cubic equation $y^2 = x^3 - x$ - this is an example of an elliptic curve, which we will study more closely in the next few lectures.
(3) The equation $x^2 - y^2 = 0$ defines an "X" shape - the union of the two lines $y = x$ and $y = -x$. We call this a curve of degree 2, even though it's the union of two lines.
(4) The equation $(y - x)^3 = 0$ defines a "tripled line" it looks like the line $y = x$, but each point has multiplicity three, because of the exponent 3.

As you see from the last two examples, the degree should be determined from the equation, and not from the shape of the curve geometrically.

[1] Some people would call this an *algebraic* plane curve, to emphasize the fact that f is a polynomial, as opposed to, say, an exponential or trigonometric function.

Definition 2. If f is a polynomial, $C_f(\mathbb{R})$ denotes the set of real points on the curve, that is

$$C_f(\mathbb{R}) = \{(x, y) \mid f(x, y) = 0, x, y \in \mathbb{R}\}$$

This is exactly the curve itself. Similarly $C_f(\mathbb{Q})$ is the set of **rational points** on the curve, that is

$$C_f(\mathbb{Q}) = \{(x, y) \mid f(x, y) = 0, x, y \in \mathbb{Q}\}$$

Clearly $C_f(\mathbb{Q}) \subseteq C_f(\mathbb{R})$, but note that $C_f(\mathbb{Q})$ may be empty - there are curves with no rational points. Equivalently, there are equations with no rational solutions, for example the double line $x^2 = 2$.

In fact, even the set of real points may be empty! For example, the equation $x^2 + y^2 + 1 = 0$ has no solutions in real numbers. In such cases (actually, in *all* cases) it is useful to consider also the set of **complex points** on the curve, which is

$$C_f(\mathbb{C}) = \{(x, y) \mid f(x, y) = 0, x, y, \in \mathbb{C}\}$$

The problem is that the set of complex points is almost always impossible to draw, so we mostly stick to real numbers in this course. Nevertheless, we have $C_f(\mathbb{Q}) \subseteq C_f(\mathbb{R}) \subseteq C_f(\mathbb{C})$.

2. Curves of Degree 2

A curve of degree 2 is called a **conic**. These are just the solutions to quadratic equations in two variables. If the equation $f(x, y) = 0$ defining the conic factors into two linear factors it is **degenerate** (this is when the conic is actually just the union of two lines); otherwise we call it **nondegenerate**.

We argued in the Pythagorean triples lecture that given one rational point P on a circle, we could produce infinitely many more by looking at lines through P with rational slope and seeing where they intersect the circle. For every such line except those lines tangent to P, the line met the circle in exactly one other point.

We also made a similar argument for an ellipse. We'd like to say that the same argument will go through for any conic. Namely, we'd like to assert the following:

"Almost True" Statement: Every line meets a conic in two (not necessarily distinct) points.

This is basically true, but there are two problems which may occur and prevent us from making this general of a claim.

Problems:

(1) The conic is degenerate. Example: $xy = 0$, the union of two lines; or $(y - x)^2 = 0$, a doubled line.
(2) The second point of intersection is at infinity. Example: The conic C, and a vertical line through the origin.

- draw pictures of the examples.
- we can bypass the second problem for the moment by considering only conics of the form $ax^2 + by^2 = c$.
- in this case, an argument just as in the Pythagorean Triples lecture shows:

Proposition 1. *If the conic $ax^2 + by^2 = c$ has any rational points, then it has infinitely many.*

Of course, such a conic may have no rational points.

Examples 2.1. (1) The conic $x^2 + y^2 + 1 =$ has no rational points, because it doesn't even have any *real* points!

(2) The conic $x^2 + y^2 = 3$ has no rational points, as can be seen by supposing there is a rational solution, and then reducing mod 3.

In fact, the above two examples are the only ways there can fail to be rational points, according to the following incredible theorem:

Theorem 1. *(Hasse's Local-Global Principle) If a conic $f(x, y) = ax^2+bxy+cy^2 = 0$ has a real solution and also has solutions mod p for every prime p, then it has a solution in rational numbers.*

The solutions in real numbers and mod p are called **local** solutions, while the solutions in rational numbers are called **global** solutions.

We won't prove this, but we may occasionally use the following much simpler criterion:

"No-Solution-Test": If an equation has no solution mod p for some prime p, then it has no solution in integers.

This is an easier statement because the conclusion is that there is no solution in integers, not necessarily rationals.

You may want to compare this with the two ways that a system of linear equations can fail to have a solution from lecture 23 - 1) a "rank problem" (no real solutions), or 2) a "divisibility problem" (no solutions mod p for some p). Thus there is a "local-global" principle for linear systems of equations, too, just like the one above for certain types of conics.

3. Curves of Degree Greater than 2

For the most part we understand rational points on conics - if there's one, there are infinitely many! The only problem we haven't addressed is the example of $y = x^2$, and the vertical line through the origin. We'll understand this better once we've talked about the projective plane.

First let's look at curves of higher degree.

Question: If we pick a point on our curve (rational or not) and take a line through that point, where will the line meet the curve (besides in the given point)?

The answer is the following theorem:

Theorem 2. *If C is a curve of degree d and L is a line, then $C \cap L$ either consists of at most d points, or else every point of L lies on the curve C, in which case we call L a* **component** *of C.*

Proof. Let C be given by the equation $f(x, y) = 0$, where f is a polynomial of two variables x, y of degree d. First assume the line L is not vertical (which corresponds to slope $= \infty$). Then L has an equation of the form $y = mx + b$ for some $m, b \in \mathbb{R}$. Points (x, y) satisfying this equation and the equation for C correspond to values of x satisfying

$$f(x, mx + b) = 0,$$

which is just a polynomial in x of degree d. The fundamental theorem of algebra[2] says that this equation has exactly[3] d solutions if x is allowed to be complex. Now some or all of these complex numbers may actually be real, but in any case, if we look at only the real number solutions, there are at most d of them.

$\square$

[2] If you haven't seen this theorem, don't worry.

[3] Some of the solutions may be repeated, for example $x^2 = 0$ has the "double root" $x = 0$.