1.2 26/06/16 - Introduction II: Pythagorean Triples and Fermat for Exponent 3

Let us now study some special cases of the *Fermat problem*, the statement of which is colloquially know as *Fermat's Last Theorem*. Recall that this says that if $n \in \mathbb{Z}_{\geq 3}$ and $X, Y, Z \in \mathbb{Z}$ are such that $X^n + Y^n = Z^n$, then the solution (X, Y, Z) is trivial, i.e., $XYZ = 0$. We will discuss and prove the cases $n = 3$ and $n = 4$.

1.2.A Pythagorean Triples

Before we do that however, it is helpful to review the case $n = 2$ of Pythagorean triples. Here we are trying to solve $X^2 + Y^2 = Z^2$ in the integers. The first thing we note is that this is a *homogeneous equation*; this means that if (X, Y, Z) is a solution, then so is $(\lambda X, \lambda Y, \lambda Z)$ for any λ. This is very helpful, and allows some geometric and arithmetic simplifications. We will develop the language of projective curves in the next lecture to deal with the geometric situation more generally, and we will deal with the arithmetic implications below, but for now it suffices to note that if $Z \neq 0$ (the solutions with $Z = 0$ are easy to understand!), then we can divide throughout by Z to get that $(x, y) := (X/Z, Y/Z)$ satisfies

$$x^2 + y^2 = 1.$$

This curve is familiar to us as the circle C. Thus, Pythagorean triples give rise to points on the circle. (This is a fundamental insight going back even to Descartes!) What is more, if X, Y, Z are integers, then x and y are rational numbers; therefore, if we have a good understanding of rational points on the circle $x^2 + y^2 = 1$, then we will have a way to find all Pythagorean triples. For instance, if we could parametrize all points on the circle, then we can hope to parametrize all Pythagorean triples.

Now the circle is a curve of degree two, i.e., a conic. This has the important consequence that a general line meets it in *two points* (some are tangent, and some might need you to allow complex numbers). This leads us to the following simple strategy: if we pick any rational point P on C and project from it to a general line L, then we will get (essentially) a bijective correspondence between points on this line and on the circle. (Make sure you believe this!) Since it is very easy to parametrize points on a line, this gives us a way of parametrizing points on the conic curve. Let's see this work out in practice.

Take the point P to be $(-1, 0)$ and the line L to be the x-axis. Given a point (x_0, y_0) on the circle C, the line joining P and (x_0, y_0) meets the line L in the point $(0, t)$ where $t = y_0/(x_0 + 1)$. Note that this does not work when $x_0 = -1$ corresponding to P itself, when the the limiting line is the tangent line $x = -1$, which does not meet L–in the affine plane. We'll figure out how to fix this next time.

Going the other way, we can parametrize points on L as $(0, t)$ for varying t. The line joining $(-1, 0)$ and $(0, t)$ has equation

$$y = tx + t,$$

which we would like to intersect with the circle. Plugging this in to the equation for the cirlce gives us the quadratic equation

$$x^2 + (tx + t)^2 = 1,$$

which factors as

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

One root is the expected $x = -1$; it's the other one which we are after, which has $x = (1 - t^2)/(1 + t^2)$. Using $y = tx + t$, this gives us the other point of intersection

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

(Those who have seen a little trigonometry might recognize this is as the formula for $(\cos\theta, \sin\theta)$ in terms of $\tan(\theta/2)$; why is that?). These operations are visibly inverses to each other and the resulting maps

$$(x, y) \mapsto y/(x + 1) \text{ and } t \mapsto \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2}\right)$$

are almost inverses to each other (check!). The only point that gives us some trouble is the point $P = (-1, 0)$ itself, which should somehow correspond to "$t = \infty$". Now, the above discussion works over any field (but see 1.2.1). In particular, if t is a rational number, then so is (x, y), and conversely. In this way, we have completely figured out all rational points on C.

Remark 1.2.1. So far we haven't really used many properties of the rational numbers. The same algebra can be carried out over more or less any field K (even if it's not easy visualize what that might mean), with some important caveats. Here are two things to think about:

(a) What (if anything) goes wrong when you work with a field of characteristic two such as $K = \mathbb{F}_2$ instead?
(b) If we work over the field $K = \mathbb{C}$, what happens to the values $t = \pm i$ of the parameter t?

Remark 1.2.2. Note that we could have "dehomogenized with respect to other variables" too. For instance, if $Y \neq 0$, then we may set $(s, t) := (X/Y, Z/Y)$ instead to get instead the equation $s^2 + 1 = t^2$, or equivalently $t^2 - s^2 = 1$. This would have led to a parametrization similar to the one above. This additional symmetry–while not useful here–can be extremely useful in other contexts; we will study this systematically next time.

Let us see what consequence this has for the arithmetic. Suppose we are interested in finding integer solutions X, Y, Z to $X^2 + Y^2 = Z^2$ with $X, Y, Z \neq 0$. We may assume by flipping signs if needed that $X, Y, Z > 0$. We may also divide throughout by the greatest common divisor of all X, Y, Z to assume that X, Y, Z have no common factors as a *triple*. This has the consequence that X, Y, Z have no common factors *pairwise*: if a prime p divides two of X, Y, Z, then the equation $X^2 + Y^2 = Z^2$ forces it to divide the third. (Therefore, things like $(6, 10, 15)$ will never work.) With these simplifications, we can show

Theorem 1.2.3. Let X, Y, Z be pairwise coprime positive integers such that $X^2 + Y^2 = Z^2$. Then there are unique coprime $m, n \in \mathbb{Z}$ of different parity with $m > n > 0$ such that exactly one of (X, Y, Z) or (Y, X, Z) is

$$(m^2 - n^2, 2mn, m^2 + n^2).$$

Proof. The point $(x, y) = (X/Z, Y/Z)$ lies on the circle $C : x^2 + y^2 = 1$ and is different from $(-1, 0)$, so the above discussion tells us that there is a unique $t \in \mathbb{Q}$ such that

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

Since x and y are positive, we conclude that $0 < t < 1$ (check!). Write $t = n/m$ for coprime positive $m, n \in \mathbb{Z}$ such that $m > n > 0$, so that

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

We claim that the fractions on the right are in their lowest terms (up to small factors of two; see below). Why is this?

Let's analyze the parities of m and n. Since they are coprime, they can't both be even. If they have different parity, then $m^2 \pm n^2$ are odd, and this forces

$$\gcd(m^2 + n^2, m^2 - n^2) = \gcd(2mn, m^2 + n^2) = 1.$$

i.e., these fractions are in lowest form (check!). This proves that (X, Y, Z) is the of the form desired.

Finally, can they both be odd? If this is the case, then the integers $m_1 = (m + n)/2$ and $n_1 := (m - n)/2$ are coprime, satisfy $m_1 > n_1 > 0$, have different parity (check!), and further we have that

$$\left(\frac{m^2 - n^2}{m^2 + n^2}, \frac{2mn}{m^2 + n^2}\right) = \left(\frac{2m_1 n_1}{m_1^2 + n_1^2}, \frac{m_1^2 - n_1^2}{m_1^2 + n_1^2}\right),$$

and so the argument from the previous paragraph implies $(Y, X, Z) = (m_1^2 - n_1^2, 2m_1 n_1, m_1^2 + n_1^2)$, finishing the proof. ∎

Exercise 1.2.4. Fill in the details in the above proof: show that if $m, n \in \mathbb{Z}$ are coprime and of different parity, then $\gcd(m^2+n^2, m^2-n^2) = \gcd(2mn, m^2+n^2) = 1$, and show that if $m > n > 0$ are both odd, then $m_1 = (m+n)/2$ and $n_1 := (m-n)/2$ are coprime and have different parity.

In the above study, we didn't really use any properties of the curve C except that it is of degree two–and that it had a rational point P to get our method started. Therefore, the same argument applies to any conic C with a rational point $P \in C(\mathbb{Q})$. This allows us to give complete parametrizations of all rational points of such objects; see, for instance, 2.1.6(a). (Get a hold on the integer points this way might be a little harder; see 2.1.6(b).)

Clearly, then we would have a complete understanding of rational solutions to all conic sections C if we could just prove that they all contained some rational point P. Sadly, this is not the case (2.1.4(a)). How can we figure out when a given conic has a rational point, and when it doesn't?

Example 1.2.5. Let's look at a few examples.

(a) Does the curve
$$4x^2 + 12xy + 9y^2 + 1 = 0$$
have any rational points? No. This is because it can be written as $(2x+3y)^2+1 = 0$, which doesn't have any real solutions (x, y), let alone rational ones. We say such curves are *locally obstructed at* ∞ or *locally obstructed over* $\mathbb{R}$.

(b) Does the curve
$$x^2 + xy + y^2 = 2$$
have any rational points? No, but this is not because it doesn't have any real points (e.g., $(\sqrt{2}, 0)$ works). This is more subtle. Suppose there is a rational solution (x, y). Write $x := X/Z$ and $y := Y/Z'$ for integers $X, Y, Z, Z' \in \mathbb{Z}$ with $Z, Z' > 0$ and $\gcd(X, Z) = \gcd(Y, Z') = 1$. Plugging this into the equation and multiplying through by $Z^2(Z')^2$ gives us
$$X^2(Z')^2 + XYZZ' + Y^2Z^2 = 2Z^2(Z')^2.$$
This equation implies that $Z \mid (Z')^2$ and so by symmetry that $Z' \mid Z^2$. This is enough to conclude that Z and Z' have the same prime factors, but this is not enough to conclude that $Z = Z'$ (consider $2^2 3^3$ vs. $2^3 3^2$). We need to be a little more careful. For this, let p be a prime dividing both Z and Z', and let $\alpha := v_p(Z)$ (resp. $\alpha' := v_p(Z')$), so $\alpha, \alpha' \geq 1$. We have that $v_p(X) = v_p(Y) = 0$, that $\alpha \leq 2\alpha'$ and $\alpha' \leq 2\alpha$. We have to show that $\alpha = \alpha'$. Suppose to the contrary that $\alpha \neq \alpha'$; we may then assume without loss of generality that $\alpha' > \alpha$ (the other case being symmetrical). In this case, the p-adic valuation of the terms on the left is $2\alpha', \alpha+\alpha'$, and 2α, which forces the sum to have p-adic valuation 2α. The right hand side has v_p equal to either $2\alpha + 2\alpha$ if $p \neq 2$, and $2\alpha + 2\alpha' + 1$ if $p = 2$, and we easily see that none of these options work. Thus, we must have $\alpha = \alpha'$. Since this is true for all p and $Z, Z' > 0$, we conclude that $Z' = Z$, and so we can remove it to get the homogeneous equation
$$X^2 + XY + Y^2 = 2Z^2 \tag{1.2.6}$$
we were hoping for. Now we are almost done. Let's consider the parities of X and Y. Well, it cannot be that both of them are odd or that precisely one is odd, because then the left side of 1.2.6 is odd. Therefore, both of them must be even, say $X = 2X_1$ and $Y = 2Y_1$. This then gives us the equation $2(X_1^2 + X_1Y_1 + Y_1^2) = Z^2$, which forces Z to be even as well–but $\gcd(X, Z) = 1$, and this is a contradiction. This proves that there couldn't have been any solutions to begin with.
We say that the curve $x^2 + xy + y^2 = 2$ is *locally obstructed at 2*. Note that this does not mean it has no solutions mod 2 (i.e., that $C(\mathbb{F}_2) = \emptyset$, which is false); rather, it means that it has no *2-adic solutions*, i.e., $C(\mathbb{Q}_2) = \emptyset$. (Don't worry if you don't know yet what that means.)

Similarly, curves can be locally obstructed at other primes; see 2.1.7

For right now, you can take local obstruction at a prime $p > 0$ to mean that "the arithmetic doesn't work out at p". Clearly, if there is a local obstruction at any prime (including the "prime" $p = \infty$ corresponding to $\mathbb{Q}_\infty = \mathbb{R}$), then there is no rational solution. But the converse is not that clear. Can we have a curve that has solutions *everywhere locally*, but none globally?

It turns out that this "failure of the local-to-global principle" cannot happen for conics, but it can happen for cubics. The fact that it cannot happen to conics is known as the Hasse-Minkowski

Theorem. This holds more generally not just for conic curves, but for quadratic forms in any number of variables. Proving this would take us a little far afield, so I will leave the simplest case of quadratic equations as an exercise (2.1.10), and refer the reader to [3, Ch. IV, Thm. 8] for the general proof.

The really interesting thing I want to draw your attention to, however, is that this result fails in general for curves of degree ≥ 3: there are cubic curves which have solutions in $\mathbb{Q}_p$ for all $p \leq \infty$, but no solutions in $\mathbb{Q}$. One of the simplest examples of this sort is the Selmer curve

$$3X^3 + 4Y^3 + 5Z^3 = 0.$$

(This is the homogeneous equation; the inhomogeneous equation is $3x^3 + 4y^3 = -5$.) We will revisit this curve later in the course, and develop tools to study such "obstructions to the Hasse-Minkowski principle". For now, let us turn to

1.2.B Fermat for Exponent 3

The next case of Fermat's problem is $n = 3$, where we are solving $X^3 + Y^3 = Z^3$. Again, this is a homogeneous equation. It is now believed that Fermat did not have a proof for this case, and that the first proof was given by Euler in 1770 (although see the discussion on the Wikipedia page). Euler's proof used the arithmetic of $\mathbb{Z}[\omega]$, where $\omega = (-1 + \sqrt{-3})/2$ is a primitive cube root of unity.

We'll pursue a totally different idea. The idea is that the curve $X^3 + Y^3 = Z^3$ is a cubic plane curve (again, given the understanding of (de)homogenization), and has a rational inflection point; this allows us to put it in *Weierstrass form*. More generally, we can consider for any nonzero α the equation

$$X^3 + Y^3 = \alpha Z^3.$$

This equation has rational point $[1 : -1 : 0]$. Let's move this a little bit, specifically to $[0 : 1 : 0]$: consider the "change of coordinates" given by

$$(X_1, Y_1, Z_1) = (12\alpha Z, 36\alpha(X - Y), X + Y). \tag{1.2.7}$$

Exercise 1.2.8. Figure out how to invert this change of coordinates: express (X, Y, Z) in terms of (X_1, Y_1, Z_1).

We will see below why this change of coordinates is a good one; we will explain later where it comes from. Given this "inverse change of coordinates", we note that the curve above transforms to the one of the form

$$Y_1^2 Z_1 = X_1^3 - 432\alpha^2 Z_1^3,$$

which after dehomogenezing (dividing throughout by Z_1^3) is the curve

$$y^2 = x^3 - 432\alpha^2.$$

This is nothing but our example from 1.1.6, with $c = -432\alpha^2$, and we are back to finding rational points on this cubic curve. The Fermat case corresponds to $\alpha = 1$, and hence $c = -432$. (This might explain whey $c = -432$ is special in 1.1.6.) This is again what we call an *elliptic curve*. We will show later using techniques for elliptic curves that the only points on here are the points $(12, \pm 36)$.

Exercise 1.2.9. Which points on $X^3 + Y^3 = Z^3$ do the points $(12, \pm 36)$ on $y^2 = x^3 - 432$ correspond to?

This finishes the proof of Fermat's last theorem modulo the claim made about the curve $y^2 = x^3 - 432$, which we will prove later. I also still need to explain where this change of coordinates comes from; that will be done sooner. (Hopefully this motivates you to keep coming to this course!) We'll deal with the case $n = 4$ next time.