3. THIRD LECTURE : 3/11

Definition A *Diophantine equation* is an equation of the form

$$p(x_1, ..., x_n) = 0, \tag{3.1}$$

where $p(\mathbf{x}) \in \mathbb{Z}[x_1, ..., x_n]$ is a polynomial with integer coefficients. By a *solution* to (3.1) we mean an integer solution, i.e.: one for which all the x_i have integer values.

The study of Diophantine equations is one of the major themes in the history of number theory. It is obviously a broad topic, so it's important to ask the right questions in order to get to interesting results. It is my intention in this course only to scratch the surface of the theory. Let me in passing, however, mention a few theorems which are deserving of the title 'great'[1] It is far beyond our remit in this course to prove any of these.

Theorem 3.1. *The question of whether an arbitrary Diophantine equation has a solution or not is* undecidable, *i.e.: there can't be found any algorithm which takes an arbitrary Diophantine equation as input and decides whether the equation has a solution or not after a finite amount of computation.*

This is really a result in mathematical logic so, once again, if you haven't seen results like this before it will probably strike you as weird. The most famous theorem of this sort concerns the so-called *Halting Problem* studied by Alan Turing. The question of decidability for solutions to Diophantine equations is known as *Hilbert's Tenth Problem*. It was definitively solved by Matyasevich in 1970, but his work built on that of several previous authors and has the character of a 'final piece in the jigsaw'. One of the other major contributors was Julia Robinson, probably the foremost female mathematician of the 20th century.

Theorem 3.2. (Faltings 1983) *Let C be a non-singular rational curve of genus at least 2. Then there are only finitely many rational points on C.*

This is a technical formulation of Faltings' Theorem, but a more concrete way to think about it is that it implies that if $p(x, y, z)$ is a homogeneous, irreducible polynomial of degree at least 4, then the Diophantine equation $p(x, y, z) = 0$ has only finitely many *primitive* solutions, i.e.: solutions with GCD$(x, y, z) = 1$. The proof of this theorem uses heavy machinery from algebraic geometry. Indeed, the modern study of Diophantine equations is dominated by algebro-geometric methods, and is often highly sophisticated. The final theorem I wish to quote, probably the most famous theorem in all of math, exemplifies this state of affairs extremely well :

Theorem 3.3. (Wiles 1994) *The Diophantine equation $x^n + y^n = z^n$ has no solutions for which $xyz \neq 0$ when $n > 2$.*

Now back to the course material proper. When it comes to Diophantine equations, one first has to figure out a good place to start one's study. A sensible choice would seem to be to start (just as one does in ordinary one-variable algebra) with linear equations, then move on to quadratics, cubics etc. It turns out that things already get extremely

[1] There are a lot of theorems which could be given the title of 'great' and no exhaustive list is remotely possible.

hard with cubics (again, maybe no surprise, given our experience with one-variable cubic equations $p(x) = 0$[2]). Quadratic equations are at just the right level of difficulty for 'serious' research, and there is a very rich theory for these, which we will once again scratch the surface of in due course. First of all, though, we must take care of linear equations. Here there is a satisfactory theory, though as we will see, there is a slight sting in the tail.

Theorem 3.4. *Let $a_0, a_1, ..., a_n$ be non-zero integers. Then the Diophantine equation*

$$a_1x_1 + \cdots + a_nx_n = a_0 \tag{3.2}$$

has a solution if and only if GCD$(a_1, ..., a_n)$ divides a_0.

Proof. The case $n = 1$ is obvious : the equation $a_1x_1 = a_0$ has the unique solution $x_1 = a_0/a_1$. For the case $n = 2$, i.e.: for the equation

$$a_1x_1 + a_2x_2 = a_0, \tag{3.3}$$

we use Euclid's lemma. Let $d := \text{GCD}(a_1, a_2)$. According to that lemma, there exist integers u_1, u_2 such that

$$a_1u_1 + a_2u_2 = d, \tag{3.4}$$

and if $0 < a_0 < d$ then there are no integer solutions to (3.3). Suppose $d|a_0$, say $a_0 = qd$. Then multiplying (3.4) through by q we have an integer solution $x_1 = qu_1$, $x_2 = qu_2$ to (3.3). On the other hand, suppose $d \dagger a_0$. Then $a_0 = qd+r$ where $0 < r < d$. Suppose there were an integer solution to (3.3), say

$$a_1v_1 + a_2v_2 = a_0. \tag{3.5}$$

Then multiplying (3.4) through by q and subtracting from (3.5) we obtain

$$a_1s_1 + a_2s_2 = r, \quad \text{where } s_1 = v_1 - qu_1, \, s_2 = v_2 - qu_2, \tag{3.6}$$

which contradicts the fact that (3.3) has no integer solution when $0 < r < d$.

This establishes the case $n = 2$ of the theorem. The general case is now obtained by induction on n (see exercise 1 on homework 1). □

Significantly, Euclid's algorithm provides an effective form of this theorem, i.e.: we can efficiently find an explcit solution to (3.3) when the conditions of Theorem 3.4 are fulfilled. We have already seen how this works when $n = 2$. For general n, one uses an inductive procedure (again see exercise 1 on homework 1).

The final part of the jigsaw which yields a highly satisfactory theory of linear equations is that it is possible to write down a formula for ALL solutions to such an equation in terms of any particular solution (which Euclid's algorithm can find). We give the formula for $n = 2$ only. As n increases, it will be more complicated to write down but can be done in principle by the same inductive reasoning (see homework 1).

[2] Note, though, that the problem of determining all RATIONAL solutions to a one-variable equation $p(x) = 0$, where $p(x) \in \mathbb{Z}[x]$, can be reduced to the integer factorisation problem. For if $p(x) = a_nx^n + \cdots + a_1x + a_0$ and $x = p/q$ is a root, then it is easy to see that we must have $q|a_n$ and $p|a_0$. In particular, if $x = p$ is an integer root, then $p|a_0$. So to find all integer roots, it suffices to factor a_0 and test its factors one-by-one.

Theorem 3.5. *Let a_0, a_1, a_2 be non-zero integers such that GCD(a_1, a_2) divides a_0. Let $d :=$ GCD(a_1, a_2), $a_0 := qd$, and let u_1, u_2 be any integers satisfying $a_1u_1 + a_2u_2 = d$. Then the general solution to (3.3) is given by*

$$x_1 = qu_1 + k\left(\frac{a_2}{d}\right), \quad x_2 = qu_2 - k\left(\frac{a_1}{d}\right), \quad k \in \mathbb{Z}. \tag{3.7}$$

Proof. It is simple to check that if (x_1, x_2) is given by (3.7) then (3.3) is satisfied. Conversely, suppose $a_1x_1 + a_2x_2 = a_0$. We also have $a_1(qu_1) + a_2(qu_2) = a_0$. Subtracting we obtain

$$a_1(x_1 - qu_1) = a_2(qu_2 - x_2). \tag{3.8}$$

Divide through by d to get

$$\left(\frac{a_1}{d}\right)(x_1 - qu_1) = \left(\frac{a_2}{d}\right)(qu_2 - x_2). \tag{3.9}$$

Now the point is that GCD$(a_1/d, a_2/d) = 1$. The two sides of (3.9) must have exactly the same prime factors (we're sort of using FTA here). Hence, all the prime factors of a_2/d, say, must appear among those of $x_1 - qu_1$, in other words, a_2/d must divide $x_1 - qu_1$. Similarly, a_1/d must divide $qu_2 - x_2$. Let $x_1 - qu_1 = k\left(\frac{a_2}{d}\right)$ and $qu_2 - x_2 = l\left(\frac{a_1}{d}\right)$. Substituting back into (3.9) yields $k = -l$, and thus (x_1, x_2) satisfy (3.7), as required. □

We seem to have a completely satisfactory theory for linear Diophantine equations. We have an explicit criterion for whether or not a solution exists, an explicit formula for all solutions when they do exist, and an efficient algorithm for testing whether the criterion is satisfied, and for finding an explicit solution when it is.

Now for the little sting in the tail. Suppose all the coefficients in (3.2) are positive and we are only interested in positive solutions. This is a fairly natural restriction for problems involving counting of some sort. It is known as the *Frobenius coin problem* : think of $a_1, ..., a_n$ as being coin denominations which one has at one's disposal, and one wants to make up a total of a_0 cents. At first, our restriction doesn't seem to cause any problems, since

Theorem 3.6. *Let $a_1, ..., a_n$ positive integers such that GCD$(a_1, ..., a_n) = 1$. Then (3.2) has a solution in non-negative integers x_i for all sufficiently large a_0.*

This is as we would expect, though the proof is not entirely trivial (see exercise 2 on homework 1). But now some strange things start to happen :

Definition Let $a_1, ..., a_n$ be positive integers satisfying GCD$(a_1, ..., a_n) = 1$. The *Frobenius number* $G(a_1, ..., a_n)$ is the largest positive integer a_0 for which (3.2) has no solution in non-negative integers.

Example If $a_1 = 3, a_2 = 5$ then $a_0 = 7$. In fact, there is a general formula when $n = 2$, namely $G(a_1, a_2) = (a_1 - 1)(a_2 - 1) - 1$. For a proof, see the homework. There are also good general estimates known when $n = 3$, but they are more complicated and no exact formula is known as far as I know. Even worse, we have

Theorem 3.7. *The problem of computing the Frobenius numbers $G(a_1, ..., a_n)$ for arbitrary inputs is NP-hard for any $n > 3$.*

To my mind, this result is at first glance very surprising, since our theory for linear Diophantine equations seems so simple and satisfactory. So you never know ... I don't know who proved this theorem.

We now make a first foray into the territory of non-linear Diophantine equations. These were already studied by the Greeks (especially Diophantus !), and the subject was enthusiastically revived by Fermat and his contemporaries in the 17th century. The latter had a number of famous results of the form : 'such and such Diophantine equation has only the following solutions (maybe no solutions)'. The methods employed were basically elementary, the key often being some clever application of FTA. In particular, the following consequence of FTA was used widely :

Fact A. *Let a, b be positive integers such that GCD$(a, b) = 1$. If ab is a k:th power, then each of a and b is itself a k:th power.*

Sometimes, Fermat and Co. got carried away in their usage of the unique factorisation idea central to FTA - we will give an example next day. Efforts by later generations to give rigorous proofs of their results laid the groundwork for the development in the 19th century of the body of knowledge nowadays known as *algebraic number theory*.

We start with a result which was perhaps already known to Pythagoras (and maybe even earlier civilisations).

Theorem 3.8. *Let x, y, z be positive integers such that GCD$(x, y, z) = 1$ and y is odd. Then the following two statements are equivalent :*

(i) x is even and $x^2 + y^2 = z^2$,
(ii) there exist positive integers $a < b$, of opposite parity and satisfying GCD$(a, b) = 1$, such that

$$x = 2ab, \quad y = b^2 - a^2, \quad z = b^2 + a^2. \tag{3.10}$$

Proof. Suppose (ii) holds. Then one checks directly that $x^2 + y^2 = z^2$. Let $d =$ GCD(x, y, z). Then $d|b^2 \pm a^2$, hence $d|2a^2$ and $d|2b^2$. Thus d also divides GCD$(2a^2, 2b^2) =$ 2. Hence $d = 1$ or 2. But d cannot be 2, since y is odd.

Now suppose (i) holds. Write the equation as $(z + y)(z - y) = x^2$. Since both y and z are odd and GCD$(y, z) = 1$, we easily deduce that
GCD$(z + y, z - y) = 2$. Hence we can write

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

and, by Fact A above, each of $\frac{1}{2}(z \pm y)$ is a perfect square, i.e.: there exist integers $a < b$ such that

$$\frac{z-y}{2} = a^2, \qquad \frac{z+y}{2} = b^2.$$

Then (ii) follows easily. □

A triple of relatively prime integers satisfying the equivalent conditions of Theorem 3.8 is called a *primitive Pythagorean triple*.

Next day, we will use this theorem to prove the case $n = 4$ of Fermat's Last Theorem.

4. Fourth Lecture : 5/11

The one case of his Last Theorem that Fermat is known to have actually proved is

Theorem 4.1. *Let x, y, z be integers such that $x^4 + y^4 = z^4$. Then $xyz = 0$.*

Proof. We consider more generally the equation

$$x^4 + y^4 = z^2 \tag{4.1}$$

and show that it has no integer solutions such that $xyz \neq 0$. The proof is by contradiction and makes use of Fermat's technique of *infinite descent*. More precisely, the idea is as follows : we suppose there exists a solution (x, y, z) of (4.1) for which $xyz \neq 0$. Then there must be a solution for which, in addition, GCD$(x, y, z) = 1$, for if (x, y, z) is a solution and $d =$ GCD(x, y, z), then $(x/d, y/d, z/d^2)$ is also an integer solution.

Now the idea is that, given any primitive solution (x, y, z), one may construct a new one (X, Y, Z) such that $|Z| < |z|$. But since there must be, among all primitive solutions, one in which $|z|$ is minimised, we thereby obtain the desired contradiction.

So let (x, y, z) be a primitive solution to (4.1). Then (x^2, y^2, z) is a primitive Pythagorean triple so, by Theorem 3.8, if we assume WLOG that x is even and y odd, then there exist relatively prime integers a, b such that

$$x^2 = 2ab, \tag{4.2}$$

$$y^2 = b^2 - a^2, \tag{4.3}$$

$$z = b^2 + a^2. \tag{4.4}$$

We can rewrite (4.3) as

$$y^2 + a^2 = b^2,$$

and so (y, a, b) is also a primitive Pythagorean triple. Since y is odd, so must a be even, and there exist relatively prime integers p, q such that

$$a = 2pq, \tag{4.5}$$

$$y = q^2 - p^2, \tag{4.6}$$

$$b = q^2 + p^2. \tag{4.7}$$

Substituting (4.5) and (4.7) into (4.2) yields

$$x^2 = 4pq(p^2 + q^2). \tag{4.8}$$

Now p and q are relatively prime, hence one sees easily that both are relatively prime to $p^2 + q^2$. Thus, the three numbers p, q and $p^2 + q^2$ are pairwise relatively prime. Since their product is, by (4.8), a perfect square, it follows from Fact A that each is a perfect square. In other words, there exist pairwise relatively prime integers X, Y, Z such that

$$p = X^2, \quad q = Y^2, \quad p^2 + q^2 = Z^2.$$

Substituting the first two of these relations into the third yields the relation

$$X^4 + Y^4 = Z^2,$$

so we have constructed the desired new primitive solution to (4.1). It remains to check that $|Z| < |z|$. But, using (4.7) and (4.4), we have

$$Z^2 = p^2 + q^2 = b < b^2 + a^2 = z \leq z^2,$$

as required. □

FTA is key to the above proof, via Fact A. Fermat seems to have used this trick in much of his work on Diophantine equations. However, this usually involved using FTA in a more 'general' context. The following example illustrates his modus operandi well :

Theorem 4.2. *The only integers* x, y *such that*

$$y^2 + 2 = x^3 \tag{4.9}$$

are $x = 3, y = \pm 5$.

Proof. Write (4.9) as

$$(y + \sqrt{-2})(y - \sqrt{-2}) = x^3. \tag{4.10}$$

We shall first verify that, for any integer y, the numbers $y + \sqrt{-2}$ and $y - \sqrt{-2}$ have no common factor in

$$\mathbb{Z}[\sqrt{-2}] = \{a + b\sqrt{-2} : a, b \in \mathbb{Z}\}.$$

For suppose $z := a + b\sqrt{-2}$ is a common factor. Then z divides

$$(y + \sqrt{-2}) - (y - \sqrt{-2}) = 2\sqrt{-2},$$

and, taking squares of absolute values (as complex numbers), we conclude that

$$a^2 + 2b^2 \mid 8,$$

as ordinary integers. The only possibilities are thus

(i) $a = \pm 1,\ b = 0.$
(ii) $a = \pm 2,\ b = 0.$
(iii) $a = 0,\ b = \pm 1.$
(iv) $a = 0,\ b = \pm 2.$

In case (i), $z = \pm 1$, and hence not a proper factor. In all other cases, there must exist integers c, d such that

$$y + \sqrt{-2} = (a + b\sqrt{-2})(c + d\sqrt{-2}).$$

From this it follows, by equating the real and imaginary parts, that

$$y = ac - 2bd, \tag{4.11}$$

$$1 = ad + bc. \tag{4.12}$$

Now (4.12) immediately rules out (ii) and (iv), since in both cases the rhs of (4.12) is even. But from (4.9) it follows already that y must be odd (otherwise the lhs of (4.9) will be even, but not divisible by 4), and then (4.11) also eliminates (iii).

Hence, we have proven that $y \pm \sqrt{-2}$ have no common factor in $\mathbf{Z}[\sqrt{-2}]$.

By (4.10), this implies that each of them must be a cube in $\mathbb{Z}[\sqrt{-2}]$.

Thus, there exist integers a, b such that

$$y + \sqrt{-2} = (a + b\sqrt{-2})^3. \tag{4.13}$$

Multiplying out the rhs of (4.13) and equating the real and imaginary parts yields the two equations

$$y = a^3 - 6ab^2, \tag{4.14}$$

$$1 = 3a^2b - 2b^3 = b(3a^2 - 2b^2). \tag{4.15}$$

Immediately, (4.15) implies that

$$b = 3a^2 - 2b^2 = \pm 1.$$

Since a, b are integers, the only possibility is $b = 1, \ \ a = \pm 1$. Substituting these possibilities into (4.14) gives $y = \pm 5$, v.s.v. □

OBS !! There is a major gap in the proof, namely the part in italics. What I state there is correct, but it requires a proof[3]. What one would actually like to prove is a generalisation of FTA to the ring $\mathbb{Z}[\sqrt{-2}]$. We will return to this issue later in the course. But, for the moment, it is worth remarking that FTA does not hold in, for example, the ring **Z**$[\sqrt{-5}]$. For in this ring one has, for example,

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}), \tag{4.17}$$

and one may check that each of the four numbers $2, 3, 1 \pm \sqrt{-5}$ is 'prime' in the ring, i.e.: has no factor other than itself and ± 1.

Remark 4.3. Eq. (4.9) is an example of a *Weierstraβ equation* and defines a so-called *elliptic curve* over $\mathbb{Q}$. The most general form of a Weierstraβ equation over $\mathbb{Q}$ (more generally over a field of characteristic other than 2 or 3) is

$$y^2 = x^3 + Ax + B, \qquad A, B \in \mathbb{Q}. \tag{4.18}$$

This defines a non-singular, so-called elliptic curve over $\overline{\mathbb{Q}}$, the algebraic closure of $\mathbb{Q}$, if and only if the *discriminant*

$$\Delta := 4A^3 + 27B^2$$

is non-zero. There is a famous theorem of Siegel that every elliptic curve over $\mathbb{Q}$ contains only finitely many integer points $(x, y) \in \mathbb{Z}^2$. For a proof, see Chapter 9 of the book

[3] There is no evidence that Fermat himself understood the nature of this gap, i.e.: that one doesn't always have unique factorisation in finite extensions of $\mathbb{Z}$, so-called rings of *algebraic integers*. Indeed, his claim to have proven his Last Theorem may well be due to a faulty application of FTA in this more general context. The equation $x^n + y^n = z^n$ can be factored as

$$(x + y)(x + \zeta y) \cdots (x + \zeta^{n-1} y) = z^n, \tag{4.16}$$

where $\zeta = e^{2\pi i/n}$ is a primitive n:th root of unity. In 1847, Lamé gave a faulty proof of FLT which had as its starting point this factorisation and the deduction that each factor on the lhs is an n:th power. It is speculated that Fermat may have had a similar argument in mind.

J. Silverman, *The Arithmetic of Elliptic Curves*, Springer New York (1986), GTM Series No. 106.

If FTA leads towards what is known as *algebraic number theory*[4], then Euclid's other main theorem (Theorem 1.2 above) leads to what is called *analytic number theory*. This subject is mainly concerned with the distribution of the prime numbers, and the methods used chiefly come from real and complex analysis (hence the name). We now begin to scratch the surface of this topic.

Proof. of Theorem 1.2. Suppose, on the contrary, that there are only finitely many primes. List them all as $p_1, ..., p_k$. Consider the number

$$N := \left(\prod_{i=1}^{k} p_i \right) + 1. \tag{4.19}$$

It is clearly not divcisible by any p_i. But it must have SOME prime factor (by Lemma 1.3), which contradicts the completeness of our list of primes. □

Definition Let $\pi : \mathbb{N} \to \mathbb{N}$ be the function given by

$$\pi(n) = \# \text{ primes up to and including } n. \tag{4.20}$$

Euclid's theorem says that $\pi(n) \to \infty$ as $n \to \infty$. The central problem of analytic number theory is to determine the asymptotic behaviour of the function $\pi(n)$. We will dip into this issue next day.

[4] Basically, the study of factorisation in arbitrary rings, especially in rings of algebraic integers.