1.1 26/06/15 - Introduction I: Diophantine Equations, Plane Curves, and Bachet's Example

One of the oldest and most natural questions in mathematics goes back at least to the 3rd century CE, and probably much more. This question is the study of solutions to *Diophantine equations*, named after Diophantus of Alexandria, who "formulated and solved many such problems" ([1, p. xv]).

A *Diophantine equation* is an equation of the form

$$f(x_1, \dots, x_n) = 0$$

where $n \in \mathbb{Z}_{\geq 1}$, and $f \in \mathbb{Q}[x_1, \dots, x_n]$ is a polynomial in n variables with rational coefficients. To *solve it* means to find all rational solutions $(x_1, \dots, x_n)$ to the equation, although what it means to "find" all solutions can indeed be quite subjective. What is usually meant and hoped for is some sort of complete description or parametrization of such solutions. Variations of this problem ask for solutions over a ring R to similar equations defined over R instead of $\mathbb{Q}$ for various rings R such as $\mathbb{Z}$, a number field K, or its ring of integers $\mathcal{O}_K$. Of course, there is a cool interaction between solutions over $\mathbb{Z}$ and $\mathbb{Q}$; see 1.1.3 for one example.

There are (at least) two measures of the complexity of a given Diophantine equation: the number n of variables, and the degree $d \in \mathbb{Z}_{\geq 1}$ of the polynomial f. Let's study some examples for small n and d.

Example 1.1.1 (Single-Variable Equations)**.** The case of $n = 1$ corresponds to equations of a single variable $x = x_1$.

(a) When $d = 1$, we are looking at linear equations in one variable, which are of the form $ax + b = 0$ for some $a, b \in \mathbb{Q}$ with $a \neq 0$. I trust the reader knows how to solve such equations.

(b) When $d = 2$, we are looking at solving quadratic equations over $\mathbb{Q}$, which look like $ax^2 + bx + c = 0$ for $a, b, c \in \mathbb{Q}$ with $a \neq 0$. Our insistence on asking for rational solutions means that there are essentially four things that can happen, depending on the discriminant $\Delta = b^2 - 4ac$; these are:

  (i) There is a unique root of "multiplicity two"; this happens iff $\Delta = 0$ (e.g., $x^2 - 4x + 4 = 0$).

  (ii) The two distinct roots are non-real complex conjugates; this happens iff $\Delta < 0$ (e.g., $x^2 - 5x + 5 = 0$).

  (iii) The two distinct roots are irrational but real Galois conjugates; this happens iff $\Delta > 0$ but Δ is not a perfect square (e.g., $x^2 - 5x + 7 = 0$).

  (iv) The two distinct roots are both rational; this happens iff Δ is a perfect square (e.g., $x^2 - 5x + 6 = 0$).

(c) When $d = 3, 4$, more complicated formulae exist for the solutions, but these are not always helpful. Sometimes we can use elementary calculus to estimate the number of real roots (see, e.g., 2.1.1). When $d \geq 5$, a famous theorem says that no such formula in terms of radicals even exists in general. However, as long as we are only interested in rational solutions, there exists a complete algorithmic procedure to find them all; this is contained in 1.1.2.

Theorem 1.1.2 (Rational Root)**.** Let $d \in \mathbb{Z}_{\geq 1}$, let $a_0, \dots, a_d \in \mathbb{Z}$, and consider the equation

$$a_0 x^d + a_1 x^{d-1} + \cdots + a_d = 0.$$

Suppose that $x = p/q$ is a rational solution to this equation, where $p, q \in \mathbb{Z}$ are coprime and $q \neq 0$. Then $p \mid a_d$ and $q \mid a_0$. In particular, if $a_0 = 1$ (i.e., the equation is monic), then every rational root is an integer.

Remark 1.1.3. Do you see why this gives a complete (algorithmic) answer (in theory, if we can find all factors of a given number) of finding all rational (and hence integral) roots to a single-variable rational polynomial?

Proof of 1.1.2. Since p/q is a solution to the equation, we have

$$a_0 p^d + a_1 p^{d-1} q + \cdots + a_d q^d = 0.$$

Therefore, $q \mid a_0 p^d$; since $(p, q) = 1$, this forces $q \mid a_0$. The part with p is similar and left to the reader. ■

Exercise 1.1.4. Finish the proof by showing that $p \mid a_d$. You might have to be slightly more careful than for q because p could be zero; what would that mean?

We didn't really use any special properties of the integers here–other than the fact that they admit unique factorization. Therefore, the same result and proof works also with $\mathbb{Z}$ replaced by any unique factorization domain (UFD) R; see 2.1.2.

Example 1.1.5 (Plane Curves)**.** The next simplest case of $n = 2$ corresponds to *plane curves*. These correspond to a single equation relating two variables $x = x_1$ and $y = x_2$; the corresponding locus can be plotted as a subset of the Euclidean plane $\mathbb{R}^2$.

(a) When $d = 1$, we are talking about linear equations of the form $ax + by + c = 0$, with $a, b, c \in \mathbb{Q}$ and not both a and b zero. The theory of rational solutions here is familiar from high-school coordinate geometry, and the theory of integer solutions is also well-understood, although it might need a little thought (2.1.3).

(b) When $d = 2$, we are looking at *conic sections*, so named because these curves can be obtained by slicing a (doubly infinite) cone by planes at various angles (see this excellent video by 3Blue1Brown illustrating this). Examples include circles (e.g., $x^2 + y^2 = 1$), parabolae (e.g., $y = x^2$), ellipses (e.g., $x^2 + 2y^2 = 1$), hyperbolae (e.g., $xy = 1$ or $3x^2 - 2y^2 = 1$), pairs of lines lines (e.g., $y^2 - x^2 = 0$), and even "doubled lines" (e.g., $x^2 = 0$) (whatever that means). Plot these on Desmos! Here, the theory of rational solutions is very well-understood. These curves may or may not have any rational points (e.g., $x^2 + y^2 = 1$ and $x^2 + 3y^2 = 2$ respectively; see 2.1.4(a)), but if there is at least one, then we understand all of them very well. We will discuss this more next time.

(c) When $d = 3$, we are looking at *cubic curves*. Again these may or may not have any rational points at all (e.g., $y^2 = x^3 + 1$ and $x^3 + 2y^3 = 4$ respectively; see 2.1.4(b)), but the theory of cubic curves, even those with rational points (these are called *elliptic curves*), is quite subtle. The object of this summer will be to study precisely these objects.

We mention in closing that the cases $d \geq 4$ or $n \geq 3$ are even more complicated; the study of such diophantine equations properly belongs to the field of math known as *arithmetic geometry*. We will not say much about such objects in general, except perhaps for some specific examples.

In 1.1.5, more or less the same discussion can be had with $\mathbb{Q}$ replaced with essentially any field K. (We will systematically make this replacement and work with general fields starting very soon.)

For another Diophantine equation, consider

Example 1.1.6 (Bachet's Example)**.** When does a cube differ from a square by 1? Phrased algebraically, this amounts to solving the equation

$$y^2 = x^3 + 1$$

in pairs of integers (x, y). Five such solutions are $(-1, 0), (0, \pm 1)$, and $(2, \pm 3)$. Are there more?

More generally, given a $c \in \mathbb{Z}$, we can ask for integer solutions to the equation

$$y^2 = x^3 + c. \tag{1.1.7}$$

The case $c = -2$ was studied extensively by French mathematicians Fermat and Bachet in the early 17^{th} century; see the discussion of the history of this problem in [2, Ch. 17, §10]. Bachet observed in 1621 that if (x_0, y_0) is a solution to (1.1.7) with $y_0 \neq 0$, then so is

$$(x_1, y_1) := \left(\frac{x_0^4 - 8cx_0}{4y_0^2}, \frac{-x_0^6 - 20cx_0^3 + 8c^2}{8y_0^3} \right). \tag{1.1.8}$$

This is quite crazy! It's very easy algebra to verify that it works, but how does one come up with such a thing?

Let's assume this for a moment. Given a "seed solution" (x_0, y_0), we can iterate the map $(x_0, y_0) \mapsto (x_1, y_1)$ to produce an infinite sequence (x_n, y_n) of solutions to 1.1.7. For instance, if $c = 1$, then starting from the seed $(x_0, y_0) = (0, 1)$ returns the sequence

$$(0, 1), \quad (0, 1), \quad (0, 1), \ldots.$$

Boring! If we start from the seed $(2,3)$, then we get the sequence

$$(2,3), \quad (0,-1), \quad (0,-1), \ldots.$$

That's slightly more interesting; it's cool that it stabilizes.

Let's take a different c. If we take $c=3$ and the seed $(x_0,y_0)=(1,2)$ gets us the sequence

$$(1,2), \quad \left(\frac{-23}{4^2}, \frac{11}{4^3}\right), \quad \left(\frac{2540833}{88^2}, \frac{4050085583}{88^3}\right), \ldots.$$

That's very interesting! These numbers seem to get pretty huge pretty quickly. (We will prove later that the number of decimal digits in the numerator and denominator in lowest terms roughly doubles with each iteration.)

When do we get infinitely many distinct solutions in the sequence (x_n,y_n)? It is a theorem of Fueter from 1930 (reproven by Mordell in 1966), which we will prove later, that if $c \notin \{1,-432\}$, then for any seed (x_0,y_0) with $x_0y_0 \neq 0$, the sequence (x_n,y_n) gives us infinitely many rational solutions.

What about integer solutions? It is a theorem of Thue from 1908 that for any nonzero c, there are only finitely many integer solutions (x,y) to 1.1.7. Therefore, even if there are infinitely many solutions produced by Bachet's formula (as for $c \notin \{1,-432\}$), only finitely many can be integral. We may or may not get to this result by the end of the course; if you're particularly interested in seeing a proof, I can try to incorporate it.

All of this still begs the question: where does Bachet's formula (1.1.8) really come from? The beautiful answer is that it really comes from *geometry*.[1]

The equation $y^2=x^3+c$ defines a *cubic curve*, say C, in the plane. This cubic curve has the property that it is *smooth*, i.e., has a well-defined tangent line at every point; equivalently, at every point of the curve, at least one of the partial derivatives $2y$ or $3x^2$ is nonzero.[2] Given a point $P_0=(x_0,y_0)$ on this cubic curve C, we can draw the tangent line L to C at P_0. Then an elementary case of *Bézout's Theorem* tells us that L will intersect C in a third point P_1 on C; this is quite clear if we parametrize L and restrict the equation of C via this parametrization to get a cubic equation in one variable whose roots correspond to the intersection points. "Two" of these intersection points are at P_0 (i.e., P_0 corresponds to a double root of this cubic equation), and since a cubic has three roots, there must be a third point–this is P_1. The coordinates of P_1 are exactly the solution (x_1,y_1) expressed in Bachet's formula. In older texts, this procedure $P_0 \mapsto P_1$ is called the "chord and tangent process".

Let's work this out explicitly. Suppose we are given a point (x_0,y_0) on the curve (1.1.7). An easy exercise in elementary geometry or calculus tells us that the slope of the curve 1.1.7 at (x_0,y_0) when $y_0 \neq 0$ is $(3x_0^2)/(2y_0)$, so the equation to the tangent line is

$$y = \left(\frac{3x_0^2}{2y_0}\right)x + \left(y_0 - \frac{3x_0^3}{2y_0}\right). \tag{1.1.10}$$

Substitute this expression for y into 1.1.7 to get the cubic equation

$$x^3 - \left[\left(\frac{3x_0^2}{2y_0}\right)x + \left(y_0 - \frac{3x_0^3}{2y_0}\right)\right]^2 + c = 0.$$

This cubic equation has a double root at $x=x_0$. Let's call its third root x_1. Then Vieta's formula allow us to compute the the product of the roots as the negative of the constant term, which gives us

$$x_0^2x_1 = \left(y_0 - \frac{3x_0^3}{2y_0}\right)^2 - c,$$

giving an expression for x_1 in terms of x_0,y_0, and c. Plugging this into 1.1.10 gives us our y_1. A little bit of simplification using that (x_0,y_0) lies on 1.1.7–left to the reader (2.1.5)–then expresses (x_1,y_1) in the promised form 1.1.8.

[1] I say *really*, because this is probably not how Bachet found it.

[2] This requires us to work over a field of characteristic other than 2 or 3, so $\mathbb{Q}$ is certainly OK.

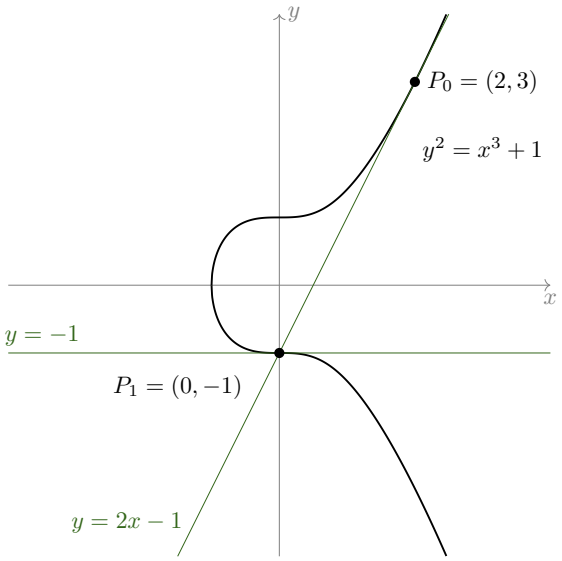

Figure 1.1.9: The "chord and tangent process" which yields Bachet's formula, illustrated for $c = 1$ and seed $P_0 = (x_0, y_0) = (2, 3)$.

The above procedure also has an analog when you are given two points, say P, Q on C: you take the line L joining P and Q, and consider the third intersection points R of P and Q with C (this is the "chord" part of the "chord and tangent process"). It is a fascinating conjecture of Poincaré from around 1901, proven by Mordell 1922 (and then generalized by Weil to number fields in 1928), which says that given any such smooth cubic C over $\mathbb{Q}$, there are finitely many "seed points" $P_i \in C(\mathbb{Q})$ such that *every* other rational point on C can be obtained by iteratively applying the chord and tangent processes to the points P_i. This is now known as the Mordell Theorem or the Mordell-Weil Theorem, and one of our goals for this summer will be to prove this. This cool interplay between algebra, number theory, and geometry is the tip of the iceberg which is the theory of elliptic curves, and what we shall study this summer.