Boston University
Summer I 2010

MA341 Number Theory
Kalin Kostadinov

Lecture 1 (18 May 2010)

Diophantine Equations: Number Theory Meets Algebra and Geometry

Diophantine Equations. ... his boyhood lasted 1/6th of his life; he married after 1/7th more; his beard grew after 1/12th more, and his son was born 5 years later; the son lived to half his father's age, and the father died 4 years after the son. This is a riddle describing the life of Diophantus. He lived in the hellenistic city of Alexandria, in nowadays Egypt, and is considered to be one of the the most influential mathematicians of all time. His treatise, "Arithmetic", contained a lot of riddles like the above one, and ways of solving them. The relations between the unknown quantities are put concisely into equations, and discussing such equations will be topic of this lecture. Let us now formally define the type of equations we will be interested in.

Definition. A Diophantine equation is a polynomial equation $P(x_1, .., x_n) = 0$, where the polynomial P has integral coefficients and one is interested in solutions for which all the unknowns take integer values.[1]

For example $x^2 + y^2 = z^2$ is a Diophantine equation, and $x = 3, y = 4, z = 5$ is one of its infinitely many solutions. Another example is $x + y = 1$, and all its solutions are given by $x = t, y = 1 - t$, where t runs through all integers. A third example is $x^2 + 4y = 3$. This Diophantine equation has no solutions, although note that $x = 0, y = \frac{3}{4}$ is a solution with rational values for the unknowns.

The ultimate question. So could we always solve a Diophantine equation? This question is too general and worse, too vague. What do we mean by 'solve'? For example, we could easily write a computer program which takes as an input an arbitrary Diophantine equation, and then prints all its solutions, if we only allow it to run infinitely long time. Just let it check, one by one, all the possible combinations of values for the variables. Another meaning we may choose for 'solve' is to find a parametrization for all solutions, like we did in the second example above. We will see how to do this in a lot of cases, but the (sad?lucky?) truth is that such a parametrization exists only for a small subset of all Diophantine equations. So, we do something which is often a helpful tactic when faced with a seemingly untractable question: temporarily give up and ask an easier question. Rather then looking for all solutions, can we at least say, given a Diophantine equation, whether it is the case that it has any solutions. This is a question of the existence of a general algorithm: can we write a computer program, which takes as an input an arbitrary equation, and then, after a finite time, prints 'YES' or 'NO', depending on whether the equation has, or has not, a solution? For example, it should print 'YES' for $x^2 + y^2 = z^2$ and 'NO' for

[1] There are variations of this definition, with different restrictions on the coefficients of the equations and unknowns. For example, Diophantus himself considered equations with rational coefficients, as in the riddle, and looked for solutions that are rational numbers.

$x^2 + 4y = 3$. The question about the existence of such an algorithm figured under number 10 in the famous list of 23 problems that David Hilbert distributed after his seminal lecture in the year 1900. The answer turned out to be negative, and this was ultimately proved 70 years later, when Yu. Matiyasevich, building on previous work of M. Davis, H. Putnam, and J. Robinson, showed that there cannot be such an algorithm. The argument of the proof, which lies beyond the scope of this lecture, is based on a special way of *counting* : it turns out that there are 'more' *Diophantine* sets than there are *computable* sets. For a nice and accessible recent discussion on this topic, I recommend looking at Bjorn Poonen's article "Undecidability in Number Theory", in the *Notices of the AMS,* vol. 55, #3 (available online).

Classification of equations. We have seen that the quest for solving all the Diophantine equations is hopeless; it is time for a 'divide and conquer' tactic. What are the features that distinguish between Diophantine equations? A thing that immediately comes to mind is the number of variables. There are equation in one variable, like $x^5 - 3x^2 = 1$, in two variables, like $y^2 = x^3 + x$, in three variables like $x^2 + y^2 = z^2$, in four variables like $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$, and so on. We look at the one-variable case first.
An equation in one variable of degree n has exactly n solutions, counting multiplicities, if we allow complex values for the variables. So a strategy for finding the solutions of a Diophantine equation in one variable is to first find the solutions in the domain of complex numbers, then inspect them to see if any among them are integers. Unfortunately, general formulas for solving equations exist only for degrees up to 4, and the general equation of degree 5 and higher is unsolvable, and this is an impediment to carrying out the first step in that strategy. A roundabout approach is to use Newton's secant method, to find approximations of the real solutions of the equation, and then check the integers that are eventually approximated by these solutions. This is now a working strategy. There is another, easier way to find the solutions of one-variable Diophantine equations, based on the following:

Criterion. Let $a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n = 0$ be an equation with integer coefficients, which has the rational number (in lowest terms) $\frac{p}{q}$ as one of its solutions. Then p divides a_n and q divides a_0.

As an illustration, let us look at the equation $2x^5+9x^4+3x^3+38x^2-11x-5 = 0$. The criterion says that the only possible rational roots are $\pm\frac{1}{2}, \pm 1, \pm\frac{5}{2}, \pm 5$. Checking them all out, gives that only $\frac{1}{2}$ and -5 are indeed roots, so the only solution of the above, as a Diophantine equation, is $x = -5$.
So in the one-variable case, we have as good an answer, as we could hope for; in more variables, the situation is different. Two-variable Diophantine equations have been a subject of extensive research, and their theory constitutes one of the most beautiful, most elaborate, parts of mathematics, which nevertheless still keeps some of its secrets for the next generation of researchers. We will get some glimpses of it today, and then revisit for a different perspective on a few more occasions during the course. And as for Diophantine equations in three and more variables, our knowledge is perhaps best characterized as

modest. For example, we know that the smallest solution of $x^3 + y^3 + z^3 = 70$ is $x = 11, y = 20, z = -21$ (found by computer search), but for the slightly different equation $x^3 + y^3 + z^3 = 33$ it is still unknown whether any integer solutions exist at all.
We now look for other features that divide the Diophantine equations into classes in addition to the number of variables.
One that easily suggest itself is the *total degree* of an equation, which is the largest number one could get by adding the degrees of a single monomial in the equation. For example, $x^5+11x^3y^2-3x^2yz^4 = 13$ has total degree $2+1+4 = 7$. The most accessible slice here are the Diophantine equations of total degree one, and this is what we talk about next.

Linear Diophantine equations. The general linear Diophantine equation is
$a_0x_0 + a_1x_1 + \cdots + a_nx_n = b$;
it has a solution if and only if every integer that exactly divides all of the coefficients $a_0, a_1, \ldots, a_n$, divides b as well. If there is one solution, then there are infinitely many of them. So a linear Diophantine equation has either no solutions, or infinitely many solutions, and we have the above important and simple criterion for distinguishing between this two extremes. For example, the linear Diophantine equation $12x+21y+15z = 7$ has no solutions, since 3 divides $12, 21,$ and 15, but does not divide 7, while $12x + 21y + 14z = 7$ has infinitely many solutions, since the only numbers that divide simultaneously $12, 21, 14$, are ± 1, and they trivially divide 7 as well. The above property of $12, 21, 14$ will occur frequently in our discussions, so we give it a special name:

Definition. We say that the integers in the set $\{k_1, k_2, ...k_n\}$ are *relatively prime* when the only numbers that divide them all are 1 and -1.

Note that relatively prime is not the same as pairwise relatively prime; in the case above $\{12, 21\}$ are not relatively prime, and neither are $\{21, 14\}$ or $\{12, 14\}$, but $\{12, 21, 14\}$ are relatively prime.
Now that we know that $12x + 21y + 14z = 7$ has infinitely many solutions, the natural question arise whether could we find them. The answer is yes, there exists an algorithm which produces the general solution of a linear Diophantine equation. We will not give the algorithm here, though. One reason being that we will talk about it in the workshop about computers and number theory, and another that it is relatively easy to come with an ad hoc solution for any given linear Diophantine equation. In our case, we could first notice that all the values of x that satisfy the equation should be divisible by 7. So any solution should come from a solution of the equation $12t + 3y + 2z = 1$, where $x = 7t$. Next, all the values of y that satisfy this new equation better be odd (Why?). Hence we simplify to
$12t + 3(2u + 1) + 2z = 1$, i.e. to $6t + 3u + z = -1$, where $y = 2u + 1$. Now we could write the general solution:
$x = 7t$, $y = 2u + 1$, $z = -1 - 6t - 3u$, where t, u run independently through all integers.
Notice that the 7 in x is the same as the greatest common divisor of 21 and 14, which were the coefficients of y and z in the original equation, and odd/even in

y has to do with 2, which was also the greatest common divisor of 12 and 14, the coefficients of x and z, respectively. With some effort, and an extra step or two, this ad hoc solution could be turned into an algorithm that will work in general. However, be warned that the parametrization of the solution is far from unique.

Pythagorean triples. We now turn to a famous non-linear, second-degree Diophantine equation in three variables: $x^2 + y^2 = z^2$. Any triple (x, y, z) that solves it gives the lengths of the legs and the hypothenuse of a right triangle, by the converse of the celebrated Pythagoras theorem, hence the name.
We will describe a parametrization for all the Pythagorean triples, and for this we need first a reduction. A Pythagorean triple (x, y, z) is called primitive, if $\{x, y, z\}$ are relatively prime. For example $(3, 4, 5)$ is a primitive Pythagorean triple (PPT for short), and while $(30, 40, 50)$ is a Pythagorean triple, since $30^2 + 40^2 = 2500 = 50^2$, it is not a PPT. Observe that is enough to find all PPTs first, since then any Pythagorean triple could obtained by multiplying each member of a specific PPT by some integer. Also note that the three integers in a PPT are pairwise relatively prime. So in a ppt (x, y, z) at least one of x and y is odd. It can't be that both x and y are odd, since then $x^2 + y^2$ is an even number, not divisible by 4, and such number is never a square. So we could assume that x is odd, y is even, z is odd. (The other ppt's are obtained by permuting x and y.)
Now time for a little algebra: Rewrite the equation as

$$y^2 = z^2 - x^2 = (z - x)(z + x) = (2u)(2v), \text{ where } u = \frac{z - x}{2}, v = \frac{z + x}{2}.$$

Under our assumptions both $z - x$ and $z + x$ are even, so u and v are in fact integers. They are also relatively prime, since if a number d divides both u and v, then it divides their sum $u+v = z$ and their difference $v-u = x$, and because z and x were assumed relatively prime, $d = \pm 1$.
Next, the product of the relatively prime numbers u and v is the perfect square $(y/2)^2$, so u and v must be squares themselves, say $u = m^2$ and $v = n^2$.
To summarize, our assumptions allowed us to deduce the existence of integers m and n such that

$$x = n^2 - m^2, y = 2mn, z = m^2 + n^2$$

Conversely, for each of the integers m, n above, (x, y, z) is a Pythagorean triple, since

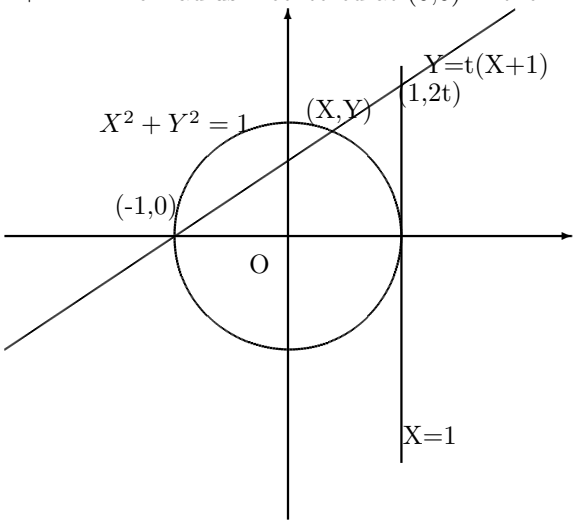
$$(n^2 - m^2)^2 + (2mn)^2 = (m^2 + n^2)^2$$

as it is easily seen after squaring and simplifying.
We remark that we get a PPT exactly when m and n are relative prime of different parity.
As an example, we show that every odd number is part of a Pythagorean triple. Indeed, take say 17, and look for the two consecutive integers that add up to 17 - they are 8 and 9. Take $m = 8, n = 9$. Then $n^2-m^2 = 9^2-8^2 = (9-8)(9+8) = 17$, while $2mn = 144$ and $m^2 + n^2 = 145$. So $17^2 + 144^2 = 145^2$.

The geometry of an equation. We have seen a lot of equations, and used some algebra, but where is the geometry promised in the title? Well, here it comes. The Cartesian geometry tells us how to go from two-variable equations to curves in the XY-plane, and vice versa. So to look for certain kind of solutions (integers, rational numbers) of a two-variable equation is the same as to look for points with certain types of coordinates (integer valued, rational numbers valued) on the associated curve. As an example of this approach we give a second solution to the equation $x^2+y^2=z^2$. First notice that with each of its primitive solutions (x,y,z) we could associate a rational solution (in lowest terms) of the equation $X^2+Y^2=1$, and vice versa, via the transformation formulas

$$X=\frac{x}{z}, Y=\frac{y}{z}, \quad \text{and} \quad z=\text{denomr}(X)=\text{denomr}(Y), x=zX, y=zY$$

For example, the solution $x=3, y=4, z=5$ of $x^2+y^2=z^2$ corresponds to the solution $X=\frac{3}{5}, Y=\frac{4}{5}$ of $X^2+Y^2=1$.
What we just observed is important and worth repeating:
There exists an one-to-one correspondence between the primitive Pythagorean triples and the rational points on the unit circle.
So we can concentrate on finding the points with rational coordinates on the circle $X^2+Y^2=1$ of radius 1 centered at (0,0) in the XY-plane.



Looking at this very imperfect picture, here is our action plan. We consider an arbitrary non-vertical straight line that passes through the point $(-1,0)$. It intersect the vertical line $X=1$ at a point $(1,2t)$, for some number real number t. Then the equation of the first line is $Y=t(X+1)$, and we could get the points where it intersects the circle as solutions of the system of equations

$$X^2+Y^2=1$$
$$t(X+1)=Y$$

Expressing Y from the second as a function of X and plugging into the first we get that the two X coordinates satisfy the equation $X^2+t^2(X+1)^2=1$

which is the same as the quadratic equation $(1+t^2)X^2+2t^2X+t^2-1=0$. We could solve this using the quadratic formula, but since we already know that one solution is $X=-1$, because the line intersects the circle at $(-1,0)$, the other solution is $X=\frac{1-t^2}{1+t^2}$, by Viète's formula. For Y we now get $Y=\frac{2t}{1+t^2}$.
Now notice that a rational value for t gives a rational value for X and Y, and vice versa, rational values for X and Y correspond to a rational t, since $t=\frac{Y}{X+1}$.
Geometrically, we get a correspondence between the rational points on the line $X=1$ and the rational points on the circle $X^2+Y^2=1$. Any pair of corresponding points is connected by a line through $(-1,0)$ with rational coefficients, and any such line connects two corresponding points.
So we conclude that all rational points on the circle are given by $(\frac{1-t^2}{1+t^2},\frac{2t}{1+t^2})$, where t runs through all rational numbers, plus the special point (-1,0), which in a way corresponds to t being 'infinite'.
This in turn gives the description of all pythagorean triples: Let $t=\frac{m}{n}$ in lowest terms, so that $X=\frac{1-(m/n)^2}{1+(m/n)^2}=\frac{n^2-m^2}{n^2+m^2}$ and $Y=\frac{2(m/n)}{1+(m/n)^2}=\frac{2mn}{n^2+m^2}$.
Then $x=n^2-m^2, y=2mn, z=n^2+m^2$ give all pythagorean triples when m,n run through all integers. Notice that this the exact same description we got in the previous section.

Elliptic curves. The geometric approach of the last section was so successful, that it is only natural to try applying it to some more complicated equations. We do exactly this now, and look at the equations $Y^2=X^3+X^2$ and $Y^2=X^3+X$, both of total degree 3.
To extend the method, let see what was essential about it. One, we had a point with rational coordinates, to project from. Two, a rational line through that point crossed the curve in exactly one other point, which then necessarily had to be with rational coordinates as well. So the method works for any two-variable second degree equation with a rational point.
As a straight line crosses a general curve of degree 3 in three points, rather than two, the method will not apply straightforward to the two equations above. We could navigate around this in the case of $Y^2=X^3+X^2$. If we project from the very special point $(0,0)$, where the curve self-intersects, then every straight line will cross the curve in exactly one more point, and we could proceed in the same fashion as we did with $X^2+Y^2=1$ and find a rational parametrization for all the solutions: we end up with $X=t^2-1$ and $Y=t(t^2-1)$.
So $Y^2=X^3+X^2$ from the point of view of Diophantine equations is not that different from $X^2+Y^2=1$. Both of them are what is called "rational curves".
Although $(0,0)$ is on $Y^2=X^3+X$ as well, it does not have the desired property. A line through it will intersect the curve in two more points, and it might very well be the case when they are both non-rational, even if the line is rational. This tells us that the curve is $Y^2=X^3+X$ is essentially different than the circle. It is an example of a class called *elliptic curves* .
The geometric method, properly modified, works for elliptic curves too, as shown in the following deep result, due to L.J. Mordell, a British mathematician from the first half of the 20th century:

Mordell's Theorem: For every elliptic curve there exist a finite number of rational points on that curve, so that starting with them, one could get to any other rational point on the curve by a sequence of operations of the following three kinds:

• reflecting with respect to the x axis a known rational point to get another rational point;

• drawing a tangent through a known rational point and adding the second intersection point to the set of known rational points;

• drawing the chord connecting two known points and adding the third intersection point of the chord with the curve to the set of known rational points.

Despite this algorithmic description, the problem of finding all rational points on an elliptic curve is still not completely solved, since we don't yet have a proven method for generating the starting set of rational points.

The theory of Elliptic curves is a beautiful part of Mathematics; it contains results of striking elegance, has supplied methods for solving practical problems, and despite the active research efforts of a generation of mathematicians, still keeps some of its deepest secrets untold.

Lecture Highlights, Further Reading, Extra Problems:

- Diophantine equations are equations with integer coefficients for which a solution is sought where all unknown variables take integer values. Diophantine equations are difficult to solve.
- Problem solving strategy: When faced with a very difficult problem, it may help to try a simplification of the problem first.
- Problem solving strategy: It helps to classify large problems into smaller pieces, and attack each piece separately.
- Coordinates give a relation between equations and curves. Using geometric methods may produce revealing solutions to algebraic problems.
- Further reading: Rational Points on Elliptic Curve, by J.H. Silverman and J. Tate, part of the Undergraduate Text in Mathematics book series.
- Solve the Diophantine equations $x^2+y^2=10z^2$ and $x^2+2xy+2y^2=5z^2$.
- Carry out details to find the parametrization of $Y^2=X^3+X^2$ described in the text. Find some points on $Y^2=X^3+1$, using Mordell's theorem.
- Try applying the procedures we used to solve the Pythagorean equation $x^2+y^2=z^2$ to the equation $x^3+y^3=z^3$. This last equation is the first case of the Fermat Last Theorem. It was shown by Euler that it has no solutions with $xyz\neq 0$. Where do the geometric and the algebraic method fail? Do they really fail?