Diophantine equations and beyond

lecture King Faisal prize 2014

Gerd Faltings

Max Planck Institute for Mathematics

31.3.2014

Introduction

In this lecture I try to give an overview over my work. I have done research in commutative algebra, algebraic geometry, and number theory. My most important achievement was the proof of the Mordell conjecture. I try to explain it and why this naturally led to work in other fields. I apologise in advance for using the words "I" and "me" more that usual in a lecture.

Numbers

Diophantine problems deal with solutions of algebraic equations in rational numbers. Recall that the natural numbers $\{1, 2, 3...\}$ are obtained by simple counting, and they suffice for some purposes. However for applications one usually has to construct more complicated numbers: First one needs the zero and negative numbers $\{0, -1, ...\}$, then rational numbers a/b where a and b are integers with b different from 0. The real numbers $\mathbb{R}$ are obtained as limits of rational numbers, as for example the squareroot $\sqrt{2}$ or the number π. Finally for complex numbers $\mathbb{C}$ one has to add a squareroot i of -1, that is they are linear combinations $a + bi$. with a and b real numbers. For many purposes the complex numbers suffice. For example any algebraic equation has a root in $\mathbb{C}$.

Finite fields

The complex numbers $\mathbb{C}$ (as well as $\mathbb{Q}$ or $\mathbb{R}$) form a field, that is nontero elements have a multiplicative inverse. Another important class of fields are finite fields and their extension. For each prime number p consider integers modulo p, that is $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. As a set it can be identified with the numbers $0, 1, ..., p-2, p-1$. To add or multiply two of these form their usual sum or product in the integers and substract suitable multiples of p to end up in the desired range. In general finite fields are extensions of $\mathbb{F}_p$ and their number of elements is a power of p. Finite fields have recently found some applications in cryptography.

p-adic numbers

Sometimes one has to replace the real numbers by different completions of the rationals. For any prime p consider two rationals a/b and a'/b as p-adically close if the integer $a - a'$ is divisible by a big power of p. Then adding p-adic limits extends the rationals $\mathbb{Q}$ to the field $\mathbb{Q}_p$ of p-adic numbers. For example the series

$$1 + 2 - 2 + 24 - , , , = 1 - \sum_{n \geq 1} (-2)^n (1 \cdot 3 \cdot (2n-1))/(1 \cdot ...n)$$

has a limit in $\mathbb{Q}_2$ which is equal to $\sqrt{5}$. $\mathbb{Q}_p$ has some similarity to the reals $\mathbb{R}$ but is also quite different in certain aspects. For example it does not suffice to add one element (like i) to it to make it algebraically closed. Instead one has to adjoin infinitely many elements and add a further completion to extend $\mathbb{Q}_p$ to a complete algebraically closed field $\mathbb{C}_p$.

diophantine equations

Diophantine geometry deals with solutions of algebraic equations in integers or rational numbers. Examples are Pythagorean triples (solutions in integers of $a^2 + b^2 = c^2$, or in rationals of $x^2 + y^2 = 1$), or the Fermat equation ($a^n + b^n = c^n$, or $x^n + y^n = 1$, $n \geq 3$). In general an algebraic variety is the set of common solutions of finitely many polynomial equations. Examples are $x^2 - y^2 = 1$ or $x^2 + y^3 + z^5 = 0$, but not $3^x - 2^y = 1$ (it involves nonalgebraic functions). The algebraic variety is defined over $\mathbb{Q}$ if the polynomials defining it have coefficient in $\mathbb{Q}$. It then has points (that is common solutions of the equations) in any overfield of $\mathbb{Q}$, that there are $\mathbb{Q}$-rational, $\mathbb{R}$-rational and $\mathbb{C}$-rational points.

Forms

The algebraic variety given by $x^2 + y^2 = -1$ is defined over $\mathbb{Q}$ but has no rational points over $\mathbb{Q}$ or even over $\mathbb{R}$. On the other hand over $\mathbb{C}$ it is isomorphic to $x^2 + y^2 = 1$ (multiply x and y by i) and has many $\mathbb{C}$-rational points. These two varieties are called twisted forms of each other. This illustrates that over the complex numbers $\mathbb{C}$ the classification of varieties is simpler. Thus even for a variety defined over a numberfield one usually first investigates its complex points.

Smoothness and dimension

An algebraic variety is called smooth if the complex points form a manifold. For example the variety given by $x^2 + y^2 = 1$ is smooth while $y^2 - x^3 = 0$ defines a nonsmooth variety. Smooth varieties (also called manifolds) are easier to investigate, while the presence of singularities poses additional problems.
An important invariant of an algebraic variety is its dimension. It roughly says on how many complex parameters the $\mathbb{C}$-rational points depend. Varieties of dimension zero are finite sets. In the next case of dimension one the varieties are called curves. Such a curve has an important invariant, the genus. Rational points on curves of genus zero can be parametrised and their study becomes comparatively easy. For example rational solutions to $x^2 + y^2 = 1$ are of the form
$x = (1 - t^2)/(1 + t^2), y = 2t/(1 + t^2)$. For curves of genus one Mordell showed that rational points form a finitely generated abelian group. That is they are much rarer than for genus zero, but there still may be infinitely many of them. He conjectured that for genus bigger than one the set of rational points is finite.

Elliptic and hyperelliptic curves

Examplesare hyperelliptic curves given by an equation

$$y^2 = (x - a_1)...(x - a_r),$$

with $a_1, ..., a_r$ distinct. They are of genus g if $r = 2g + 1$ or $r = 2g + 2$. For $g = 1$ ($r = 3, 4$) they are called elliptic curves, These are algebraic groups, that is they support a commutative group law where the addition is given by algebraic equations. For $g > 1$ one can define an algebraic group of dimension g associated to them, called the Jacobian. The group of rational points on this Jacobian is finitely generated. The Mordell conjecture says that for $g > 1$ only finitely many of these points lie on the curve itself.

Heights

Important concepts in the study of diophantine equations are height and the notion of good reduction. The height of a rational number a/b is the maximum of the sizes of the numerator a and the denominator b (assumed to be coprime). The height of an n-tuple $(x_1, ..., x_n)$ of rational numbers is the maximum of the height of the coordinates x_i. The height is important for proving finiteness theorems because if we have an upper bound for the height of solutions we can easily find all of them by checking a finite list of possibilities. Of course obtaining such a bound tends to be difficult for interesting problems. For most objects of interest in diophantine geometry there exists a natural height function which measure their complexity. To achieve a good function one usually has to modify the naive height by a bounded function. This does not affect its main finiteness property. For example for elliptic curves one gets the Néron-Tate height which is a quadratic function on the rational points,

Bad reduction

Good reduction at a prime p means in the simplest case that for a rational number a/b p does not divide the denominator. This can fail only for finitely many primes which are called the primes of bad reduction. Another example: An equation defines a smooth algebraic variety if some discriminant is nonzero. If the equation has integers as coefficients this discriminant is also an integer, and thus only divisible by finitely many primes. These are the primes of bad reduction for this property. In general for any diophantine object there exists a finite set of primes of bad reduction. For example elliptic curves can be defined as solutions to the equation

$$y^2 = (x - a)(x - b)(x - c).$$

Such a curve is smooth if a, b, and c are pairwise different. If they are integers we can reduce them modulo a prime p and obtain a curve over $\mathbb{F}_p$. It is smooth (that is the elliptic curve has good reduction at p) if the three numbers are different modulo p, that is if p is not among the finitely many prime divisors of $(a - b)(a - c)(b - c)$.

Number fields

In more generality one may enlarge the rationals by adding solutions of algebraic equations to obtain algebraic numberfields. An example is the field $\mathbb{Q}(\sqrt{2})$ which consists of all linear combinations $a + b\sqrt{2}$ with rationals a, b. Note that this field has a symmetry by sending $a + b\sqrt{2}$ to $a - b\sqrt{2}$, quite similar to complex conjugation on $\mathbb{C}$. So if an equation has coefficients in $\mathbb{Q}$ its solutions in $\mathbb{Q}(\sqrt{2})$ admit this symmetry. Another example exists by adjoining all cube roots of 2, that is $2^{1/3}, \zeta 2^{1/3}$ and $\zeta^2 2^{1/3}$, with $\zeta = e^{2\pi i/3}$. Here the symmetry consists of all permutations of the roots. A version of it exists for all numberfields and is called the Galois-group. One of the most powerful tools in diophantine geometry consists in passing from rational points to representations of the Galois-group.

Frey curves

As an example we illustrate this in the case of the Fermat equation $a^n + b^n = c^n$: We want to show that it has no nontrivial solutions in integers if $n \geq 3$. It suffices to consider the cases where $n = 4$ or where n is an odd prime. Then we consider the auxiliary Frey elliptic curve, that is solutions of the equation

$$y^2 = x(x - a^n)(x + b^n).$$

If we add a point at infinity the (say) complex solution (x, y complex numbers) form a commutative group, that is we can define an addition of two such points. The group law is determined by the fact that the three intersection points of the curve with any straight line (in the (x, y)-plane) add up to zero. The height of such an elliptic curve is given by the size of the discriminant which is $4a^n b^n c^n$, and the primes of bad reduction are the primes which divide a, b or c. The n-division points (adding the point n-times to itself gives zero) form a subgroup of order n^2. As the group law is given by algebraic equation with coefficients in $\mathbb{Q}$ these division points lie in algebraic numberfields and thus admit an action of the Galois-group.

Galois representations for Fermat

The Galois action associated to the Frey curves has certain peculiar properties. Namely at the primes of bad reduction one would expect that the representation notices them (it "ramifies"). In our case (the Frey-curve for a Fermat equation) the primes of bad reduction are those dividing one of the numbers a, b, c. However because they occur as n-th powers these primes of bad reduction are not seen by the n-torsion points. In Wiles solution of the Fermat problem the key step is to show that this implies that the elliptic curve is "congruent modulo n" to an elliptic curve with no primes of bad reduction. Actually one has to work with a generalisation of elliptic curves (modular forms), but nevertheless one shows that no such object without primes of bad reduction exists, and thus the Fermat equation has no nontrivial solution.
For more complicated diophantine equations one sometimes can still associated to solutions a Frey elliptic curve. However this construction would help only if one could bound the height of this curve. To achieve this one has to solve the "abc-conjecture", one of the most important open problems in the theory.

The Mordell conjecture over function-fields

For more general diophantine equations we cannot hope for no solutions, but only for qualitative statements like finiteness of solutions, for special types of algebraic varieties. To achieve this for curves (that is to show the Mordell conjecture) the method of Parshin-Arakelov associated to each solution a new curve and its Jacobian, which is a generalisation of the elliptic curves which we encountered for the Fermat problem. They and Szpiro could show the Mordell conjecture over function fields, which are similar to numberfields but where additional tools are available. Also Szpiro emphasized the importance of Arakelov theory which allows to carry over some techniques from function to numberfields. However one important tool (Kodaira-Spencer classes) was missing.

The Mordell conjecture over numberfields

Quite unexpectedly this difficulty was resolved. Namely associated to the auxiliary curves are their Jacobians and the Galois action on the torsion points of the Jacobians. This Galois action again has primes of bad reduction, and this set of primes is not empty but al least predetermined by the curve and not by the rational point on it. By the general theory (the Weil conjectures) there are only finitely many such representations. If we show that only finitely many points can give rise to the same representation we derive finiteness of rational points.
Now if two points give rise to the same representation the corresponding Jacobians are not the same but at least they are similar. The technical term is that they are isogenous. I could show in 1983 that in a given isogeny class the height of the Jacobian (a measure for its complexity) is bounded, and derives that there can be only finitely many Jacobians in this isogeny class. Thus there are only finitely many rational points.

Arakelov geometry

Arakelov geometry has been developed to transfer techniques from functionfields to numberfields. For example a vectorbundle on a curve C is given by a vectorspace over the generic point of C and suitable adic metrics at all closed points. Similarly one defines an Arakelov vectorbundle over the integers $\mathbb{Z}$ is a projective $\mathbb{Z}$-module with a positive definite quadratic form on the corresponding real vectorspace. It has a welldefined degree. So in short Arakelov geometry amounts to endowing all objects with suitable hermitian metrics. One can extend many results from classical algebraic geometry to this context, for example the Riemann-Roch theorem. For example good height functions are usually given by Arakelov degrees for metricised linebundles.

Toroidal compactifications

The proof of the Mordell conjecture led me to further work in two fields where the necessary results could be obtained in an adhoc manner but where a fully satisfactory treatment required further work. One was the need to define heights for abelian varieties. These correspond to rational points on a certain moduli space, but to get a good theory one has to compactify that space, that is to add certain degenerate abelian varieties. Over the complex numbers compactifications had been defined by Baily-Borel, Shimura, and Mumford, but they had no arithmetic interpretation, nor did the construction allow to bound the primes of bad reduction. However Mumford also had found a construction of degenerate abelian varieties, and I could show that Mumford's construction gives all such degenerations, and allows to define local coordinates at the boundary of the compactifications. The resulting arithmetic compactification is called the toroidal compactification. I wrote a book about that in collaboration with C.L.Chai.

p-adic Galois representations

Another development was the local theory of p-adic Galois representations. What was necessary for Mordell had already been done by Tate in the one dimensional case. It turned out that his method generalises to higher dimensions and yields a "p-adic Hodge decomposition". Moreover the method allowed to attack the comparison between etale and crystalline cohomology which had been conjectured by J.M.Fontaine. The whole theory now has been put on a conceptual basis by P.Scholze.

Cohomology Theories

In classical topology one associated to a topological space homology and cohomology groups, which are algebraic inveriants which are less complex than topological spaces and thus easier to handle. For algebraic varieties over $\mathbb{Q}$ one can evaluate them on the complex points, but this forgets the rational structure and especially the Galois-action. Grothendieck has define etale cohomology which exists for algebraic varieties even in characteristic p, but to get good properties one has to chose coefficients which are torsion of order prime to p. For example the etale homology of an elliptic curve is given by its torsion points.
To get invariants modulo p Grothendieck introduced the crystalline cohomology. Over p-adic fields both etale and crystalline cohomology make sense, and Fontaine's theory describes the relation between those two.

Diophantine approximation

Another (in fact historically earlier) approach to diophantine eqution has been via diophantine approximation. We illustrate that for the example of Roth's theorem: If α is an algebraic number (that is it satisfies a polynomial equation) then α cannot be too well approximated by a rational numbers. Namely for any exponent $r > 2$ there are only finitely many rationals a/b with $|\alpha - a/b| \leq 1/b^r$. Thue was the first to apply this method to diophantine geometry. For example if $F(x, y)$ is a homogeneous polynomial of degree ≥ 3 with integral coefficients, integers solutions to $F(x, y) = 1$ must have the property that x/y is a good approximation to one of the roots of $F(t, 1)$, and one derives that the number of such solutions is finite.
Siegel extended this to more complicated integral points. However only Vojta succeeded much later to give a proof for the Mordell conjecture along these lines. Trying to understand his proof I developed a geometric approach (the "product theorem") which allowed to generalise it to higher dimensional varieties.

diophantine approximation and algebraic geometry

For $r = 2$ the theory of continuous fractions gives (for real irrational numbers) infinitely many solutions. For the proof of Roth's theorem one assumes that the assertion is wrong and derives a contradiction. If the assertion is wrong there are infinitely many fractions a/b satisfying the inequality. Among them the denominators b can become arbitrarily large, so one can find e sequence a_i/b_i with the b_i rapidly increasing (which can be made precise but this requires some technicalities). Then one constructs an auxiliary polynomial F in variables $T_1, ..., T_s$ of multidegree $(d_1, ..., d_s)$ (with d_i approximately proportional to the inverse of $log(b_i)$) which vanishes to high order at $(\alpha, ..., \alpha)$. Finally one derives a contradiction if s is big enough (depending on how close r is to 2) and if the d_i increase rapidly enough.

vectorbundles

On the side I have been interested in vectorbundles on curves, first by a paper of Giesecker on vectorbundles on Mumford curves and then by lectures of Witten in Princeton. The Gieseker paper inspired a description of semistable vectorbundles on Mumford curves, and the Witten lectures a proof of the Verlinde conjecture. I also discovered a characterisation of semistable bundles by the fact that some twist has trivial cohomology, and how to generalise the Hitchin fibration to arbitrary semisimple groups.