

Mordell, past and present

MIT conference

Gerd Faltings

Max Planck Institute for Mathematics

8.7.2024

I give a short history of the subject. It is entirely my personal view and does not claim historical accuracy. Of course diophantine equations have a long history, starting with the Egyptians and Diophantus, and including the Fermat problem.

The 1920's

In 1922 Mordell proved that on a cubic over \mathbb{Q} the rational points form a finitely generated abelian group, and remarked that for curves of higher genus there might always be only finitely many. He was a pure number theorist and did not speculate about other number fields. In 1928 Weil generalised this to the groups of rational points on Jacobians of curves., over arbitrary number fields This implies the same for all abelian varieties, but their algebraic theory had not evolved enough at this time. (Later Weil himself fixed that). Probably many tried to derive from this the Mordell conjecture, somehow showing that the image of a curve in its Jacobians avoids almost all elements in a finitely generated subgroup of its points. But nobody claimed success, until Vojta managed to do this in 1989. Before that Chabauty showed in 1941 finiteness of rational points if the rank of the Mordell-Weil group is smaller than the genus. In this case rational points lie on a proper p -adic subgroup of the Jacobian. Also Siegel proved in 1929 finiteness of integral points on an affine subset of an elliptic curve.

The method used in this time was diophantine approximation: One can measure the complexity of rational points P by the (logarithmic) height function $h(P)$. If we bound $h(P)$ there are only finitely many P 's left, so if there are infinitely many 'bad P 's their heights are unbounded. So we can find P_1 and P_2 with $h(P_1)$ big and the ratio $h(P_2)/h(P_1)$ big (what this means can be described by explicit constants). But then one constructs a function $f(P, Q)$ in two variables which is bihomogeneous of degrees d_1 and d_2 roughly proportional to the inverses of $h(P_i)$, which vanishes to a high order (better high index) at (P_1, P_2) . Finally one shows (and this is the subtle point, for example the content of Roth's lemma) that this is not possible.

Shafarevich made his conjecture on finiteness of curves with given genus and locus of bad reduction in his talk on the ICM in Stockholm 1962. In this decade Mordell was shown for the function field of a complex curve B . The first proof was due to Manin in 1963 and brought us the Gauss-Manin connection. Another one was done by Grauert in 1965. In all these developments one has to exclude constant (or more precisely isotrivial) curves. That is done by requiring that the map from B to the moduli space of curves (or to the moduli space of abelian varieties) is nonconstant. Equivalently the pullback of an ample line bundle on the moduli space should have positive degree. As ample line bundle one chooses the determinant of the space of differentials on the relative curve. Finally in 1968 Parshin proved Mordell and Shafarevich for curves with everywhere good reduction:

Firstly Shafarevich implies Mordell via the Kodaira construction which associates to a curve C and a point $P \in C$ a new curve D which is a covering (necessarily nonabelian) of C only ramified in P . Secondly for a nonconstant curve $C \rightarrow B$ with everywhere good reduction the relative differentials ω are ample. Namely it is not difficult to show that ω has non negative degree on each effective curve. Secondly Weierstrass points define a non trivial map from the determinant of relative differentials on C to some positive power of ω . Finally Kodaira vanishing implies that the map from B into the moduli space cannot be deformed, and Hodge theory bounds the degree of relative differentials and thus of this map. Namely 2-forms inject into singular cohomology which has bounded dimension.

The Tate conjecture

In 1966 Tate proved his conjecture for abelian varieties A over a finite field \mathbb{F}_q . That is Frobenius linear endomorphisms of the Tate module $T_l(A)$ (l prime to the characteristic) are induced (up to a factor \mathbb{Z}_l) by endomorphisms of the abelian variety. For the proof note that the graph of the endomorphism induces a sequence of subgroups $G_n \subset A \times A$. If we show that infinitely many of the quotients $A \times A/G_n$ are isomorphic this provides us with plenty of endomorphisms, enough to prove the result. Finiteness is clear for (say) principally polarised abelian varieties because they are parametrised by the \mathbb{F}_q -points of a quasicompact moduli space, but Tate had to struggle to get rid of polarisations. Soon this became much easier.

In 1971 Arakelov generalised Parshin's result to curves with bad reduction. Also in 1974 he invented Arakelov theory hoping to generalise from function fields to number fields. Namely a relative curve over the integers of a number field corresponds to a relative curve over an affine subset $B^0 \subset B$, so that somehow the fibers over the remaining points of B are missing, and intersection numbers have little meaning. Arakelov's idea was to add metrics at the (missing) infinite places of the number field, and then define meaningful intersection numbers. Although finally this theory was not used it still provided valuable insight about what to do at infinite places.

In 1974 Zarhin extended Tate's theorem to abelian varieties over function fields of curves. He invented two major improvements: Firstly he eliminated polarisations by using that for any abelian variety A the product $A^4 \times A^t$ is principally polarised. Secondly he bounded the heights of the quotients $A \times A/G_n$, thus deriving finiteness. For this he used Mumford's theory of theta functions. Today we use the differentials of top order on A . These are invariant under the isogeny of l -power order. The relation between those two approaches is that theta null values are modular forms of weight $1/2$.

In 1979 Szpiro extended Parshin's result to base curves B of positive characteristic. Here a new difficulty arises because if the base curve is defined over a finite field pullbacks of C via powers of the Frobenius on B have the same locus of bad reduction but the numerical invariants tend to infinity. This was fixed by requiring that the Kodaira-Spencer class of the relative curve C (the differential of the map into moduli space) does not vanish (otherwise pullback by Frobenius would give counterexamples). This class is a class in $H^1(C, \omega_C^{-1})$, so this cohomology group is nonzero. Using variants of Kodaira vanishing in characteristic p this bounded numerical invariants and thus implied the Shafarevich conjecture. Still for number fields there seems to be no Kodaira-Spencer class, so something was missing.

Finally I got personally involved. Szpiro was a friend of my advisor Nastold and I visited him in Paris and learned about his point of view. Especially I first heard about Arakelov theory and found this very interesting and inspiring. In 1982 I extended the Parshin-Arakelov method to families of abelian varieties over function fields of characteristic zero and tried to prove the Shafarevich conjecture for them. I got half of it, namely the boundedness of numerical invariants, but also found out that sometimes families could be deformed. On the way I learned about the importance of the sheaf of relative differentials on the moduli space of abelian varieties. Also I remember that the referee suggested to replace the relevant arguments by results from hyperbolic geometry, but (fortunately) this was unknown territory for me and I refused. Also Szpiro inspired me to look into Arakelov theory.

Back in Germany I thought about the Tate-conjecture. for abelian varieties. Following Zarhin I computed the change of heights under an isogeny. For the height I used the degree of differentials metrised by square integration, which was the obvious choice prescribed by Arakelov theory. This change of heights was given by the sum of two terms, namely one determined by ramification, and the other (easier) by the change of square integration. I then realised that for p -divisible groups Tate had already computed the first term and shown how to read this off from the Galois representation, and finally the Weil conjectures finished the job, that is the heights of the quotients became eventually constant. Also results of Raynaud replacing Tate worked for finite subgroups (not related to p -divisible groups), and so I obtained finiteness in a given isogeny class. As the number of such classes could also be bounded, by applying Cebotarev to Galois representations, I had shown Tate, Shafarevich and Mordell at once and my career take a definitive turn upwards. Also one long term consequence is the present conference. The missing tools from function fields (Hodge theory, Kodaira-Spencer class) had been replaced by finiteness of possible Galois representations.

In 1989 Vojta gave a new proof for Mordell using diophantine approximation. He also was the first he could make use of Mordell-Weil for that: Namely he divided the real vector space spanned by the Mordell-Weil group into cones where all the vectors in a cone have approximately the same direction (I think this idea is due to Mumford). He then set up the machinery of diophantine approximation and showed that in a fixed cone the heights of points were bounded. As people were a little bit hesitant I felt like a boyscout and decided to do a good deed and look into the paper. I was amply rewarded because I found it not only to be correct but also that it could be generalised to abelian varieties. And finally I understood how to prove Roth's lemma!

Recent results

A new trend has been the use of p -adic methods. For example in 2005 Kim gave a new proof of Siegel's theorem, using unipotent bundles. On a curve C we consider the universal unipotent crystals of unipotency degree $\leq n$. These form a projective system, admit a Hodge filtration, and if C has good reduction at p they are Frobenius crystals. \mathbb{Z}_p -points on C lie in a finite union of p -adic open disks on which we fix their reduction modulo p . It is known that \mathbb{Q}_p -crystals are constant on such disks. The Hodge filtration then defines a p -adically analytic map from the open disk into a (projective limit of) flag varieties. It can be shown that its image is Zariski dense. Also p -adic Hodge theory assigns to these data a Galois representation of the local field. If the point in the open disk is a rational point over a number field the Galois representation is induced from a global Galois representation for which there are strong restrictions, and thus the image lies in a proper Zariski closed subset and thus is finite.

Similar reasonings lead Lawrence and Venkatesh 2019 to a new proof of Mordell. In it there appear no height functions. Instead the main tool is finiteness of possible Galois representations. Namely they define an analytic map from the curve into a parameter space of Galois representations, via a variant of the Parshin-Kodaira construction. This map has Zariski dense image. Also in the parameter space reducible representations (even reducible locally at p) lie in a proper Zariski closed subspace. Finally the irreducible ones belong to a finite list.

Another development is the search for effective methods to really determine all rational points. This has led to effective bounds for the number of rational points, but not for their maximal height. That this is possible in principle was already remarked by Parshin. On the other hand if the Chabauty method is applicable (Chabauty-Coleman) it gives effective results. Recently the method has been enhanced (quadratic Chabauty) and now covers more cases, for example the infamous cursed curve.