Undecidability in number theory[1]

Bjorn Poonen

Does the equation $x^3+y^3+z^3=29$ have a solution in integers? Yes: $(3,1,1)$, for instance. How about $x^3+y^3+z^3=30$? Again yes, although this was not known until 1999: the smallest solution[2] is $(-283059965, -2218888517, 2220422932)$. And how about $x^3+y^3+z^3=33$? This is an unsolved problem.

Of course, number theory does not end with the study of cubic equations in three variables: one might ask also about

$$x^{1729}y^{1093}z^{196884}-163xyzt^{262537412640768000}=561.$$

D. Hilbert, in the list of 23 problems he published after a famous lecture in 1900, asked his audience to find a method that would answer all such questions. More precisely, *Hilbert's tenth problem* (hereafter denoted H10) asks for an algorithm that takes as input a multivariable polynomial $f(x_1,\dots,x_n)$ with integer coefficients and outputs YES or NO according to whether there exist integers $a_1,a_2,\dots,a_n$ such that $f(a_1,\dots,a_n)=0$.

In 1970, Yu. Matiyasevich, building on earlier work of M. Davis, H. Putnam, and J. Robinson, showed that *no such algorithm exists.*

The purpose of this article is to discuss

- some of the concepts in the proof,
- a few by-products of the proof, and
- current research on related problems that are still open, such as the analogue for rational number solutions.

H10 and the DPRM theorem

The notion of algorithm. To make sense of the negative answer to H10, we need a precise notion of algorithm. In 1900 such a notion had not yet been developed. But in the 1930s, several rigorous models of computation were proposed and were shown to be equivalent; one of these was the *Turing machine.* The equivalence made believable the *Church-Turing thesis*, which is the assertion that every purely mechanical procedure can be carried out by a Turing machine.[3] Because of this, "algorithm" is taken to mean "Turing machine".

An informal description of a Turing machine may be more enlightening than a mathematically precise definition. A Turing machine is equivalent to a finite-length program running on a physical computer, except that the computer has unlimited time and memory and is not subject to physical errors (such as data loss from power outages). The memory is sometimes modelled as an infinite tape, initialized to the binary representation of the nonnegative integer input. The computer reads and writes 0s and 1s from and to the memory tape during its operation, and may or may not print characters on a separate output tape, following the rules of its program. It might run forever, or it might halt when some condition specified by the program is satisfied.

Turing machines may accept any objects as input if we fix an encoding of these objects as nonnegative integers. For example, a polynomial with integer coefficients could be represented by the concatenation of the ASCII codes of the characters in a TEX string for the polynomial. The exact encoding does not matter as long as a Turing machine can convert between the proposed encodings.

Diophantine, listable, and computable sets. Davis, Putnam, Robinson, and Matiyasevich deduced the negative answer to H10 from a

[1] This survey article has been accepted for publication in the *Notices of the Amer. Math. Soc.*, and is expected to appear in the March 2008 issue.

[2] Discovered by E. Pine, K. Yarbrough, W. Tarrant, and M. Beck following an approach suggested by N. Elkies.

[3] Quantum computers might seem at first not to fit this framework. But they can be simulated by classical Turing machines in exponential time, and H10 asks for *any* algorithm without being fussy about its running time. When one ignores running time, quantum computers are no more powerful than classical ones.

stronger theorem having many more implications. To explain it, we need a few definitions.

Definition 1. A set $A \subseteq \mathbb{Z}$ is *diophantine* if there exists a polynomial $p(t, \vec{x}) \in \mathbb{Z}[t, x_1, \ldots, x_n]$ such that

$$A = \{a \in \mathbb{Z} : (\exists \vec{x} \in \mathbb{Z}^n)\ p(a, \vec{x}) = 0\}.$$

One should think of p as defining a family of polynomial equations, depending on a parameter t; then A is the set of values of the parameter for which the resulting equation in the remaining variables $x_1, \ldots, x_n$ has a solution. Equivalently, if B is the set of solutions to $p(t, \vec{x}) = 0$ in $\mathbb{Z}^{1+n}$, then A is the projection of B onto the first coordinate. The definition can be extended in an obvious way to subsets of $\mathbb{Z}^m$ for $m > 1$.

Example 2. The subset $\mathbb{N} := \{0, 1, 2, \ldots\}$ of $\mathbb{Z}$ is diophantine since for $a \in \mathbb{Z}$, we have

$$a \in \mathbb{N} \iff (\exists x_1, \ldots, x_4 \in \mathbb{Z})\ x_1^2 + \cdots + x_4^2 = a.$$

Definition 3. A set $A \subseteq \mathbb{Z}$ is *listable* (or *recursively enumerable*) if there is an algorithm that prints A, i.e., a Turing machine such that A is the set of integers it prints out when left running forever.

Example 4. The set of integers expressible as a sum of three cubes is listable. (Print out $x^3 + y^3 + z^3$ for all $|x|, |y|, |z| \leq 10$; then print out $x^3 + y^3 + z^3$ for $|x|, |y|, |z| \leq 100$; and so on.) A similar argument shows that any diophantine subset of $\mathbb{Z}$ is listable.

Definition 5. A set $A \subseteq \mathbb{Z}$ is *computable* (or *recursive*) if there is an algorithm for deciding membership in A, i.e., an algorithm that takes as input an integer a and outputs YES or NO according to whether $a \in A$.

Any computable set is listable, since given an algorithm for deciding membership in A, one can apply it successively to 0, 1, -1, 2, -2, ...and print each number for which the membership test returns YES.

But it is not obvious that every listable set is computable. An algorithm that prints A does not immediately let one test whether 33 is in A, say: if after running the algorithm for a while the number 33 is not printed, it may be hard to decide whether it will appear later on.

In fact, the next section shows that there exists a listable set that is not computable.

The halting problem. The negative answer to H10 was proved by relating it to undecidability results in logic and computability theory from the 1930s. These undecidability results were proved using diagonalization arguments reminiscent of G. Cantor's famous proof of the uncountability of $\mathbb{R}$.

One such result concerns the *halting problem*, which asks for an algorithm that takes as input a computer program p and an integer x, and outputs YES or NO, according to whether program p run on input x eventually halts (instead of entering an infinite loop, say).

Theorem 6 (Turing 1936)**.** *The halting problem is undecidable; that is, no Turing machine can solve it.*

Sketch of proof. Fix an encoding of programs as nonnegative integers; identify programs with their integer codes. Suppose that there were an algorithm for deciding when program p halts on input x. Using this we could build a new program H such that for any x,

$$\begin{aligned} &H \text{ halts on input } x \\ \iff\ &\text{program } x \text{ does not halt on input } x. \end{aligned}$$

Taking $x = H$, we find a contradiction: H halts on input H if and only if H does not halt on input H. □

Corollary 7. *There exists a listable set that is not computable.*

Proof. Let A be the set of numbers $2^p 3^x$ such that program p halts on input x. By Theorem 6, A cannot be computable. On the other hand, here is a program that prints A: loop over $N = 1, 2, \ldots$, and during iteration N, for each $p, x \leq N$, run program p on input x for N steps, and print $2^p 3^x$ if the program halts within these N steps. □

The DPRM theorem. We are now ready to state the following remarkable theorem.[4]

DPRM theorem (Davis, Putnam, Robinson, Matiyasevich 1970)**.** *A subset of $\mathbb{Z}$ is listable if and only if it is diophantine.*

To prove their theorem, these four authors essentially built a computer out of diophantine equations! They showed that diophantine equations are rich enough to simulate any computer in the sense that given a computer program, one can

[4] Historically, the notions of diophantine, listable, and computable and the DPRM theorem were stated for subsets of $\mathbb{N}$ instead of $\mathbb{Z}$. This makes little difference, however: reductions in both directions are possible because of Example 2 and the equality $\mathbb{Z} = \mathbb{N} \cup (-\mathbb{N})$.

construct a polynomial equation that has an integer solution if and only if the program halts. The proof of the DPRM theorem looks curiously like the construction of a complicated computer program, with high-level routines built out of more elementary ones, except that instead of routines one has diophantine equations everywhere. An improved version of the original proof may be found in Chapters 1–5 of [Mat93].

A brief history of the DPRM theorem. The DPRM theorem was conjectured in 1949 by Davis, who also carried out the first reductions towards its proof. In 1961, Davis, Putnam, and Robinson proved its analogue for *exponential diophantine equations* over $\mathbb{N}$ (such as $2x^{3y^x z+x^2} = 5x^2 + yz$). This meant that it remained to show that exponentiation was diophantine, i.e., that $\{(a,b,c) \in \mathbb{N}^3 : c = a^b\}$ was a diophantine set. Earlier, in 1952, Robinson had proved that the diophantineness of exponentiation would follow from the existence of a 2-variable diophantine relation of "exponential growth". Finally, in 1970, Matiyasevich used properties of Fibonacci numbers F_n to prove that the relation $m = F_{2n}$ was diophantine; this gave what Robinson needed, and completed the proof of the DPRM theorem.

For more history, see the references at the end of this article, including the film [Csi08] and the website [H10web].

Negative answer to H10. The DPRM theorem easily implies a negative answer to H10, as we now explain. The undecidability of the halting problem gave us a listable set that is not computable. By the DPRM theorem, having this is the same as having a diophantine set that is not computable. By definition, this means that we have a polynomial $p(t, \vec{x})$ such that there is no algorithm for deciding for which values $a \in \mathbb{Z}$ the equation $p(a, \vec{x}) = 0$ has a solution in integers $x_1, \ldots, x_n$. Thus there cannot be an algorithm for deciding the existence of integer solutions to all polynomial equations.

Remark. H10 was not the first problem outside logic and computability theory to be proved undecidable. In 1947 A. A. Markov and E. Post independently found a finitely presented semigroup for which the word problem is undecidable, and in 1955 P. S. Novikov did the same for a finitely presented group. (The *word problem* for a finitely presented semigroup G with finite set of generators A is the problem of deciding, given two finite sequences of elements of A, whether the product of the first sequence equals the product of the second sequence in G.) The word problem for groups had been motivated by topology, and it was not long afterward that fundamental problems in topology itself were found to be undecidable: for instance, Markov in 1958 proved that the problem of deciding whether two finite simplicial complexes are homeomorphic is undecidable.

Other fun consequences of DPRM

Undecidability for polynomials of fixed degree in a fixed number of variables. The proof of the previous section shows that there is a pair (n, d) of positive integers such that there is no algorithm for deciding the existence of integer solutions to n-variable polynomial equations of total degree d. In the 1960s, before the DPRM theorem was proved, the fact that it would imply that equations of bounded degree in a bounded number of variables suffice to represent all diophantine sets was considered by some as evidence that the theorem could not be true!

After 1970, several authors, including Yu. Matiyasevich, J. Robinson, and Z. W. Sun, proved undecidability results for explicit small values of n and d. For instance, it is now known that there is no algorithm for deciding the existence of integer solutions to polynomial equations in 11 variables. In the positive direction, it is known only that there is an algorithm for polynomials in one variable! It is likely that the problem is decidable also for polynomials in two variables, but so far the elaborate machinery developed by arithmetic geometers is too weak to prove even this.

As for degree, a trick discovered by T. Skolem in the 1920s shows that any polynomial equation in integers is equivalent to one of degree at most 4 (and the equivalence is constructive): for instance, $y^2 = x^5 + 7$ is solvable if and only if

$$(u - x^2)^2 + (v - u^2)^2 + (y^2 - xv - 7)^2 = 0$$

is. Thus there is no algorithm for equations of degree 4. In the positive direction, there is an algorithm for equations of degree at most 2 in any number of variables. The situation for degree 3 is still unknown.

Number of solutions.

Theorem 8 (Davis 1972)**.** *Let A be a nonempty proper subset of* $\mathbb{N} \cup \{\aleph_0\}$*. There is no algorithm* *that takes as input $f(\vec{x}) \in \mathbb{Z}[x_1, \ldots, x_n]$ and outputs YES or NO according to whether the cardinality of $\{\vec{a} \in \mathbb{Z}^n : f(\vec{a}) = 0\}$ belongs to A.*

The proof, which is very short, shows that an algorithm for any A as above could be used to give an algorithm for H10.

Simple equations whose smallest solution is huge.

Theorem 9. *There is a polynomial $p(t, \vec{x})$ such that for any function $F \colon \mathbb{Z} \to \mathbb{N}$ that is computable and defined on all of $\mathbb{Z}$, there exists $a \in \mathbb{Z}$ such that $p(a, \vec{x}) = 0$ has a solution $\vec{x} \in \mathbb{Z}^n$ but no solution with $\max |x_i| < F(a)$.*

Proof. Use the same p as in the proof of the negative answer to H10. If there were a computable bound on the size of the smallest solution when a solution existed, then one could decide for which $a \in \mathbb{Z}$ the equation $p(a, \vec{x}) = 0$ was solvable simply by searching up to that bound. This contradicts the choice of p. □

Prime-producing polynomials. Before the DPRM theorem was proved, Putnam observed that it would imply the following theorem.

Theorem 10. *There exists a polynomial $F(x_1, \ldots, x_n) \in \mathbb{Z}[x_1, \ldots, x_n]$ such that the positive integers in its range (as a function $\mathbb{N}^n \to \mathbb{Z}$) are exactly the prime numbers.*

Proof. The natural number version of the DPRM theorem gives a polynomial $p(t, \vec{x})$ such that for $a \in \mathbb{N}$, the equation $p(a, \vec{x}) = 0$ is solvable in natural numbers if and only if a is prime. Define $F(t, \vec{x}) := t(1 - p(t, \vec{x})^2)$. It can be positive only when $p(t, \vec{x}) = 0$, and in this case, t is prime and $F(t, \vec{x}) = t$. Conversely, every prime arises this way. □

A reasonably simple prime-producing polynomial in 26 variables was constructed in a paper by J. P. Jones, D. Sato, H. Wada, and D. Wiens: see [Mat93, p. 55]. Later Matiyasevich constructed a 10-variable example.

Riemann hypothesis. The DPRM theorem gives an explicit polynomial equation that has a solution in integers if and only if the Riemann hypothesis (RH) is false. Indeed, one can write a computer program that searches for a counterexample to RH (e.g., by applying the argument principle and numerical integration to rectangles with corners in $\mathbb{Q}[i]$ lying in the strip $1/2 < \operatorname{Re} s < 1$, or by testing an equivalent formulation of RH as in [DMR76, p. 335] or [Mat93, §6.4]); then one can use the DPRM theorem to simulate the program with a polynomial equation.

M. Baker half-jokingly observed that one might try to prove RH by showing that the equation has no solutions modulo 17, say! As one might expect, however, things are not so easy: the equation produced by the DPRM theorem will have solutions modulo any fixed positive integer.

H10 over other rings

Even before 1970, researchers began asking Hilbert's question for rings other than $\mathbb{Z}$.

Definition 11. Let R be a commutative ring. Then *Hilbert's tenth problem over R* (H10 over R) asks for an algorithm that takes as input $f(\vec{x}) \in R[x_1, \ldots, x_n]$ and outputs YES or NO according to whether there exists $\vec{a} \in R^n$ such that $f(\vec{a}) = 0$.

Technically, to make sense of this, we need to fix an encoding of elements of R suitable for input into a Turing machine. In cases where this is not possible (e.g., if R is uncountable), then it is understood that we restrict the possible inputs by requiring that the coefficients of f belong to some "large" countable subring R_0 of R. For instance, if $R = \mathbb{C}$, we might take R_0 to be the subfield of algebraic numbers.

The question of whether H10 over R has a positive answer now depends on the ring R (and possibly also R_0). The remainder of this article will focus on rings R that are of interest to number theorists. For more information about these problems, see [DLPVG00, Shl07].

H10 over rings of algebraic integers. The ring of Gaussian integers, $\mathbb{Z}[i] := \{a + bi : a, b \in \mathbb{Z}\}$, shares many properties with $\mathbb{Z}$, so one might expect a negative answer for H10 over $\mathbb{Z}[i]$. More generally, inside any number field k (i.e., finite extension of $\mathbb{Q}$), one has the *ring of integers $\mathcal{O}_k$*, defined as the set of $\alpha \in k$ satisfying $f(\alpha) = 0$ for some monic $f(x) \in \mathbb{Z}[x]$.

Conjecture 12. *For any number field k, H10 over $\mathcal{O}_k$ has a negative answer.*

Through work of J. Denef, L. Lipshitz, T. Pheidas, A. Shlapentokh, and the author spanning about 30 years, the following is known:

Theorem 13. *For a number field k, H10 over $\mathcal{O}_k$ has a negative answer if any of the following hold:*

(i) *k is totally real (i.e., every homomorphism $k \to \mathbb{C}$ has image contained in $\mathbb{R}$).*

(ii) *k is a quadratic extension of a totally real number field.*

(iii) *k has exactly one conjugate pair of nonreal embeddings.*

(iv) *There exists an elliptic curve E over $\mathbb{Q}$ such that $E(\mathbb{Q})$ and $E(k)$ have the same positive rank.*

To make sense of (iv), recall the *Mordell-Weil theorem*, which states that for any elliptic curve E over a number field k, the abelian group $E(k)$ of points on E with coordinates in k is finitely generated. Condition (iv) is probably satisfied for every number field k, but this seems extremely difficult to prove.

The reason that the proof of the negative answer for $\mathbb{Z}$ cannot be adapted directly to arbitrary $\mathcal{O}_k$ is that it uses the fact that the integer solutions to Pell's equation $x^2 - dy^2 = 1$ for a fixed nonsquare $d \in \mathbb{N}$ form an abelian group of rank 1. It is only for number fields like those in (i)–(iii) of Theorem 13 that something close enough to this holds over $\mathcal{O}_k$.

In contrast with Conjecture 12, if $\overline{\mathbb{Z}}$ is the ring of all algebraic integers, i.e., $\{\alpha \in \mathbb{C} : f(\alpha) = 0 \text{ for some monic } f(x) \in \mathbb{Z}[x]\}$, then H10 over $\overline{\mathbb{Z}}$ has a positive answer, as shown by R. Rumely.

H10 over $\mathbb{Q}$. H10 over $\mathbb{Q}$ is equivalent to one of the big open problems in arithmetic geometry, namely whether there is a general algorithm for deciding whether a variety X over $\mathbb{Q}$ has a rational point[5].

Reductions. Might one deduce a negative answer to H10 over $\mathbb{Q}$ from the negative answer to H10 over $\mathbb{Z}$? Given a polynomial equation over $\mathbb{Q}$, one can construct an equivalent system of polynomials over $\mathbb{Z}$ by replacing each rational variable by a ratio of two new integer variables, clearing denominators, and adding auxiliary equations to force the denominator variables to take nonzero values in any solution (such auxiliary equations exist since the subset $\mathbb{Z} - \{0\}$ of $\mathbb{Z}$ is diophantine). Since a system of polynomial equations $f_1 = \cdots = f_n = 0$ over $\mathbb{Z}$ is equivalent to a single polynomial equation $f_1^2 + \cdots + f_n^2 = 0$ over $\mathbb{Z}$, the previous sentence shows that H10 over $\mathbb{Q}$ can be embedded as a subproblem of H10 over $\mathbb{Z}$. Unfortunately, this goes the wrong way: the subproblem might still be decidable even though the whole problem is not.[6]

One way to get a reduction in the useful direction would be to show that $\mathbb{Z}$ is *diophantine over* $\mathbb{Q}$, i.e., that there is a polynomial $p(t, \vec{x}) \in \mathbb{Q}[t, x_1, \ldots, x_n]$ such that $\mathbb{Z}$ equals the set of $a \in \mathbb{Q}$ such that $p(a, \vec{x}) = 0$ has a solution $\vec{x} \in \mathbb{Q}^n$. Indeed, we could use this to embed H10 over $\mathbb{Z}$ as a subproblem of H10 over $\mathbb{Q}$: given a polynomial equation to be solved in integers, we could consider the same equation over $\mathbb{Q}$ together with auxiliary equations that force the rational variables to take integer values (this is where we need $\mathbb{Z}$ to be diophantine over $\mathbb{Q}$).

Actually, something a little weaker would suffice for the desired reduction. It would suffice to have a *diophantine model* of the ring $\mathbb{Z}$ over $\mathbb{Q}$, i.e., a diophantine set $S \subseteq \mathbb{Q}^n$ that "looks like $\mathbb{Z}$" in the sense that it is equipped with a bijection $\phi\colon \mathbb{Z} \to S$ such that the graphs of + and × (subsets of $\mathbb{Z}^3$) correspond under ϕ to diophantine subsets of $S^3 \subseteq \mathbb{Q}^{3n}$.

Even more generally, it would suffice to have a *diophantine interpretation* of $\mathbb{Z}$ over $\mathbb{Q}$: this is like a diophantine model, except that $\mathbb{Z}$ is identified not with a diophantine subset of some $\mathbb{Q}^n$, but with a diophantine subset modulo a diophantine equivalence relation.

Remark. It has been suggested that one might try to build a diophantine model of $\mathbb{Z}$ over $\mathbb{Q}$ using an elliptic curve E with $E(\mathbb{Q}) \simeq \mathbb{Z}$. Such elliptic curves are easy to find, and under the bijection $\mathbb{Z} \to E(\mathbb{Q})$ the graph of + on $\mathbb{Z}$ corresponds to a diophantine subset; unfortunately it is not clear whether the same is true for the graph of ×.

Mazur's conjecture. B. Mazur has proposed a conjecture that, if true, would rule out some of these approaches towards a negative answer to H10 over $\mathbb{Q}$. If X is a variety over $\mathbb{Q}$, then the set $X(\mathbb{R})$ of real points on X inherits a topology from the topology of $\mathbb{R}^n$.

Conjecture 14 (Mazur 1992)**.** *For any variety X over $\mathbb{Q}$, the topological closure of $X(\mathbb{Q})$ in $X(\mathbb{R})$ has at most finitely many connected components.*

A deep theorem of G. Faltings can be used to prove Mazur's conjecture for a curve X. But our

[5] Readers unfamiliar with the notion of variety will lose little generality, for our purposes, in thinking of X as a system of polynomial equations, and a rational point as a simultaneous solution in rational numbers.

[6] On the other hand, if H10 over $\mathbb{Z}$ had had a *positive* answer, it would have implied a positive answer to H10 over $\mathbb{Q}$. It has been argued that this, together with the fact that Hilbert asked his question for $\mathbb{Z}$ instead of $\mathbb{Q}$, suggests that Hilbert expected a positive answer to his tenth problem.

almost complete lack of understanding of rational points on higher-dimensional varieties makes it difficult to gather much evidence for or against the conjecture in general. See [Maz94] for further discussion.

Mazur's conjecture, together with some elementary topology, implies that for any set $S \subseteq \mathbb{Q}^n$ that is diophantine over $\mathbb{Q}$, the closure of S in $\mathbb{R}^n$ has at most finitely many connected components. In particular, it implies that $\mathbb{Z}$ is not diophantine over $\mathbb{Q}$. (This was Mazur's reason for introducing his conjecture.) A more complicated argument of G. Cornelissen and K. Zahidi involving the DPRM theorem shows that Mazur's conjecture implies also that there is no diophantine model of $\mathbb{Z}$ over $\mathbb{Q}$.

On the other hand, it is not known whether Mazur's conjecture rules out also a diophantine interpretation of $\mathbb{Z}$ over $\mathbb{Q}$.

Subrings of $\mathbb{Q}$. Given that we have a negative answer for $\mathbb{Z}$ and do not know the answer for $\mathbb{Q}$, we might ask about rings in between. Every such ring is $\mathbb{Z}[S^{-1}]$ for some subset S of the set $\mathcal{P}$ of all primes: $\mathbb{Z}[S^{-1}]$ consists of the rational numbers whose denominators are divisible only by primes in S. How large can we make S and still prove a negative answer for H10 over $\mathbb{Z}[S^{-1}]$?

If S is finite, work of Robinson on diophantine definitions of valuation rings in $\mathbb{Q}$ implies that $\mathbb{Z}$ is diophantine over $\mathbb{Z}[S^{-1}]$, so the negative answer for $\mathbb{Z}$ implies a negative answer for $\mathbb{Z}[S^{-1}]$. If S is infinite, we may measure its size by defining the *natural density* of S as

$$\lim_{X\to\infty} \frac{\#\{p \in S : p \leq X\}}{\#\{p \in \mathcal{P} : p \leq X\}},$$

if the limit exists.

In 2003 the author proved

Theorem 15. *There exists a computable set $S \subseteq \mathcal{P}$ of density 1 such that*

(i) *There exists a curve E such that $E(\mathbb{Z}[S^{-1}])$ is an infinite discrete subset of $E(\mathbb{R})$. (So the analogue of Mazur's conjecture for $\mathbb{Z}[S^{-1}]$ is false.)*

(ii) *There is a diophantine model of $\mathbb{Z}$ over $\mathbb{Z}[S^{-1}]$.*

(iii) *H10 over $\mathbb{Z}[S^{-1}]$ has a negative answer.*

The proof takes E to be an elliptic curve of rank 1 (minus its point at infinity), and shows that by choosing S carefully, we can control the subset $E(\mathbb{Z}[S^{-1}])$ of $E(\mathbb{Q})$ sufficiently well to obtain a discrete set that looks enough like $\mathbb{Z}$ to serve as a diophantine model.

Unfortunately, the complement of S in $\mathcal{P}$, while sparse, is still infinite, so Theorem 15 implies nothing about H10 over $\mathbb{Q}$.

First-order sentences

In terms of logic, H10 asks for an algorithm to decide the truth of *positive existential sentences*

$$(\exists x_1 \exists x_2 \cdots \exists x_n)\; f(x_1, \ldots, x_n) = 0$$

in the language of rings, where the variables run over integers. More generally, one can ask for an algorithm to decide the truth of arbitrary *first-order sentences*, in which any number of quantifiers and boolean operations are permitted: a typical such sentence is

$$(\exists x)(\forall y)(\exists z)(\exists w)\; (x \cdot z + 3 = y^2) \vee \neg(z = x + w).$$

Long before DPRM, the work of K. Gödel, A. Church, and A. Turing in the 1930s made it clear that there was no algorithm for solving the harder problem of deciding the truth of first-order sentences over $\mathbb{Z}$.

First-order sentences over $\mathbb{Q}$. Though it is not known whether $\mathbb{Z}$ is diophantine over $\mathbb{Q}$, we have

Theorem 16 (Robinson 1949)**.** *One can characterize $\mathbb{Z}$ as the set of $t \in \mathbb{Q}$ such that a particular first-order formula of the form*

$$(\forall \vec{x})(\exists \vec{y})(\forall \vec{z})(\exists \vec{w})\; p(t, \vec{x}, \vec{y}, \vec{z}, \vec{w}) = 0$$

is true, when the variables range over rational numbers.

Combining this with the non-existence of an algorithm for first-order sentences over $\mathbb{Z}$, Robinson obtained

Corollary 17. *There is no algorithm to decide the truth of a first-order sentence over $\mathbb{Q}$.*

How complicated must a class of first-order sentences be, in order that we are able to prove that no algorithm can decide the truth of all sentences in the class? Using quaternion algebras, the author in 2007 improved Robinson's result by defining $\mathbb{Z}$ in $\mathbb{Q}$ by a formula with 2 universal quantifiers followed by 7 existential quantifiers:

Theorem 18. *The set $\mathbb{Z}$ equals the set of $t \in \mathbb{Q}$ such that*

$$\begin{gathered}(\forall a, b)(\exists x_1, x_2, x_3, x_4, y_2, y_3, y_4)\\ (a + x_1^2 + x_2^2 + x_3^2 + x_4^2)(b + x_1^2 + x_2^2 + x_3^2 + x_4^2)\\ \cdot \Big[\big(x_1^2 - ax_2^2 - bx_3^2 + abx_4^2 - 1\big)^2\\ + \prod_{n=0}^{2309} \big((n - t - 2x_1)^2 - 4ay_2^2 - 4by_3^2 + 4aby_4^2 - 4\big)^2\Big]\\ = 0\end{gathered}$$

is true, when the variables range over rational numbers.

Corollary 19. *There is no algorithm for deciding, given an algebraic family of morphisms of varieties, whether there exists one that is surjective on rational points.*

Cornelissen and Zahidi obtained an even better result conditional on the truth of a plausible conjecture about elliptic curves.

If we could eliminate the two universal quantifiers in Theorem 18, we would have a negative answer to H10 over $\mathbb{Q}$. But we cannot see how to eliminate even one of them.

Status of knowledge. The table below summarizes what is known regarding the questions

- Is there an algorithm for H10 over R?
- Is there an algorithm to decide the truth of arbitrary first-order sentences over R?

over various rings R, listed roughly in order of increasing arithmetic complexity[7]:

Ring	H10	1st order
$\mathbb{C}$	YES	YES
$\mathbb{R}$	YES	YES
$\mathbb{F}_q$	YES	YES
p-adic fields	YES	YES
$\mathbb{F}_q((t))$	?	?
$\overline{\mathbb{Z}}$	YES	YES
number field	?	NO
$\mathbb{Q}$	?	NO
global function field	NO	NO
$\mathbb{F}_q(t)$	NO	NO
$\mathbb{C}(t)$	?	?
$\mathbb{C}(t_1, \ldots, t_n)$, $n \geq 2$	NO	NO
$\mathbb{R}(t)$	NO	NO
$\mathcal{O}_k$	?	NO
$\mathbb{Z}$	NO	NO

For $\mathbb{C}$ the positive answers are a consequence of 19th century elimination theory. For $\mathbb{R}$ they come from A. Tarski's elimination theory for *semialgebraic sets*, subsets of $\mathbb{R}^n$ defined by polynomial equations and polynomial inequalities. For finite fields $\mathbb{F}_q$, the answers are trivially positive! By a p-adic field, we mean a finite extension of the field $\mathbb{Q}_p$ of p-adic numbers; A. Macintyre developed an elimination theory for these, though the positive answers were given before this, in work of J. Ax, Yu. Ershov, S. Kochen, and A. Nerode. It is surprising that the answers for the closely analogous field $\mathbb{F}_q((t))$ of formal Laurent series over a finite field are not known.

We have already mentioned Rumely's positive answer for H10 over $\overline{\mathbb{Z}}$; this was extended to first-order sentences by L. van den Dries. The negative answers for first-order sentences over a number field k and its ring of integers $\mathcal{O}_k$ are due to Robinson.

By *global function field* we mean the field $\mathbb{F}_q(t)$ of rational functions with coefficients in a finite field, or a finite extension of $\mathbb{F}_q(t)$. Such fields are studied both because they are closely tied to algebraic geometry and because they are analogous to number fields in many ways. The breakthrough giving the negative answer to H10 for $\mathbb{F}_q(t)$ for odd q was due to T. Pheidas. The extension to all global function fields (and even finite extensions of $\mathbb{F}_q(t_1, \ldots, t_n)$ for $n \geq 2$) was completed by C. Videla, A. Shlapentokh, and K. Eisenträger. The proofs use the Frobenius endomorphism in an essential way, however, and hence cannot be adapted to number fields.

The negative answer to H10 over $\mathbb{C}(t_1, \ldots, t_n)$ for $n \geq 2$ is due to K. H. Kim and F. W. Roush; this result should be better known among algebraic geometers than it is since it implies that there is no algorithm for the general problem of deciding whether a rational map of varieties $X \dashrightarrow \mathbb{P}^n$ over $\mathbb{C}$ for fixed $n \geq 2$ admits a rational section. The analogue with $\mathbb{P}^n$ replaced by an arbitrary fixed variety Y of dimension at least 2 was proved by K. Eisenträger using work of L. Moret-Bailly. Although the answers for $\mathbb{C}(t)$ are unknown, the answers for $\mathbb{R}(t)$ are negative, as shown by J. Denef.

Our list of results is by no means complete: for instance, we have said nothing about rings of holomorphic or meromorphic functions, function

[7] There is no formal definition of arithmetic complexity, but for fields k we can look at the size of the absolute Galois group $\operatorname{Gal}(k_s/k)$, where k_s is a separable closure of k. Domains may be considered more complex than their fraction fields, since they have "extra structure" coming from the divisibility relation.

fields over an algebraically closed field of positive characteristic, etc. There remain many open problems for anyone who is interested.

Acknowledgements

I have borrowed extensively from many excellent earlier expositions of the subject; some of these are listed below. I thank M. Davis, E. Frenkel, Yu. Matiyasevich, and A. Shlapentokh for many comments.

Short list of references

[Csi08] George Csicsery, *Julia Robinson and Hilbert's tenth problem*, 2008. Film in progress, Zala Films, www.zalafilms.com.

[DMR76] Martin Davis, Yuri Matiyasevich, and Julia Robinson, *Hilbert's tenth problem: Diophantine equations: positive aspects of a negative solution*, Mathematical developments arising from Hilbert problems (Proc. Sympos. Pure Math., Vol. XXVIII, Northern Illinois Univ., De Kalb, Ill., 1974), Amer. Math. Soc., Providence, R. I., 1976, pp. 323–378. (loose erratum). MR 0432534 (55 #5522)

[DLPVG00] Jan Denef, Leonard Lipshitz, Thanases Pheidas, and Jan Van Geel (eds.), *Hilbert's tenth problem: relations with arithmetic and algebraic geometry*, Contemporary Mathematics, vol. 270, American Mathematical Society, Providence, RI, 2000. Papers from the workshop held at Ghent University, Ghent, November 2–5, 1999. MR **1802007 (2001g:**00018)

[H10web] *Hilbert's tenth problem page.* Website created by Maxim Vsemirnov under the supervision of Yuri Matiyasevich, http://logic.pdmi.ras.ru/Hilbert10.

[Mat93] Yuri V. Matiyasevich, *Hilbert's tenth problem*, Foundations of Computing Series, MIT Press, Cambridge, MA, 1993. Translated from the 1993 Russian original by the author; With a foreword by Martin Davis. MR **1244324 (94m:**03002b)

[Maz94] B. Mazur, *Questions of decidability and undecidability in number theory*, J. Symbolic Logic **59** (1994), no. 2, 353–371. MR **1276620 (96c:**03091)

[Shl07] Alexandra Shlapentokh, *Hilbert's tenth problem. Diophantine classes and extensions to global fields*, New Mathematical Monographs, vol. 7, Cambridge University Press, Cambridge, 2007. MR 2297245