

BOOK REVIEW

APPEARED IN BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 32, Number 1, January 1995, Pages 114-119
©1995 American Mathematical Society
0273-0979/95 \$1.00 + \$.25 per page

Hilbert's Tenth Problem, by Yuri V. Matiyasevich. The MIT Press, Cambridge, MA, and London, 1993, xxii + 264 pp., \$45.00. ISBN 0-262-13295-8

It was the reading of a translation of Diophantus's books on arithmetic that led Pierre de Fermat to found modern number theory and the study of what are now called Diophantine equations. Diophantine equations are nothing more than equations between polynomials in several variables, their *Diophantineness* lying not in the nature of the equations but in that of the solutions being sought. Diophantus and algebraic geometers like rational solutions, while Fermat and his successors prefer integral solutions. Fermat himself is associated with two important Diophantine equations, namely: the Fermat equation,

$$x^n + y^n = z^n,$$

for which he claimed to have only the obvious solutions for $n > 2$; and the so-called Pell equation,

$$x^2 - Dy^2 = 1, \quad D \text{ not a perfect square,}$$

the name of which originates in an error of attribution on Leonhard Euler's part.

The Pell equation demonstrates dramatically the difference between the original Diophantine programme of searching for positive rational solutions and the newer Fermatian programme of searching for integral solutions. In response to Fermat's challenge, Lord Brouncker quickly gave a parametric description of the rational solutions but had a much harder time of it in the integral case. Eventually, he came up with an algorithm for the generation of the integral solutions (an algorithm already known, incidentally, to the Hindu mathematicians a few centuries earlier), but he was never able to prove the existence of solutions, i.e., that his algorithm always provided a solution. This was first done by Joseph Louis Lagrange, today's preferred solution due in fact to Peter Gustav Lejeune Dirichlet. Lagrange further went on to found the theory of quadratic forms, a subject that found its highest expression in 1801 with Carl Friedrich Gauß's publication of his *Disquisitiones arithmeticae*.

Following Gauß there were attempts to extend the work to quadratic forms of more variables and to cubic forms. The former subject was first brought to completion in 1972 by Carl Ludwig Siegel, and the latter subject is still inchoate. Success beyond the third degree has been largely sporadic, with few general results until recently. In 1900 at the International Congress of Mathematicians in Paris, David Hilbert, looking forward to the coming century, proposed twenty-three problems

with which twentieth century mathematicians would have to contend. The tenth on the list, commonly simply termed “Hilbert’s Tenth Problem”, called for a general method to determine the solvability or unsolvability in integers of Diophantine equations. With our current knowledge of the unsolvability of this problem, it would be interesting to know why Hilbert felt it had a positive solution. Did it look like the Gaußian theory could extend indefinitely to more and more variables and higher and higher degrees? Did he think one could cut through all the details and give an abstract proof, in the way he finished off the theory of invariants? Or was it just a manifestation of his faith in the mathematician’s ability to solve all problems he posed for himself—a faith on which he was quite explicit? The actual statement of Hilbert’s Tenth Problem is rather brief and uninformative:

10. Determination of the solvability of a Diophantine equation. Given a diophantine equation with any number of unknown quantities and with rational integral numerical coefficients: *To devise a process according to which it can be determined by a finite number of operations whether the equation is solvable in rational integers.*

The twentieth century saw a great deal of work on Diophantine equations, if not directly on Hilbert’s Tenth Problem. Among the early outstanding work was that of Thue, Siegel, and Roth giving whole classes of equations that could have only finitely many solutions. The bounds given, however, were on the number of solutions and not the sizes, whence Thue *et al.* could not answer the fundamental question of the existence of at least one solution. This was not remedied until 1968 when Alan Baker gave effective (if unfeasibly large) upper bounds on the sizes and, at least in principle, solved the existence problem.

One might also mention Thoralf Skolem and his p -adic method. Skolem wrote a short monograph on Diophantine equations in the 1930s. This was described by one prominent logician giving a series of talks on the history of logic as the first systematic exposition of the subject. This is not exactly true; one can point, for example, to a Dover reprint of two books by Carmichael for an attempt to present a coherent account of the subject using the fact that the solutions to certain Diophantine equations (such as the Pell equation) naturally form multiplicative groups. Such partial success at coherency aside, the true nature of the field was nicely summed up by Eric Temple Bell in his *Development of mathematics*, published in the later 1930s:

Diophantus contented himself with special solutions of his problems; the majority of his numerous successors have done likewise, until diophantine analysis today is choked by a jungle of trivialities bearing no resemblance to cultivated mathematics. It is long past time that the standards of Diophantus be forgotten though he himself be remembered with becoming reverence.

Of course, Bell should never be relied upon for historical facts, and he did love to express extreme views; but his overall summary in the present case is not that far off target. One has but to leaf through the second volume of L. E. Dickson’s *History of the theory of numbers* (available in Chelsea reprint) to see the “jungle” that so appalled Bell. Even Mordell’s book on the subject, with its general results, has the appearance of a mixed bag of tricks.

The 1930s saw another development, initially unrelated to Diophantine analysis and Hilbert’s Tenth Problem. This was the birth of the Theory of Algorithms or

Effective Computability. The decade saw a number of researchers offering analyses and definitions of computable functions. Because these various definitions were routinely proven equivalent, logicians soon felt confident enough to prove the unsolvability of various problems by effective means. Initially, these were problems in the theory of algorithms itself, then in logic, and in the 1940s in logical areas of algebra—semi-Thue systems and word problems for semigroups. In the 1950s, the word problem for groups was shown unsolvable. For the most part, Diophantine equations were ignored by logicians. Thoralf Skolem, who made contributions to both the theory of algorithms and of Diophantine equations, when asked why he did not work on the problem, wrote that it seemed an interesting one but that he had not gotten around to it yet. He never would work on it.

The logical assault on Hilbert's Tenth Problem began around 1950, the first tentative papers appearing in the ensuing decade, the first major breakthrough appearing in print in 1961, and the ultimate solution being published in 1970. The first contributions were made by Julia Robinson and Martin Davis.

Robinson defined a relation R on natural numbers to be *Diophantine* if it could be written in the form

$$R(x_0, \dots, x_{n-1}) : \exists y_0 \cdots y_{m-1} P(x_0, \dots, x_{n-1}, y_0, \dots, y_{m-1}) = 0,$$

where P is a polynomial with integral coefficients and y_0, \dots, y_{m-1} range over *natural* numbers. (Logicians prefer their Diophantine equations to have nonnegative integral solutions, an inessential reformulation of the usual Fermatian Diophantine problem.) Finding she could not exhibit many demonstrably Diophantine relations, she allowed exponentiation to enter into P to form *exponential Diophantine* relations. She was able to show several interesting relations to be exponential Diophantine, and she reduced the general problem of showing all exponential Diophantine relations to be Diophantine to that of showing any relation of roughly exponential growth to have a Diophantine graph. In this reduction, she used the sequence of solutions to the special Pell equations

$$x^2 - (a^2 - 1)y^2 = 1, \quad a \geq 2.$$

Davis took a more logical approach. The theory of algorithms recognises two basic types of sets of natural numbers, namely: *recursive* sets, for which an algorithm determining membership exists, and *recursively enumerable* sets, for which an algorithmic enumeration exists. There are recursively enumerable sets which are not recursive. If every recursively enumerable set could be shown to be Diophantine, then Hilbert's Tenth Problem would have no effective solution. The techniques Gödel developed in proving his famous Incompleteness Theorems readily show that every recursively enumerable set can be written in the form

$$\exists y_0 Q_1 y_1 \cdots Q_{m-1} y_{m-1} P(x, y_0, \dots, y_{m-1}) = 0,$$

where each Q_i is either an existential quantifier or a *bounded* universal quantifier, i.e., a quantifier of the form $\forall y_i \leq y_0$. Davis simplified this representation to the *Davis Normal Form*

$$\exists y \forall z \leq y \exists w_0 \cdots w_{m-1} \leq y P(x, y, z, w_0, \dots, w_{m-1}) = 0.$$

Within a few years, Robinson's husband Raphael showed one could take $m = 4$.

Towards the end of the 1950s, Hilary Putnam joined Davis. Together they proved—modulo the unproved assumption of the existence of arbitrarily long arithmetic progressions of prime numbers—the unsolvability of the exponential Diophantine problem over the natural numbers. With Julia Robinson's help, the unproven

conjecture was bypassed. Together, Davis, Putnam, and Robinson applied Robinson's exponential Diophantine relations to eliminate the single bounded universal quantifier from the Davis Normal Form. Their proof was published in 1961.

With the Davis-Putnam-Robinson Theorem, Robinson's reduction of the problem of representing exponential Diophantine relations as Diophantine relations to the special problem of giving a Diophantine representation of a single relation of roughly exponential growth assumed a greater importance. The 1960s saw no progress in the construction of such a relation, merely a profusion of further reductions based on it; for example, on the eve of the final solution, Robinson showed it sufficient to prove the Diophantine nature of any infinite set of prime numbers. In March 1970 the world of logic learned that the then twenty-two-year-old Yuri Matiyasevich had shown the relation

$$y = F_{2x}$$

to be Diophantine, where F_0, F_1, \dots is the Fibonacci sequence. Very quickly, a number of researchers adapted Matiyasevich's proof to give a direct proof of the Diophantineness of the sequences of solutions to the special Pell equations

$$x^2 - (a^2 - 1)y^2 = 1, \quad a \geq 2,$$

cited earlier.

Matiyasevich's completion of the logical conquest of Hilbert's Tenth Problem was not the end of the story. There was more work to do, particularly, show the unsolvability of Diophantine equations and exponential Diophantine equations in a small number of variables, improve on Raphael Robinson's refinement of the Davis Normal Form, search for newer and better coding techniques, and so on. In all of this, Matiyasevich has been at the centre of research.

Matiyasevich's most important contribution since solving the problem has to be his introduction of new exponential Diophantine coding techniques. With such, he improved the initial Matiyasevich-Robinson small-number-of-variables result from the algorithmic unsolvability of the general 13-variable Diophantine problem to the algorithmic unsolvability of the 9-variable problem. He has also shown the unsolvability of the general exponential Diophantine problem in three variables. In further applying his new technique, he has given a direct proof of the Davis-Putnam-Robinson Theorem and a proof of the existence of *singlefold* exponential Diophantine representations of recursively enumerable sets.

A singlefold representation of a recursively enumerable set X would be one of the form

$$\begin{aligned} x \in X &\Leftrightarrow \exists y_0 \cdots y_{m-1} P(x, y_0, \dots, y_{m-1}) = 0 \\ &\Leftrightarrow \exists! y_0 \cdots y_{m-1} P(x, y_0, \dots, y_{m-1}) = 0, \end{aligned}$$

where $\exists!$ asserts "there are unique". The existence of singlefold exponential Diophantine representations of nonrecursive, recursively enumerable sets means that there are classes of exponential Diophantine equations for which we can—à la Thue, Siegel, and Roth—give effective bounds on the number of solutions but for which there are no effective bounds on the sizes thereof. Thus, the behaviour Baker ruled out for Thue et al. can and does occur in the exponential Diophantine case. It remains an open problem whether such behaviour can actually occur in the ordinary Diophantine case.

There has been enough progress in the field in the last twenty-odd years to merit a systematic exposition of the subject. The reviewer gave such an exposition a couple of years ago in the second chapter of his *Logical number theory I* (Springer-Verlag, Heidelberg, 1991), and now Matiyasevich himself has produced a volume on the subject. Of the two, Matiyasevich's exposition ought to be accessible to a wider mathematical audience.

Following a brief historical preface by Martin Davis, the book starts off with an introductory chapter giving the basic definitions and a few examples of Diophantine relations. The real work begins in Chapter 2 with Matiyasevich's proof that the graph of the exponential function is a Diophantine relation. The proof uses the sequence of solutions to the Pell equation but not the Pell equation itself: Let $\beta_0(a), \alpha_0(a), \beta_1(a), \alpha_1(a), \dots$ denote the sequence of pairs of nonnegative solutions to the special Pell equation for $a \geq 2$; i.e., let

$$\beta_n(a)^2 - (a^2 - 1)\alpha_n(a)^2 = 1.$$

The sequence of successive y -values, i.e., of successive elements of the α -sequence, are characterised by another equation

$$x^2 - axy + y^2 = 1.$$

The nonnegative solutions to this equation are just the pairs

$$x = \alpha_n(a) \text{ and } y = \alpha_{n+1}(a) \quad \text{or} \quad x = \alpha_{n+1}(a) \text{ and } y = \alpha_n(a).$$

Using this new equation allows one to ignore the β -sequences altogether.

Chapter 3, entitled "Diophantine Coding", discusses a few tricks for encoding finite sequences. There are the Cantor pairing function, Gödel's use of the Chinese Remainder Theorem to code finite sequences, and b -adic or "positional coding" of finite sequences. All three schemes are shown to be Diophantine in nature; i.e., the graphs of their projection functions are shown to be Diophantine relations. Via the b -adic coding the relation $z = \binom{x}{y}$, defining the binomial coefficients, is concluded to be Diophantine. It is at this point that the obligatory prime-enumerating polynomial can be exhibited. More important, however, Matiyasevich uses the binomial coefficients and an almost forgotten yet almost standard result of Eduard Kummer to show the Diophantine nature of the relation asserting numbers coding sequences in different bases to code the same sequence. Under older techniques, one would have to introduce a bounded quantifier at this point and destroy the Diophantine nature of the representation.

Chapter 4 gives some universal Diophantine equations, the existence of which is usually established recursion theoretically. The constructions of this chapter are purely number theoretical. From the existence of such an equation, a simple diagonal argument yields the existence of a Diophantine set, the complement of which is not Diophantine—a number-theoretic analogue (actually: reformulation) of the existence of nonrecursive, recursively enumerable sets.

It is only at this point that mathematical logic has to enter explicitly into the exposition. In Chapter 5 Matiyasevich introduces Turing machines to give formal definitions of recursiveness and recursive enumerability and shows the Diophantine relations to be precisely the recursively enumerable ones. For the logicians, I note that the Diophantine simulation of Turing machines uses Matiyasevich's special b -adic coding and avoids the introduction of the bounded universal quantifiers of the more familiar Gödelian techniques. The closure of the class of Diophantine relations

under bounded quantification is established in Chapter 6 by three different techniques, namely: demonstrating the closure of the class of recursively enumerable sets under such quantification by the construction of an appropriate Turing machine, the Davis-Putnam-Robinson exponential Diophantine (hence: Diophantine) simulation of a single bounded universal quantifier, and a new technique Matiyasevich calls summation. This new technique is a bit daunting, but it is the core of the proof of the unsolvability of exponential Diophantine equations in a small number of variables (and is, in fact, used later in the book for just this purpose), and the reader is advised not to skip it.

The rest of the book is devoted to the exploitation of the basic results, the most important being the construction of the singlefold exponential Diophantine representations of recursively enumerable sets, the three-variable exponential Diophantine result just cited, and the existence of algorithmically unsolvable problems in the calculus. Just about every application anyone has made of Matiyasevich's Theorem is cited if not proven, and the bibliography is fairly exhaustive.

CRAIG SMORYŃSKI
TEXAS TECH UNIVERSITY