DOI 10.1007/s10559-023-00605-y
Cybernetics and Systems Analysis, Vol. 59, No. 5, September, 2023

INCOMPLETENESS OF ARITHMETIC FROM THE VIEWPOINT OF DIOPHANTINE SET THEORY

A. M. Gupal[1†] **and O. A. Vagis**[1‡]

UDC 19.217.2

Abstract. *The authors analyze Diophantine sets and show that all recursively enumerable sets are Diophantine. Based on the classical results from the theory of recursive functions, a simple version of the theorem on the incompleteness of arithmetic is provided: there is a polynomial that has no positive integer solutions, and for which it is impossible to prove the absence of positive roots.*

Keywords: *Diophantine set, recursively enumerable sets, incompleteness of arithmetic.*

INTRODUCTION

The well-known Gödel's incompleteness theorem is a result releted to formal arithmetic, where the computability theory and logic are interrelated. The proof of incompleteness has a number of auxiliary statements and is still a challenge. Based on the classical results from the theory of recursive functions and Diophantine sets, a simple proof of the incompleteness of arithmetic is given.

In 1971, an important result was obtained about the undecidability of Hilbert's tenth problem, namely, whether there exists an algorithm that, given a polynomial $p(x_1,\dots,x_n)$ with integer coefficients, recognizes the existence of solutions to the equation $p=0$ in integers. Mathematician Yu. V. Matiyasevich showed that such an algorithm does not exist [1].

The main technical result obtained during the proof of the undecidability of Hilbert's tenth problem is a theorem on the coincidence of the class of Diophantine sets and of the class of recursively enumerable sets. The following result testifies to the computational capabilities of polynomials: it is possible to explicitly specify a polynomial of many variables with integer coefficients such that the set of all positive values that it takes for integer values of the variables is exactly the set of prime numbers.

DIOPHANTINE SETS

Diophantine equations are equations of the form

$$D(a_1,\dots,a_n,\ x_1,\dots,x_m)=0, \tag{1}$$

where D is a polynomial with integer coefficients with respect to all the variables $a_1,\dots,a_n,\ x_1,\dots,x_m$, divided into two parts: parameters $a_1,\dots,a_n$ and unknowns $x_1,\dots,x_m$. When the parameter values are fixed, specific Diophantine equations are obtained.

When choosing different parameter values, equations that have solutions and equations that do not have solutions are obtained. Parameters of the Diophantine equation (1) determine some set $\mathfrak{M}$, which consists of all the sets of the values of parameters $a_1,\dots,a_n$, for which there exist values of variables $x_1,\dots,x_m$ that satisfy Eq. (1):

$$\langle a_1,\dots,a_n\rangle\in\mathfrak{M}\Leftrightarrow\exists\, x_1,\dots,x_m\,[D(a_1,\dots,a_n,\ x_1,\dots,x_m)=0]. \tag{2}$$

[1]V. M. Glushkov Institute of Cybernetics, National Academy of Sciences of Ukraine, Kyiv, Ukraine, †*gupalanatol@gmail.com*; ‡*valexdep135@gmail.com*. Translated from Kibernetyka ta Systemnyi Analiz, No. 5, September–October, 2023, pp. 16–21. Original article submitted February 27, 2023.

 1060-0396/23/5905-0698

The number n is the dimension of the set $\mathfrak{M}$, and equivalence (2) is the Diophantine representation of the set $\mathfrak{M}$.

We need to solve the following problem: given a set consisting of n-tuples of natural numbers, establish whether this set is Diophantine, and if so, find any Diophantine representation for it. Sometimes, the Diophantinity of a set is trivial; for example, the set of all even numbers is obviously Diophantine. In other cases, it is technically difficult to detect Diophantinity, for example, for prime numbers.

Union of two Diophantine sets of the same dimension is a Diophantine set. Indeed, if $D_1(a_1,\ldots,a_n,\ x_1,\ldots,x_{m_1})=0$ and $D_2(a_1,\ldots,a_n,x_1,\ldots,x_{m_2})=0$ are Diophantine representations of two sets, then $D_1(a_1,\ldots,a_n,\ x_1,\ldots,x_{m_1})\times D_2(a_1,\ldots,a_n,\ x_1,\ldots,x_{m_2})=0$ is the Diophantine representation of their union.

Intersection of two Diophantine sets of the same dimension is also Diophantine and is determined by the equation

$$D_1^2(a_1,\ldots,a_n,\ x_1,\ldots,x_{m_1})+D_2^2(a_1,\ldots,a_n,\ x_1,\ldots,x_{m_2})=0.$$

It is shown that the complement of the set of n-tuples of natural numbers to the set of all n-tuples of natural numbers is not necessarily a Diophantine set, and this fact is not trivial at all.

It became possible to specify the type of equations in Diophantine representations by the set of natural numbers (i.e., in the case $n=1$). It is easy to understand that the equation

$$D(a,x_1,\ldots,x_m)=0 \tag{3}$$

has a solution in the unknowns $x_1,\ldots,x_m$ if and only if the equation

$$(x_0+1)(1-D^2(x_0,\ldots,x_m))-1=a \tag{4}$$

has a solution in the unknowns $x_0,\ldots,x_m$. Indeed, from an arbitrary solution of Eq. (3), it is possible to obtain solution (4) for $x_0=a$. On the other hand, the coefficient $(1-D^2(x_0,\ldots,x_m))$ must be positive, which is only possible if $D(x_0,\ldots,x_m)=0$. In this case, it follows from (4) that $x_0=a$, and this means that (3) also holds.

Thus, a set of natural numbers is Diophantine if and only if it is the set of all natural values acquired by some polynomial with integer coefficients for the natural values of the variables.

LOGICAL TERMINOLOGY

Instead of using sets, it is more convenient to use the properties and relations (predicates) between numbers. A property P of natural numbers is the Diophantine property if the set of all numbers that have the property P is Diophantine. Respectively, the equivalence of the form

$$\mathrm{P}(a)\Leftrightarrow\exists\, x_1,\ldots,x_m\,[D(a,x_1,\ldots,\ x_m)=0]$$

is called the Diophantine representation of the property P.

Similarly, relation R between n natural numbers is called a Diophantine relation if the set of n-tuples of natural numbers contained in the ratio R is Diophantine; respectively, equivalence of the form

$$R(a_1,\ldots,a_n)\Leftrightarrow\exists\, x_1,\ldots,x_m\,[D(a_1,\ldots,a_n,\ x_1,\ldots,x_m)=0]$$

is called the Diophantine representation of the relation R.

According to this terminology, union of sets is associated with the construction of a new property (relation) using the logical operation “or” (disjunction). In other words, if R_1 and R_2 are two Diophantine relations (or properties), then relation R (property) such that $R(a_1,\ldots,a_n)\Leftrightarrow R_1(a_1,\ldots,a_n)\vee R_2(a_1,\ldots,a_n)$ holds for all $a_1,\ldots,a_n$, is also Diophantine.

Similarly, the logical operation “and” (conjunction) corresponds to intersection of sets. The equivalence of the type $R(a_1,\ldots,a_n)\Leftrightarrow R_1(a_1,\ldots,a_n)\&\,R_2(a_1,\ldots,a_n)$ is also considered as a general Diophantine representation of the relation R if R_1 and R_2 were proved to be Diophantine.

The concept of a Diophantine function as a function whose graph is a Diophantine set is an important method of establishing Diophantinity. Respectively, a Diophantine representation of the function F is the equivalence of the type

$$a = F(b_1,\ldots,b_n) \Leftrightarrow \exists\, x_1,\ldots,x_m [D(a,b_1,\ldots,b_n,\ x_1,\ldots,x_m) = 0],$$

where D is a polynomial with integer coefficients.

EXAMPLES OF DIOPHANTINE SETS, RELATIONS, AND FUNCTIONS

Let us give some simple examples: Diophantinity of inequalities $\le$ and $<$

$$a \le b \Leftrightarrow \exists\, x[a + x = b],\ \ a < b \Leftrightarrow \exists\, x[a + x + 1 = b].$$

Diophantinity of the Set of all Even Numbers: $\exists\, x[a = 2x]$. Diophantinity of the division relation $a \mid b \Leftrightarrow \exists\, x[a\,x = b]$ makes it possible to give a general representation of the Diophantinity of the function rem (b, c), which is a remainder of the division of b by c: $a = \mathrm{rem}(b,c) \Leftrightarrow a < c \,\&\, c \mid b - a$.

Based on the Diophantinity of rem, we can give a general Diophantine representation of the function $b \operatorname{div} c$, which is the integer part of the division of b by c:

$$a = b \operatorname{div} c \Leftrightarrow a\,c + \mathrm{rem}(b,c) = b$$

and the general Diophantine representation of the ternary comparison positive modulo $a = b \pmod c$ $= \mathrm{rem}(a,c) = \mathrm{rem}(b,c)$. The function $b \operatorname{div} c$ plays an important role in establishing the Diophantinity of the power function b^c.

The general Diophantine representation of the function *GCD* (the greatest common divisor of positive integers) is obtained:

$$a = GCD(b,c) \Leftrightarrow bc > 0 \,\&\, a \mid b \,\&\, \exists\, x\, y[a = bx - cy].$$

LANGUAGES FOR DESCRIPTION OF DIOPHANTINE SETS

Language $\mathcal{A}_0 = \{+, \times, =, \exists\}$, where $\exists$ is the existence quantifier, makes it possible to represent any polynomial $p(a,x)$ and make conclusions in the form $\exists\, x p(a,x) = 0$, where $a = \{a_1,\ldots,a_k\}$, $x = \{x_1,\ldots,x_m\}$. Then it is possible to characterize Diophantine sets as such and only such that are representable by the language $\mathcal{A}_0$. However, the capabilities of language $\mathcal{A}_0$ are limited and it cannot be used to express many sets interesting from the number-theoretic point of view. For example, a set of prime numbers can be expressed by the formula

$$p > 1 \,\&\, \forall\, y \le p\, \forall z \le p[y\,z \ne p \vee y = 1 \vee z = 1],$$

which includes the limited universal quantifier $\forall_{\le}$.

Yu. V. Matiyasevich studied languages $\mathcal{A}_1 - \mathcal{A}_5$, the capabilities of each subsequent language being increased. Collectively, they contain symbols $+, \times, \uparrow$ (the exponentiation operator), $=, \ne, >, \ge, \mid$ (division operation), limited universal quantifier $\forall_{\le}$ (mod), $\&$, $\vee$, $\exists$, as well as some functions. Language $\mathcal{A}_5$ can be obtained by adding the limited universal quantifier $\mathcal{A}_5 = \mathcal{A}_4 \bigcup \{\forall_{\le}\}$.

Proving the Diophantinity of the function of two arguments b^c and of the limited universal quantifier turned out to be technically difficult. Based on the Diophantinity of the function b^c, the Diophantinity of the binomial coefficients and of the factorial was obtained, and the Diophantine representation of prime numbers in a different form was given:

$$Prime(a) \Leftrightarrow a > 1 \,\&\, GCD(a,\ (a-1)!) = 1.$$

Matiyasevich proved the equal volumes of the languages $\mathcal{A}_0$ and $\mathcal{A}_5$, i.e., that any set expressed in one language can also be expressed in the other language. This implies that the set of all prime numbers is Diophantine.

On the other hand, in the theory of recursive functions, it is established that the class of sets described by the language $Я_5$ coincides with the class of recursively enumerable (r.e.) sets. Thus, based on the equivalent volume of languages $Я_0$ and $Я_5$, it is proved that the class of Diophantine sets coincides with the class of r.e. sets. One of the classical results in the theory of recursive functions is the theorem on the existence of a r.e. set (given below) for which there is no means that allows to recognize, in a finite number of steps, the presence or absence of an arbitrary number in this set. This result, together with the Diophantinity of r.e. sets, gives a negative solution to Hilbert's tenth problem [1].

TERMINOLOGY FROM THE THEORY OF RECURSIVE FUNCTIONS

Let us provide some necessary information from the theory of recursive functions. Without loss of generality, we will consider the case of one-parameter Diophantine sets.

Let $p(x, y_1, \ldots, y_m)$ be a polynomial with integer coefficients. Then a predicate (relation) $M(x)$ given by the formula

$$M(x) = \exists\, y_1, \ldots,\ \exists\, y_m\, (p(x, y_1, \ldots, y_m) = 0)$$

is called a Diophantine predicate, and the domain of quantifiers $\exists\, y_1, \ldots, \exists\, y_m$ is a set of natural numbers.

THEOREM 1. Diophantine predicates are partially solvable.

Proof. Predicate $p(x, y_1, \ldots, y_m) = 0$ is solvable because it can be checked by substituting the values of the variables x and $y_1, \ldots, y_m$ into the polynomial. There is a theorem [2] that shows that a predicate of the form

$$M(x) = \exists\, y_1, \ldots,\ \exists\, y_m\, (p(x, y_1, \ldots, y_m) = 0)$$

is partially solvable.

It is clear that among the partially solvable predicates, Diophantine ones are those that can be represented in a relatively simple form. For a long time, unsolvable Diophantine predicates were unknown to exist. This problem is related to Hilbert's tenth problem. That is why, the result proved by Matiyasevich in 1971 turned out to be a remarkable achievement.

THEOREM 2. Every partially solvable predicate is Diophantine.

As is known from the theory of recursive functions, problem $x \in W_x$ is unsolvable, where W_x is the domain of definition of the partial computable function ϕ_x, which is executed by the program P_x with subscript x. This program terminates at the input x. Program P_x consists of a finite set of commands that can be encoded with a natural number.

The characteristic function of this problem, which is given by the formula

$$f(x) = \begin{cases} 1 & \text{if } x \in W_x, \\ 0 & \text{if } x \notin W_x \end{cases}$$

is not a computable function. Otherwise, the computable function $g(x)$, which differs from any computable function, can be constructed using the diagonal method [2]:

$$g(x) = \begin{cases} 1 & \text{if } x \notin W_x, \\ \text{not defined} & \text{if } x \in W_x. \end{cases}$$

Definition 1. Let A be a subset of the set of natural numbers N. The characteristic function of A is the function

$$C_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Then the set A is called recursive if C_A is a computable function or if a predicate $x \in A$ is solvable.

Thus, the set $K = \{x \mid x \in W_x\}$ is not recursive. For $K = \{x \mid x \in W_x\}$, there is no program that would allow to recognize, in a finite number of steps, the presence or absence of an arbitrary number $n = 0, 1, 2, \ldots$ in this set.

Although the predicate $x \in W_x$ has a non-computable characteristic function, the next function related to this problem is computable

$$f(x) = \begin{cases} 1 & \text{if } x \in W_x, \\ \text{not defined} & \text{if } x \notin W_x. \end{cases}$$

If we assume that one encodes the answer "yes," then any algorithm that calculates f is a procedure that returns "yes" when $x \in W_x$ but continues its operation indefinitely if the statement $x \in W_x$ does not hold. Such a procedure is called partially solvable for the problem $x \in W_x$; moreover, this problem or this predicate are considered to be partially solvable.

Definition 2. A predicate $M(x)$ is called partially solvable if the function

$$f(x) = \begin{cases} 1 & \text{if } M(x) \text{ is true}, \\ \text{not defined} & \text{if } M(x) \text{ is false} \end{cases}$$

is computable. This function is called partial characteristic function of $M(x)$.

Definition 3. Let A be a subset of the set N. Then the set A is recursively enumerable if the function f defined by the formula

$$f(x) = \begin{cases} 1 & \text{if } x \in A, \\ \text{not defined} & \text{if } x \notin A \end{cases}$$

is computable (or if the predicate $x \in A$ is partially solvable).

The set A is shown to be r.e. if and only if it is the definition domain of a computable function [2].

It is much easier to establish r.e. of any set than to discover the Diophantinity of this set. For example, the set of prime numbers is naturally r.e.; however, it is not obvious that it is Diophantine. This r.e. set turned out to be the set of positive values of some polynomial (it has 26 variables [3]). This statement was considered highly implausible until Matiyasevich proved his theorem.

INCOMPLETENESS OF THE ARITHMETIC AND DIOPHANTINE SETS

It is clear that a Diophantine set is recursively enumerable; therefore, all recursively enumerable sets from the Matiyasevich theorem are Diophantine.

Since the predicate $x \in K$ is partially solvable, the set $K = \{x \mid x \in W_x\}$ is recursively enumerable. The set K is the definition domain of the r.e. set W_i of some computable function ϕ_i.

Let us now choose the polynomial $p(a, y_1, \ldots, y_m)$ such that

$$a \in W_a \Leftrightarrow \exists\, y_1, \ldots, \exists\, y_m \; (p(a, y_1, \ldots, y_m) = 0). \tag{5}$$

This can be done as a result of the Diophantinity of the r.e. set K. In formula (5), the Diophantine predicate $\exists\, y_1, \ldots, \exists\, y_m \; (p(a, y_1, \ldots, y_m) = 0)$ is a formal analog of the predicate $x \in K$, and this is a key result obtained by Matiyasevich. In Gödel's construction, a significant part of Gödel's proof is focused on obtaining a formal analog of the predicate $x \in K$ [4].

Consider a function defined as follows:

$$F(a) = \begin{cases} 1 & \text{if } \exists\, y_1, \ldots, \exists\, y_m \; (p(a, y_1, \ldots, y_m) = 0), \\ 0 & \text{otherwise.} \end{cases}$$

If there existed a solution procedure for Hilbert's tenth problem, it would be possible to efficiently calculate the value of the function F, i.e., the problem $a \in W_a$ would be solvable; however, it is impossible.

A simple version of Gödel's theorem on the incompleteness of arithmetic follows from the theory of Diophantine sets. Since Diophantine sets are r.e., all the equations $p(x_1, \ldots, x_k) = 0$ that have integer solutions can be enumerated recursively.

Statements about the absence of solutions for the equations

$$\forall x_1, \ldots, \forall x_k \, p(x_1, \ldots, x_k) \neq 0 \tag{6}$$

remain enumerable.

If we assume that for some consistent system of axioms it is possible to prove all the facts of the form (6), this would mean the existence of an algorithm that enumerates theorems (6). Since any proofs are finite, all of them form a recursively enumerable set [2], i.e., this algorithm enumerates the complements to the set of polynomials that have positive solutions. According to the well-known theorem [2], a set A is recursive (solvable) if and only if the sets A and $\overline{A}$ are r.e. Therefore, we get the solvability of the set of polynomials that have positive roots (in this case, the set of polynomials (6) is also solvable). Thus, we have a contradiction with the negative solution of Hilbert's tenth problem.

THEOREM 3. For any consistent theory containing arithmetic, there exists a polynomial that has no positive integer solutions and for which it is impossible to prove the absence of natural roots.

CONCLUSIONS

An analysis of Diophantine sets has shown that all recursively enumerable sets are Diophantine. Based on the classical results from the theory of recursive functions, it is possible to present a simple version of the theorem on the incompleteness of arithmetic: there is a polynomial that has no positive integer solutions and for which the absence of natural roots cannot be proved.

REFERENCES

1. Yu. V. Matiyasevich, "Diophantine sets," Uspekhi Mat. Nauk, Vol. 2, Iss. 5, 185–222 (1971).
2. N. J. Cutland, Computability: An Introduction to Recursive Function Theory, Cambridge Univ. Press (1980).
3. Yu. V. Matiyasevich, Hilbert's Tenth Problem [in Russian], Nauka, Moscow (1993).
4. E. Mendelson, Introduction to Mathematical Logic, Book World Promotions (1964).