PANORAMAS OF MATHEMATICS
BANACH CENTER PUBLICATIONS, VOLUME 34
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 1995

DIAGONAL REASONINGS IN MATHEMATICAL LOGIC

ZOFIA ADAMOWICZ

Institute of Mathematics of the Polish Academy of Sciences
Śniadeckich 8, 00-950 Warszawa, Poland
E-mail: zosiaa@impan.gov.pl

First we show a few well known mathematical diagonal reasonings. Then we concentrate on diagonal reasonings typical for mathematical logic.

1. Examples of mathematical diagonal reasonings.

Theorem 1 (Cantor's Theorem). *The set of reals is uncountable.*

To prove the theorem we show that the set of sequences of zeros and ones, that is, the set of functions f such that $f : N \longrightarrow \{0, 1\}$, is uncountable.

Indeed, for every sequence of functions (f_n) there is a function f which is not a term of the sequence. We define f as follows:

$$f(n) = \begin{cases} 0 & \text{if } f_n(n) = 1 \\ 1 & \text{if } f_n(n) = 0 \end{cases} \tag{1}$$

Hence it follows that all such functions cannot be arranged in a sequence.

Cantor's construction of the reals

A real is here an appropriate equivalence class of a Cauchy sequence f. If we are given a sequence of sequences

$f_0 :\ \ (f_0)_0, (f_0)_1 \ldots$
$f_1 :\ \ (f_1)_0, (f_1)_1 \ldots$
$\ldots,$

which itself is a Cauchy sequence, then it is convergent to a certain Cauchy sequence which roughly is the diagonal of the above matrix.

1991 *Mathematics Subject Classification*: Primary 03B25.

Lecture given at the Banach Center Colloquium on 13th January 1994.

The paper is in final form and no version of it will be published elsewhere.

THEOREM 2 (Baire's theorem). *A first category set in a complete (compact) space is meager.*

Outline of a proof. Let A be a first category set, $A = \bigcup_n A_n$, where A_n are nowhere dense. We have to show that in every ball K there is an element x such that $x \notin A$. Let $K = K_0$. Let $K_1 \subseteq K$ be disjoint with A_1. Let $x_1 \in K_1$. Let $K_2 \subseteq K_1$ be disjoint with A_2. We take $x_2 \in K_2$. We continue. At the same time we ensure that (x_n) is a Cauchy sequence — the balls are chosen in such a way that their radii converge to zero. We take $x = \lim x_n$. Then $x \notin A$.

We may treat the above proof as a diagonal reasoning — in the nth step we guarantee that $x \notin A_n$.

Another example of a diagonal reasoning:

THEOREM 3. *There is a function from N to N which is not definable.*

Here we have to make precise what is meant by definability.

We are given the set of positive integers N with the functions $+, \cdot$ and relations $=, <$ and with the distinguished elements $0, 1$; i.e. we are given the relational structure $\mathbb{N} = \langle N, +, \cdot, =, <, 0, 1\rangle$.

On the other hand we are given the language: the variables $x_1, x_2, x_3, \ldots$, the relation and function symbols $+, \cdot, =, <$, the constants $0, 1$ (the symbols $+, \cdot, =, <, 0, 1$ are used here in two different meanings — as functions, relations and numbers and as symbols of the language), the connectives $\vee, \wedge, \neg$ and the quantifiers $\exists, \forall$. Now we define a formula of this language. By the *terms* of this language we mean the symbols of the form such as e.g.:

$$(((x_{i_1} + x_{i_2}) \cdot x_{i_3} + x_{i_4}) \cdot x_{i_5} \cdot x_{i_6}) + x_{i_7}. \tag{2}$$

A formula may be atomic, of the form $t_1 = t_2$, $t_1 < t_2$, where t_1, t_2 are terms, or more complex, e.g. $\exists x_1\ (x_1 + 0 = x_1)$. More complex formulas are obtained by joining the simpler ones with the use of the connectives or by adding quantifiers to the simpler ones.

A set $A \subseteq N^k$ is definable if there is a formula $\phi(x)$ such that

$$A = \{\langle n_1, \ldots, n_k\rangle \in N^k :\ \phi(n_1, \ldots, n_k)\}, \tag{3}$$

e.g.

$$A = \{n \in N :\ \exists m\ (n = m + m)\} \tag{4}$$

— the set of even numbers,

$$f = \{\langle n, m\rangle :\ m \cdot m < n < (m+1) \cdot (m+1)\ \vee\ m \cdot m = n\} \tag{5}$$

— the function $f(n) = [\sqrt{n}]$.

Now we show that there is a nondefinable function from N to N.

Since the language is countable, there are countably many definitions in it (that is countably many of the appropriate formulas ϕ). Thus there are countably many definable functions. Let us arrange all such functions in the sequence:

f_0: $f_0(0), f_0(1), \ldots$
f_1: $f_1(0), f_1(1), \ldots$
$\ldots$

and define $f(n) = f_n(n) + 1$. Then the function f is not a term of this sequence, and thus is not definable.

Similarly, we show that there is an ordinal number which is not definable. Consider the language of set theory. Here we have two relation symbols $=, \in$. As before, the language is countable, and thus there are countably many definable ordinal numbers. Let us denote these numbers by $\alpha_1, \alpha_2, \ldots$. Let α be the least ordinal number greater than all these numbers. Then α is different from all the α_i, and thus is not definable.

Here we have obtained one of the well known "paradoxes" of the beginning of our century — on one hand α is not definable, and on the other we just have defined it.

The case of our function f is similar — we have given its definition.

We explain this paradox in section 3.

2. Universal relations. Consider the family of open sets in the Baire space N^N. As the basis of the topology we take the sets B_s determined by finite sequences $s = \langle\langle n_1, m_1\rangle, \ldots, \langle n_k, m_k\rangle\rangle$ of pairs of natural numbers:

$$B_s = \{f \in N^N : \ f(n_1) = m_1 \& \ldots \& f(n_k) = m_k\}. \tag{6}$$

The basis is countable, we may enumerate it $B_1, B_2, \ldots$.

Let $f \in N^N$. Let A_f denote the open set $\bigcup_i B_{f(i)}$.

Now consider the set $A = \{\langle f, g\rangle : \ g \in A_f\}$. It is easy to see that A is open in $N^N \times N^N$ with the product topology. We can look at A as at a plane set

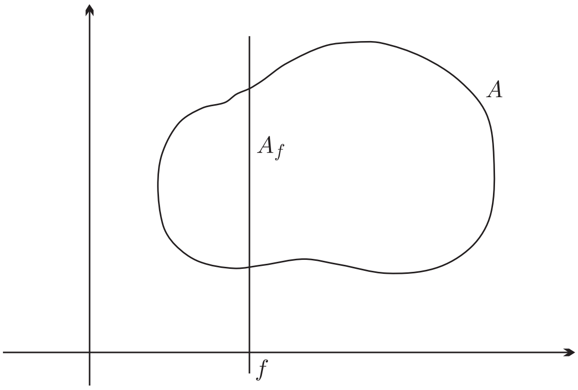


Fig. 1

where at the axes we put N^N. Then every vertical section of A (as on the picture) determines a certain open set A_f and conversely, every open set in N^N is a certain such section.

We say that A is a universal relation for open sets in N^N.

In this case there is a universal relation for open sets in N^N which itself is an open set.

Similarly we may define a universal relation for Borel sets in N^N. As we shall see, this relation is no more Borel.

We have the following property:

Theorem 4. *If we are given a universal relation for a certain family of sets then it determines a set which is not in the family.*

For example, consider again our relation $A(f,g)$ universal for open sets. Let the set B be defined as follows:

$$f \in B \Leftrightarrow \neg A(f,f). \tag{7}$$

We show that B is not open. Indeed, suppose that B is open. Then there exists g such that $B = A_g$. We have

$$g \in A_g \Leftrightarrow \neg A(g,g) \Leftrightarrow g \notin A_g, \tag{8}$$

contradiction. Thus the set B is not open (it is closed).

Let now $A(f,g)$ be a universal relation for Borel sets. Let B be defined as above. Similarly as before we show that B is not Borel. But notice that if A was Borel then B would also be Borel (here we make use of the fact that the family of Borel sets is closed under complementation — unlike for open sets). Hence A is not Borel.

It can be shown that the relation A can be chosen in such a way that it is a continuous image of a Borel set. Hence it follows that a continuous image of a Borel set is not necessarily Borel.

Here we have an opportunity to mention a famous mistake of Lebesgue — in one of his papers Lebesgue studied continuous images of Borel sets and claimed that they were Borel. This was one of those mistakes in the history of mathematics which turned out to inspire its development — in this case the development of the theory of the analytic sets — exactly continuous images of Borel sets.

Again one has to refer to Lebesgue when speaking about universal relations — this notion occurred for the first time in the paper of Lebesgue of 1905, in which he investigated universal relations for particular classes of Borel sets.

To end this section we show that the proof of the theorem about the nonexistence of the set of all sets can be presented as an application of the above method.

We show that the class $A = \{x : x \text{ is a set } \& \ x \notin x\}$ is not a set (Russel's paradox). Consider the universal relation $\phi(x,y)$ for relations $x(y)$ defined as $y \in x$, where x is a set. We have

$$\phi(x,y) \Leftrightarrow y \in x. \tag{9}$$

Then $A = \{x : \neg\phi(x,x)\}$. In view of what we have already shown, A does not lie in the domain of the universal relation ϕ, and thus is not a set.

3. Universal formulas. Instead of universal relations we may speak about universal formulas — definitions of those relations. Let us come back to arithmetic. There are countably many formulas of the language of arithmetic, thus we may enumerate them with numbers, and moreover we may do it in an effective way. We may even, up to this enumeration, identify formulas with the appropriate numbers. Let us ask whether there exists a universal relation for sets definable in $\mathbb{N}$. That is, whether there exists such a relation $A(\varphi, x)$ that the appropriate vertical section A_φ is the set defined by φ (cf. Fig. 1). That is, we look for a relation $A \subseteq N \times N$ satisfying the condition:

$$A(\varphi, x) \Leftrightarrow x \in A_\varphi \Leftrightarrow \varphi(x). \tag{10}$$

Of course, there is a set A with the above property, defined as above. However, we may ask whether A itself is definable. Let us pose the following question:

Is there a formula $\phi(\varphi, x)$ such that

$$\phi(\varphi, x) \Leftrightarrow \varphi(x) \tag{11}$$

for all the formulas φ?

Here we enter the question of the existence of universal formulas for classes of formulas, i.e. the existence of formulas ϕ having the property $\phi(\varphi, x) \Leftrightarrow \varphi(x)$, where φ runs over a certain class of formulas. We may also consider universal formulas for classes of sentences, i.e. formulas having the property $\phi(\varphi) \Leftrightarrow \varphi$, where φ runs over a certain class of sentences. This is a kind of *speaking about speaking.* Let us recall a famous example of Tarski. We may say

It is snowing

and we may also say

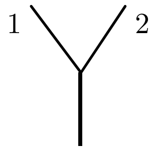
The sentence "it is snowing" is true.

Each of these sentences is true if it is really snowing.

If ϕ is a universal formula for sentences φ, then the formulation of the sentence φ corresponds to the sentence "It is snowing" and the formulation of the sentence $\phi(\varphi)$ corresponds to the sentence "The sentence 'it is snowing' is true".

Digression — a story about brothers.

At a splitting of roads



there live two brothers A and B. The brother A always tells truth, and the brother B always lies. A traveller goes to a town M. He stops at the splitting, he meets one of the brothers (he does not know which one) and he is allowed to ask just one question to learn the correct way.

It turns out that the appropriate question requires a reference to "speaking about speaking". Namely, the question is

Which way would your brother show me?

It is easy to check that no matter what answer the traveller gets he should choose the other way.

Let us try to interpret this story. Let p_i $(i = 1, 2)$ be the sentence "You should take the way i". Let $\phi_A(p)$ be the formula "A says the sentence p", and $\phi_B(p)$ "B says the sentence p".

We have $\phi_A(p) \Leftrightarrow p$ (i.e. ϕ_A is a universal formula for the sentences p) and $\phi_B(p) \Leftrightarrow \neg p$.

If the answer to the question is p_i and the brother met is A, then we have $\phi_A(\phi_B(p_i))$, and thus $\phi_B(p_i)$, i.e. $\neg p_i$. If the brother met is B then we have $\phi_B(\phi_A(p_i))$, and thus $\neg\phi_A(p_i)$, i.e. $\neg p_i$.

We have the following theorem

THEOREM 5. *There is no universal formula for all formulas (of one variable). There is no universal formula for all sentences.*

Proof. Suppose that ϕ is a universal formula for all formulas. Then we have

$$\phi(\varphi, x) \Leftrightarrow \varphi(x) \tag{12}$$

for all formulas $\varphi(x)$. Consider the formula $\psi(x)$: $\neg\phi(x, x)$. Then we have

$$\neg\phi(\psi, \psi) \Leftrightarrow \psi(\psi) \Leftrightarrow \phi(\psi, \psi)), \tag{13}$$

contradiction.

The second part of Theorem 5 immediately follows from the theorem of Gödel:

THEOREM 6 (Gödel's diagonal lemma). *For any formula $\psi(x)$ there is a sentence φ such that φ is true if and only if $\psi(\varphi)$ is true.*

The lemma says that for any property $\psi(x)$ there is a sentence φ which has the meaning "I have the property ψ".

Suppose now that $\phi(x)$ is a universal formula for all sentences. Ley ψ be the sentence from the Gödel diagonal lemma for the formula $\neg\phi$. Then we have

$$\neg\phi(\psi) \Leftrightarrow \psi \Leftrightarrow \phi(\psi), \tag{14}$$

contradiction.

From the Gödel diagonal lemma we also easily infer the following theorem:

THEOREM 7 (Tarski's theorem on nondefinability of truth). *The set of sentences of the language of arithmetic that are true in $\mathbb{N}$ is not definable in $\mathbb{N}$ by a formula of this language.*

Proof. Suppose that $\phi(x)$ defines the set of sentences true in $\mathbb{N}$. Thus we have

$$\phi(\varphi) \Leftrightarrow \varphi \tag{15}$$

for all sentences φ.

Let now ψ be defined as in the previous proof, that is ψ holds if and only if $\neg\phi(\psi)$ holds. If ψ is true, then on one hand $\neg\phi(\psi)$ holds, by the choice of ψ, and on the other hand $\phi(\psi)$ holds, since ϕ defines the set of the true sentences. Similarly, if ψ is false, then on one hand $\phi(\psi)$ holds, by the choice of ψ, and on the other hand $\phi(\psi)$ does not hold, since ψ does not belong to the set of true sentences. We obtain a contradiction.

The above theorem holds not only for arithmetic, but it is quite general. It holds for most of the mathematical theories, in particular for set theory.

Therefore, we cannot express in a given language the notion of truth for sentences of the language. In particular we are not able to express the fact that the number n belongs to the set defined by the formula $\varphi(x)$ — that $\varphi(n)$ is true. Thus, there is no universal formula for the family of definable sets — the answer to the question posed at the beginning of this section is negative. In particular, the function diagonalizing the definable functions and the ordinal number defined in section 1 are not defined in that language to which the notion of definability there considered refers.

4. Tarski's truth definition. Up to now we have said about a sentence that it is "true" or about a formula $\phi(x)$ that it "holds" for a number n, in an intuitive way. The notion of the satisfiability of a formula $\phi(x_1, \ldots, x_k)$ in a given relational structure by the sequence $\langle n_1, \ldots, n_k\rangle$ of elements of the universe of the structure may be defined in a precise way. Again, let us do it for arithmetic, for another language or another structure this can be done similarly.

If t is a term, for instance the term considered in section 1

$$t = (((x_{i_1} + x_{i_2}) \cdot x_{i_3} + x_{i_4}) \cdot x_{i_5} \cdot x_{i_6}) + x_{i_7}, \tag{16}$$

then by the value of this term at the sequence $\langle n_{i_1}, \ldots, n_{i_k}\rangle$, $t(n_{i_1}, \ldots, n_{i_k})$, we mean the number

$$(((n_{i_1} + n_{i_2}) \cdot n_{i_3} + n_{i_4}) \cdot n_{i_5} \cdot n_{i_6}) + n_{i_7}. \tag{17}$$

The atomic formula $t_1 = t_2$ or $t_1 < t_2$ is satisfied in $\mathbb{N}$ by the sequence $\langle n_{i_1}, \ldots, n_{i_k}\rangle$ if respectively

— the natural number $t_1(n_{i_1}, \ldots, n_{i_k})$ is equal to the number $t_2(n_{i_1}, \ldots, n_{i_k})$

or

— the number $t_1(n_{i_1}, \ldots, n_{i_k})$ is less than $t_2(n_{i_1}, \ldots, n_{i_k})$.

Further on we proceed inductively.

— $\neg\psi(x_{i_1}, \ldots, x_{i_k})$ is satisfied by $\langle n_{i_1}, \ldots, n_{i_k}\rangle$ if ψ is not satisfied by $\langle n_{i_1}, \ldots, n_{i_k}\rangle$.

— $\psi_1 \vee \psi_2(x_{i_1}, \ldots, x_{i_k})$ is satisfied by $\langle n_{i_1}, \ldots, n_{i_k}\rangle$ if ψ_1 is satisfied or ψ_2 is satisfied by $\langle n_{i_1}, \ldots, n_{i_k}\rangle$.

— $\psi_1 \wedge \psi_2(x_{i_1}, \ldots, x_{i_k})$ is satisfied by $\langle n_{i_1}, \ldots, n_{i_k}\rangle$ if ψ_1 is satisfied and ψ_2 is satisfied by $\langle n_{i_1}, \ldots, n_{i_k}\rangle$.

— $\exists x\ \psi(x, x_{i_1}, \ldots, x_{i_k})$ is satisfied by $\langle n_{i_1}, \ldots, n_{i_k}\rangle$ if there exists a number n in N such that ψ is satisfied by $\langle n, n_{i_1}, \ldots, n_{i_k}\rangle$.

As we see, at one side of these definitions there occur symbols of our language — the one under consideration, about which we speak, and at the other side the words "not, or, and, there exists" of the language in which we speak (called metalanguage). As we showed before it is not possible to express the above definition in the language under consideration — truth can be defined only from outside.

5. First and second Gödel's theorems. Consider the declaration "I am lying". Observe that it is neither true nor false — if I am telling truth then I am lying, and if I am lying then I am telling truth.

Is the sentence "I am lying" expressible in the language of arithmetic?

We are looking for a sentence φ such that φ was equivalent with the sentence "φ is not true". However the property "is not true" cannot be expressed in our language — since we cannot express the property "is true". Indeed, by the Tarski theorem on the nondefinability of truth, there is no arithmetical formula $\phi(\varphi)$ meaning "φ is true". We cannot express the sentence "I am lying" as a mathematical sentence. However, we may express a slightly different sentence, namely the sentence "I am not provable". There is an arithmetical formula T such that $T(\varphi)$ has the meaning "φ has a proof in arithmetic (is a theorem of arithmetic)". Now let us outline the construction of the formula T.

First, let us make precise what theory is meant by arithmetic. Let this theory be denoted by P (from Peano). The axioms of the arithmetic P are the sentences:

$$\forall x, y\ \ x + y = y + x, \ \ \forall x, y\ \ x \cdot y = y \cdot x \tag{18}$$

$$\forall x, y, z\ \ (x + y) + z = x + (y + z), \ \ \forall x, y, z\ \ (x \cdot y) \cdot z = x \cdot (y \cdot z) \tag{19}$$

$$\forall x, y, z\ \ (x + y) \cdot z = x \cdot z + y \cdot z \tag{20}$$

$$\forall x\ \ x + 0 = x, \quad x \cdot 1 = x \tag{21}$$

$$\forall x, y, z\ \ (x + z = y + z \Rightarrow x = y) \tag{22}$$

$$\forall x\ \ (x \neq 0 \Leftrightarrow \exists y\ x = y + 1) \tag{23}$$

$$\forall x, y\ \ (x < y \Leftrightarrow \exists z\ x + z + 1 = y) \tag{24}$$

and all the sentences:

$$(\varphi(0) \wedge \forall\ \ x(\varphi(x) \Rightarrow \varphi(x + 1))) \Rightarrow \forall x\ \ \varphi(x), \tag{25}$$

where φ is a formula of the language.

Here we have used the connectives $\Rightarrow$ and $\Leftrightarrow$ which were not introduced in the definition of the language — one has to replace them by the appropriate combinations of the connectives $\neg, \vee, \wedge$.

Thus, the arithmetic P is a certain (infinite) set of sentences. It is easy to see that this set of sentences is definable in $\mathbb{N}$ — it is a set of sentences of a particular form which can be described in the language of arithmetic. Let $P(x)$ denote the formula defining this set of sentences in $\mathbb{N}$.

Let now $d = \langle \psi_1, \ldots \psi_n \rangle$ be a sequence of formulas. Sequences of numbers can be treated as numbers — we identify them with their numbers under a certain effective enumeration of sequences. We say that d is a proof of the sentence φ in the theory P, if ψ_n is the sentence φ, and every ψ_i is either an axiom ($P(\psi_i)$ holds) or there are $j, k < i$ such that ψ_k is the formula $\psi_j \Rightarrow \psi_i$ — that is, ψ_i can be derived from the previous formulas by the *modus ponens* rule. It is easy to see that the above description can be carried out in arithmetic — thus there is a formula $D(d, \varphi)$ expressing the meaning "d is a proof φ in P".

Now we can define our formula $T(\varphi)$ as $\exists d\ D(d, \varphi)$.

THEOREM 8 (First Gödel's theorem). *There is an arithmetical sentence φ independent from arithmetic such that both φ and $\neg\varphi$ have no proof in arithmetic.*

Proof. Let φ be the sentence from the Gödel diagonal lemma for the formula $\neg T$. Then we have:

$$\varphi \text{ holds if and only if } T(\varphi) \text{ does not hold.}$$

Thus φ has the meaning "I am not provable".

Suppose that φ has a proof. Then $T(\varphi)$ holds, contradiction.

Suppose now that $\neg\varphi$ has a proof. Then $\neg\varphi$ is true. In this case φ is false, and thus $T(\varphi)$ holds. Thus φ has a proof in P. Hence both $\neg\varphi$ and φ have proofs in P, contradiction.

Again, this theorem concerns not only arithmetic, but almost every mathematical theory. In particular it is true for set theory. This means that there are sentences independent from set theory. Moreover, even if we add such a sentence to set theory as an axiom, then we obtain a theory for which again the first Gödel theorem holds, and thus again there are sentences independent from that theory.

We see that this theorem puts bounds to our ability of knowing — there are true sentences which we cannot prove — we cannot grasp the whole truth.

Notice that we are able to formulate in arithmetic a sentence with the meaning "the arithmetic is consistent". Indeed, let $Cons(P)$ be the sentence $\forall d\ \neg D(d, \text{“}0 = 1\text{”})$ — the contradiction "$0 = 1$" has no proof in P.

THEOREM 9 (Second Gödel's theorem). *There is no proof of the sentence $Cons(P)$ in P.*

Similarly as before, this theorem concerns not only the theory P, but most theories. It can be read as:

In a given theory it is not possible to prove the consistency of this theory.

PROBLEM. *Is it possible to prove Gödel's first or second theorem without the diagonal lemma? Is it possible to prove them without diagonalizing at all?*

A partial answer has recently been given by H. Kotlarski.