THE TEACHING OF MATHEMATICS
2001, Vol. IV, 1, pp. 1–34

SELECTED CHAPTERS FROM ALGEBRA

I. R. Shafarevich

Abstract. This paper is the fifth part of the publication "Selected chapters from algebra", the first four having been published in previous issues of the Teaching of Mathematics, Vol. I (1998), 1–22, Vol. II, 1 (1999), 1–30, Vol. II, 2 (1999), 65–80, Vol. III, 1 (2000), 15–40 and Vol. III, 2 (2000), 63–82.

AMS Subject Classification: 00 A 35

Key words and phrases: Real numbers, limits, infinite sums, decimal representations, real roots of polynomials, Sturm's theorem.

CHAPTER V. REAL NUMBERS AND POLYNOMIALS

1. Axioms of real numbers

In the present chapter we shall try to make our idea of real numbers more precise. Our tendency will not be towards very rigorous reasoning, but we shall only try to give enough accuracy to our notions and reasoning in this field, so that we are able to *prove* statements about real numbers.

If we choose an origin and a unit on a line, we can represent real numbers as points on the line. Thus, if we make our idea of real numbers more precise, we give at the same time a more precise description of a line and points lying on it. In the sequel we shall often, as an illustration, use this bijective correspondence between real numbers and points on a line.

Let us try to take geometry as an example and bring the precision of definitions and arguing to the level which already exists in the school geometry courses. There, some axioms appear as the basis of all the construction, and starting from these axioms all other statements are proved. Axioms themselves are not proved: we take them on the basis of experiment or intuition.

In order to be more concrete, let us look at the construction of *plane geometry* based on axioms. We can distinguish three types of logical notions. First of all, there are basic geometrical notions—points and lines. Then, there are basic relations: a point lies on a line; a point lies on a line between other two points. Neither of these are defined. We think as if a "list" of all points and all lines exists

---

This paper is an English translation of: И. Р. Шафаревич, *Избранные главы алгебры. Глава V. Действительные числа и многочлены*, Математическое образование, N 2(5), апр.--инь 1998, Москва, стр. 31--66. In the opinion of the editors, the paper merits wider circulation and we are thankful to the author for his kind permission to let us make this version.

somewhere, and we know which points lie on which lines or which triples of points A, B, C on the line l are such that B lies between A and C. And only in third place there are axioms, i.e., statements about basic notions and relations among them. For instance: each two distinct points belong to exactly one line. Or: among three distinct points on a line, there is exactly one lying between other two.

There is a complete analogy with real numbers. The basic notions here are real numbers themselves. This means that, for the moment, we do not assume anything more about real numbers, but only that they constitute a certain set. Basic relations between real numbers are of two different types: operations and inequalities. Let us describe them in more detail.

1) Operations with real numbers

For every two real numbers a and b we define a third number c, called the *sum* of a and b. We write this as: $a + b = c$.

For every two real numbers a and b we define a third number d, called the *product* of a and b. We write this as: $ab = d$.

2) Inequalities between real numbers

For some pairs of real numbers a and b we have that a is less than b. We write this as: $a < b$. The same relation is also written as $b > a$. If we want to say that $a < b$ or $a = b$, we write $a \leqslant b$ (or $b \geqslant a$).

Before we pass to the formulation of axioms connecting basic notions with basic relations among them, let us emphasize once more the analogy with geometry. Write analogous notions in the table:

Algebra	Geometry
Basic notions	
Real numbers	Point, line, ...
Basic relations	
sum: $a + b = c$	A point lies on a line.
product: $ab = d$	Point C lies between
inequality: $a < b$	points A and B.
	...
Axioms	
...	...

There is no need to list geometrical axioms here; and axioms on real numbers shall be listed now. They will be formulated in terms of basic notions and relations between them, listed in the table. We group the axioms according to the basic relations they deal with.

I (axioms of addition)

I_1. Commutative law: $a + b = b + a$ for arbitrary real numbers a and b.

I_2. Associative law: $a + (b + c) = (a + b) + c$ for arbitrary real numbers a, b and c.

I_3. There exists a number called *zero*, denoted by 0, such that $a + 0 = a$ is valid for each real number a.

(Remark. There exists exactly one such number. If $0'$ were another number with the same property, we would have $0' + 0 = 0'$, by the definition of 0, $0' + 0 = 0 + 0'$ by the commutative law and $0 + 0' = 0$, by the definition of $0'$. Finally, we obtain $0' = 0' + 0 = 0 + 0' = 0$, i.e., $0' = 0$.)

I_4. For each real number a there exists a number called opposite, denoted by $-a$, such that $a + (-a) = 0$.

(Remark. For the given number a there exists exactly one such number. If a' were another number with the same property: $a + a' = 0$, we would have $(a + (-a)) + a' = 0 + a' = a'$. Also, $(a + (-a)) + a' = ((-a) + a) + a'$, and by the associative law, $((-a) + a) + a' = (-a) + (a + a')$. By the property of number a', $a + a' = 0$ and $(-a) + 0 = -a$. Taking these equalities together, we obtain that $a' = -a$.)

II (axioms of multiplication)

II_1. Commutative law: $ab = ba$ for arbitrary real numbers a and b.

II_2. Associative law: $a(bc) = (ab)c$ for arbitrary real numbers a, b and c.

II_3. There exists a number called *unit*, denoted by 1, such that $a \cdot 1 = a$ for an arbitrary real number a.

(Remark. There exists only one such number. It can be proved in the same way as the remark following axiom I_3—we only have to replace addition by multiplication, and 0 by 1.)

II_4. For each real number a, different from 0, there exists a number called inverse, denoted by a^{-1}, such that $a \cdot a^{-1} = 1$.

(Remark. For each real number a different from 0, there exists only one such number. The proof is exactly the same as in the remark following axiom I_4.)

III (axiom of addition and multiplication)

III_1. Distributive law: $(a + b)c = ac + bc$ for arbitrary real numbers a, b and c.

IV (axioms of order)

IV_1. For any two real numbers a and b exactly one of the following three relations holds: $a = b$ or $a < b$ or $b < a$.

IV_2. If for some three real numbers a, b and c we have $a < b$ and $b < c$, then $a < c$.

IV_3. If $a < b$, then $a + c < b + c$ for arbitrary three real numbers a, b and c.

IV_4. If $a < b$ and $c > 0$, then $ac < bc$ for arbitrary three real numbers a, b and c.

V (real and rational numbers)

Rational numbers are contained among real numbers, and operations and inequalities, defined for real numbers, when applied to rational ones, give usual operations and inequalities.

VI (axiom of Archimedes)

For each real number a there exists a natural number n such that $a < n$.

VII (axiom of embedded segments)

Let $a_0,\ a_1,\ a_2,\ \ldots$ and $b_0,\ b_1,\ b_2,\ \ldots$ be two sequences of real numbers, satisfying $a_0 \leqslant a_1 \leqslant a_2 \leqslant \cdots$, $b_0 \geqslant b_1 \geqslant b_2 \geqslant \cdots$ and $b_n \geqslant a_n$ for each n. Then there exists a real number c, such that $b_m \geqslant c$ and $c \geqslant a_n$ for all m and n.

If we use representation of real numbers on a line, then numbers x satisfying the condition $a \leqslant x$ and $x \leqslant b$ ($a \leqslant x \leqslant b$ for short) are represented by the set which is called a *segment* and denoted by $[a, b]$. So, the premises of the last axiom state that the segments $I_n = [a_n, b_n]$ are embedded one into another: $I_0 \supset I_1 \supset I_2 \supset \cdots$. The axiom states that there exists a point (i.e., a number) which is common for all these embedded segments (hence the name of the axiom).

All the usual properties of real numbers easily follow from the listed axioms. It would be too boring to devote several pages to these completely obvious arguments. Hence, we shall only formulate some assertions which we shall need later—and give just some remarks in connection with their proofs (see also problems 2, 3, 4).

It follows from the axioms of group II that for each number a different from 0 and each number b, the number $c = a^{-1}b$ is the unique solution of the equation $ax = b$. It is called the *quotient* of b and a and denoted by $\frac{b}{a}$. All the usual rules about dealing with parentheses and fractions follow from the axioms.

Since for a natural number n the equality $n = 1 + \cdots + 1$ (n summands) is valid, it follows from the axioms of group III that for each number a, the number na (product of n and a) is equal to the sum $a + \cdots + a$ (n summands).

Axiom IV_3 implies that if $a < b$ and $c < d$, then $a + c < a + d < b + d$. If $a < 0$, then $-a > 0$ (because from $-a < 0$ it would follow $0 < 0$). As a result we conclude that each real number is either positive ($a > 0$), has the form $-b$, where $b > 0$, when we say that it is negative, or it is equal to 0. Multiplication obeys the usual "rule of signs". As usual, we write $|x| = x$ if $x \geqslant 0$ and $|x| = -x$ when $x < 0$.

Axiom of embedded segments (axiom VII) is particularly useful when the length of segment I_n (i.e., the difference $b_n - a_n$) becomes arbitrary small when n increases. In other words, if for an arbitrary real number $\varepsilon > 0$ there exists an index N such that $b_n - a_n < \varepsilon$ for all $n \geqslant N$. In such a case one can conclude more than just what is said in the axiom:

Lemma 1. *If differences $b_n - a_n$ become arbitrary small with increasing of the index n, then number c, whose existence is guaranteed by axiom VII, is unique.*

Proof. Suppose that there exist two such numbers: c and c' and, for example, $c < c'$. Then $a_n < c < c' < b_n$ and $c' - c = b_n - a_n - (c - a_n) - (b_n - c') \leqslant b_n - a_n$. We obtain (for n sufficiently large) that $c' - c < \varepsilon$ for an arbitrary given number $\varepsilon > 0$. For instance, such a relation has to be valid for $\varepsilon = \frac{c'-c}{2}$, whence $\frac{1}{2}(c' - c) < 0$, but this contradicts the fact that $c' - c > 0$, $\frac{1}{2} > 0$.

We meet exactly this situation when we intend to measure the given real number approximately, with deficiency or excess, using rational numbers. In that case a_n and b_n are rational numbers. An example is the construction of $\sqrt{2}$ we spoke

about in Section 1 of Chapter I. Thus, axiom VII formulates what we intuitively have in mind when we speak about "better and better measuring". Together with the preceding Lemma it gives us the possibility of *constructing* real numbers with the prescribed properties. We shall often use this observation later.

Concerning axioms V and VI we just remark that we assume here natural and, more generally, rational numbers to be known. We shall not analyse these notions in detail.

Let us remark at the end that the given axioms are not *independent*. This means that some of them could be proven as theorems, relying on other axioms (see, e.g., problem 6). We have just gathered those properties of real numbers which we are used to and which are intuitively convincing. Taking greater number of axioms we obtained the right to skip not very interesting proofs of some intuitively obvious facts.

Problems

1. Which of the axioms I–VII are also valid in the set of rational numbers, and which are specific for real numbers?

2. Prove, using axioms I–III, that for each real number a, $0a = 0$.

3. Prove that for arbitrary real numbers a and b the equation $a + x = b$ has a solution and that it is unique.

4. Prove that for arbitrary real numbers $a \neq 0$ and b the equation $ax = b$ has a solution and that it is unique.

5. Consider the set of rational numbers as a subset of the set of real numbers—on the basis of axiom V. Prove that rational number 0 coincides with the real number 0 whose existence is based on axiom I_3. Do the same for rational number 1 and the real number 1 whose existence is based on axiom II_3.

6. Not using axiom V, prove that numbers 0, 1, $1+1$, $\ldots$, $1+1+\cdots+1$ (n summands) are different for all natural n. Here 1 denotes the number whose existence is guaranteed by axiom II_3. Hence, prove that natural numbers are contained amongst the reals, and that operations and inequalities, defined for real numbers, when applied to natural ones, give usual operations and inequalities. Prove after that the assertion of axiom V. In that way, this axiom is in fact superfluous in our list, since it could be proven on the basis of other axioms.

7. Instead of the operation of multiplication, given by definition for real numbers, define a new operation $\odot$ given by the formula $a \odot b = a + b + ab$. Does it obey the axioms of group II?

2. Limits and infinite sums

In order to illustrate the role of axiom of embedded segments as a method of construction of new real numbers, we shall introduce several notions which will also be useful later.

We met in Chapter IV sequences which were bounded as well as sequences which increased unboundedly. Consider now sequences which are decreasing. For the sake of simplicity, consider first sequences of positive numbers and call such a sequence *unboundedly decreasing* if its terms unboundedly approach zero. The exact definition can be made analogously to the definition of unboundedly increasing sequences, given in Section 2, Chapter IV.

A sequence a_n of nonnegative real numbers is said to *approach zero unboundedly* if for each arbitrary small positive number ε there exists a natural number N such that $a_n < \varepsilon$ for all $n > N$. In such a case we also say that the sequence a_n *tends* to 0 and denote it by: $a_n \to 0$ when $n \to \infty$ ("when n tends to infinity").

A typical example of such a sequence is the sequence $a_n = \frac{1}{n}$.

Consider now a less obvious example.

Lemma 2. *If a is an arbitrary positive number smaller than 1, then the sequence $a_n = a^n$ unboundedly approaches 0, i.e., $a^n \to 0$ when $n \to \infty$.*

Really, put $a = 1/A$. Then $A > 1$ and it can be written in the form $A = 1 + x$ with $x > 0$. Using binomial formula, $A^n = (1 + x)^n = 1 + nx + y$, where y is a sum of positive numbers, so $y > 0$. Thus, $A^n > 1 + nx$ and so for each $\varepsilon > 0$ there exists such N that $A^n > 1/\varepsilon$ for all $n \geqslant N$ (this N can be found explicitly). Hence, $a^n < \varepsilon$ which means that $a^n \to 0$ when $n \to \infty$.

We can generalize the previous definition to sequences $(a_1, a_2, \dots, a_n, \dots)$ whose terms can also be negative. Then the numbers $|a_1|$, $|a_2|$, $\dots$, $|a_n|$, $\dots$ are nonnegative and we can apply the previous definition to them. We shall say that the sequence a_n approaches zero unboundedly, if the sequence of numbers $|a_n|$ unboundedly approaches 0. In that case one also writes $a_n \to 0$ when $n \to \infty$.

Now we have come to our main definition. If for a sequence $a = (a_1, a_2, \dots, a_n, \dots)$ there exists a number α, such that $a_n - \alpha \to 0$ when $n \to \infty$, then α is called the *limit* of the sequence a. One also says that the sequence a_n *tends* to α and one writes $a_n \to \alpha$ when $n \to \infty$.

Not every sequence has a limit. For example, if a sequence has a limit, then it is bounded. Really, let $\alpha_n \to \alpha$ when $n \to \infty$. Then there exists an N, such that $|\alpha_n - \alpha| < 1$ for $n > N$. Since $\alpha_n = \alpha + (\alpha_n - \alpha)$, it follows that $|\alpha_n| \leqslant |\alpha| + 1$ for $n > N$ and therefore $|\alpha_n| \leqslant C$ for all n, where C is the maximum of numbers $|\alpha_1|$, $\dots$, $|\alpha_N|$, $|\alpha| + 1$. But even if a sequence is bounded, it can have no limit. An example is the sequence $(0, 1, 0, 1, \dots)$ where 0 and 1 alternate. If it had a limit α, we could take in the definition of the limit $\varepsilon = \frac{1}{2}$ and we would have $|a_n - \alpha| < \frac{1}{2}$ for all $n > N$. But among a_n's with $n > N$ there are both 0 and 1. Therefore we would have $|\alpha| < \frac{1}{2}$ and $|1 - \alpha| < \frac{1}{2}$. Clearly, such a number α does not exist.

But if a sequence has a limit, this limit is unique. Namely, suppose that a sequence $(a_1, a_2, \dots, a_n, \dots)$ has two limits: α and β, $\alpha \neq \beta$. Then for each ε there exist numbers N and N', such that for $n > N$ it is $|a_n - \alpha| < \varepsilon$ and for $n > N'$ it is $|a_n - \beta| < \varepsilon$. Let $n > N$ and $n > N'$; then $|a_n - \alpha| < \varepsilon$ and $|a_n - \beta| < \varepsilon$, wherefrom $|\alpha - \beta| < 2\varepsilon$. But ε in our reasoning is an arbitrary positive number, and we can choose it so that $\varepsilon < \frac{1}{2}|\alpha - \beta|$, hence a contradiction.

As not every bounded sequence has a limit, considering just such sequences would not lead us to the construction of new real numbers. Our main result will be that there is a simple special type of sequences which always have limits and therefore they will give us a method of constructing new real numbers.

A sequence $(a_1, a_2, \dots, a_n, \dots)$ is called *increasing*, if $a_n \leqslant a_{n+1}$ for all n, i.e., $a_1 \leqslant a_2 \leqslant a_3 \leqslant a_4 \leqslant \cdots$.

Theorem 1. *Each bounded and increasing sequence of positive numbers has a limit.*

The proof will follow the logic of an anecdote which was popular when I was a student (i.e., before the war). The story was about different ways to catch a lion in a desert. There was a French method, method of NKVD-investigators, mathematician's method, ... Mathematician's method went like this. He divides the desert into two parts. The lion is situated in one of these parts. He divides this part again in two parts—and he continues like this till the lion appears in a part of the desert whose dimensions are less than the dimensions of the cage. It remains to put the cage around it. This was a parody to a way of proving existence theorems, one of which we are going to demonstrate now.

Let $a = (a_1, a_2, \dots, a_n, \dots)$ be an increasing sequence of positive numbers. By the assumption it is bounded, so there exists a constant C such that all $a_n < C$. Divide the segment $I_1 = [0, C]$ into two equal parts by the number $C/2$. Then one of the following is valid. Either there exists an m, such that $a_m \geqslant C/2$, and then all a_n with $n \geqslant m$ are contained in the segment $[C/2, C]$ (since the sequence is increasing); or $a_n \leqslant C/2$ for all n, and then all terms of the sequence belong to the segment $[0, C/2]$. Denote by I_2 one the segments, $[0, C/2]$ or $[C/2, C]$, namely the one which contains all the terms of sequence a, starting from some place. After that, divide the new segment into two parts. Obviously, we can continue the process unboundedly and we will obtain a sequence of embedded segments $I_1 \supset I_2 \supset I_3 \supset \cdots \supset I_m \supset \cdots$, where segment I_k has the length $C/2^k$, and which possesses the property that each segment I_k contains all the terms of sequence a, starting from some place. By the axiom of embedded segments (axiom VII) there exists a real number α, belonging to all the segments I_k. It is indeed the limit of sequence a. Really, as we have seen, all the terms of sequence a, starting from some place, belong to segment I_k, This means that for each natural number k there exists an N such that $a_n \in I_k$ for all $n > N$. But also $a \in I_k$. Since the length of segment I_k is equal to $C/2^k$, it follows that $|a_n - \alpha| < C/2^k$ for $n > N$. This gives us the property which appears in the definition of the limit, if we choose k so that $C/2^k < \varepsilon$. In particular, note that such a choice is always possible (the sequence $\left(1, \frac{C}{2}, \frac{C}{4}, \frac{C}{8}, \dots\right)$ tends to 0).

Theorem 1 is particularly useful when the sequence $a = (a_1, a_2, \dots, a_n, \dots)$ is the sequence of sums of a sequence of nonnegative numbers $c = (c_1, c_2, \dots, c_n, \dots)$ $(c_n \geqslant 0)$, i.e., when $a_1 = c_1$, $a_2 = c_1 + c_2$, ... , $a_n = c_1 + c_2 + \cdots + c_n$. In such a case, obviously, the sequence a is increasing. But it has to be checked (and it could by no means be easy) whether it is bounded. For example, if in the sequence c all

$c_n = 1$, then $a_n = n$ and the sequence a is unbounded. We considered a less trivial example in Section 2 of Chapter IV: in the sequence c all $c_n = 1/n$. We saw that in that case the sequence a is also unbounded. But if we can check that the sequence a of sums is bounded, then according to Theorem 1 it has a unique limit α. This limit is called the *sum* of the sequence $(c_1, c_2, \dots, c_n, \dots)$, which is denoted by

$$c_1 + c_2 + \cdots + c_n + \cdots = \alpha.$$

Sometimes the infinite sum c is called a *series* and its sum—the *sum* of the series.

If the sequence of sums a_n is bounded, then, as we have seen, the sum of the series $c_1 + c_2 + \cdots + c_n + \cdots$ exists. If it is unbounded, then we say that the sum of the series does not exist. Hence, Lemma 1, Section 2, Chapter IV, states that the sum of the series $1 + \dfrac{1}{2} + \dfrac{1}{3} + \cdots$ does not exist.

Consider an example. Let a nonnegative number a, less than 1, be given, and let $c = (1, a, a^2, \dots, a^n, \dots)$. Then $a_n = 1 + a + a^2 + \cdots + a^{n-1}$ (in the n-th place in the sequence c there appears a^{n-1}). The sum $1 + a + a^2 + \cdots + a^{n-1}$ can be evaluated using the formula for the sum of a geometric progression—formula (12) of Chapter I:

$$a_n = 1 + a + a^2 + \cdots + a^{n-1} = \frac{1 - a^n}{1 - a} = \frac{1}{1 - a} - \frac{a^n}{1 - a}. \tag{1}$$

We have seen that $a^n \to 0$ when $n \to \infty$, wherefrom it follows immediately that $\dfrac{a^n}{1 - a} \to 0$ when $n \to \infty$. Thus, formula (1) gives that $a_n \to \dfrac{1}{1 - a}$. We can write this as:

$$1 + a + a^2 + \cdots + a^n + \cdots = \frac{1}{1 - a} \quad \text{for } a < 1. \tag{2}$$

The series on the left-hand side of relation (2) is called an *infinite geometric progression*, and formula (2) itself—the formula for the sum of an infinite geometric progression.

But there are examples of series where existence of sums is not hard to prove, but the explicit evaluation of the sums is much harder. For example, in Section 2 of Chapter IV we proved that the sums $\dfrac{1}{1^2} + \dfrac{1}{2^2} + \cdots + \dfrac{1}{n^2}$ are bounded. This means that the sum of the series $\dfrac{1}{1^2} + \dfrac{1}{2^2} + \cdots + \dfrac{1}{n^2} + \cdots$ exists. But what is its value? This problem attracted mathematicians in the middle of XVIII century. It was Euler who solved it, when he found an interesting equality

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} + \cdots = \frac{\pi^2}{6}. \tag{3}$$

This was one of the most sensational Euler's discoveries. Euler went even further, evaluating the sum of the series $1 + \dfrac{1}{2^k} + \dfrac{1}{3^k} + \cdots + \dfrac{1}{n^k} + \cdots$ for arbitrary *even* k. It appeared that these sums were connected with the numbers of Bernoulli, which we described in the Appendix of Chapter II. Namely, the following formula is valid for each even k:

$$1 + \frac{1}{2^k} + \frac{1}{3^k} + \cdots + \frac{1}{n^k} + \cdots = \pi^k (-1)^{\frac{k}{2} - 1} \frac{B_k}{2}\, k!. \tag{4}$$

We know nearly nothing about analogous sums with odd k. It was proved only recently (in 1978) that the sum $1+\frac{1}{2^3}+\frac{1}{3^3}+\cdots+\frac{1}{n^3}+\cdots$ is an irrational number. This remains probably the only known fact about these sums for odd values of k.

Let us remark that just knowing the fact that a series $c_1+c_2+\cdots+c_n+\cdots$ has a sum, one can deduce useful corollaries even if the value of the sum is not known.

Lemma 3. *If the sum of the series $c_1+c_2+\cdots+c_n+\cdots$ exists, then the sequence of numbers $d_n=c_n+c_{n+1}+\cdots$ unboundedly approaches 0.*

We shall use an easy property of the limit. Suppose that a sequence a_1, a_2, $\ldots$, a_n, $\ldots$ has a limit α, i.e., $a_n\to\alpha$ when $n\to\infty$. Then for each number β the sequence $\beta-a_1$, $\beta-a_2$, $\ldots$, $\beta-a_n$, $\ldots$ has the limit $\beta-\alpha$. Really, the difference $\beta-\alpha-(\beta-a_n)=a_n-\alpha$, and the difference $a_n-\alpha\to 0$, hence $\beta-\alpha-(\beta-a_n)\to 0$ when $n\to\infty$. Denote the sum of the series $c_1+c_2+\cdots+c_n+\cdots$ by α and the number $c_1+c_2+\cdots+c_m$ by a_m. By the definition of the sum of an infinite series, the sum α of the series $c_1+c_2+\cdots+c_n+\cdots$ is equal to the limit of the sequence a_1, a_2, $\ldots$, a_m, $\ldots$. In the same way the sum d_n of the sequence $c_{n+1}+c_{n+2}+\cdots$ is equal to the limit of the sequence $a_{n+1}-a_n$, $a_{n+2}-a_n$, $\ldots$ $a_{n+k}-a_n$, $\ldots$. By the remark from the beginning of the proof, the last limit is equal to $\alpha'-a_n$, where α' is the limit of the sequence a_{n+1}, a_{n+2}, $\ldots$, a_{n+k}, $\ldots$ (for fixed n). But the limit of the sequence a_{n+1}, a_{n+2}, $\ldots$ is the same as the limit of the sequence a_1, a_2, $\ldots$, i.e., $\alpha'=\alpha$. We obtain that $d_n=\alpha-a_n$. But, by the definition of limit, $\alpha-a_n\to 0$, i.e., $d_n\to 0$ when $n\to\infty$.

As an example, put $d_n=\frac{1}{n^2}+\frac{1}{(n+1)^2}+\cdots$. We see that $d_n\to 0$ when $n\to\infty$.

Considering limits of infinite sums leads us away from algebra, which is mainly concerned with finite expressions. These questions are closely related with another branch of mathematics, called analysis. That is why we are not going to consider them in more detail. Let us remark only that the most interesting results—such as formulas (3) and (4)—appear on borders of these areas.

Problems

1. Prove that if the sum of the series $c_1+c_2+\cdots+c_n+\cdots$ exists, then $c_n\to 0$ when $n\to\infty$.

2. Prove that if $a_n<C$ for each n and $a_n\to\alpha$ when $n\to\infty$, then $\alpha\leqslant C$. Give an example when equality is obtained.

3. Let $a_n\to\alpha$ when $n\to\infty$. Put $b_n=a_{2n}$. Does the sequence b_1, b_2, $\ldots$ have a limit and what is its value? Is it possible, from the existence of the limit of this sequence, to conclude that the sequence a_1, a_2, $\ldots$ itself has a limit? If it does have a limit, what is its value?

4. Does there exist a limit of the sequence $a_1, a_2, \dots$ where

$$a_n = \frac{1}{2} - \frac{1}{3} + \dots + \frac{(-1)^n}{n}\,?$$

Hint. Group consecutive terms in pairs.

5. Let $f(x)$ be a polynomial of degree d. Prove that $a_n \to 0$ when $n \to \infty$, where $a_n = f(n)/n^{d+1}$.

6. Find the sum of the series $b + ba + ba^2 + ba^n + \dots$, where $|a| < 1$ and b is arbitrary. Usually, the sequence $(b, ba, ba^2, \dots)$ is also called an infinite geometric progression.

7. In a square with side b, centres of the sides are joined by segments. In the new square which is obtained in that way the same procedure is done, etc. Find the sum of areas of all squares that can be obtained in this way.

8. Find the sum of the series $\dfrac{1}{1\cdot 2} + \dfrac{1}{2\cdot 3} + \dots + \dfrac{1}{n\cdot(n+1)} + \dots$.

9. Construct a sequence of positive rational numbers smaller than 1, such that a_n has the denominator n and which does not have a limit.

10. Prove that if the sequence $a_1, a_2, \dots$ has a limit α, and the sequence $b_1, b_2, \dots$ has a limit β, then the sequence of sums $a_1 + b_1, a_2 + b_2, \dots$ has the limit $\alpha + \beta$.

3. Decimal representation of real numbers

In Section 1 we described real numbers using a system of axioms. Now we are going to show how real numbers can be given concretely. Here we shall not say anything new—we shall speak about justification of the well known representation of real numbers by infinite decimal fractions. But now we shall show how the existence of such a representation can be deduced from axioms listed in Section 1.

We shall use the usual representation in which integer part can be either positive or negative, while fractional part (sometimes called the mantissa) is always nonnegative.

Let A be an arbitrary integer (of either sign) and $a_1, a_2, \dots, a_n, \dots$ an infinite sequence of numbers, each of which can take one of 10 values: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. All this together will be denoted by $A, a_1a_2a_3\dots$ and called an infinite decimal fraction. For the time being it is just an infinite sequence, written in a different way. Now we are going to show how a real number can be corresponded to it. We define, for each index n, a number

$$\alpha_n = A + \frac{a_1}{10} + \dots + \frac{a_n}{10^n}. \tag{5}$$

Obviously, the sequence $\alpha_1, \alpha_2, \dots, \alpha_n, \dots$ is increasing. Let us prove that it is bounded. Really, since all $a_i \leqslant 9$, we have

$$\frac{a_1}{10} + \frac{a_2}{10^2} + \dots + \frac{a_n}{10^n} \leqslant \frac{9}{10}\left(1 + \frac{1}{10} + \dots + \frac{1}{10^{n-1}}\right).$$

We apply the formula about the sum of geometric progression:

$$1+\frac{1}{10}+\cdots+\frac{1}{10^{n-1}}=\frac{1-\frac{1}{10^n}}{1-\frac{1}{10}}<\frac{10}{9}$$

and as a result we obtain that

$$\frac{a_1}{10}+\frac{a_2}{10^2}+\cdots+\frac{a_n}{10^n}<1 \tag{6}$$

so that $\alpha_n < A+1$.

By Theorem 1, the sequence $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ has a limit α. Real number α will be called the number *corresponded* to the infinite decimal fraction, and this will be denoted by

$$\alpha = A, a_1 a_2 \ldots a_n \ldots \tag{7}$$

Sometimes it is said that α is *equal* to the decimal fraction $A, a_1 a_2 \ldots a_n \ldots$. This simply means that α is equal to the sum of the infinite series $A+\frac{a_1}{10}+\cdots+\frac{a_n}{10^n}+\cdots$.

Our next goal is to explore this correspondence between decimal fractions and real numbers. Is it bijective? In other words: can a real number correspond to two different decimal fractions? And is each real number corresponded to some decimal fraction?

Consider the first question. First of all, remark that the answer is sometimes positive. Take, e.g., the infinite decimal fraction $0,9999\ldots$, where each decimal after the comma is equal to 9. Which real number does it represent? According to general definition we have to consider the sequence $\alpha_n=\frac{9}{10}+\frac{9}{10^2}+\cdots+\frac{9}{10^n}$. This sum is easy to evaluate: according to the formula about the sum of geometric progression (formula (12) in Chapter I) it is equal to

$$\frac{9}{10}\left(1+\frac{1}{10}+\cdots+\frac{1}{10^{n-1}}\right)=\frac{9}{10}\frac{1-\frac{1}{10^n}}{1-\frac{1}{10}}=\frac{9}{10}\frac{1-\frac{1}{10^n}}{\frac{9}{10}}=1-\frac{1}{10^n}.$$

Obviously, the limit of the sequence $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ is equal to 1, so that $1 = 0,9999\ldots$. But, on the other hand, surely $1 = 1,00\ldots$, where in front of the comma there is just 1, and after it all zeros. In such a way, the same real number 1 is corresponded to two distinct infinite decimal fractions.

It is clear that one can construct a lot of examples of the same kind. In general, such an example has the following form. Let an infinite decimal fraction has the form $A, a_1 \ldots a_k 99\ldots$, i.e., suppose that starting form some place (in our case from the $(k+1)$-st one) all the decimals are equal to 9. We can assume that $a_k \neq 9$, i.e., k-th is the first place after which all the 9's follow. Then, literally repeating previous reasoning, one can conclude that this fraction is equal to the same number as the fraction $A, a_1 \ldots a_{k-1}(a_k+1)000\ldots$, in which all the decimals after the k-th one are equal to 0. A fraction having all the decimals 9, starting from some place, is said to have 9 as a period. We have seen that for such fractions one-to-one correspondence between fractions and real numbers is violated.

It is a bit of a surprise that such violation appears only in those cases.

Theorem 2. *Two distinct infinite decimal fractions, neither of which has 9 as a period, are corresponded to distinct real numbers.*

The proof can be obtained easily if we connect our construction of a real number, defined by a decimal fraction, with the usual measuring of numbers with accuracy of $1/10^m$, with deficiency and excess. One has to divide the line into segments of the length $1/10^m$, whose endpoints are rational numbers with denominator 10^m. Then each point from the line, that is, each real number, falls in one of the segments. The endpoints of the segment give a measure of the number, with deficiency and excess and accuracy of $1/10^m$. However, violating of one-to-one correspondence appears because of the endpoints of segments themselves. To which of the segments, left or right, is each of these points corresponded? This is the same problem which appears in connection with number 9 in the period. We are going to show that our choice (without 9 in periods) corresponds to the case when the endpoints of segments are always attached to segments on the right-hand side. In other words, the constructed numbers α_m and the number α which they define are connected by the relation

$$\alpha_m \leqslant \alpha < \alpha_m + \frac{1}{10^m}. \tag{8}$$

(The fact that numbers α_m are rational with the denominators of the form 10^m follows from their form (5).)

Remember that number α was defined as the limit of the sequence $\alpha_1, \alpha_2, \dots, \alpha_n, \dots$. All numbers α_n with $n \geqslant m$, obviously satisfy the condition $\alpha_n \geqslant \alpha_m$. Hence, such an inequality is valid for their limit α. Really, from the assumption $\alpha < \alpha_m$ we could deduce that $\alpha_n - \alpha = (\alpha_n - \alpha_m) + (\alpha_m - \alpha) \geqslant \alpha_m - \alpha$ for all $n \geqslant m$. But, by the definition of limit, the absolute value of the number $\alpha_n - \alpha$ is smaller than an arbitrary given positive number for n large enough. This contradicts the fact that it is not smaller than the fixed positive number $\alpha_m - \alpha$ (see Problem 2 in Section 2).

In this way the left-hand inequality in (8) is proved. The right-hand one can be proved similarly, if the sign $<$ is replaced by $\leqslant$. Namely, for each $n > m$ we have

$$\alpha_n = \alpha_m + \frac{a_{m+1}}{10^{m+1}} + \dots + \frac{a_n}{10^n} \leqslant \alpha_m + \frac{1}{10^m}\left(\frac{a_{m+1}}{10} + \dots + \frac{a_n}{10^{n-m}}\right) \tag{9}$$

and applying inequality (6) we conclude that $\alpha_n < \alpha_m + \frac{1}{10^m}$. Repeating the previous reasoning we obtain that $\alpha \leqslant \alpha_m + \frac{1}{10^m}$.

But, if we want to obtain the right-hand inequality in (8) with the sign $<$, we have to use the fact that the fraction $A, a_1 a_2 \dots$ does not have 9 as a period. The proof is only a bit more complicated. Let us prove the right-hand inequality in (8) for fixed index m. We shall use the fact that the decimal fraction does not have 9 as a period. That means that somewhere after a_m there has to appear a digit a_k different from 9. For an arbitrary $n > k$ we can write

$$\alpha_n = \alpha_m + (a_{m+1}/10^{m+1} + \dots + a_k/10^k) + (a_{k+1}/10^{k+1} + \dots + a_n/10^n).$$

As before, we see that

$$a_{k+1}/10^{k+1} + \cdots + a_n/10^n \leqslant 1/10^k$$

and so

$$\alpha_n \leqslant \alpha_m + (a_{m+1}/10^{m+1} + \cdots + (a_k + 1)/10^k).$$

Since $a_k \neq 9$, the digit $a_k + 1$ is one of the digits 1, 2, ... , 9. Put

$$c = a_{m+1}/10 + \cdots + (a_k + 1)/10^{k-m}.$$

We can repeat our reasoning once more and obtain that $c < 1$. Number c depends only on the choice of m and k, and not on n. Hence, replacing α_n by its limit α, we obtain, as before, $\alpha \leqslant \alpha_m + c/10^m < \alpha_m + 1/10^m$.

That proves inequality (8).

It follows right away from the inequality (8) that to each two distinct decimal fractions, not having 9 as a period, there correspond two distinct real numbers. Let, to the contrary, the same number α corresponds to fractions $A, a_1a_2\ldots$ and $A', a_1'a_2'\ldots$. Then together with inequalities (8) we have relations

$$\alpha_m' \leqslant \alpha < \alpha_m' + \frac{1}{10^m},$$

where $\alpha_m' = A' + \dfrac{a_1'}{10} + \cdots + \dfrac{a_m'}{10^m}$. Let $\alpha_m' \neq \alpha_m$ and $\alpha_m' > \alpha_m$. From these relations it follows that $\alpha_m' < \alpha_m + \frac{1}{10^m}$, i.e., $\alpha_m' - \alpha_m < \frac{1}{10^m}$. But this contradicts the fact that α_m and α_m' are distinct rational numbers having the same denominator 10^m. Hence, $\alpha_m' = \alpha_m$ for all m. But numbers a_m are uniquely determined by the numbers α_m, since $\alpha_m - \alpha_{m-1} = a_m/10^m$. Thus, they coincide in both fractions, too.

We pass now to the second question: does every real number correspond to some infinite decimal fraction? As well as the answer, the method of proof is already known to us. We just want to convince ourselves that the reasoning can be based on the axioms we formulated.

First of all, let us remark that each real number α is situated between two consecutive integers, i.e., there exists an integer A, such that $A \leqslant \alpha < A + 1$. Let, for start, α be positive. Applying Archimedes' axiom, we conclude that there is an integer n such that $\alpha < n$. Obviously, $n > 0$, and since there exist only a finite number of natural numbers not exceeding n, there also exists the last (the smallest) one with that property. Denote this number by m. Then $\alpha < m$, but $m - 1$ does not possess this property; that means $m - 1 \leqslant \alpha < m$ and $A = m - 1$ has the desired properties. If α is negative, we put $\alpha' = -\alpha$. Then $\alpha' > 0$ and we can apply our procedure: there exists n such that $n \leqslant \alpha' < n + 1$. Axiom IV_3 implies that $-(n + 1) < \alpha \leqslant -n$. If $\alpha' \neq n$, we can put $A = -(n + 1)$ and $A < \alpha < A + 1$. If $\alpha' = -n$, then we have to put $A = -n$. And so, for each real number α there exists an integer A such that $A \leqslant \alpha < A + 1$, hence α can be represented as $\alpha = A + \varepsilon$, where $0 \leqslant \varepsilon < 1$.

Now observe that if some three numbers a_1, a_2, a_3 satisfy $a_1 < a_2$ and $a_2 < a_3$, then for each α satisfying conditions $a_1 \leqslant \alpha < a_3$, one of the following conditions

must be satisfied: either $a_1 \leqslant \alpha < a_2$ or $a_2 \leqslant \alpha < a_3$. The fact is demonstrated in Fig. 1 where the interval $[a_1, a_3)$ is simply the union of the intervals $[a_1, a_2)$ and $[a_2, a_3)$. Formally, it is a consequence of the fact that for each α exactly one of the relations $\alpha < a_2$, $a_2 < \alpha$ and $a_2 = \alpha$ holds.

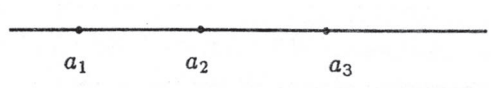


Fig. 1

Consider a more general case. Let the following conditions be satisfied for n numbers $\alpha_1, \ldots, \alpha_n$: $\alpha_1 < \alpha_2$, $\alpha_2 < \alpha_3$, $\ldots$, $\alpha_{n-1} < \alpha_n$. Then for each number α, satisfying $\alpha_1 \leqslant \alpha < \alpha_n$, one of the conditions $\alpha_{i-1} \leqslant \alpha < \alpha_i$ $(i = 2, 3, \ldots, n)$ is valid. In order to prove it one just has to apply the previous assertion to the case of three numbers α_1, α_2, α_n. Then either $\alpha_1 \leqslant \alpha < \alpha_2$ (and our statement is valid for $i = 2$), or $\alpha_2 \leqslant \alpha < \alpha_n$. In the latter case consider numbers α_2, α_3, α_n, etc. For some i we come to the desired condition $\alpha_{i-1} \leqslant \alpha < \alpha_i$.

We can return now to our original question. We have already proved that each real number α can be represented in the form $A + \varepsilon$, where A is an integer and $0 \leqslant \varepsilon < 1$. Consider now numbers $\dfrac{k}{10}$, $k = 0, 1, \ldots, 10$. According to the previous result, we can conclude that $\dfrac{k}{10} \leqslant \varepsilon < \dfrac{k+1}{10}$ for some k, $0 \leqslant k < 10$. Denoting this number by a_1, we can write $\varepsilon = \dfrac{a_1}{10} + \varepsilon_1$, where $0 \leqslant \varepsilon_1 < \dfrac{1}{10}$. Hence, $\alpha = A + \dfrac{a_1}{10} + \varepsilon_1$. Continuing the process, we obtain numbers $a_1, \ldots, a_n, \ldots$, where always $0 \leqslant a_i \leqslant 9$, and the sequence $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$, where $\alpha_n = A + \dfrac{a_1}{10} + \cdots + \dfrac{a_n}{10^n}$, has the limit α, i.e., the number α is corresponded to the infinite decimal fraction $A, a_1 a_2 \ldots a_n \ldots$.

Summing up, one can say that *forming infinite decimal fractions for real numbers does not establish a one-to-one correspondence between infinite decimal fractions and real numbers, but such a correspondence becomes one-to-one if we exclude those decimal fractions which have 9 as a period.*

Problems

1. Prove that a real number α corresponds to an infinite decimal fraction having 0 as a period if and only if α is a rational number a/b where a and b are integers such that just 2 and 5 can be prime factors of b.

2. When finding the infinite decimal fraction which corresponds to a rational number a/b, it is enough to find the mantissa, so we can assume that $0 < a < b$.

Let $\alpha_n = \dfrac{a_1}{10} + \dfrac{a_2}{10^2} + \cdots + \dfrac{a_n}{10^n}$, where $0, a_1 a_2 \ldots$ is the infinite decimal fraction corresponding to the number a/b. Prove that $\dfrac{a}{b} - \alpha_n = \dfrac{r_n}{10^n b}$, where $0 \leqslant r_n < b$ and the numbers r_n are connected by the relation $10r_{n-1} = ba_n + r_n$, i.e., a_n is the quotient and r_n the remainder when $10r_{n-1}$ is divided by b. Convince yourself that this method of successive evaluation of digits a_n of a decimal fraction agrees with the usual division algorithm.

3. Prove that the infinite decimal fraction corresponding to a rational number is periodic, i.e., it has the form $(**\cdots)(\mathcal{P})(\mathcal{P})\cdots$, where $(**\cdots)$ denotes a certain finite group of symbols, after which the group of symbols $(\mathcal{P})$, called the *period*, repeats. *Hint.* Use Problem 2 (i.e., the division algorithm) and note that the possible number of remainders when $10r_{n-1}$ is divided by b is finite (not greater than b).

4. Prove that if the denominator b of the fraction a/b is relatively prime with 10, then the period begins immediately after the comma.

5. Under the assumptions of Problem 4, prove that the number of digits in the period is equal to the smallest number k for which $10^k - 1$ is divisible by b.

6. Under the assumptions of Problems 4 and 5, prove that the number of digits in the period is not greater than the number of natural numbers not exceeding b and relatively prime with b. This number is given by formula (25) of Chapter III.

7. Prove that each periodic infinite decimal fraction corresponds to a rational number A. Namely, if $A, a_1 a_2 \ldots a_n$ stays in front of the period $(p_0, p_1, \ldots, p_{m-1})$, and $A + \dfrac{a_1}{10} + \cdots + \dfrac{a_n}{10^n} = Q$, $p_0 10^{m-1} + p_1 10^{m-2} + \cdots + p_{m-1} = \mathcal{P}$, then the rational number corresponding to the given fraction is $Q + \dfrac{\mathcal{P}}{10^n(10^m - 1)}$.

8. Prove that the infinite decimal fraction $0, 1010010001\ldots$, where the number of zeros between two consecutive 1's increases by 1 each time, corresponds to an irrational number.

4. Real roots of polynomials

Having made a firmer basis for the theory of real numbers, we can now obtain some new results about real roots of polynomials with real coefficients. In order to do this, we have to investigate first the behaviour of a polynomial $f(x)$ in the neighbourhood of a value $x = a$.

Theorem 3. *For each polynomial $f(x)$ and each number a there exists a constant M, such that the inequality*

$$|f(x) - f(a)| \leqslant M|x - a| \tag{10}$$

is valid for all x such that $|x - a| \leqslant 1$.

Remember that $|A|$ (read as "absolute value of number A"), by the definition, is equal to A if $A \geqslant 0$ and to $-A$ if $A < 0$. It follows that $|A|$ is always a nonnegative

number. From school courses it is known that

$$|A+B| \leqslant |A| + |B| \tag{11}$$

$$|A+B| \geqslant |A| - |B| \tag{12}$$

$$|AB| = |A| \cdot |B|. \tag{13}$$

Theorem 3 gives a quantitative estimate of how much $f(x)$ differs from $f(a)$ if x slightly differs from a. In order to prove the theorem, put $y = x - a$, i.e., $x = a + y$ and substitute this value into the polynomial $f(x)$. Each term $a_k x^k$ of the polynomial $f(x)$, after the substitution, gives the expression $a_k(a+y)^k$, which can be written as a sum of powers of y and then similar terms in $f(a+y)$ can be reduced. As a result we obtain that $f(a+y)$ is a polynomial in y, which we denote by $g(y) = c_0 + c_1 y + \cdots + c_n y^n$. Then $f(x) = f(a+y) = g(y)$, $f(a) = f(a+0) = g(0) = c_0$, $x - a = y$ and inequality (10) which we intend to prove becomes

$$|g(y) - g(0)| \leqslant M|y| \tag{14}$$

for all y satisfying $|y| \leqslant 1$.

In the transformed form, the expression $g(y) - g(0)$ acquires a simple form $c_1 y + \cdots + c_n y^n$ (since $g(0) = c_0$). Inequality (11) can be applied also to a sum with an arbitrary number of summands (which can be proved directly by induction) and, in particular, to our sum $c_1 y + \cdots + c_n y^n$. We obtain that

$$|g(y) - g(0)| = |c_1 y + \cdots + c_n y^n| \leqslant |c_1 y| + \cdots + |c_n y^n|.$$

Using equality (13) (also applied to an arbitrary number of factors), $|c_k y^k| = |c_k| \cdot |y|^k$, so that

$$|g(y) - g(0)| \leqslant |c_1||y| + \cdots + |c_n||y|^n.$$

Since, by the assumption, $|y| \leqslant 1$, we have $|y|^k \leqslant |y|$ and

$$|g(y) - g(0)| \leqslant (|c_1| + \cdots + |c_n|)|y|$$

for $|y| \leqslant 1$.

It is enough to put $M = |c_1| + \cdots + |c_n|$ to obtain inequality (14), which also means inequality (10).

Now we are able to prove an important property of polynomials.

Theorem 4. (Bolzano's theorem) *If a polynomial for $x = a$ and $x = b$ takes values with opposite signs, then it takes the value 0 somewhere between a and b.*

In other words, if for a polynomial $f(x)$ values $f(a)$ and f(b) are numbers of opposite signs and $a < b$, then there exists c, such that $a < c < b$ and $f(c) = 0$.

Theorem 4 appears rather obvious if one looks at the graph of the polynomial $f(x)$ (Fig. 2). It states that the graph cannot "jump" across the x-axis without intersecting it. On the other hand, it is completely possible to *draw* such a graph (Fig. 3). So, we have to prove that such a graph cannot be the graph of a polynomial. For more general functions it is connected with a rather involved property

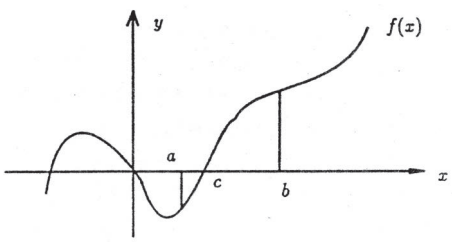


Fig. 2

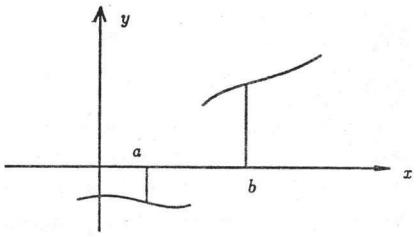


Fig. 3

which is called *continuity*. In the case of polynomials it is enough to use the easy inequality (10), proved in Theorem 3.

The proof is based on the same principle of "catching a lion in a desert", we have already used for proving Theorem 1.

Suppose, for example, that $f(a) > 0$, $f(b) < 0$. Consider the segment $[a, b]$ (i.e., the set of real numbers x satisfying $a \leqslant x$ and $x \leqslant b$). Denote this segment by I_1 and divide it into two segments of equal length by the point $r = \frac{a+b}{2}$. If $f(r) = 0$, then the theorem is proved ($c = r$). If $f(r) \neq 0$ and, for example, $f(r) > 0$, then the polynomial $f(x)$ takes values of opposite signs for $x = r$ and $x = b$. Denote then by I_2 the segment $[r, b]$. If that $f(r) < 0$, then the segment $[a, r]$ will be denoted by I_2. In any case we obtain a segment I_2 contained in I_1, having two times smaller length, and having again the property that the polynomial $f(x)$ has values of opposite signs at its endpoints—namely, positive at the left-hand end and negative at the right-hand one.

This process can be continued. Either we shall at some moment reach a root of the polynomial $f(x)$ (and the theorem will be proved), or the process shall continue unboundedly. It remains to consider the latter case. We obtain an infinite sequence of embedded segments $I_1 \supset I_2 \supset \cdots \supset I_n \supset \cdots$, $I_n = [a_n, b_n]$, such that each of them is of half-a-length of the previous one, and the polynomial $f(x)$ takes values of opposite signs at the endpoints a_n and b_n of each segment I_n, more precisely, $f(a_n) > 0$, $f(b_n) < 0$. Now we are going to use the more precise definition of real numbers we gave in Section 1. Segments I_n satisfy the prepositions of Axiom VII (axiom of embedded segments) and Lemma 1 of Section 1. Really, segments I_n are embedded one into another, by their construction, and since I_n is half-of-length of segment I_{n-1}, its length is equal to $\dfrac{b-a}{2^{n-1}}$, and so this length becomes unboundedly small when n increases. Hence, according to Axiom VII and Lemma 1, there exists a unique number c, belonging to all segments I_n, i.e., such that

$$a_n \leqslant c \leqslant b_n. \tag{15}$$

In this way we have constructed the number c which we searched for. Namely, we now prove that $f(c) = 0$.

Consider the values $f(a_n)$ of the polynomial $f(x)$ at the left-hand endpoints of segments I_n. By the assumption, all $f(a_n) > 0$. Inequality (11) implies that the sequence $a_1, a_2, \ldots$ approaches the number c unboundedly: really, $a_n \leqslant c \leqslant b$ and

$0 \leqslant c - a_n \leqslant b_n - a_n$, where, by the assumption, $b_n - a_n = \dfrac{b-a}{2^{n-1}}$. Therefore the inequality $|a_n - c| < \varepsilon$ will be satisfied if $\dfrac{b-a}{2^{n-1}} < \varepsilon$, and this will be valid for each $\varepsilon > 0$ if n is chosen large enough. Let us prove that it follows from this that the values $f(a_n)$ approach the value $f(c)$ unboundedly. Really, in order to prove that $|f(a_m) - f(c)| \leqslant \varepsilon$ for m large enough, we can use inequality (10) from Theorem 3. Since a_m approaches c unboundedly, we have $|a_m - c| < 1$ for m large enough, and we can apply inequality (10). We see that $|f(a_m) - f(c)| < M|a_m - c|$ and so $|f(a_m) - f(c)| < \varepsilon$ if $M|a_m - c| < \varepsilon$, i.e., if $|a_m - c| < \varepsilon/M$. But we have convinced ourselves that this inequality is valid for m large enough (since ε/M can again be denoted by ε!).

What can be said about the number $f(c)$, which is known to be the limit of the sequence of positive numbers $f(a_n)$? Clearly, $f(c) \geqslant 0$. Really, if $f(c)$ were negative, than for positive $f(a_n)$ we would have $f(a_n) - f(c) \geqslant -f(c)$, and hence $|f(a_n) - f(c)| \geqslant -f(c)$, but this would contradict the fact that $|f(a_n) - f(c)| < \varepsilon$ if $\varepsilon < -f(c)$.

We have thus proved that $f(c) \geqslant 0$. Following exactly the same arguments, considering numbers b_n satisfying $f(b_n) < 0$, we can prove that $f(c) \leqslant 0$. Therefore, for the number $f(c)$ only one possibility remains—$f(c) = 0$. The theorem is proved.

One should pay attention to a completely new way of reasoning in proving this theorem. We have proved in fact (under certain conditions) the existence of a root of the polynomial $f(x)$. But we have not done it using any kind of formula (as, for example, when solving a quadratic equation) but using the axiom of embedded segments. But, at the same time, it is by no means a pure "theorem of existence", where we know only that a certain quantity exists—and nothing more than that. For example, we can in fact find the root c with deficiency and excess and with an arbitrary prescribed accuracy, constructing numbers a_n and b_n such that c lies between them (inequality (15)) and which get closer and closer to each other.

Bolzano's theorem gives us the possibility to know a lot about concrete polynomials. Consider, for example, the polynomial $f(x) = x^3 - 7x + 5$ and make a table of its values for integer values of x, with small absolute values (Table 1). One can see from the table that the polynomial $f(x)$ takes values of opposite signs at the ends of the segments $[2, 3]$, $[0, 1]$ and $[-3, -2]$. By Bolzano's theorem it has a root in each of these segments. Hence, the polynomial $f(x)$ has at least three roots. But its degree is equal to 3 and by Theorem 3 of Chapter II it cannot have more than 3 roots. We have proved that the polynomial $f(x)$ has exactly 3 roots and they lie in segments $[2, 3]$, $[0, 1]$ and $[-3, -2]$.

x	-3	-2	-1	0	1	2	3
$f(x)$	-1	11	11	5	-1	-1	11

Table 1.

There are some other polynomials for which Bolzano's theorem gives the precise answer, too. An important case is the polynomial $x^n - a$ whose roots are called

"roots of a of degree n" (denoted as $\sqrt[n]{a}$). Consider first the case when $a > 0$. Then the polynomial $f(x) = x^n - a$ takes for $x = 0$ negative value $-a$. On the other hand, it is easy to find a value $x = c$ such that $f(c) > 0$ Really, by Archimedes' axiom (Axiom VI) there exists a natural number m such that $m > a$. Then $m^n > m$ and $m^n - a > m - a > 0$. Using Bolzano's theorem, we can state that there is a root of the polynomial in the segment $[0, m]$. If, on the other hand, $a < 0$ and n is even, then such polynomials obviously do not have roots: $x^n \geqslant 0$ as an even power of a real number, and $x^n - a > 0$. If n is odd, then putting $x = -y$ we obtain that $x^n - a = -y^n - a = -(y^n + a)$. The polynomial $y^n + a$ (for $a < 0$), as we have just proved, has a root, and so the same is true for the polynomial $x^n - a$. In school courses these arguments are usually omitted (because of the lack of a precise theory of real numbers), but it is proved (very easily) that for n odd the polynomial $x^n - a$ does not have more than one root (as we have seen—it has exactly one) and that for n even and $a > 0$—not more than two roots which differ only in the sign (which means it has exactly two roots).

But in the case of other polynomials, it can happen that Bolzano's theorem does not give anything. Take as an example the polynomial $x^2 - x + 2$. Using the formula for solutions of quadratic equation we can conclude that this polynomial has no real roots. But if we tried to give values $0, \pm 1, \pm 2, \ldots$ to the argument x, we would obtain only positive values, and Bolzano's theorem wouldn't give us anything. Therefore, we will try now to explore polynomials more thoroughly.

Theorem 3 estimates values of a polynomial for values of x being close to a certain value a. We shall prove now a similar assertion about values of the polynomial for large (by absolute value) values of x.

Theorem 5. *For the polynomial $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ there exists a constant $N > 0$ such that*

$$|a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}| < |a_n x^n| \tag{16}$$

for all values of x such that $|x| > N$.

The theorem states that for sufficiently large values of x, the absolute value of the leading term exceeds the absolute value of the sum of all other terms. In order to prove this, we use inequality (11) (for an arbitrary number of summands) and equality (13). It follows from them that $|a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}| \leqslant |a_0| + |a_1||x| + \cdots + |a_{n-1}||x|^{n-1}$, and $|a_n x^n| = |a_n||x|^n$. In order to prove inequality (16) it is enough to convince oneself that $|a_0| + |a_1||x| + \cdots + |a_{n-1}||x|^{n-1} \leqslant |a_n||x|^n$, and this will be proved if we show that

$$|a_k||x|^k < \frac{1}{n}|a_n||x|^n \tag{17}$$

for each $k = 0, 1, \ldots, n-1$ and $|x| > N$ for N large enough. Then, summing up all the inequalities (17) for $k = 0, 1, \ldots, n-1$ we obtain the inequality we needed.

Inequality (17) can be solved in the usual way. It is equivalent to

$|x|^{n-k} > \dfrac{n|a_k|}{|a_n|}$, i.e.,

$$|x| > \sqrt[n-k]{n\frac{|a_k|}{|a_n|}}. \tag{18}$$

Therefore, it is enough to choose for N an arbitrary number larger than all the numbers $\sqrt[n-k]{n\dfrac{|a_k|}{|a_n|}}$, $k = 0, 1, \ldots, n-1$, and it will satisfy the assertion of Theorem 5.

Theorem 5 has a lot of useful corollaries. Note first that under the assumptions of the theorem (i.e., for $|x| > N$) we always have $|f(x)| > 0$, which follows immediately from inequality (12)

$$|f(x)| = |a_0 + a_1x + \cdots + a_nx^n| \leqslant |a_nx^n| - |a_1 + a_1x + \cdots + a_{n-1}x^{n-1}|.$$

But this means that the polynomial $f(x)$ does not have roots x with $|x| > N$. In other words, roots of a polynomial (if they exist) have to be contained in the segment $|x| \leqslant N$, where, as we have shown (inequality (18)) N can be chosen as the greatest of the numbers $\sqrt[n-k]{n\dfrac{|a_k|}{|a_n|}}$. One calls such a number N the *bound of roots* of the polynomial. So, for the polynomial $x^3 - 7x + 5$ one can take for N an arbitrary number greater than $\sqrt[3]{3 \cdot 5}$ and $\sqrt{3 \cdot 7}$. For example, $N = 4{,}6$ satisfies the conditions. This means that all roots of the polynomial are distributed between $-4{,}6$ and $4{,}6$. We have convinced ourselves earlier that they are in fact contained between -3 and $+3$ (Table 1).

Theorem 5 implies more that just the assertion that $f(x) \neq 0$ if $|x| > N$, for the found value of N. To evaluate the value of $a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + a_nx^n$ means to sum up two real numbers $a_0 + a_1x + \cdots + a_{n-1}x^{n-1}$ and a_nx^n, first of which is smaller (by absolute value) than the other (for $|x| > N$). But then the sign is determined by the sign of the second summand. We come to the following conclusion:

Corollary 1. *For $|x| > N$, where N is the bound of roots defined in Theorem 5, values of the polynomial $f(x)$ have the same sign as the leading term a_nx^n.*

Suppose that the degree n of the polynomial is odd. Then the sign of the leading term a_nx^n for $x > 0$ agrees with the sign of the coefficient a_n, and for $x < 0$ it is opposite. Corollary 1 shows that for $x > N$ and $x < -N$ the polynomial itself acquires values of opposite signs (namely, the signs of a_n and of $-a_n$). Bolzano's theorem implies that between these values there is at least one root of the polynomial. We obtained the following proposition:

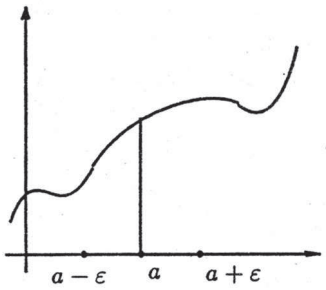
Corollary 2. *Each polynomial of odd degree has at least one root.*

This is really an unexpected result. In fact, you know that a polynomial of the second degree may have no roots (e.g., the polynomial $x^2 + 1$). One may think that the same could happen to polynomials of greater degrees: 3, etc. But here, according to the corollary, a polynomial of the third degree always has a root. The

situation appears more complicated; it depends not on how large the degree of the polynomial is, but on its parity.

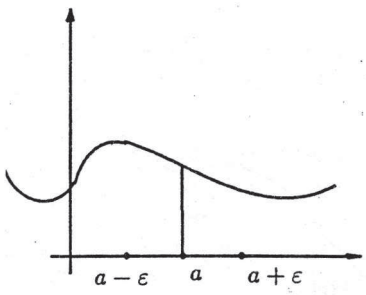
Finally, consider one more property of polynomials, which can make investigations in some cases much easier. Theorem 3 gave us information about the absolute value of the difference $f(x) - f(a)$ when the difference $x - a$ is small. We shall investigate now the *sign* of the difference $f(x) - f(a)$. Here we shall exclude the cases when the value $x = a$ appears to be a root of the derivative $f'(x)$ of the polynomial $f(x)$. These special values of a could be investigated easily in the same manner, but we will not need this at the moment.

Theorem 6. *Let a polynomial $f(x)$ be given and take a value $x = a$ which is not a root of its derivative $f'(x)$ (i.e., $f'(a) \neq 0$). If $f'(a) > 0$, then the values $f(x)$ for x close, but to the left of a, are smaller than $f(a)$, and for x close, but right of a, are greater than $f(a)$. If $f'(a) < 0$, then the situation is opposite.*



$f'(a) > 0$

Fig. 4



$f'(a) < 0$

Fig. 5

This means that there exists sufficiently small $\varepsilon > 0$ (depending on $f(x)$ and on a), such that when $f'(a) > 0$, for $a - \varepsilon < x < a$ we have $f(x) < f(a)$, and for $a < x < a + \varepsilon$, we have $f(x) > f(a)$. If, however, $f'(a) < 0$, then for $a - \varepsilon < x < a$ we have $f(x) > f(a)$, and for $a < x < a + \varepsilon$, we have $f(x) < f(a)$ (see graphs of $f(x)$ on Figs. 4 and 5).

The proof is quite easy. We know by Bezout's theorem that the polynomial $f(x) - f(a)$ is divisible by $x - a$. Therefore

$$f(x) - f(a) = (x - a)g(x, a), \tag{19}$$

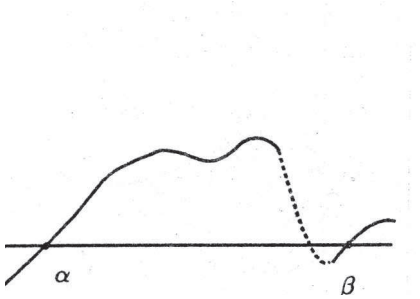
where the coefficients of the polynomial $g(x, a)$ depend on a. For $x = a$ the polynomial $g(x, a)$ takes the value $f'(a)$ (this was just our definition of the derivative of a polynomial, see formula (13) of Chapter II). By the assumption, $f'(a) \neq 0$, and so $g(a, a) = f'(a) \neq 0$. Denote by ε an arbitrary number, smaller than the distance from a to the nearest root of the polynomial $g(x, a)$ (here, a is fixed and x is the unknown), so that the polynomial $g(x, a)$ does not vanish on the segment $[a - \varepsilon, a + \varepsilon]$. Then it preserves the same sign on this segment as it has for $x = a$: if it acquired two values of opposite signs, then by Bolzano's theorem it would vanish

somewhere inside the segment, which would contradict the choice of the number ε. This contains in fact the assertion of Theorem 6. Let, for example, $f'(a) > 0$. Then $g(a, a) = f'(a) > 0$, too, and according to what was said, $g(x, a) > 0$ for $a - \varepsilon < x < a + \varepsilon$. The other factor $x - a$ in formula (19) also behaves in the known way: $x - a < 0$ for $a - \varepsilon < x < a$ and $x - a > 0$ for $a < x < a + \varepsilon$. Multiplying, we obtain from formula (19) that $f(x) - f(a) < 0$ for $a - \varepsilon < x < a + \varepsilon$ and $f(x) - f(a) > 0$ for $a < x < a + \varepsilon$. This is really the assertion of the theorem. The case $f'(a) < 0$ is treated completely analogously.

The theorem we have just proved has an interesting corollary.

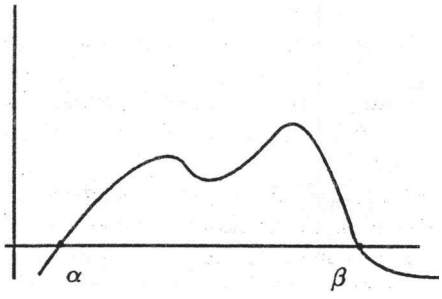
Theorem 7. (Rolle's theorem) *Between two adjacent roots of a polynomial, not having multiple roots, there is always a root of its derivative.*

We assume that our polynomial does not have multiple roots only to make argument shorter. Anyway, this will be the only case that we shall need later.



$f'(\alpha) > 0$, $f'(\beta) > 0$ – impossible

Fig. 6



$f'(a) > 0$, $f'(\beta) < 0$ – possible

Fig. 7

Let α and β, $\alpha < \beta$, be two adjacent roots of the polynomial $f(x)$, so it has no roots lying between them. Since we have assumed that the polynomial has no multiple roots, α and β are not multiple roots and by Theorem 5 of Chapter II, $f'(\alpha) \neq 0$, $f'(\beta) \neq 0$. Let, for example, $f'(\alpha) > 0$. Let us prove that then $f'(\beta) < 0$. Really, if $f'(\beta) > 0$, then by the preceding theorem we would have $f(x) > f(\alpha) = 0$ for $\alpha + \varepsilon > x > \alpha$ and $f(y) < f(\beta) = 0$ for $\beta - \varepsilon < y < \beta$. Then, for arbitrary x satisfying $\alpha + \varepsilon > x > \alpha$ and for arbitrary y, satisfying $\beta - \varepsilon < y < \beta$, we would have $f(x) > 0$ and $f(y) < 0$. Then Bolzano's theorem would imply that the polynomial f had a root lying between x and y, i.e., in the segment $[\alpha, \beta]$. But this would contradict the fact that α and β, as we assumed, were adjacent roots of the polynomial $f(x)$. We see that there remains the only possibility that $f'(\beta) < 0$, but then by Bolzano's theorem the polynomial $f'(x)$ has a root between α and β. On Figs. 6 and 7 an impossible and a possible case of signs for $f'(\beta)$ (if $f'(\alpha) > 0$) are demonstrated. The case when $f'(\alpha) < 0$ can be considered literally in the same way.

At the end of this Section we shall show that the theorems we have proved are already sufficient to solve completely the question about the number of roots

for a polynomial of the third degree. In Section 3 of Chapter II we saw that each equation of the third degree can be replaced by an equivalent equation of the form $x^3 + ax + b = 0$. We shall investigate such a form in the sequel.

First of all let us solve the question about multiple roots. We proved in section 2 of Chapter II that multiple roots of a polynomial are in fact joint roots of the polynomial and its derivative. According to formula (15) of Section II, for the polynomial $f(x) = x^3 + ax + b$ the derivative is equal to $f'(x) = 3x^2 + a$. If $a > 0$, then the derivative has no roots and this means that the polynomial $f(x)$ has no multiple roots. If $a < 0$, then denote by δ the positive root of the polynomial $3x^2 + a$ (i.e., $\delta = \sqrt{-a/3}$). Then the polynomial $f(x)$ can have as a multiple root only one of the numbers δ or $-\delta$. Since the polynomial $f(x)$ can be written in the form $f(x) = (x^2 + a)x + b$ and for $x = \pm\delta$, $x^2 = -a/3$ and $x^2 + a = 2a/3$, then the condition that $f(x)$ has a multiple root takes the form $\pm\delta\frac{2a}{3} = -b$, i.e., $\delta^2\frac{4a}{9} = b^2$, and since $\delta^2 = -a/3$, the condition becomes $-\frac{4a^3}{27} = b^2$, i.e., $4a^3 + 27b^2 = 0$. If this condition is satisfied, then the polynomial has a multiple root α and may be represented in the form $f(x) = (x - \alpha)^2 g(x)$. Here the polynomial $g(x)$ has to be of the first degree which means that it has a single root β. Thus, the polynomial $f(x)$ has two roots equal to α, and one root equal to β.

Consider now the remaining case when the polynomial $f(x)$ does not have multiple roots, i.e., $4a^3 + 27b^2 \neq 0$. According to Corollary 2 of Theorem 5, the polynomial $f(x)$ has at least one root α. If it has another root β, then it must be divisible by $(x - \alpha)(x - \beta)$, i.e., it has the form $f(x) = (x - \alpha)(x - \beta)g(x)$, where $g(x)$ is a polynomial of the first degree and therefore it has a root γ. In such a way, the polynomial $f(x)$ has three roots: α, β and γ. It cannot have more than three roots. We conclude that only two things can happen: either the polynomial $f(x)$ has 1 root or the polynomial $f(x)$ has 3 roots. Our problem is to find out which of the cases takes place (for given coefficients a and b).

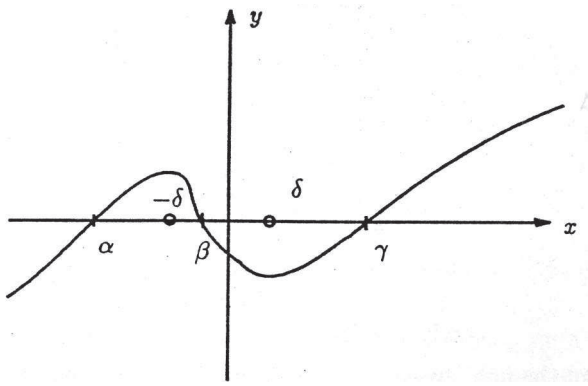


Fig. 8

Suppose that the polynomial $f(x)$ has three roots: α, β and γ, where $\alpha < \beta < \gamma$. This means that the polynomial does not have roots smaller than α and larger

than γ. But according to Corollary 1 of Theorem 5 there exists a number N such that for x large enough (more precisely, for $x \geqslant N$), the values of the polynomial have the same sign as the values of the leading term x^3—i.e., they are positive, and for $x \leqslant -N$ they are negative, for the same reason. Hence, for $x < \alpha$ it is always $f(x) < 0$, and for $x > \gamma$ it is always $f(x) > 0$ (Fig. 8).

Since we have $f(x) < 0$ for $\alpha - \varepsilon < x < \alpha$ and arbitrary $\varepsilon > 0$, according to Theorem 6, $f'(\alpha) > 0$ and so $f(x) > 0$ for $\alpha < x < \alpha + \varepsilon$. Since $f(x)$ has no roots between α and β, by Bolzano's theorem its values are of the fixed sign, so $f(x) > 0$ for $\alpha < x < \beta$. Analogously, we obtain that $f(x) < 0$ for $\beta < x < \gamma$. According to Theorem 7, between the roots α and β, and also between the roots β and γ, there is a root of the derivative $f'(x)$ of the polynomial $f(x)$. Since $f'(x) = 3x^2 + a$, for $a > 0$ the derivative has no roots and such a case (existence of three roots of the polynomial $f(x)$) is impossible. For $a = 0$, $f(x) = x^3 + b$. As we have seen earlier, such a polynomial has only one root. Finally, if $a < 0$, the derivative $f'(x) = 3x^2 + a$ has two roots: $\delta > 0$ and $-\delta < 0$ (here, $\delta = \sqrt{-a/3}$). Obviously, $\alpha < -\delta < \beta < \delta < \gamma$.

Since the polynomial takes positive values on the interval (α, β), and negative values on the interval (β, γ), we have

$$f(-\delta) > 0, \qquad f(\delta) < 0 \tag{20}$$

(under the preposition that the polynomial $f(x)$ has three roots).

Conversely, if conditions (20) are satisfied, then by Bolzano's theorem the polynomial $f(x)$ has a root lying between $-\delta$ and δ. Denote this root by β. Besides, according to Corollary 1 of Theorem 5, for x sufficiently large, the polynomial takes positive values, and for x sufficiently small it takes negative values. Bolzano's theorem implies then that the polynomial has a root smaller than $-\delta$, and also a root greater than δ. Denote these roots by α and γ, respectively. Thus, conditions (20) imply that the polynomial has 3 roots: α, β and γ. In other words, conditions (20) are *necessary and sufficient* for the polynomial $f(x)$ to have 3 roots. In all other cases it has 1 root.

The assertions we have just proved solve our problem. We will only transform conditions (20) into a simpler form. Since $f(x) = (x^2 + a)x + b$ and $3\delta^2 + a = 0$, $\delta^2 = -a/3$, we have $f(\pm\delta) = (\delta^2 + a)(\pm\delta) + b = \pm\delta\frac{2a}{3} + b$ and so conditions (20) acquire the form

$$-\frac{2a}{3}\,\delta + b > 0, \qquad \frac{2a}{3}\,\delta + b < 0,$$

i.e., $\dfrac{2a}{3}\,\delta < b < -\dfrac{2a}{3}\,\delta$. These inequalities are equivalent to just one: $b^2 < \dfrac{4a^2}{3^2}\,\delta^2$. Since $\dfrac{4a^2}{3^2}\,\delta^2 = -\dfrac{4a^3}{27b^2}$, conditions (20) are equivalent to the inequality $4a^3 + 27b^2 < 0$. This is in fact the final answer: if $4a^3 + 27b^2 < 0$, then the polynomial $x^3 + ax + b$ has 3 roots, if $4a^3 + 27b^2 = 0$, it has two equal roots and one other root, and if $4a^3 + 27b^2 > 0$, then it has only 1 root.

Clearly, all that has been said applies only to a polynomial of the third degree.

For polynomials of arbitrary degrees analogous investigations can be done, but arguments are a bit more complicated, so we shall leave them for the Appendix.

Problems

1. We proved at the end of Chapter I that the polynomial $x^3 - 7x^2 + 14x - 7$ has no rational roots, so its roots—if they exist—are irrational numbers. Determine the number of roots of this polynomial, their signs and also, for each of the roots, two consecutive integers such that this root is lying between them.

2. Prove that the polynomial $x^4 + ax + b$ either has no roots, or it has two roots and find conditions (on coefficients a and b) such that the first or the latter case takes place.

3. Prove that the number of roots of a polynomial of even degree is even and of odd degree is odd.

4. Prove that the polynomial $x^n + ax + b$, for n even, has 0 or 2 roots, and for n odd—1 or 3. Determine conditions (on coefficients a and b) such that the first or the latter case takes place.

5. Determine the number of roots of the polynomial $x^n + ax^{n-1} + b$ (depending on n, a and b).

6. Prove that each polynomial $f(x)$ takes arbitrarily large values (by absolute value), for sufficiently large values of x (by absolute value).

7. Prove that as a bound of roots N the number $\dfrac{M}{|a_n|} + 1$ can be taken, where M is the largest of the numbers $|a_0|, \ldots, |a_{n-1}|$. *Hint.* Use the inequality $|a_0 + \cdots + a_{n-1}z^{n-1}| \leqslant M(1 + |z| + \cdots + |z|^n)$.

8. Prove that the polynomial $f(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + a_nx^n$, where $a_n > 0$, $a_i \leqslant 0$ for $i = 1, \ldots, n-1$, $a_0 < 0$, has exactly one positive root. *Hint.* Write $f(x)$ in the form $a_nx^n \left(1 + \dfrac{a_{n-1}}{a_nx} + \cdots + \dfrac{a_0}{a_nx^n}\right)$ and find whether the expressions $\dfrac{a_{n-k}}{a_nx^n}$ increase or decrease when x increases, remaining positive.

9. Let a polynomial $f(x)$ have all the coefficients at even powers of x equal to 0, and all the coefficients at odd powers positive. Prove that it has a unique root.

APPENDIX

Sturm's Theorem

We shall present now a method allowing to determine for each polynomial $f(x)$ the number of its roots lying in a given segment $[a, b]$.

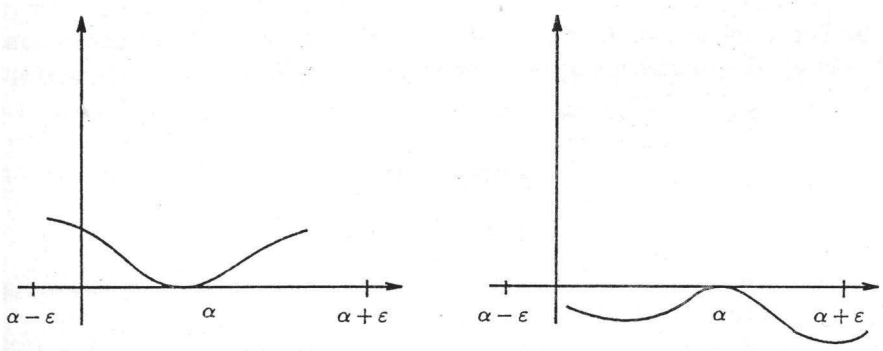
The idea of the method is based on the fact that, although for a single polynomial $f(x)$ there is no simple method which could connect its properties with some properties of polynomials with smaller degree, for a pair of polynomials $f(x)$, $g(x)$

such a method is well known: it consists of divison with remainder of the polynomial $f(x)$ by $g(x)$: $f(x) = g(x)q(x) + r(x)$, and passing from the pair of polynomials (f, g) to the pair of polynomials (g, r). Repeating this process leads us to the algorithm of Euclid for finding the greatest common divisor of polynomials f and g. For example, the question of the existence of common roots of polynomials f and g can be reduced to the question of the existence of common roots of the polynomials of smaller degree g and r and, as a result, to the question of the existence of roots of the polynomial of smaller degree g. c. d.(f, g). The method can be applied to the case of the pair of a polynomial and its derivative and then we obtain the answer to the question of the existence of multiple roots of the polynomial. That is how we proceeded in Chapter II, and we shall also proceed like that now: we shall first consider a certain property of roots of the pair of polynomials (f, g), which can be treated using division with remainder. Applying then this property to the pair consisting of a polynomial and its derivative, we shall find the answer to our question.

Let us start with a simple observation, related to a single polynomial $F(x)$. Let $x = \alpha$ be its root and let this root have the multiplicity k. Then we can write down (by the definition of the multiplicity of roots, given in Section 2 of Chapter II)

$$F(x) = (x - \alpha)^k G(x), \tag{1}$$

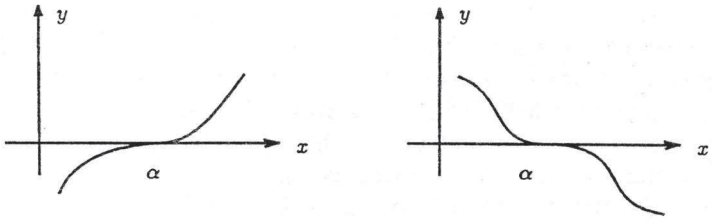
where $G(\alpha) \neq 0$. Thus, if a number ε is smaller than the distance from α to the nearest root of the polynomial $G(x)$, then $G(x)$ takes the values of the same sign in the segment $[\alpha - \varepsilon, \alpha + \varepsilon]$. Really, if for any two numbers x and y lying in this segment the polynomial G had values $G(x)$ and $G(y)$ of opposite signs, then, by Bolzano's theorem, there would exist a root of the polynomial between x and y. But this would contradict the way how ε had been chosen—that there had been no root of the polynomial G lying in the segment $[\alpha - \varepsilon, \alpha + \varepsilon]$. In particular, all the values of the polynomial $G(x)$ for x in the segment $[\alpha - \varepsilon, \alpha + \varepsilon]$ have the same sign as $G(\alpha)$. Formula (1) implies now that if multiplicity k is even, then the values of the polynomial $F(x)$ for x lying in the segment $[\alpha - \varepsilon, \alpha + \varepsilon]$ have the same sign as $G(\alpha)$. The graph could be situated as in Fig. 9.



$G(\alpha) > 0$ $G(\alpha) < 0$

Fig. 9

If, on the other hand, the multiplicity k is odd, then for $G(\alpha) > 0$ we have $F(x) < 0$ for $\alpha - \varepsilon \leqslant x < \alpha$ and $F(x) > 0$ for $\alpha < x \leqslant \alpha + \varepsilon$, and for $G(\alpha) < 0$—the opposite: $F(x) > 0$ for $\alpha - \varepsilon \leqslant x < \alpha$ and $F(x) < 0$ for $\alpha < x \leqslant \alpha + \varepsilon$. In the former case (i.e., for $G(\alpha) > 0$) α is a *root with increasing*, and in the latter (for $G(\alpha) < 0$)—*root with decreasing*. Possible graphs of the polynomial $F(x)$ in both cases are displayed in Fig. 10.



$G(\alpha) > 0$ $G(\alpha) < 0$

Fig. 10

Definition. Let $F(x)$ be a polynomial having as roots neither a nor b. *Characteristics of the polynomial $F(x)$ on the segment $[a, b]$* is the difference between the number of its roots with increasing and the roots with decreasing, lying in that segment. Here, roots having even multiplicity are not counted. The characteristics is denoted by $[F(x)]_a^b$. For example, the polynomial represented in Fig. 11 has 3 roots with increasing and 2 roots with decreasing, so we have $[F]_a^b = 1$.

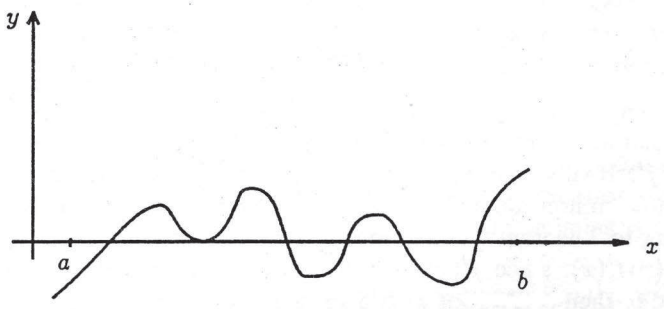


Fig. 11

Since after each root with increasing there must follow a root with decreasing (roots with even multiplicity do not count), the characteristics is determined by the signs of numbers $F(a)$ and $F(b)$, namely:

$[F(x)]_a^b = 0$	if $F(a)$ and $F(b)$ are of the same sign
$[F(x)]_a^b = 1$	if $F(a) < 0$, $F(b) > 0$
$[F(x)]_a^b = -1$	if $F(a) > 0$, $F(b) < 0$

Table 1

Thus, the characteristics of the polynomial $F(x)$ on the given segment is determined by its signs at the endpoints of the segment and so it can be evaluated easily, although by the definition it is connected with its roots which are ususally hard to find.

Our situation can be visually demonstrated as if a passenger is travelling, crossing several times the border between two states, say France and Germany. What is the difference between the number of crossings the border from France to Germany and from Germany to France? Obviously, it is equal to 0 if the passenger started and finished his travel in the same state; it is equal to 1 if he started in France and finished in Germany and to -1 if he started in Germany and finished in France. His itinerary can be demonstrated as a line similar to the graph in Fig. 3, where France is the area below the x-axis and Germany is above.

Consider now two polynomials, f and g, and assume that, first of all, they have no common roots, and, secondly, that the former (i.e., f) does not vanish at $x = a$, nor at $x = b$. *The characteristics of the polynomial f with respect to the polynomial g on the segment* $[a, b]$ is the difference between the number of roots of the polynomial f contained in the segment $[a, b]$ and being roots with increasing of the polynomial fg and the number of its roots being roots with decreasing for fg. The characteristics is denoted by $(f, g)_a^b$.

The main example, which was the reason to introduce this notion is given by the following proposition.

Theorem 1. *If a polynomial $f(x)$ has no multiple roots and neither it nor its derivative vanishes at the endpoints a and b of the segment $[a, b]$, then the characteristics $(f, f')_a^b$ is equal to the number of roots of the polynomial f contained in the segment $[a, b]$.*

The theorem is an easy consequence of Corollary of Theorem 4, Section 3. We simply state that all the roots of the polynomial $f(x)$ are roots with increasing of the polynomial ff'. Really, according to Theorem 5 of Chapter II, the polynomials f and f' have no common roots. If α is a root of the polynomial $f(x)$ with $f'(\alpha) > 0$, then according to Theorem 4 of Section 3 α is a root with increasing for $f(x)$, and so also for $f(x)f'(x)$, since $f'(x) > 0$ in a neighbourhood of α. If, on the other hand, $f'(\alpha) < 0$, then α is a root with decreasing for $f(x)$, and so again a root with increasing for $f(x)f'(x)$, since $f'(x) < 0$ in a neighbourhood of α.

The characteristics $(f, g)_a^b$ is in fact an expression which can be evaluated using division with remainder. Note first the following simple properties:

a) $(f, -g) = -(f, g)$.

This is obvious since when multiplying the polynomial g by -1, the roots with increasing and the roots with decreasing of the polynomial fg interchange.

b) If $g(a) \neq 0$ and $g(b) \neq 0$, then $(f, g)_a^b + (g, f)_a^b = [fg]_a^b$.

This is also obvious since, by the assumption, the polynomials f and g have no common roots. Hence, the roots of the polynomial fg split into the roots of the polynomial f and those of the polynomial g. The number of roots with increasing (and similarly for roots with decreasing) of the polynomial fg is equal to the sum

of the numbers of such roots of the polynomial f and of the polynomial g, which gives us the equality b).

c) If polynomials g and h take the same values at the roots of a polynomial f (i.e., if $g(\alpha) = h(\alpha)$ whenever $f(\alpha) = 0$), then

$$(f, g)_a^b = (f, h)_a^b.$$

Really, if $g(\alpha) = h(\alpha)$, then a root α of the polynomial $f(x)$ is at the same time a root with increasing (decreasing) for the polynomials fg and fh.

d) If a polynomial f is divisible by a polynomial g, then

$$(f, g)_a^b = [fg]_a^b.$$

Really, the polynomial g has no roots, since its roots would be common roots for the polynomials f and g. Therefore, $(g, f)_a^b = 0$ and from the property b) it follows that $(f, g)_a^b = [fg]_a^b$.

We shall describe now the process of evaluating the characteristics $(f, g)_a^b$. Divide f by g with remainder:

$$f = gq + r. \tag{2}$$

According to property b), we have $(f, g)_a^b = -(g, f)_a^b + [fg]_a^b$. On the other hand, it follows from relation (2) that $f(\alpha) = r(\alpha)$ whenever $g(\alpha) = 0$. Hence, by property c) we obtain that $(f, g)_a^b = (g, r)_a^b$. The obtained equalities together show that

$$(f, g)_a^b = -(g, r)_a^b + [fg]_a^b. \tag{3}$$

As a matter of fact, relation (3) solves our problem, since it reduces the evaluation of the characteristics $(f, g)_a^b$ to the evaluation of the characteristics $(g, r)_a^b$ for the polynomials g and r of smaller degree, because the expression $[fg]_a^b$ is determined by the values of the polynomials f and g at the endpoints a and b of the segment $[a, b]$ (see Table 1).

Our process of passing from the pair (f, g) to a pair of polynomials with smaller degree is the same as in the process of determining the greatest common divisor of the polynomials f and g. In such a case the characteristics is determined by property d).

We intend to improve our result in two directions. Firstly, we shall present in a unified form the final answer which can be obtained after passing from the pair (f, g) to (g, r) and then executing all the divisions in the consecutive steps of the Euclid's algorithm. Secondly, our inductive reasoning needs that conditions imposed on the polynomials f and g ($f(a) \neq 0$, $f(b) \neq 0$) are then imposed to the polynomials g, r etc. We shall show how one can get rid of these additional restrictions.

First of all, we shall transform a bit the answer we have obtained (formula (3)). We start with changing the notation. The polynomial f will be denoted by f_1, g by f_2 and $-r$ by f_3. Taking into account condition a) of the characteristics, formula (3) obtains the form

$$(f_1, f_2)_a^b = (f_2, f_3)_a^b + [f_1 f_2]_a^b, \tag{4}$$

and the formula of division with remainder (formula (2)) the form

$$f_1 = f_2 q_1 - f_3$$

(we have denoted here q by q_1). Now it is clear how to apply formula (4), reducing degrees of polynomials considered. Starting from f_1 and f_2 define polynomials f_i by induction:

$$f_{i-1} = f_i q_{i-1} - f_{i+1}, \tag{5}$$

where the degree of f_{i+1} is smaller than the degree of f_i (assuming that f_{i-1} and f_i are already defined). Clearly, f_{i-1} are just those plynomials which appear as remainders in the Euclid's algorithm, only with the changed signs. After several steps we come to a polynomial f_k, differing eventually only by sign with the $\gcd(f_1, f_2)$.

Applying formula (4) to f_2 and f_3 instead to f_1 and f_2, we obtain that $(f_2, f_3)_a^b = (f_3, f_4)_a^b + [f_2 f_3]_a^b$. Substituting this value for $(f_2, f_3)_a^b$ into formula (4), we get

$$(f_1, f_2)_a^b = (f_3, f_4)_a^b + [f_1 f_2]_a^b + [f_2 f_3]_a^b.$$

Repeating this process k times and noting that $[f_k f_{k+1}]_a^b = 0$ as a result we obtain:

$$(f_1, f_2)_a^b = [f_1 f_2]_a^b + [f_2 f_3]_a^b + \cdots + [f_{k-1} f_k]_a^b. \tag{6}$$

However, in order that we have the right to apply formula (4), we have to assume that $f_i(a) \neq 0$, $f_i(b) \neq 0$ for all $i = 1, 2, \ldots, k$.

Consider carefully the expression $[fg]_a^b$ which can be evaluated using Table 1 for $F = fg$. In our case it can be rewritten as

$$[fg]_a^b = \begin{cases} 0, & \text{if } f(a)g(a) > 0 \text{ and } f(b)g(b) > 0, \text{ or } f(a)g(a) < 0 \text{ and } f(b)g(b) < 0, \\ 1, & \text{if } f(a)g(a) < 0 \text{ and } f(b)g(b) > 0, \\ -1, & \text{if } f(a)g(a) > 0 \text{ and } f(b)g(b) < 0. \end{cases}$$

Table 2

If two numbers A and B, distinct from 0, are given, then one says that in the pair (A, B) there exists one change of sign if A and B are of opposite signs, and that there is no change of sign if they are of the same sign. Using this terminology, one can reformulate information of Table 2, denoting by n the number of changes of sign in the pair $(f(a), f(b))$ and by m the number of changes of sign in the pair $(f(b), g(b))$. Table 2 obtains the form:

$[fg]_a^b$	m	n
0	0	0
0	1	1
1	1	0
-1	0	1

We see that in all the cases we have $[fg]_a^b = m - n$.

We shall apply now the last remark to formula (6). Denote by m_i the number of changes of sign in the pair $(f_i(a), f_{i+1}(a))$, and by n_i the number of changes of sign in the pair $(f_i(b), f_{i+1}(b))$. As a consequence of the remark, formula (6) obtains the form

$$(f_1, f_2)_a^b = m_1 - n_1 + m_2 - n_2 + \cdots + m_k - n_k. \tag{7}$$

What is the meaning of the number $m_1 + m_2 + \cdots + m_k$? One has just to write down the numbers $f_1(a)$, $f_2(a)$, ... , $f_k(a)$ and find out how many changes of sign are there in this sequence—the number of these changes will be $m_1 + m_2 + \cdots + m_k$. In general, if a sequnce of numbers A_1, ... , A_k, distinct from 0, is given, then by the *number of changes of sign* in this sequence, we shall mean the number of places where numbers of opposite signs stay. For example, in the sequence 1, -1, 2, 1, 3, -2 there are 3 changes of sign. We can say that $m_1 + m_2 + \cdots + m_k$ is the number of changes of sign in the sequence $f_1(a)$, $f_2(a)$, ... , $f_k(a)$, and that $n_1 + n_2 + \cdots + n_k$ is the number of changes of sign in the sequence $f_1(b)$, $f_2(b)$, ... , $f_k(b)$. Formula (7) can be interpreted now in the following way:

Theorem 2. *If none of the terms f_1, ... , f_k of Sturm's sequence of polynomials f_1, f_2 vanishes, either in a, or in b, and the polynomials f_1, f_2 have no common roots, then the characteristics $(f, g)_a^b$ is equal to the difference between the numbers of changes of sign in the sequences of values of polynomials in Sturm's sequence at the points a and b.*

We have now to get rid of the restrictions $f_i(a) \neq 0$, $f_i(b) \neq 0$ for $i = 1, \ldots, k$, which can be uncomfortable in applications: we shall assume just that $f_1(a) \neq 0$ and $f_1(b) \neq 0$. In order to do that we have to generalize a bit the notion of the number of changes of sign. If some of the terms in the sequence A_1, ... , A_k are equal to 0, then the number of changes of sign in it is defined as the number of changes of sign in the sequence which is obtained by deleting all the zeros in the given sequence. For example, deleting zeros in the sequence 1, 0, 2, -1, 0, 3, 1, the sequence 1, 2, -1, 3, 1 is obtained, and the latter has two changes of sign. Hence, the given sequence has two changes of sign, by definition.

Denote now by ε the distance from a to the nearest root (distinct from a) of any of the polynomials $f_i(x)$. Thus, $f_i(x) \neq 0$ for $a < x < a + \varepsilon$. Choose any such value a', $a < a' < a + \varepsilon$. A value b' is chosen analogously. Let us state a lemma.

Lemma. *The number of changes of sign in the sequence $f_1(a)$, ... , $f_k(a)$ is equal to the number of changes of sign in the sequence $f_1(a')$, ... , $f_k(a')$. The same is true when a and a' are replaced by b and b'.*

First of all, let us show that the Lemma can really help us to extend Theorem 2 to arbitrary polynomials f_1, f_2 with the only conditions that $f_1(a) \neq 0$, $f_1(b) \neq 0$ and that f_1 and f_2 have no common roots.

Really, by the assumption, the polynomial f_1 has no roots in the segments $[a, a']$ and $[b', b]$. Hence, all of its roots contained in the segment $[a, b]$, are already contained in the segment $[a', b']$. Therefore, $(f_1, f_2)_a^b = (f_1, f_2)_{a'}^{b'}$. Theorem 2 can now be applied to the characteristics $(f_1, f_2)_{a'}^{b'}$. The number of changes of sign in

the sequence $f_1(a'), \ldots, f_k(a')$, as well as in the sequence $f_1(b'), \ldots, f_k(b')$, is determined by the Lemma. Thus, we obtain the wanted result:

Theorem 3. *If polynomials f_1 and f_2 have no common roots, $f_1(a) \neq 0$ and $f_1(b) \neq 0$, then the characteristics $(f_1, f_2)_a^b$ is equal to the difference between the numbers of changes of sign in the sequence $f_1(a), \ldots, f_k(a)$ and in the sequence $f_1(b), \ldots, f_k(b)$, where $f_1(x), \ldots, f_k(x)$ is Sturm's sequence corresponding to the pair of polynomials f_1, f_2.*

We shall show now that the Lemma is valid. Consider, for example, the value $x = a$. Suppose that $f_i(a) = 0$ for some $i = 1, \ldots, k$. By the assumption, $i \neq 1$ since $f_1(a) \neq 0$. Also, $i \neq k$ since the polynomial $f_k(x)$ can differ from $\gcd(f_1, f_2)$ only by sign and so it is a number distinct from 0. Note that then $f_{i-1}(a) \neq 0$ and $f_{i+1}(a) \neq 0$. Really, if we had, for example, $f_i(a) = 0$, $f_{i+1}(a) = 0$, then it would follow from formula (5) that $f_{i-1}(a) = 0$. In exactly the same way, this would imply that $f_{i-2}(a) = 0$ etc., and finally $f_1(a) = 0$, which would contradict the original assumption. But we can say even more—not only that the numbers $f_{i-1}(a)$ and $f_{i+1}(a)$ are distinct from 0, but they have opposite signs—it follows immediately by substituting $x = a$ into equality (5) and taking into account the assumption that $f_i(a) = 0$.

Compare now the sequences $f_1(a), \ldots, f_k(a)$ and $f_1(a'), \ldots, f_k(a')$. Let $f_i(a) = 0$. Then, as we have seen, $f_{i-1}(a) \neq 0$ and $f_{i+1}(a) \neq 0$, and $f_{i-1}(a)$ and $f_{i+1}(a)$ have opposite signs. But then $f_{i-1}(a') \neq 0$ and $f_{i+1}(a') \neq 0$, and $f_{i-1}(a')$ has the same sign as $f_{i-1}(a)$, while $f_{i+1}(a')$ has the same sign as $f_{i+1}(a)$. This follows from the fact that the polynomials f_{i-1} and f_{i+1} have no roots in the segment $[a, a']$, and so (by Bolzano's theorem) they can have no values of opposite signs. Write down the respective parts of our sequences. Suppose that $f_{i-1}(a) > 0$. Then we obtain the following table:

	$f_{i-1}(x)$	$f_i(x)$	$f_{i+1}(x)$
$x = a$	$+$	0	$-$
$x = a'$	$+$	$?$	$-$

The characteristics $(f_1, f_2)_{a'}^{b'}$ depends on the number of changes of sign in the lowest row. But we see that it coincides with the number of changes of sign in the row above it—whatever the unkown sign, denoted by ?, is, there will be exactly one change of sign in each of the rows. The case when $f_{i-1}(a) < 0$ can be treated exactly in the same way. The Lemma is proved.

Combining Theorem 3 with Theorem 1 we obtain the basic result:

Theorem 4. (Sturm's Theorem) *If a polynomial $f(x)$ has no multiple roots and does not vanish for $x = a$ and $x = b$, then the number of its roots in the segment $[a, b]$ is equal to the difference between the number of changes of sign of the values of polynomials in the Sturm's sequence, formed for the polynomials $f(x)$ and $f'(x)$ at $x = a$ and $x = b$.*

One has only to note that the lack of multiple roots of the polynomial $f(x)$ is equivalent to the lack of common roots of the polynomials $f(x)$ and $f'(x)$—this is just the assertion of Theorem 5 of Chapter II. Therefore we can apply Theorem 1 to the polynomial $f(x)$ and then Theorem 3 to the pair of polynomials $f(x)$ and $f'(x)$.

Sturm's theorem gives a possibility to answer the basic questions about distribution of roots of a polynomial. First of all, using the theorem, the number of roots can be determined. In order to do that, it is enough to remember Theorem 3 of Section 3, which indicates a number N such that all the roots of the polynomial lie between $-N$ and N. After that it is sufficient to apply Sturm's theorem to the segment $[-N, N]$. However, it is remarkable that in order to determine the number of roots it is neither necessary to evaluate the number N (using Theorem 3), nor to evaluate the values of polynomials in Sturm's sequence for $x = -N$ and $x = N$. Really, for applying Sturm's theorem it is not necessary to know the values $f_i(\pm N)$ themselves, but only their *signs*. That is why it is sufficient to choose a number N large enough, such that the segment $[-N, N]$ contains not only all the roots of the polynomial $f_1(x)$, but also all the roots of all the polynomials $f_i(x)$ of the Sturm's sequence (i.e., we can choose a respective number N_i for each polynomial $f_i(x)$ and take for N the largest of them). According to Corollary 1 of Theorem 3, Section 3, the sign of the value $f_i(N)$, resp. $f_i(-N)$, coincides with the sign of the leading term of the polynomial $f_i(x)$ for $x = N$, resp. $x = -N$. They are determined by the sign of the leading coefficient of the polynomial $f_i(x)$ and by the parity of its degree. Therefore, there is no need to evaluate N and the values $f_i(N)$ and $f_i(-N)$.

When the number of roots is determined, it is possible to indicate segments, each of which contains exactly one root. In order to do that it is already necessary to evaluate the number N, indicated in Theorem 3 of Section 3. After that the segment $[-N, N]$ is divided into two equal parts and using Sturm's theorem the number of roots in each part is found. Then the same is done with the segments $[-N, 0]$ and $[0, N]$ and the process is continued till each of the segments contains only one root.

If it is known that a segment $[a, b]$ contains exactly one root of the polynomial $f(x)$ and the polynomial has no multiple roots, then the values $f(a)$ and $f(b)$ must be of opposite signs. Really, if the root is equal to α, then, according to Theorem 4 of Section 3, for ε small enough, the values $f(\alpha - \varepsilon)$ and $f(\alpha + \varepsilon)$ have the same sign. But $f(\alpha - \varepsilon)$ and $f(a)$ have to be of the same sign—otherwise the polynomial would have one more root in the segment $[\alpha - \varepsilon, \alpha]$. The same is true for the values $f(\alpha + \varepsilon)$ and $f(b)$. Thus, $f(a)$ has the same sign as $f(\alpha - \varepsilon)$, $f(b)$ the same as $f(\alpha + \varepsilon)$, and $f(\alpha - \varepsilon)$ and $f(\alpha + \varepsilon)$ have opposite signs. Hence, $f(a)$ and $f(b)$ have opposite signs. Knowing that, it is possible to evaluate the root α with arbitrary level of accuracy. It is sufficient to divide the segment $[a, b]$ into two parts by a point c and evaluate $f(c)$. Either $f(a)$ and $f(c)$, or $f(c)$ and $f(b)$ have opposite signs. In the former case α is contained in the segment $[a, c]$, and in the latter—in the segment $[c, b]$. After that we continue the process with the segment containing α until we include α in a segment of arbitrary small length. This means that we have evaluated it with arbitrary level of accuracy.

Consider, for example, the polynomial $f(x) = x^3 + 3x - 1$. Applying the criterion from Section 3, we have to evaluate the expression $4a^3+27b^2 = 4\cdot 27+27$. Since it is positive, the polynomial has one root. Applying Theorem 3 of Section 3, we find the value $N = 3$. Therefore, the root is contained between -3 and 3, where $f(-3) < 0$, $f(3) > 0$. Since $f(0) < 0$, the root is contained between 0 and 3. Since $f(1) = 3$ and $f(2) = 13$, the root is contained between 0 and 1. In order to find its first decimal, we have to determine in which of the 10 segments (between 0 and 1/10, 1/10 and 2/10, ... , 9/10 and 1) it lies. Put first $x = 1/2$, then $f(x) = 5/8$. Since $f(0)$ and $f(1/2)$ are of opposite signs, the root is contained between 0 and 1/2. Put now $x = 3/10$. Since $f(\frac{3}{10}) = \frac{27}{1000}+\frac{9}{10}-1 = \frac{27}{1000}-\frac{1}{10} < 0$, the root is contained between 3/10 and 5/10. Finally, $f(\frac{4}{10}) = \frac{64}{1000} + \frac{12}{10} - 1 > 0$. Hence, the root lies between 3/10 and 4/10 and it has the form $\alpha = 0{,}3\ldots$.

Since Sturm's theorem has an elegant formulation and a lot of applications, it became widely known immediately after it had been proved. Jacques Sturm, a French mathematician who had proved it, when teaching about the theorem in his lectures, used to say: "Now I will prove a theorem, the name of which I have the honor to bare".

Problems

1. Construct Sturm's sequence for the polynomials $f(x)$ and $f'(x)$ if $f(x) = x^2 + ax + b$ or $f(x) = x^3 + ax + b$. Using Sturm's theorem deduce again the results about the numbers of roots of these polynomials, obtained already at the end of Section 3. *Hint.* In the case of $f(x) = x^3 + ax + b$ consider separately different cases of possible signs for a and $D = 4a^3 + 27b^2$.

2. Determine, using Sturm's theorem, the number of roots of the polynomial $x^n + ax + b$, depending on n (more precisely, on its parity), a and b.

3. Find the number of roots of the polynomial $x^5 - 5ax^3 + 5a^2x + 2b$. *Hint.* The answer depends on the sign of the expression $a^5 - b^9$.

4. Let a be a root of the derivative $f'(x)$ of a polynomial $f(x)$. Put $f_1(x) = f(x)$, $f_2(x) = f'(x)/(x-a)$. Let $f(x)$ has no multiple roots, and $f_1(x), \ldots , f_k(x)$ is Sturm's sequence for the polynomials $f_1(x)$ and $f_2(x)$. Express the number of roots of the polynomial $f(x)$ in terms of the number of changes of sign in the sequences $f_i(N)$, $f_i(a)$ and $f_i(-N)$, $i = 1, \ldots, k$ where N is a sufficiently large number.

5. Let two polynomials f_1 and f_2 be given, with degrees n and $n - 1$, respectively, and suppose that in their Sturm's sequence the degree of the polynomial $f_i(x)$ is $n - i + 1$ and its leading coefficient is positive. Prove that the polynomial $f_1(x)$ has n roots. Moreover, each of the polynomials $f_i(x)$ has $n-i+1$ roots, and between each two adjacent roots of the polynomial $f_i(x)$ there lies a root of the polynomial $f_{i+1}(x)$.

6. Let a polynomial $f(x)$ of degree n has n roots. Prove that in the Sturm's sequence (for the polynomials f and f') each polynomial has the degree which is smaller exactly by 1 than the degree of the previous one, and all the leading coefficients are positive. Prove that these conditions are sufficient in order that a polynomial of degree n has n roots.

I. R. Shafarevich, Russian Academy of Sciences, Moscow, Russia