Little Mathematics Library

A. G. KUROSH

ALGEBRAIC EQUATIONS OF ARBITRARY DEGREES

Mir Publishers · Moscow

ИЗДАТЕЛЬСТВО · МИР ·

А. Г. Курош

АЛГЕБРАИЧЕСКИЕ УРАВНЕНИЯ ПРОИЗВОЛЬНЫХ СТЕПЕНЕЙ

Издательство «Наука»

A.G.Kurosh

ALGEBRAIC EQUATIONS OF ARBITRARY DEGREES

Translated from the Russian
by
V. Kisin

MIR PUBLISHERS
MOSCOW

First published 1977
Revised from the 1975 Russian edition

На английском языке

Contents

Preface

This booklet is a revision of the author's lecture to high school students taking part in the Mathematics Olympiad at Moscow State University. It gives a review of the results and methods of the general theory of algebraic equations with due regard for the level of knowledge of its readers. No proofs are included in the text since this would have required copying almost half of a university textbook on higher algebra. Despite such an approach, this booklet does not make for light reading. Even a popular mathematics book calls for the reader's concentration, thorough consideration of all the definitions and statements, check of calculations in all the examples, application of the methods described to his own examples, etc.

A secondary school course of algebra is diversified but equations are its focus. Let us restrict ourselves to equations with one unknown, and recall what is taught in secondary school.

Any pupil can solve *first degree* equations: if an equation

$$ax + b = 0$$

is given, in which $a \neq 0$, then its single root is

$$x = -\frac{b}{a}$$

Furthermore, a pupil knows the formula for solving *quadratic* equations

$$ax^2 + bx + c = 0$$

where $a \neq 0$:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

If the coefficients of this equation are real numbers and if the number under the radical sign is positive, i. e. $b^2 - 4ac > 0$, then this formula yields two different real roots. But if $b^2 - 4ac = 0$, our equation has a single root; in this case the root is called a *multiple* one. For $b^2 - 4ac < 0$ the equation has no real roots.

Finally, a pupil can solve certain types of *third* and *fourth-degree* equations whose solution is easily reduced to that of quadratic equations. For example, the third-degree (cubic) equation

$$ax^3 + bx^2 + cx = 0$$

which has one root $x = 0$, and, after factoring out x, is transformed into a quadratic equation

$$ax^2 + bx + c = 0$$

A fourth-degree (quartic) equation

$$ay^4 + by^2 + c = 0$$

called *biquadratic*, may also be reduced to a quadratic equation by setting $y^2 = x$, calculating the roots of the resulting quadratic equation and then extracting their square roots.

Let us emphasize once again that these are only very special types of cubic and quartic equations. Secondary school algebra gives no methods of solving arbitrary equations of these degrees, and all the more so of higher degrees. However, we encounter higher degree algebraic equations in different branches of engineering, mechanics and physics. The theory of algebraic equations of an arbitrary degree n, where n is a positive integer, has required centuries to develop and now constitutes one of the main parts of *higher algebra* taught at universities and pedagogical institutes.

1. Complex Numbers

The theory of algebraic equations is essentially based on the theory of complex numbers taught at high school. However, students often doubt the justification for introducing these numbers and their actual existence. When complex numbers were introduced, even mathematicians doubted their actual existence, hence the term "imaginary numbers" which still survives. However, modern science sees nothing mysterious in the complex numbers, and they are no more "imaginary" than negative or irrational numbers.

The necessity for complex numbers was caused by the fact that it is impossible to extract a square root of a negative real number and still remain in the field of real numbers. As a result some quadratic equations have no real roots; the equation

$$x^2 + 1 = 0$$

is the simplest of such equations. Is there a way to expand the realm of numbers so that these equations also possess roots?

In his study of mathematics at school the student sees the system of numbers at his disposal constantly extended. He starts with *integral positive numbers* in elementary arithmetic. Very soon *fractions* appear. Algebra adds negative numbers, thus forming the system of all *rational numbers*. Finally, introduction of irrational numbers results in the system of all *real* numbers.

Each of these consecutive expansions of the store of numbers makes it possible to find roots for some of the equations which previously had no roots. Thus, the equation

$$2x - 1 = 0$$

acquires a root only after fractions are introduced, the equation

$$x + 1 = 0$$

has a root after the introduction of negative numbers, and the equation

$$x^2 - 2 = 0$$

has a root only after irrational numbers are added.

All this completely justifies one more step on the way to enlarge the store of numbers. We shall now proceed to a general outline of this last step.

It is known that if a positive direction is fixed on a given straight line, if the origin O is marked, and if a unit of scale is chosen (Fig. 1), then each point A on this line can be put in

FIG. 1

correspondence with its *coordinate*, i. e. with a real number which expresses in the chosen units of scale the distance between A and O if A lies to the right of the point O, or the distance taken with a minus sign, if A lies to the left of O. In this manner all the points on the line are put in correspondence with different real numbers, and it can be proved that each real number will be used in the process. Therefore we can assume that the points of our line are images of the real numbers corresponding to them, i. e. these numbers are as if aligned along the straight line. Let us call our line the *line of numbers.*

But is it possible to expand the store of numbers in such a way that new numbers can be represented in just as natural a manner by the points of a plane? So far we have not constructed a system of numbers wider than that of real numbers.

We shall start by indicating the "material" with which this new system of numbers is to be "constructed", i. e. what objects will act as new numbers. We must also define how to carry out algebraic operations – addition and multiplication, subtraction and division – on these objects, i. e. on these future numbers. Since we want to construct numbers which can be represented by all the points of the plane, the simplest way is to consider the points of the plane themselves as the new numbers. To seriously consider these points as numbers, we must merely define how to carry out algebraic operations with them, i. e. which point is to be the sum of two given points of the plane, which one is to be their product, etc.

Just as the position of a point on a straight line is completely defined by a single real number, its coordinate, the position of an arbitrary point on a plane can be defined by a pair of real numbers. To do this let us take two perpendicular straight lines, intersecting on the plane at the point O and on each of them fix the positive direction and set off a unit of scale (Fig. 2). Let

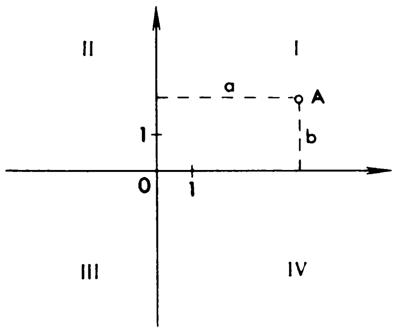

FIG. 2

us call these lines *coordinate axes*, the horizontal line the *abscissa axis*, and the vertical line the *ordinate axis*. The coordinate axes divide the entire plane into four quadrants, which are numbered as shown in Fig. 2.

The position of any point A in the first quadrant (see Fig. 2) is completely defined by two positive real numbers, e. g. the number a which gives in the selected units of scale the distance from this point to the ordinate axis (the *abscissa* of point A), and the number b which gives in the selected units of scale its distance from the abscissa axis (the *ordinate* of the point A). Conversely, for each pair (a, b) of positive real numbers we can indicate a single precisely defined point in the first quadrant with a as its abscissa and b as its ordinate. Points in other quadrants are defined in a similar manner. However, to ensure a mutual one-to-one correspondence between all the points of the plane and the pairs of *coordinates* (a, b), i.e. in order to avoid the same pair of coordinates (a, b) corresponding to several distinct points on the plane, we assume the abscissas of points in quadrants *II* and *III* and the ordinates of points in quadrants *III* and *IV* to be negative. Note that points on the abscissa axis are given by coordinates of the type $(a, 0)$, and those on the ordinate axis by coordinates of the type $(0, b)$, where a and b are certain real numbers.

We are now able to define all the points on the plane by pairs of real numbers. This enables us to talk further not of a point A, given by the coordinates (a, b), but simply of a point (a, b).

Let us now define addition and multiplication of the points on the plane. At first these definitions may seem extremely artificial. However, only such definitions will make it possible to realize our goal of taking square roots of negative real numbers.

Let the points (a, b) and (c, d) be given on the plane. Until now we did not know how to define the sum and the product of these points. Let us call their *sum* the point with the abscissa $a + c$ and the ordinate $b + d$, i. e.

$$(a, b) + (c, d) = (a + c, b + d)$$

On the other hand, let us call the *product* of the given points the point with the abscissa $ac - bd$ and the ordinate $ad + bc$, i. e.

$$(a, b)(c, d) = (ac - bd, ad + bc)$$

It can easily be checked that the above-defined operations on the points on the plane possess all the familiar properties of operations on numbers: addition and multiplication of points on the plane are commutative (i. e. both addends and co-factors can be interchanged); associative (i. e. the sum and the product of three points are independent of the position of the brackets) and distributive (i. e. the brackets can be removed). Note that the law of association for addition and multiplication of points makes it possible to introduce in an unambiguous way the sum and the product of any finite number of points on the plane.

Now we can also perform the operations of subtraction and division of points on the plane, inverse to addition and multiplication, respectively (in the sense that in any system of numbers, the difference between two numbers can be defined as the number which when added to the subtrahend yields the minuend, and the quotient of two numbers as the number which when multiplied

3—762

by the divisor yields the dividend). Thus,

$$(a,\ b) - (c,\ d) = (a - c,\ b - d)$$

$$\frac{(a,\ b)}{(c,\ d)} = \left(\frac{ac + bd}{c^2 + d^2},\ \frac{bc - ad}{c^2 + d^2}\right)$$

The reader will easily see that the product (as defined above) of the point on the right-hand side of the last equality by the point (c, d) is indeed equal to the point (a, b). It is even simpler to verify that the sum of the point on the right-hand side of the first equality and the point (c, d) is indeed equal to the point (a, b).

By applying our definitions to the points on the abscissa axis, i. e. to the points of the type $(a, 0)$, we obtain:

$$(a,\ 0) + (b,\ 0) = (a + b,\ 0)$$

$$(a,\ 0)(b,\ 0) = (ab,\ 0)$$

i. e. addition and multiplication of these points reduce to addition and multiplication of their abscissas. The same is valid for subtraction and division:

$$(a,\ 0) - (b,\ 0) = (a - b,\ 0)$$

$$\frac{(a,\ 0)}{(b,\ 0)} = \left(\frac{a}{b},\ 0\right)$$

If we assume that each point $(a, 0)$ of the abscissa axis represents its abscissa, i. e. the real number a, in other words, if we identify the point $(a, 0)$ with the number a, then the abscissa axis will be simply turned into a line of numbers. We can now assume that the new system of numbers, constructed from the points on the plane, contains in particular all the real numbers as points of the abscissa axis.

The points on the ordinate axis, however, cannot be identified with real numbers. For example, let us consider the point $(0, 1)$ which lies at a distance 1 upward of the point O. Let us denote this point by the letter i:

$$i = (0,\ 1)$$

and let us find its square in the sense of multiplication of the points on the plane:

$$i^2 = (0,\ 1)(0,\ 1) = (0 \cdot 0 - 1 \cdot 1,\ 0 \cdot 1 + 1 \cdot 0) = (-1,\ 0)$$

However, the point $(-1, 0)$ lies on the abscissa axis, not on the ordinate axis, and thus represents a real number -1, i. e.

$$i^2 = -1$$

Hence, we have found in our new system of numbers a number whose square is equal to a real number -1, i. e. we can now find the square root of -1. Another value of this root is given by the point $-i = (0, -1)$. Note that the point $(0, 1)$, which we denoted as i, is a precisely defined point on the plane, and the fact that it is usually referred to as an "imaginary unity" does not in the least prevent it from actually existing on the plane.

The system of numbers we have just constructed is more extensive than that of real numbers and is called the system of *complex numbers*. The points on the plane together with the operations we have defined are called *complex numbers*. It is not difficult to prove that any complex number can be expressed by real numbers and the number i by means of these operations. For example, let us take point (a, b). By virtue of the definition of addition, the following is valid:

$$(a, b) = (a, 0) + (0, b)$$

The addend $(a, 0)$ lies on the abscissa axis and is therefore a real number a. By virtue of the definition of multiplication, the second addend can be written in the form

$$(0, b) = (b, 0)(0, 1)$$

The first factor on the right-hand side of this equality coincides with a real number b, and the second factor is equal to i. Therefore,

$$(a, b) = a + bi$$

where addition and multiplication are understood as operations on the points of the plane.

By means of this standard notation of complex numbers we can immediately rewrite the above formulas for operations with complex numbers:

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

$$(a + bi) - (c + di) = (a - c) + (b - d)i$$

$$\frac{a + bi}{c + di} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2} i$$

3*

It should be noted that the above definition of multiplication of the points of the plane is in perfect agreement with the law of distribution: if on the left-hand side of the second of the above equations we calculate the product by the rule of binomial multiplication (which itself stems from the law of distribution), and then apply the equality $i^2 = -1$ and reduce similar terms, we shall arrive precisely at the right-hand side of the second equation.

2. Evolution. Quadratic Equations

Having complex numbers at our disposal, we can extract square roots not only of the number -1, but of any negative real number, always obtaining two distinct values. If $-a$ is a negative real number, i. e. $a > 0$, then

$$\sqrt{-a} = \pm \sqrt{a}\, i$$

where $\sqrt{a}$ is the positive value of the square root of the positive number a.

Returning to the solution of the quadratic equation with real coefficients, we can now say that where $b^2 - 4ac < 0$, this equation also has two distinct roots, this time complex.

Now we are able to take square roots of any complex numbers, not only real ones. For instance, if a complex number $a + bi$ is given, then

$$\sqrt{a+bi} = \sqrt{\frac{1}{2}(a + \sqrt{a^2+b^2})} + i\sqrt{\frac{1}{2}(-a + \sqrt{a^2+b^2})}$$

where the positive value of the radical $\sqrt{a^2 + b^2}$ is taken in both terms. Of course, the reader will see that for any a and b both the first term on the right-hand side and the coefficient in i will be real numbers. Each of these two radicals possesses two values which are combined with each other according to the following rule: if $b > 0$, then the positive value of one radical is added to the positive value of the other, and the negative one to the negative value of the other; if, on the contrary, $b < 0$, then the positive value of one radical is added to the negative value of the other.

Example. Extract the square root of the number $21 - 20i$. Here

$$\sqrt{a^2 + b^2} = \sqrt{441 + 400} = 29$$

$$\sqrt{\frac{1}{2}(a + \sqrt{a^2 + b^2})} = \sqrt{\frac{1}{2}(21 + 29)} = \pm 5$$

$$\sqrt{\frac{1}{2}(-a + \sqrt{a^2 + b^2})} = \sqrt{\frac{1}{2}(-21 + 29)} = \pm 2$$

Since $b = -20$, i. e. $b < 0$, we must combine the values of the last two radicals with opposite signs, i. e.

$$\sqrt{21 - 20i} = \pm (5 - 2i)$$

Knowing how to take square roots of complex numbers, we can now solve quadratic equations with arbitrary complex coefficients. Indeed, the derivation of the formula for solving quadratic equations still holds for complex coefficients, and the calculation of the square root in this formula can be reduced, as shown above, to evolution of square roots of two positive real numbers. Hence, a quadratic equation with arbitrary complex coefficients has two roots, which may in fact coincide, i. e. yield a single multiple root.

Example. Solve the equation

$$x^2 - (4 - i)x + (5 - 5i) = 0$$

Applying the formula, we obtain

$$x = \frac{(4 - i) \pm \sqrt{(4 - i)^2 - 4(5 - 5i)}}{2} = \frac{(4 - i) \pm \sqrt{-5 + 12i}}{2}$$

Calculating the square root in this expression by the method described in the preceding section, we find that

$$\sqrt{-5 + 12i} = \pm (2 + 3i)$$

and thus

$$x = \frac{(4 - i) \pm (2 + 3i)}{2}$$

Therefore, the roots of our equation are the numbers

$$x_1 = 3 + i, \quad x_2 = 1 - 2i$$

It can easily be checked that each of these numbers indeed satisfies the equation.

Let us now turn to the problem of extracting roots of an arbitrary positive integral index n from complex numbers. It can be proved that for any complex number α there exist exactly n distinct complex numbers such that raised to the power n (i. e. if we take a product of n factors equal to this number), each yields the number α. In other words, the following extremely important theorem holds:

A root of order n of any complex number has exactly n distinct complex values.

This theorem is equally applicable to real numbers, which are a particular case of complex numbers: the nth root of a real number a has precisely n distinct values which in a general case are complex. We know that among these values there will be two, one or no real numbers, depending on the sign of the number a and parity of the index n.

Thus, the cube root of one has three values:

$$1, \ -\frac{1}{2} + i\frac{\sqrt{3}}{2} \text{ and } -\frac{1}{2} - i\frac{\sqrt{3}}{2}$$

It is easily verified that each of these three numbers, raised to the power three, yields unity. The values of the root of the order four of unity are the numbers

$$1, \ -1, \ i \text{ and } -i$$

In section 1 we gave the formula for taking the square root of a complex number $a + bi$. This formula reduces calculation of the root to extracting square roots of two positive real numbers. Unfortunately, for $n > 2$ no formula exists which would express the nth root of a complex number $a + bi$ in terms of real values of radicals of certain auxiliary real numbers; it was proved that no such formula can ever be derived. Roots of order n of complex numbers are usually taken by representing the complex numbers in the so-called *trigonometric* form; however, this subject will not be treated in this booklet.

3. Cubic Equations

The formula for solving quadratic equations is also valid for complex coefficients. For third-degree equations, usually called cubic equations, we can also derive a formula, which, although more complicated, expresses with radicals the roots of these equations in terms of coefficients. This formula is also valid for equations with arbitrary complex coefficients.

Let an equation

$$x^3 + ax^2 + bx + c = 0$$

be given. We transform this equation, setting

$$x = y - \frac{a}{3}$$

where y is a new unknown. Substituting this expression of x into our equation, we obtain a cubic equation with respect to y, which is simpler, since the coefficient of y^2 will be zero. The coefficient of the first power of y and the absolute term will be, respectively, the numbers

$$p = -\frac{a^2}{3} + b, \quad q = \frac{2a^3}{27} - \frac{ab}{3} + c$$

i. e. the equation can be written as

$$y^3 + py + q = 0$$

If we subtract $\frac{a}{3}$ from the roots of this new equation, then we obtain the roots of the original equation.

The roots of our new equation are expressed in terms of its coefficients by the following formula:

$$y = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

We know that each of the three cube radicals has three values. However, these values cannot be combined in an arbitrary manner. It so happens that for each value of the first radical, the value of the second must be chosen, such that their product equals the number $-\frac{p}{3}$. These two values of the radicals must be added together to obtain a root of the equation. Thus we obtain the three roots of our equation. Therefore, each cubic equation with numerical coefficients has three roots, which in a general case are complex; obviously, some of these roots may coincide, i. e. constitute a multiple root.

The practical significance of the above formula is extremely small. Indeed, let the coefficients p and q be real numbers. It can be shown that if the equation

$$y^3 + py + q = 0$$

has three distinct real roots, then the expression

$$\frac{q^2}{4} + \frac{p^3}{27}$$

will be negative. Since the expression is under the square root sign in the formula, extracting this root will yield a complex number under each of the two cube root signs. We mentioned above that extraction of cube roots of complex numbers requires trigonometric notation, but this can be done only approximately, by means of tables.

Example. The equation

$$x^3 - 19x + 30 = 0$$

does not contain a square of the unknown, and therefore we apply the above formula to this equation with no preliminary transformations. Here $p = -19$, $q = 30$, and hence

$$\frac{q^2}{4} + \frac{p^3}{27} = -\frac{784}{27}$$

i. e. the result is negative. The first of the cube radicals in the formula yields

$$\sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} = \sqrt[3]{-15 + \sqrt{-\frac{784}{27}}} = \sqrt[3]{-15 + i\sqrt{\frac{784}{27}}}$$

We cannot express this cube radical in terms of the radicals of real numbers and thus cannot find the roots of our equation by this formula. But a direct verification demonstrates that these roots are the integers 2, 3 and -5.

In practice the above formula for solving cubic equations yields the roots of equations only when the expression $\frac{q^2}{4} + \frac{p^3}{27}$ is positive or equal to zero. In the first instance the equation has one real and two complex roots; in the second instance all the roots are real but one of them is multiple.

Example. We want to solve the cubic equation

$$x^3 - 9x^2 + 36x - 80 = 0$$

Setting

$$x = y + 3$$

we obtain the "reduced" equation

$$y^3 + 9y - 26 = 0$$

Applying the formula, we obtain

$$\frac{q^2}{4} + \frac{p^3}{27} = 196 = 14^2$$

and therefore

$$\sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} = \sqrt[3]{13 + 14} = \sqrt[3]{27}$$

One of the values of this cube radical is the number 3. We know that the product of this value and the corresponding value of the second cube radical in the formula must be equal to the number $-\frac{p}{3}$, i.e. it must be equal to the number -3. Therefore the value of the second radical will be the number -1, and one of the roots of the reduced equation is

$$y_1 = 3 + (-1) = 2$$

Knowing one of the roots of the cubic equation, we can obtain the other two in many different ways. For instance, we could find the other two values of $\sqrt[3]{27}$, calculate the corresponding values of the second radical and add up the mutually corresponding values of the radicals. Or we may divide the left-hand side of the reduced equation by $y - 2$, after which we only need to solve a quadratic equation. Either of these methods will demonstrate that the other two roots of our reduced equation are the numbers

$$-1 + i\sqrt{12} \text{ and } -1 - i\sqrt{12}$$

Therefore, the roots of the original cubic equation are the numbers

$$5,\ 2 + i\sqrt{12} \text{ and } 2 - i\sqrt{12}$$

Of course, calculation of radicals is not always as easy as in the carefully selected example discussed above; much more often they have to be calculated approximately, yielding only approximate values of the roots of an equation.

4. Solution of Equations in Terms of Radicals and the Existence of Roots of Equations

Quartic equations also allow a formula to be worked out which expresses the roots of these equations in terms of their coefficients. Involving still more "multi-storied" radicals, this formula is much more complicated than the one for cubic equations, and its practical application is very limited. However, this formula shows that any quartic equation with numerical coefficients has four complex roots, some of which may be real.

The formulas for solving third- and fourth-degree equations were found as early as the 16th century, and attempts were begun to find a formula for solving equations of the fifth and higher degrees. Note that a general form of an equation of degree n, where n is a positive integer, is

$$a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n = 0$$

The search continued unsuccessfully until the beginning of the 19th century, when the following spectacular result was proved:

For any n, greater than or equal to five, no formula can be found which would express the roots of any equation of degree n by its coefficients in terms of radicals.

In addition, for any n, greater than or equal to five, an equation can be written of degree n with *integral* coefficients, whose roots are not expressible in radicals, however complicated, if the radicands involve only integral or fractional numbers. Such is the equation

$$x^5 - 4x - 2 = 0$$

It can be proved that this equation has five roots, three real and two complex, but none of these roots can be expressed in radicals, i. e. this equation is "unsolvable in terms of radicals". Therefore, the store of numbers, both real and complex, which form the roots of equations with integral coefficients (such numbers are called *algebraic* as opposed to *transcendent* numbers which are not the roots of any equations with integral coefficients), is much greater than that of the numbers which can be written in terms of radicals.

The *theory of algebraic numbers* is an important branch of algebra; Russian mathematicians E. I. Zolotarev (1847-1878), G. F. Voronoi (1868-1908), N. G. Chebotarev (1894-1947) made valuable contributions in this field.

Abel (1802-1829) proved that deriving general formulas for solving equations of degrees $n \geqslant 5$ in terms of radicals was impossible. Galois (1811-1832) demonstrated the existence of equations with integral coefficients, unsolvable in terms of radicals. He also found the conditions under which the equation can be solved in terms of radicals. This required a new, profound theory, namely the *group theory*. The concept of a group made it possible to finally settle this problem. Later it found numerous other applications in various branches of mathematics and other sciences and became one of the most important objects of study in algebra. We shall not present the definition of this concept but shall only mention that presently Soviet mathematicians are pioneering in the development of group theory.

As far as practical determination of roots of equations is concerned, the absence of formulas for solving nth degree equations where $n \geqslant 5$ causes no serious difficulties. Numerous methods of approximate solution of equations suffice, and even for cubic equations these methods are much quicker than the application of a formula (where applicable), followed by approximate evolution of real radicals. However, the existence of the formulas for quadratic, cubic and quartic equations makes it possible to prove that these equations possess two, three or four roots, respectively. But what about the roots of nth degree equations?

If there were equations with numerical coefficients, either real or complex, which possessed no real or complex roots, the store of real numbers would have to be extended. However, this is unnecessary since complex numbers are sufficient to solve any equation with numerical coefficients. The following theorem holds:

Any equation of degree n with any numerical coefficients has n roots, complex or, in certain cases, real; some of these roots may coincide, i. e. form multiple roots.

This theorem is called the *basic theorem of higher algebra.* It was proved by D'Alembert (1717-1783) and Gauss (1777-1855) as early as the 18th century, although these proofs were perfected to complete rigorousness only in the 19th century; at present there exist several dozen different proofs of this theorem.

The concept of a multiple root, mentioned in the basic theorem, means the following. It can be proved that if an nth degree equation

$$a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n = 0$$

has n roots $\alpha_1, \alpha_2, \dots, \alpha_n$, then the left-hand side of the equation can be factored in the following manner:

$$a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n = a_0(x - \alpha_1)(x - \alpha_2)\dots(x - \alpha_n)$$

Conversely, if such a factorization is given for the left-hand side of our equation, the numbers $\alpha_1, \alpha_2, \dots, \alpha_n$ will be the roots of this equation. Some of the numbers among $\alpha_1, \alpha_2, \dots, \alpha_n$ may happen to be equal to one another. If, for example, $\alpha_1 = \alpha_2 = \dots$ $\dots = \alpha_k$, but $\alpha_1 \neq \alpha_l$ for $l = k + 1,\ k + 2, \dots, n$, i. e. in the factorization in question the factor $x - \alpha_1$ appears k times, then for $k > 1$, the root α_1 is called a *multiple* one, or, more precisely, a k-fold root.

5. The Number of Real Roots

The basic theorem of higher algebra has important applications in theoretical research, but it provides no practical method for solving the roots of equations. However, many technical problems require information about the roots of equations with real coefficients. Usually a precise knowledge of these roots is not necessary, since the coefficients themselves are the results of measurements and thus are known only approximately, their accuracy depending on the accuracy of the measurements.

Let an nth degree equation be given

$$a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n = 0$$

having real coefficients. We already know that it has n roots. Are any of them real roots? If so, how many and approximately where are they located? We can answer these questions as follows. Let us denote the polynomial on the left-hand side of our equation by $f(x)$, i. e.

$$f(x) = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n$$

The reader familiar with the concept of *function* will understand that we treat the left-hand side of the equation as a function of the variable x. Taking for x an arbitrary numerical value α and substituting it into the expression for $f(x)$, after performing all the operations, we arrive at a certain number which is called the *value* of the polynomial $f(x)$ and is denoted as $f(\alpha)$. Thus, if

$$f(x) = x^3 - 5x^2 + 2x + 1$$

and $\alpha = 2$, then

$$f(2) = 2^3 - 5\cdot 2^2 + 2\cdot 2 + 1 = -7$$

Let us plot a *graph* of the polynomial $f(x)$. To do this we choose the coordinate axes on the plane (see above) and, having selected for x a value α and calculated a corresponding value $f(\alpha)$ of the polynomial $f(x)$, we mark a point on the plane with the abscissa α and the ordinate $f(\alpha)$, i. e. a point $(\alpha, f(\alpha))$. If it were possible to do this for all α, then the points marked on the plane would form a curve. The points at which this curve intersects the abscissa axis or is tangent to it yield the values of α for which $f(\alpha) = 0$, i. e. the real roots of the equation in question.

Unfortunately, since there are an infinite number of the values of α one cannot hope to find the points $(\alpha, f(\alpha))$ for all of them and must be satisfied with a finite number of points. For the sake of simplicity we can first select several positive and negative integral values of α in succession, mark on the plane the points corresponding to them and then draw through them as smooth a curve as possible. It can be shown that it is sufficient to take only the values of α which lie between $-B$ and B, where the bound B is defined as follows: if $|a_0|$ is the absolute value of the coefficient with x^n (we remind the reader that $|a| = a$ for $a > 0$

and $|a| = -a$ for $a < 0$) and A is the greatest of the absolute values of all the other coefficients $a_1, a_2, \ldots, a_{n-1}, a_n$, then

$$B = \frac{A}{|a_0|} + 1$$

However, it is often apparent that these bounds are too wide.

Example. Plot a graph of the polynomial

$$f(x) = x^3 - 5x^2 + 2x + 1$$

Here $|a_0| = 1$, $A = 5$, and thus $B = 6$. Actually, for this particular example we can restrict ourselves to only those values of α, which fall between -1 and 5. Let us compile a table of values of the polynomial $f(x)$ and plot a graph (Fig. 3).

α	$f(\alpha)$
-1	-7
0	1
1	-1
2	-7
3	-11
4	-7
5	11

The graph demonstrates that all the three roots α_1, α_2 and α_3 of our equation are real and that they are located within the following bounds:

$$-1 < \alpha_1 < 0, \quad 0 < \alpha_2 < 1, \quad 4 < \alpha_3 < 5$$

We notice that plotting the graph was not really necessary: its intersections with the abscissa axis are found between such neighbouring values of α for which the numbers $f(\alpha)$ have opposite signs, and thus it was sufficient just to look at the table of values of $f(\alpha)$.

If in our example we found less than three points of intersection of the graph with the abscissa axis, we might think that owing to the imperfection of our graph (we traced the curve knowing only seven of its points), we could overlook several additional roots of the equation. However, there are methods which make it possible to determine exactly the number of real roots of the equation and even the number of roots located between any given numbers a and b, where $a < b$. These methods will not be stated here.

Sometimes the following theorems are useful since they give some information on the existence of real and even positive roots.

Any equation of an odd degree with real coefficients has at least one real root.

If the leading coefficient a_0 and the absolute term a_n in an equation with real coefficients have opposite signs, the equation has at least one positive root. In addition, if our equation is of an even degree, it also has at least one negative root.

Thus, the equation

$$x^7 - 8x^3 + x - 2 = 0$$

has at least one positive root, while the equation

$$x^6 + 2x^5 - x^2 + 7x - 1 = 0$$

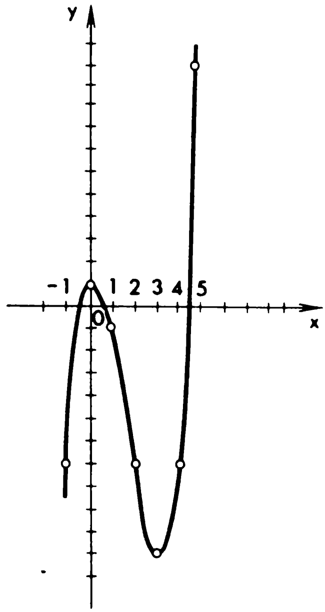

FIG. 3

has both a positive and a negative root. All this is readily verified by means of a graph.

6. Approximate Solution of Equations

In the previous section we found those neighbouring integers between which the real roots of the equation

$$x^3 - 5x^2 + 2x + 1 = 0$$

are located. The same method allows the roots of this equation to be found with greater accuracy. For example, let us take the root α_2, located between zero and unity. By calculating the values of the left-hand side of our equation $f(x)$ for $x = 0.1;\ 0.2, \ldots;$

0.9, we can find between which two of these successive values of x the graph of the polynomial $f(x)$ intersects the abscissa axis, i. e. we can now calculate the root α_2 to the accuracy of one-tenth.

Proceeding further, we can find the value of the root α_2 to the accuracy of one-hundredth, one-thousandth or, theoretically, to any accuracy we want. However, this approach involves cumbersome calculations which soon become practically unmanageable. This has led to the development of various methods of calculating approximate values of real roots of equations much quicker. Below we present the simplest of these methods and immediately apply it to the calculation of the root α_2 of the cubic equation considered above. But first it is useful to find bounds for this root narrower than the ones we already know, $0 < \alpha_2 < 1$. For this purpose we shall calculate our root to the accuracy of one-tenth. If the reader calculates the values of the polynomial

$$f(x) = x^3 - 5x^2 + 2x + 1$$

for $x = 0.1; 0.2; ...; 0.9$, he will obtain

$$f(0.7) = 0.293, \quad f(0.8) = -0.088$$

and therefore, since the signs of these values of $f(x)$ are different,

$$0.7 < \alpha_2 < 0.8$$

The method is as follows. An equation of degree n is given, whose left-hand side is denoted by $f(x)$; it is already known that one real root α of this equation (not a multiple root) lies between a and b, $a < b$. If the bounds $a < \alpha < b$ are already sufficiently close, then definite formulas make it possible to find for the root α the new bounds c and d, which are much closer, i. e. which define much more precisely the position of this root. The result will be either $c < \alpha < d$ or $d < \alpha < c$.

The bound c is calculated by means of the formula

$$c = \frac{bf(a) - af(b)}{f(a) - f(b)}$$

In this case $a = 0.7$, $b = 0.8$, and the values of $f(a)$ and $f(b)$ are given above. Therefore

$$c = \frac{0.8 \cdot 0.293 - 0.7 \cdot (-0.088)}{0.293 - (-0.088)} = \frac{0.2344 + 0.0616}{0.381} = 0.7769...$$

The formula for the bound d requires the introduction of a new concept which will play only an auxiliary role here; in essence it belongs to a different branch of mathematics called differential calculus.

Let a polynomial of degree n be given

$$f(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + ... + a_{n-2}x^2 + a_{n-1}x + a_n$$

Let us call the polynomial of degree $(n - 1)$

$$f'(x) = na_0x^{n-1} + (n-1)a_1x^{n-2} + (n-2)a_2x^{n-3} + ... \\ ... + 2a_{n-2}x + a_{n-1}$$

a *derivative* of this polynomial, and denote it as $f'(x)$. This polynomial is derived from $f(x)$ by the following rule: each term a_kx^{n-k} of the polynomial $f(x)$ is multiplied by the exponent $n - k$ of x, while the exponent itself is reduced by unity; moreover, the absolute term a_n disappears, since we can consider that $a_n = a_nx^0$.

We can again take the derivative of the polynomial $f'(x)$. This will be a polynomial of degree $(n - 2)$, which is called the *second derivative* of the polynomial $f(x)$ and is denoted as $f''(x)$.

Thus, for the above polynomial $f(x) = x^3 - 5x^2 + 2x + 1$ we obtain

$$f'(x) = 3x^2 - 10x + 2$$

$$f''(x) = 6x - 10$$

The bound d is now calculated by one of the following formulas:

$$d = a - \frac{f(a)}{f'(a)}, \quad d = b - \frac{f(b)}{f'(b)}$$

The following rule indicates which of these two formulas must be chosen. If the bounds a, b are chosen sufficiently close to each other, the second derivative $f''(x)$ will usually have the same sign for $x = a$ and $x = b$, while the signs of $f(a)$ and $f(b)$ will be different, as we know. If the signs of $f''(a)$ and $f(a)$ are the same, d must be calculated with the first formula, i. e. the one in which the bound a is used, however if the signs of $f''(b)$ and $f(b)$ coincide, the second formula, involving the bound b, must be used.

In the above example the second derivative $f''(x)$ is negative both for $a = 0.7$ and for $b = 0.8$. Therefore, since $f(a)$ is positive and $f(b)$ negative, the second formula for the bound d must be used. Since $f'(0.8) = -4.08$, we obtain

$$d = 0.8 - \frac{-0.088}{-4.080} = 0.8 - 0.0215... = 0.7784...$$

Thus for the root α_2 we have found the following bounds, narrower than those we knew before:

$$0.7769... < \alpha_2 < 0.7784...$$

or, if we widen these bounds somewhat,

$$0.7769 < \alpha_2 < 0.7785$$

It follows, therefore, that if we take for α_2 the arithmetic mean, i. e. half the sum, of the calculated bounds,

$$\alpha_2 = 0.7777$$

the error will not exceed 0.0008, equal to half the difference of these bounds.

If the resulting accuracy is insufficient, we could once again apply the above method to the new bounds of the root α_2. However, this would require much more complicated calculations.

Other methods of approximate solution of equations are more accurate. The best method, that permits the approximate calculation of not only the real but also the complex roots of equations, was devised by the great Russian mathematician N. I. Lobachevsky (1793-1856), the creator of non-Euclidean geometry.

7. Fields

The problem of roots of algebraic equations, which we have already encountered above, can be considered in more general terms. To do so we must introduce one of the most important concepts of algebra.

Let us first consider the following three systems of numbers: the set of all rational numbers, the set of all real numbers, and the set of all complex numbers. Without leaving their respective bounds, we can add, multiply, subtract and divide (except for division by zero) in each of these systems of numbers. This distinguishes them from the system of all integers where division is not always possible (for example, the number 2 cannot be divided by 5 without a remainder), as well as from the system of all positive real numbers, where subtraction is not always possible.

The reader is already familiar with performing algebraic operations not on numbers such as addition and multiplication of polynomials, and also addition of forces encountered in physics. Incidentally, in defining complex numbers we also had to consider addition and multiplication of points of the plane.

In general terms, let a set P be given, consisting either of numbers, of geometrical objects, or of some arbitrary objects which we shall call the *elements* of the set P. The operations of addition and multiplication are defined in P if for each pair of elements a, b from P one precisely defined element c from P is indicated, and called their *sum*:

$$c = a + b$$

and a precisely defined element d from P, called their *product*:

$$d = ab$$

The set P with the operations of addition and multiplication defined within it is called a *field*, if these operations possess the following five properties:

I. Both operations are commutative, i. e. for any a and b

$$a + b = b + a, \quad ab = ba$$

II. Both operations are associative, i. e. for any a, b and c

$$(a + b) + c = a + (b + c), \quad (ab)\,c = a\,(bc)$$

III. The law of distribution of multiplication with respect to addition holds, i. e. for any a, b and c

$$a\,(b + c) = ab + ac$$

IV. *Subtraction* can be carried out, i. e. for any a and b a unique root of the equation

$$a + x = b$$

can be found in P.

V. *Division* can be carried out, i. e. for any a and b, provided a does not equal zero, a unique root of the equation

$$ax = b$$

can be found in P.

Condition V mentions zero. Its existence can be derived from Conditions I-IV. Indeed, if a is an arbitrary element of P, then because of Condition IV a definite element exists in P which satisfies the equation

$$a + x = a$$

(a itself is taken for b). Since this element may depend upon the choice of the element a, we designate it by 0_a, i. e.

$$a + 0_a = a \tag{1}$$

If b is any other element of P, then again there exists one such unique element 0_b for which

$$b + 0_b = b \tag{2}$$

If we prove that $0_a = 0_b$ for any a and b, then the existence in the set P of an element, which plays the role of zero for all the elements a at the same time, will be immediately proved.

Let c be the root of the equation

$$a + x = b$$

which exists because of Condition IV; hence,

$$a + c = b$$

We now add to both sides of equation (1) the element c, which does not violate the equality due to the uniqueness of the sum:

$$(a + 0_a) + c = a + c$$

The right-hand side of this equation equals b, and the left-hand side, due to Conditions I and II, equals $b + 0_a$. Therefore,

$$b + 0_a = b$$

Comparing this to equation (2) and remembering that according to IV there exists only one solution of the equation $b + x = b$, we finally reach the equality

$$0_a = 0_b$$

This now proves that in any field P there is a *zero* element, i. e. such an element 0 that for *all* a in P the equality

$$a + 0 = a$$

holds and therefore Condition V becomes completely meaningful.

We already have three examples of fields – the field of rational numbers, that of real numbers, and that of complex numbers – while the sets of all integers and of positive real numbers do not constitute fields. Besides these three, an infinite number of other fields exist. For instance, many different fields are contained within the fields of real numbers and of complex numbers; these are the so-called *numerical fields*. In addition some fields are larger than that of complex numbers. The elements of these fields are no longer called numbers, but the fields formed by them are used in mathematical research. Here is one example of such a field.

Let us consider all possible polynomials

$$f(x) = a_0x^n + a_1x^{n-1} + ... + a_{n-1}x + a_n$$

with arbitrary complex coefficients and of arbitrary degrees; for instance, zero-degree polynomials will be represented by complex numbers themselves. Even if we add, subtract and multiply polynomials with complex coefficients by the rules we already know, we still will not obtain a field, since division of a polynomial by another polynomial with no remainder is not always possible.

Now let us consider ratios of polynomials

$$\frac{f(x)}{g(x)}$$

or *rational functions* with complex coefficients, and let us agree to treat them in the way we treat fractions. Thus,

$$\frac{f(x)}{g(x)} = \frac{\varphi(x)}{\psi(x)}$$

if and only if

$$f(x)\ \psi(x) = g(x)\ \varphi(x)$$

Then,

$$\frac{f(x)}{g(x)} \pm \frac{u(x)}{v(x)} = \frac{f(x)v(x) \pm g(x)u(x)}{g(x)v(x)}$$

$$\frac{f(x)}{g(x)} \cdot \frac{u(x)}{v(x)} = \frac{f(x)u(x)}{g(x)v(x)}$$

The role of zero is played by fractions whose numerator is equal to zero, i. e. fractions of the type

$$\frac{0}{g(x)}$$

Obviously, all fractions of this type are equal to one another. Finally, if a fraction $\frac{u(x)}{v(x)}$ does not equal zero, i. e. $u(x) \neq 0$, then

$$\frac{f(x)}{g(x)} : \frac{u(x)}{v(x)} = \frac{f(x)v(x)}{g(x)u(x)}$$

It can easily be checked that the above operations with rational functions satisfy all the requirements of the definition of a field, so that we can speak of a *field of rational functions* with complex coefficients. The field of complex numbers is totally contained in this field, since a rational function whose numerator and denominator are zero-degree polynomials is simply a complex number, and any complex number can be presented in this form.

One should not think that any field is either contained in the field of complex numbers or contains it within itself: some of the different fields consist only of a finite number of elements.

Whenever fields are used, we have to consider equations with coefficients from these fields, and inevitably the existence of roots of such equations poses a problem. Thus, in some problems of geometry we encountered equations with coefficients from the field of rational functions; the roots of these equations are called *algebraic functions*. As applied to equations with numerical coefficients, the basic theorem of higher algebra can no longer be used for equations with coefficients from an arbitrary field and is replaced by the following general theorems.

Let P be some field and let

$$a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n = 0$$

be an equation of degree n with coefficients from this field. It turns out that this equation cannot have more than n roots either in the field P or in any other greater field. At the same time the field P can be enlarged to a field Q in which our equation will have n roots (some of which may be multiple). Even the following theorem holds:

Any field P can be enlarged to such a field $\overline{P}$ that any equation with coefficients from P or even from $\overline{P}$ have roots in $\overline{P}$, and the number of roots is equal to the degree of the equation.

This field $\bar{P}$ is called *algebraically closed.* The basic theorem of higher algebra shows that the field of complex numbers belongs to the set of algebraically closed fields.

8. Conclusion

Throughout this booklet we always discussed equations of a certain degree with one variable. The study of first-degree equations is followed by that of quadratic equations in elementary algebra. In addition elementary algebra proceeds from a study of one first-degree equation with one variable to a system of two first-degree equations with two variables and a system of three equations with three variables. A university course in higher algebra continues these trends and teaches the methods for solving any system of n first-degree equations with n variables, and also the methods of solving such systems of first-degree equations in which the number of equations is not equal to the number of variables. The theory of systems of first-degree equations and some related theories including the *theory of matrices*, constitute one special branch of algebra, viz. *linear algebra*, which is widely used in geometry and other areas of mathematics, as well as in physics and theoretical mechanics.

It must be remembered, that at present both the theory of algebraic equations and linear algebra are to a large extent parts of science. Because of the requirements of adjacent branches of mathematics and physics, the study of sets, in which algebraic operations are defined, is most important. Aside from the *theory of fields*, which includes the theory of algebraic numbers and algebraic functions, the *theory of rings* is being currently developed. A *ring* is a set with operations of addition and multiplication, in which Conditions I-IV from the definition of a field are valid; the set of all integers may be cited as an example. We already mentioned another very significant branch of algebra, the *group theory.* A *group* is a set with one algebraic operation, multiplication, which must be associative; division must be carried out without restrictions.

We often encounter *noncommutative* algebraic operations, i. e. ones in which the product is changed by commutation of co-factors, and sometimes *nonassociative* operations, i. e. those in which the product of three factors depends on the location of brackets. Those groups which are used to solve equations in radicals are noncommutative.

Systematic presentation of the fundamentals of the theory of algebraic equations and of linear algebra can be found in textbooks on higher algebra. The following textbooks are most frequently recommended:

A. G. Kurosh, *Higher Algebra*, Mir Publishers, 1975 (in English).

L. Ya. Okunev, *Higher Algebra*, "Prosveshchenie", 1966 (in Russian).

An elementary presentation of the simplest properties of rings and fields, mostly numerical, can be found in:

I. V. Proskuryakov, *Numbers and Polynomials*, "Prosveshchenie", 1965 (in Russian).

An acquaintance with group theory may begin with:

P. S. Aleksandrov, *Introduction to Group Theory*, "Uchpedgiz", 1951 (in Russian).

TO THE READER

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is:

USSR, 129820, Moscow I-110, GSP
Pervy Rizhsky Pereulok, 2
Mir Publishers

Printed in the Union of Soviet Socialist Republics

Other Books by MIR PUBLISHERS
from LITTLE MATHEMATICS LIBRARY Series

G. E. Shilov

CALCULUS OF RATIONAL FUNCTIONS

An introduction to the principal concepts of mathematical analysis (the derivative and the integral) within the comparatively limited field of rational functions, employing the visual language of graphs. Meaty. Needs no more than "O" level mathematics on the part of the reader. Useful for sixth-form reading.

Contents. Graphs. Derivatives. Integrals. Solutions to Problems.

I. S. Sominsky

THE METHOD OF MATHEMATICAL INDUCTION

One of the popular lectures in mathematics, widely used in school extracurricular maths circles. Will interest sixth-formers doing "A" level mathematics, and first-year students whose courses include mathematics. Describes the method of mathematical induction so widely used in various fields of mathematics from elementary school courses to the most complicated investigations. Apart from being indispensible for study of mathematics, the ideas of induction have general educational value, and will interest a broad readership not familiar with mathematics.

Contents. Proofs of Identities; Problems of an Arithmetical Nature. Trigonometric and Algebraic Problems. Problems on the Proof of Inequalities. Proofs of Some Theorems in Elementary Algebra by the Method of Mathematical Induction.

Mir Publishers
Moscow