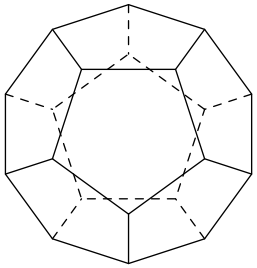
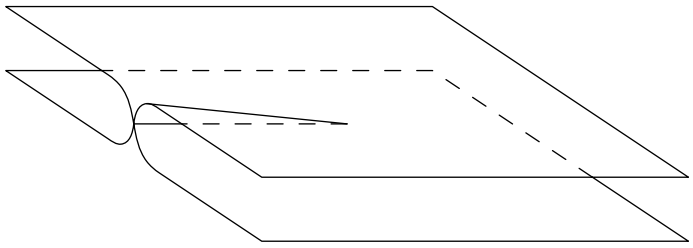
Abel's Theorem Through Problems

V. B. Alekseev





Abel's Theorem Through Problems

V. B. Alekseev

Translated by Yuliya Gorlina
and Igor Pavlovsky

Preface and Introduction translated
by Leonid Grinberg

Edited by Julian Gilbey

Contents

Translators' Foreword

This book by V. B. Alekseev was originally published in Russian in 2001. It follows a course given by Prof. V. I. Arnold in 1963 to a class of 14–16 year old students at a new high school in Moscow. A short article by Arnold about the purpose of this course can be found below.

The book was translated into English by Mathcamp staff for use at Mathcamp 2002. That summer, a self-selected group of about fifteen students worked on the problems over the course of the five-week camp, presenting their solutions to their classmates. Following the methodology of R. L. Moore, the students themselves critiqued their classmates' answers and arguments, and we only moved on when they were all happy with the presented work. The post-docs who were running the class (Mira Bernstein, David Savitt and Julian Gilbey) only stepped in to guide the pace of work and to explain points which had been missed by the whole group. On the very last day of classes, the group succeeded in solving problem number 352, thereby completing the proof of Abel's Theorem. The class developed a strong sense of the nature of mathematical argument, and developed a facility with both group theory and Riemann surfaces, as well as thoroughly enjoying the opportunity to fully engage with real mathematics.

The original book has full solutions included; we did not have the time to translate these. A translation of the whole book, including solutions and an extensive appendix on differential Galois theory, has been published by Kluwer under the title *Abel's Theorem in Problems and Solutions: Based on the Lectures of Professor V. I. Arnold.*

Julian Gilbey
London, December 2007

Foreword by Prof. V. I. Arnold

Abel's Theorem and Modern Mathematics

This foreword is an edited version of Prof. Arnold's article, which can be found at `http://www.institut.math.jussieu.fr/seminaires/singularites/abel.pdf`.

The myth of the difficulty of mathematics and of its inaccessibility to school children has caused harm similar to that of the physicists' typical opinion, originating with Landau, which regards mathematics as the continuation of the art of long multiplication.

In contrast to this opinion, the greatest mathematicians are distinguished by their inclination to replace blind calculations by clear ideas.[†] An interesting and sophisticated example of the use of clear ideas in mathematics is provided by the topological proof of Abel's Theorem on the impossibility of solving algebraic equations of degree five by radicals and rational functions.

It is very difficult to decide which parts of mathematics should and should not be taught at a standard high school. In 1963, Kolmogorov asked me to provide a one semester long mathematical course on any subject of my choice at the new high school he had established in Moscow. This was an attempt to introduce something new to replace the traditional Euclidean curriculum which was followed in all high schools at the time. I chose to teach Abel's Theorem, transforming its topological proof into a series of accessible problems for beginners, trying to provide them with several important ideas which usually remain outside the mathematical knowledge of both schoolchildren and university students (and even professors).

Abel's ideas presented in this way included, for instance, the geometrical theory of complex numbers and the topology of Riemann surfaces, including notions such as the fundamental group and ramified covering monodromies, normal subgroups and exact sequences, and also the symmetry groups of the regular polyhedra (including that of the dodecahedron and the Kepler cubes inscribed into it). All this "modern" science might more commonly be found in quantum physics lectures than in formal mathematics education; schoolchildren are more able to transform the difficult trigonometric expressions with which they have been tortured in their mathematics lessons than to think geometrically. My 14–16 year old students understood these lectures on Abel's ideas very well, solving many of the problems and exercises, and

[†]In the Russian translation of Bourbaki's "History of Mathematics", these words of Dirichlet appeared in the form "to replace clear ideas by blind calculations"; the Editor (Kolmogorov) explained to me that this (wrong) translation is a better description of Bourbaki's activity than the original words of Dirichlet.

were very successful in the final examination in which they were asked to think originally. I gave them far higher evaluations than typical university students, who are only trained to apply the standard formal rules.

This formal application of formal rules, comprising the majority of standard mathematics education, reminds me of Marat's comment that the best mathematician was Laplace, since he calculated the solution to every problem, substituting numbers in existing formulæ. Napoleon, however, dismissed Laplace from the ministerial post he had earlier granted him, accusing him of "introducing the spirit of infinitesimals into the state administration". (My French collaborators explained to me that Laplace tried to verify every cent in all of the accounts.)

Abel, unlike Laplace, had never been close to government. His proof of his theorem, presented by him to the Paris Academy of Sciences, had not been understood, for it contained too many new ideas, including geometrical and topological ones, as well as algebraic and combinatorial thinking. Moreover, his manuscript had been lost or hidden by Cauchy and was published only many years after Abel's death. According to an article in a French newspaper, this mathematician was so unfortunate that following the year in Paris "he returned by foot to his part of Siberia, called Norway, over the Atlantic Ocean ice." Abel's own description of French mathematics was astonishingly enduring and modern: "Everyone wishes to teach and no one wishes to study anything new, being expert in only his own particular domain (be it heat theory or elasticity theory, celestial mechanics or number theory), and having no interest in any more general questions."

Abel's approach to mathematics was very different. Early in his life, he had studied the geometric theory of complex numbers with Wessel, later extending it himself to the theory of Riemann surfaces and Abelian integrals (which are integrals along curves on Riemann surfaces). (In this context, the explicit integrability of the square root of quadratic polynomials depends on the topological fact that the Riemann surface of a quadratic is a sphere, yielding both the technique of "Euler' substitutions", which reduces these integrals to integrals of rational functions, and Pythagorean triples such as $3^2 + 4^2 = 5^2$ and $12^2 + 5^2 = 13^2$, whose relationship to the Riemann surface's topology deserves to be shown at high school.) The unity of all the mathematics appearing in these facts is similar to the discovery of the unity of the theories of electricity and of magnetism in physics.

In a similar manner, Abel also understood the reason for the impossibility of explicit integration of square roots of cubic polynomials (for instance, of the explicit calculation of the length of an ellipse) as deriving from the toroidal topology of the corresponding Riemann surfaces. (These integrals had already been studied by Newton in his attempt to explain the ellipticity of the planets' orbits from Kepler's law of areas swept out.)

The influence of Abel's ideas, containing, among other things, the main points of the the theory of Riemann surfaces (which were also developed simultaneously by Jacobi), including their relationship to integrals of algebraic functions (today

known as Abelian integrals), is still exceptionally high both in mathematics and in mathematical physics, which had both followed the approach proposed by Abel. It is a pity to recall that some mathematicians, more successful than Abel socially, proclaimed just the opposite opinion on his works. For example, Hardy wrote that "Abel, Riemann, and Poincaré have contributed nothing essential for mankind." My opinion is that the ideas of Abel, of Riemann and of Poincaré are the real basis of most of modern mathematics, on which are based most results of modern physics and technology, and all the successes of modern human civilisation, be they intellectual or material. This includes, for instance, the aerodynamics of aeroplanes and the space shuttles, celestial mechanics and wave propagation theory, ecology and quantum field theory. The quality and quantity of modern studies continuing these works of Abel, Riemann and Poincaré are enormous, and the ideas contained in their works influenced both the theoretical sciences and their useful applications far more than the technical details of ready-made formulæ.

In my 1964 high school course on the topological proof of Abel's Theorem, I had attributed to Abel the topological proofs of some results, including the following version of what I called "Topological Galois Theory". My proof of Abel's Theorem extends his result on the unsolvability of degree 5 equations in radicals to all complex functions which are topologically equivalent to the dependence of the roots of the degree 5 equations on their coefficients. I considered it as the *topological unsolvability* in radicals of equations of degree 5 (or higher). And I then attributed to Abel the topological unintegrability theorem for Abelian integrals, including elliptic ones: neither elliptic integrals nor the elliptic functions are topologically equivalent to any elementary function (even when combined with any univalent function).

In 1964, I proposed to my students that they should publish the proofs of these important results and it seems that forty years later they are finally attaining this goal.

The return of mathematics education (both in high schools and in university) from the sophisticated application of ready-made formulæ to real mathematical thinking and to simple ideas seems to me the only way to preserve the past century's mathematical culture and to generate new discoveries and new applications in the future.

Preface

During secondary school, algebraic equations in one variable of degree 1 (linear equations) and of degree 2 (quadratic equations) are thoroughly studied. It is discovered that for such equations, there exist general formulæ which express the roots of the equation using its coefficients, using only the arithmetic operations and radicals (square roots, cube roots and so forth). However, whether or not there exist similar formulæ for the solution of algebraic equations of higher degrees is known to far fewer people. It turns out that for equations of degree 3 or 4 such formulæ also exist. We will examine methods of solving such equations in the Introduction. But on examining a general algebraic equation in one variable with degree greater than 4, one discovers that it is not solvable in radicals; that is, there does not exist a formula which expresses the roots of such an equation in terms of its coefficients, using only the arithmetic operations and radicals. That is the result known as Abel's Theorem.

One of the goals of this book is to introduce the reader to the proof of Abel's theorem. We will not examine in detail the results obtained somewhat later by the French mathematician Évariste Galois, who analysed not general, but specific algebraic equations with fixed, numerical coefficients, and found for such equations a rule by which the roots of the equation could be found using arithmetic operations and radicals. For those who want to better acquaint themselves with Galois' results, the books of M. M. Postnikov (*Foundations of Galois Theory*) and B. L. van der Waerden (*Algebra*) can be recommended.[†]

From Galois' general results, one can deduce Abel's theorem. However in this book, we will follow a different route, which will allow the reader to acquaint himself with two very different branches of modern mathematics: Group Theory and the Theory of Functions of Complex Variables. The reader will learn what groups and fields are, and what properties they possess. From there, we move on to discovering what complex numbers are, with a brief discussion of their name. We will also find out what a Riemann surface is and use them to prove The Fundamental Theorem of Algebra.

The author will accompany and guide the reader on this path, but will allow him a wide opportunity to test out his own abilities. For this purpose, the reader will be presented with a large number of problems. These are, essentially, the main contents of the text of this book. The problems are consistently numbered throughout the book.

The book contains many terms which may be new to the reader. In order for the reader to more easily navigate through them, there is an index of terms with the page number on which they are defined.

[†]Ian Stewart's book *Galois Theory* is also highly recommended (ed.)

The book is based on lectures presented over a number of years by Professor Vladimir Igorovich Arnold and the author at the Faculty of Mechanics and Mathematics of the Moscow State University. The author wishes to express his gratitude to Prof. V. I. Arnold, who offered a large number of valuable comments during the preparation of this book.

Introduction

We begin this book by examining the question of how equations in one variable of degree from 1 to 4 are solved. The methods of solving algebraic equations of degree 1 and 2 were known even to mathematicians of the ancient world; methods of solving algebraic equations of degree 3 and 4 were only devised in the 16th century.

A *general algebraic equation in one variable of degree n* is an equation of the form

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$$

such that $a_n \neq 0$.[†]

When $n = 1$, we obtain the linear equation

$$a_1 x + a_0 = 0, \qquad a_1 \neq 0.$$

This equation obviously has the one solution

$$x = -\frac{a_0}{a_1}$$

for any values of the coefficients.

If $n = 2$, we get the quadratic equation

$$ax^2 + bx + c = 0, \qquad a \neq 0$$

(where instead of a_2, a_1, a_0, we write a, b and c, as is customary in school). Dividing both sides of this equation by a, and setting $p = b/a$, $q = c/a$, we get the quadratic equation

$$x^2 + px + q = 0. \tag{1}$$

We can manipulate this to obtain

$$x^2 + px + \frac{p^2}{4} = \frac{p^2}{4} - q$$

so that

$$\left(x + \frac{p}{2}\right)^2 = \frac{p^2}{4} - q. \tag{2}$$

In the course of middle school, only the case of $\frac{p^2}{4} - q \geqslant 0$ is examined further. If $\frac{p^2}{4} - q < 0$, then it is said that the right hand side of equation (2) cannot be square-rooted, and that equation (1) has no real roots. In order to avoid such exceptions, it will be more convenient for us to study algebraic equations not in the field of real numbers, but in the field of complex numbers.

We will examine complex numbers in detail (including their definition) in Chapter 2. For now, it is sufficient for the reader to know or take on faith the following statements about complex numbers:

[†]The coefficients a_n, a_{n-1}, …, a_0 can for now be considered to be arbitrary real numbers.

- The set of complex numbers is an extension of the set of real numbers; that is, real numbers are a subset of complex numbers in the same way that, for example, whole numbers are a subset of real numbers.
- Complex numbers can be added, subtracted, multiplied, divided and raised to a natural power, and these operations have the same properties as the corresponding operations with real numbers.
- If z is a non-zero complex number and n is a natural number, then there exist exactly n roots of the nth degree of z; that is, n complex numbers w such that $w^n = z$. If $z = 0$, then $\sqrt[n]{0} = 0$. If w_1 and w_2 are the square roots of z, then $w_2 = -w_1$.

Below, we will be interested not only in real and complex roots of equations, but also in the nature of the coefficients of such equations. At the same time, the above discussion on linear and quadratic equations remains valid, because of the second property of complex numbers just described.

Let us continue the examination of the quadratic equation. In the field of complex numbers, equation (2), for any values of p and q, is equivalent to the equation

$$x + \frac{p}{2} = \pm\sqrt{\frac{p^2}{4} - q},$$

where $\sqrt{p^2/4 - q}$ is understood to mean some particular value of the square root.

In this way the roots are

$$x_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}. \tag{3}$$

Introducing a, b and c, we obtain the well-known

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{4}$$

For the sequel, we will need two facts relating to quadratic equations:

- Viète's Theorem[†]: Complex numbers x_1 and x_2 are roots of the equation $x^2+px+q = 0$ if and only if $x_1+x_2 = -p$ and $x_1x_2 = q$. Indeed, if x_1 and x_2 are roots of the equation $x^2+px+q = 0$, then equation (3) is satisfied. From here, $x_1 + x_2 = -p$ and $x_1x_2 = q$. Conversely, if $x_1 + x_2 = -p$ and $x_1x_2 = q$, then substituting for p and q in the equation $x^2 + px + q = 0$, we obtain $x^2 - (x_1 + x_2)x + x_1x_2 = (x - x_1)(x - x_2) = 0$, and therefore x_1 and x_2 are the roots of the equation $x^2 + px + q = 0$.
- The quadratic equation $ax^2 + bx + c = 0$ is a perfect square (that is, $ax^2 + bx+c = \left(\sqrt{a}(x-x_0)\right)^2$ for some complex number x_0) if and only if the roots of the equation $ax^2 + bx + c = 0$ are both equal to x_0. This happens if and only if (see equation (4)) $b^2 - 4ac = 0$. The expression $b^2 - 4ac$ is known as the *discriminant* of the quadratic equation.

[†]François Viète (1540–1603): French mathematician

Let us now consider the monic cubic equation

$$x^3 + ax^2 + bx + c = 0. \tag{5}$$

(The general cubic equation can be converted to this form on division by a_3.) Substituting $x = y + d$, where d will be assigned a value later, we obtain

$$(y + d)^3 + a(y + d)^2 + b(y + d) + c = 0.$$

Multiplying everything out, we get the equation,

$$y^3 + (3d + a)y^2 + (3d^2 + 2ad + b)y + (d^3 + ad^2 + bd + c) = 0.$$

The coefficient of y^2 in this equation is equal to $3d + a$. Therefore, if we set $d = -\frac{a}{3}$, the substitution $x = y - \frac{a}{3}$ brings the equation into the form

$$y^3 + py + q = 0 \tag{6}$$

where p and q are polynomials in a, b and c.

Let y_0 be a root of equation (6). Writing this root as $y_0 = \alpha + \beta$ (where the values of α and β are for now unknown) and expanding y_0^3, we obtain

$$\alpha^3 + 3\alpha\beta(\alpha + \beta) + \beta^3 + p(\alpha + \beta) + q = 0,$$

which gives

$$\alpha^3 + \beta^3 + (\alpha + \beta)(3\alpha\beta + p) + q = 0 \tag{7}$$

Let us see if we can add another restriction to α and β:

$$\alpha\beta = -\frac{p}{3}$$

In this case, we get two simultaneous equations for α and β:

$$\alpha + \beta = y_0$$
$$\alpha\beta = -\frac{p}{3}$$

By Viète's theorem, for any y_0, such α and β do actually exist (possibly as complex numbers) and are the roots of the quadratic equation

$$w^2 - y_0 w - \frac{p}{3} = 0.$$

If we take such α and β (for now, with still-unknown values), then equation (7) takes the form

$$\alpha^3 + \beta^3 + q = 0. \tag{8}$$

Raising $\alpha\beta$ to the third power and combining the obtained equation with equation (8), we obtain

$$\alpha^3 + \beta^3 = -q$$
$$\alpha^3\beta^3 = -\frac{p^3}{27}$$

whence by Viète's theorem, α^3 and β^3 are the roots of the quadratic equation

$$w^2 + qw - \frac{p^3}{27} = 0.$$

In this way,

$$\alpha^3 = -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}} \qquad \text{and} \qquad \beta^3 = -\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}$$

where again, $\sqrt{q^2/4 + p^3/27}$ is understood to mean a particular value of the square root. From here, the roots of equation (6) are expressed by the formula

$$y_{1,2,3} = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

such that for each of the three values of the first cube root[†], it is necessary to take the corresponding value of the second root, such that the rule $\alpha\beta = -\frac{p}{3}$ holds.

This formula is called *Cardano's Formula*[‡]. Replacing p and q by their values in terms of a, b and c, and subtracting $\frac{a}{3}$, we obtain the formula for the roots of equation (5). Then replacing $a = \frac{a_2}{a_3}$, $b = \frac{a_1}{a_3}$ and $c = \frac{a_0}{a_3}$, we obtain the formula for the roots of the general cubic equation.

Let us now examine the monic quartic equation

$$x^4 + ax^3 + bx^2 + cx + d = 0. \tag{9}$$

(Again, the general equation reduces to the monic one upon dividing by a_4.) After performing the substitution $x = y - \frac{a}{4}$, as in the cubic case, we reduce equation (9) to the form

$$y^4 + py^2 + qy + r = 0, \tag{10}$$

where p, q and r are polynomials in a, b, c and d.

We will solve equation (10) via a method known as Ferrari's Method[§]. Let us rewrite equation (10) in the following way. Since

$$\left(y^2 + \frac{p}{2}\right)^2 + qy + \left(r - \frac{p^2}{4}\right) = 0$$

it follows that

$$\left(y^2 + \frac{p}{2} + \alpha\right)^2 - \left(2\alpha\left(y^2 + \frac{p}{2}\right) + \alpha^2 - qy + \frac{p^2}{4} - r\right) = 0 \tag{11}$$

where α is an arbitrary number. Let us now attempt to pick α in such a way that the quadratic in y in the second set of brackets

$$2\alpha y^2 - qy + \left(\alpha p + \alpha^2 + \frac{p^2}{4} - r\right)$$

[†]see the third of the above-mentioned properties of complex numbers

[‡]G. Cardano (1501–1576), an Italian mathematician

[§]L. Ferrari (1522–1565), an Italian mathematician and student of Cardano

is actually a perfect square. As was mentioned earlier, in order for it to be a perfect square, it is necessary and sufficient for the discriminant to be equal to 0, that is

$$q^2 - 8\alpha\left(\alpha p + \alpha^2 + \frac{p^2}{4} - r\right) = 0. \tag{12}$$

Expanding the brackets in this equation yields a cubic equation in α, which we now know how to solve. If we then substitute α with one of the roots of equation (12) back into (11), then the expression in brackets in (11) will be a perfect square. In that case, the left side of equation (11) will be the difference of two squares, which can thus be expressed as the product of two quadratic polynomials in y. Solving these and using $x = y - \frac{a}{4}$ gives us the solutions of the original quartic equation.

In this way, a quartic (degree 4 polynomial) can always be solved, and, furthermore, as in the case of cubics, it is possible to obtain a formula expressing the roots of the general quartic equation in terms of the coefficients of the equation using only the operations of addition, subtraction, multiplication, division, raising to a natural power and extracting natural roots.

For a long time, mathematicians attempted to find a method of solving a general quintic (degree 5) equation using only these same operations. However, in 1824, the Norwegian mathematician Niels Henrik Abel (1802–1829) proved the following theorem:

Abel's Theorem. *The general algebraic equation in one variable of degree greater than four is unsolvable in radicals. That is, there does not exist a formula expressing the roots of a general equation of degree greater than four using only the coefficients and the operations of addition, subtraction, multiplication, division, raising to a natural power and extracting natural roots.*

We will be able to prove this theorem by the end of the book. However, in order to achieve this, we will need to develop the mathematical concepts of groups, solvable groups, functions of a complex variable, Riemann surfaces and so on. We will acquaint the readers with these ideas in the remaining pages of this book. We begin with the very important mathematical concept of a group.

CHAPTER 1

Groups

Investigation of polynomial equations in the beginning of the nineteenth century led mathematicians to a new concept—the concept of a group. This new concept turned out to be so fruitful that not only did it enter and influence almost all areas of mathematics, but also started playing an important role in other fields of science, for example quantum mechanics and crystallography. Investigations connected to the concept of a group grew to form a new branch of modern mathematics—Group Theory. What is a "group" in mathematics? To answer that question, let us begin with some examples.

§1. Examples

In arithmetic, we use operations where two numbers are put into correspondence with a third number. For example, the operation of addition assigns the number 8 to the pair $(3, 5)$ and the number 4 to the pair $(2, 2)$. The operation of subtraction, when viewed as acting on all integers, also places pairs of integers into correspondence with single integers. In this situation, the order of the numbers is also important. For example, the pair $(5, 3)$ corresponds to 2 under subtraction, while the pair $(3, 5)$ corresponds to -2. Thus the pairs $(5, 3)$ and $(3, 5)$ must be viewed as different.

Pairs for which the order of elements is relevant will be called *ordered pairs*.

Definition 1. Let M be a set. A *binary operation* on M is an assignment of an element of M to each ordered pair of elements of M.

Examples of binary operations include addition on the set of natural numbers or integers, and subtraction on the set of integers. Subtraction is not a binary operation on the set of natural numbers because, for example, the ordered pair $(3, 5)$ does not correspond to a natural number.

1. Consider the operations: (a) addition of numbers; (b) subtraction of numbers, and (c) multiplication of numbers. For which of the following subsets of the set of integers are they binary operations? (1) The set of all even integers; (2) the set of all odd integers; (3) the set of all negative integers, and (4) the set of all positive integers.[†]

[†]Some of the suggested problems are of practical nature and serve to improve one's understanding of the new concepts using examples. Other problems are theoretical and their results are used later. Therefore, if the reader is not able to solve a problem, he should familiarise himself with its solution when it is discussed in class.

Let us look at several more examples of binary operations. We will frequently return to these examples.

Example 1. Let A, B and C be the vertices of an equilateral triangle ABC (fig. 1). Let's rotate the triangle 120° anticlockwise around its centre O. Then the vertex A

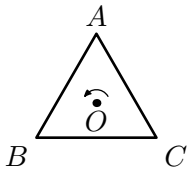


FIGURE 1.

will go to B, B to C and C to A. This way, the triangle will be superimposed on itself (if we ignore the names of the vertices), i.e., the 120° rotation around point O is a transformation that takes this triangle to itself. We will denote this transformation by a. It can be written as $a = \left(\begin{smallmatrix} A & B & C \\ B & C & A \end{smallmatrix}\right)$, where the top row contains the list of all vertices and the bottom row shows where each of them goes. A 240° rotation in the same direction around point O is also a transformation which sends the triangle to itself. We denote this transformation by b, so $b = \left(\begin{smallmatrix} A & B & C \\ C & A & B \end{smallmatrix}\right)$. There exists another rotation which sends the triangle to itself and which is different from a and b—that is the 0° rotation. Let us call that transformation e, with $e = \left(\begin{smallmatrix} A & B & C \\ A & B & C \end{smallmatrix}\right)$. It is easy to see that there are only three different rotations in the plane[†] which send the equilateral triangle ABC to itself: e, a and b.

Let g_1 and g_2 be arbitrary transformations of the triangle. Then by $g_1 \cdot g_2$ (or simply $g_1 g_2$) we mean the transformation g_3 which is obtained by first applying transformation g_2 and then g_1; we will call g_3 the *composition* of transformations g_2 and g_1.

We can make a *multiplication table* (table 1), where each row and each column corresponds to a rotation which sends the triangle ABC to itself. At the intersection

	e	a	b
e			
a			e
b			

TABLE 1.

of the row corresponding to g_1 and the column corresponding to g_2 we will put the transformation equal to $g_1 \cdot g_2$. For example, in the highlighted cell of table 1, we

[†]We mean that these are rotations which do not move into three dimensions, i.e., rotations around only those axes which are perpendicular to the plane.

put the transformation $a \cdot b$ which we get if we turn the triangle 240° and then 120°. Therefore $a \cdot b$ is the 360° turn, so it is the same as e. We can get the same result by reasoning this way: transformation b sends A to C and transformation a sends C to A. This way, the transformation $a \cdot b$ sends vertex A to A. In exactly the same way, we can see that vertex B goes to B and C goes to C. So $ab = \left(\begin{smallmatrix} A & B & C \\ A & B & C \end{smallmatrix}\right)$, i.e., $ab = e$.

2. Complete table 1.

Any transformation of a figure to itself, preserving the distances between all of its points, is called a *symmetry* of that figure. In example 1, the rotations of the triangle are its symmetries.

Example 2. In addition to rotations, an equilateral triangle has three other symmetries: reflections about the axes l_1, l_2 and l_3 (fig. 2).

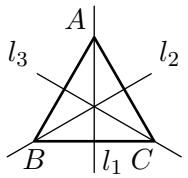


FIGURE 2.

We will call these transformations respectively c, d and f, so that $c = \left(\begin{smallmatrix} A & B & C \\ A & C & B \end{smallmatrix}\right)$, $d = \left(\begin{smallmatrix} A & B & C \\ C & B & A \end{smallmatrix}\right)$ and $f = \left(\begin{smallmatrix} A & B & C \\ B & A & C \end{smallmatrix}\right)$. In this situation, there are two different ways to view the composition of two transformations. Let us look, for example, at the composition of transformations $c \cdot d$. We could assume that when we apply d, the axis l_1 moves to a new position (the position of the old axis l_3), and view transformation c as a reflection with respect to the new position of l_1 (i.e., with respect to the old l_3). Alternatively, we could assume that the axes are not connected rigidly to the figure, and do not transform with it, and therefore after applying d, we should apply c as a reflection about the old axis l_1. The latter is the way we will view compositions of transformations. When using this approach, reasoning analogous to that discussed immediately before problem 2 turns out to work, and we discover that $cd = a$. This type of reasoning is useful for calculating compositions of transformations.

3. Construct the multiplication table for all symmetries of an equilateral triangle.

Example 3. Let e, a, b and c represent rotations of a square by 0°, 180°, 90° and 270° anticlockwise, respectively (fig. 3).

4. Construct the multiplication table for rotations of the square.

Example 4. Let d, f, g and h represent reflections of the square about the axes labelled in fig. 4.

5. Construct the multiplication table of all symmetries of the square.

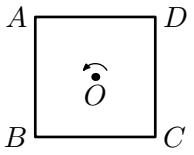


FIGURE 3.

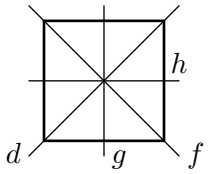


FIGURE 4.

Example 5. Let $ABCD$ be a rhombus that is not a square.

6. Find all symmetries of the given rhombus and construct their multiplication table.

Example 6. Let $ABCD$ be a rectangle that is not a square.

7. Find all symmetries of the given rectangle and construct their multiplication table.

§2. Groups of transformations

Let X and Y be two sets of elements and let every element x in X correspond to precisely one element y in Y. Then we say that we have a *map* φ from the set X to the set Y (written $\varphi : X \to Y$).[†] The element y is called the *image* of element x, while x is called a *preimage* of the element y, and we write $\varphi(x) = y$. Note that each element x in X has a unique image, but an element y in Y may have any number of preimages.

Definition 2. The map $\varphi : X \to Y$ is called a *surjective* map if for every element y in Y there exists an element x in X such that $\varphi(x) = y$, i.e., if every y in Y has a preimage in X.

8. Let φ be the map from towns in the UK to the letters of the alphabet, which takes every town to the first letter of its name. What would it mean for this map to be surjective?

Definition 3. The map $\varphi : X \to Y$ is called a *bijective* map if for every element y in Y, there exists a *unique* preimage in X.

[†]A map is also known as a *function*.

9. Consider the following maps from the set of all integers to the set of all non-negative integers:[†]

(a) $\varphi(n) = n^2$,
(b) $\varphi(n) = |n|$,
(c) $\varphi(n) = \begin{cases} 2n & \text{if } n \geqslant 0 \\ 2|n| - 1 & \text{if } n < 0. \end{cases}$

Which of these maps are surjective and which of them are bijective?

Let M be a set. A bijective map from the set M to itself, $g : M \to M$, is called a *transformation* of the set M.

Two transformations g_1 and g_2 will be considered equal if $g_1(A) = g_2(A)$ for every element A in M. Instead of the term transformation, the term *permutation* is often used. In this book, we will only use this term when dealing with finite sets. In such a case, the permutation can be written as

$$\begin{pmatrix} A_1 & A_2 & \dots & A_n \\ A_{i_1} & A_{i_2} & \dots & A_{i_n} \end{pmatrix},$$

where the first row lists all elements of the set and the second row shows where the permutation takes each of these elements.

Because a transformation is a bijective map, for every transformation g there exists an *inverse transformation* g^{-1}, which is defined as follows: if $g(A) = B$, then $g^{-1}(B) = A$. Thus in example 1, where $a = \left(\begin{smallmatrix} A & B & C \\ B & C & A \end{smallmatrix}\right)$, it follows that $a^{-1} = \left(\begin{smallmatrix} A & B & C \\ C & A & B \end{smallmatrix}\right)$, i.e., $a^{-1} = b$.

10. Find the inverse transformations of all symmetries of an equilateral triangle (examples 1 and 2).

11. Let g be the transformation of the set of real numbers defined by $g(x) = 2x$. Find the inverse transformation g^{-1}.

The *composition* of the transformations g_1 and g_2, written as $g_1 g_2$, is defined by $(g_1 g_2)(A) = g_1(g_2(A))$: first apply transformation g_2, then g_1. If g_1 and g_2 are transformations of a set M, then $g_1 g_2$ is also a transformation of the set M.

Definition 4. Let G be a non-empty set of transformations of a set M having the following properties: (1) if transformations g_1 and g_2 are in G, then their composition $g_3 = g_1 g_2$ is in G; (2) if transformation g is in G, then the inverse transformation g^{-1} is in G. Such a set of transformations G will be called a *group of transformations*.

It is easy to check that the sets of transformations discussed in examples 1–6 are groups of transformations.

12. Prove that every group of transformations contains an identity transformation e, such that $e(A) = A$ for every element A of the set M.

[†]For any number x, the number $|x|$ is the *absolute value* of x, meaning x with any minus sign stripped off. For example, $|-3| = |3| = 3$.

13. Prove that $eg = ge = g$ for every transformation g.

14. Prove that for any three transformations g_1, g_2 and g_3, the following equality holds:[†]

$$g_1(g_2g_3) = (g_1g_2)g_3.$$

§3. Groups

When solving problems 6 and 7, we constructed multiplication tables of the symmetries of a rhombus and of a rectangle. Check that it is possible to label these symmetries in such a way that these tables are the same. For many purposes, it is natural to consider such groups of transformations as being "the same".

Because of this, we will turn our attention away from the nature of the elements in the set (in this case, transformations) and the nature of the binary operation[‡] (in this case, composition of transformations). Instead, we will view a group simply as an arbitrary set with a binary operation, as long as the main properties of a group of transformations hold, as we now explain. We will usually call the binary operation *multiplication*, and if a pair (a, b) corresponds to c, then we will call c the *product* of a and b, and we write $a \cdot b = c$ or simply $ab = c$. In some cases, this operation will be given a different name, such as composition, addition, etc.

Definition 5. A *group* is a set G of elements of arbitrary nature, together with a binary operation $a \cdot b$ defined on G, such that the following conditions hold:

(1) the operation is *associative*, that is, $(ab)c = a(bc)$ for every a, b and c in G;
(2) there exists an element e in G, such that $ea = ae = a$ for every element a in G; this element e is called the *identity element* of the group G;
(3) for each element a in G, there exists an element a^{-1} in G such that $aa^{-1} = a^{-1}a = e$; this element is called the *inverse* of a.

From the results of problems 12–14, we see that any group of transformations is a group (in some sense, the converse is also true—see problem 55).[§]Because of this, we already have some examples of groups. All of these groups have a finite number of elements, and such groups are called *finite* groups. The number of elements in a finite group is called the *order* of the group. Groups that contain an infinite number of elements are called *infinite* groups.

Let us look now at some examples of infinite groups.

Example 7. Consider the set of all integers. We will use addition as our binary operation, which gives us a group. The identity element here is 0, because $0 + n =$

[†]This equality holds not only for transformations, but for any three maps g_1, g_2, g_3 such that $g_3 : M_1 \to M_2$, $g_2 : M_2 \to M_3$, $g_1 : M_3 \to M_4$.

[‡]For the definition of a binary operation, see page 1.

[§]Actually, we also need to show that $gg^{-1} = g^{-1}g = e$ for every g in the group of transformations. This is straightforward, though.

$n + 0 = n$ for every integer n. Also, for every n, there exists an inverse $-n$ (which is often called the additive inverse or negative when dealing with addition), because $n + (-n) = (-n) + n = 0$. Associativity in this case follows from the laws of arithmetic. The group we get this way is called *the group of integers under addition*.

15. Do the following sets form groups under multiplication? (a) All real numbers, and (b) all non-zero real numbers?

16. Do the positive real numbers form a group under multiplication?

17. Do the natural numbers (0, 1, 2, ...) form a group: (a) under addition, (b) under multiplication?

18. Prove that in any group, there is a unique identity element.

19. Prove that for any element a in a group, there is a unique inverse of a.

20. Prove that (a) $e^{-1} = e$; (b) for any element a in a group, $(a^{-1})^{-1} = a$.

If a and b are elements of some group, then, by the definition of a binary operation, $a \cdot b$ is an element of the group. Therefore, expressions of the form $(a \cdot b) \cdot c$, $a \cdot (b \cdot c)$, $(a \cdot b) \cdot (c \cdot d)$ and so on also give elements of the group. Any two elements defined through multiplication can also be multiplied to get yet another element of the group, and so on. To tell which operation was performed last, every time we multiply two elements, we will place parentheses around them (expressions consisting of only one element do not need parentheses, and neither do we need parentheses around the final product). Those expressions that can be formed with this method will be called *well-formed products*. For example, $(a \cdot b) \cdot (c \cdot (a \cdot c))$ is a well formed product, while $(a \cdot b) \cdot c \cdot (a \cdot c)$ is not, because the order in which the operations are performed is not clear. When we look at products $a_1 \cdot a_2 \cdot \ldots \cdot a_n$ of several real numbers a_1, a_2, ..., a_n, we do not use parentheses at all, because it turns out that the result does not depend on the order in which the operations are performed, i.e., with any placement of parentheses that gives a well formed product, the result will be the same. It turns out that this property is true for any group, which follows from the result of the following problem.

21. Let a binary operation $a \cdot b$ be associative, i.e., $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for any three elements a, b, c. Prove that every well formed product that consists from left to right of the elements a_1, a_2, ..., a_n gives the same element as the product

$$(\ldots((a_1 \cdot a_2) \cdot a_3) \cdot \ldots \cdot a_{n-1}) \cdot a_n.$$

Thus, if a_1, a_2, ..., a_n are elements of a group, any well-formed product made using these elements in this order will give the same element regardless of the placement of the parentheses. We will denote this product by $a_1 \cdot a_2 \cdot \ldots \cdot a_n$ or just $a_1 a_2 \ldots a_n$ (without parentheses).

When multiplying real numbers, we have another very important property, namely that the product $a_1 \cdot a_2 \cdot \ldots \cdot a_n$ does not change when we change the order of the terms. But in an arbitrary group, this property may not hold.

Definition 6. Two elements a and b of a group are said to *commute* if $ab = ba$. If all elements of a group commute with each other, then the group is called *commutative* or *abelian*.[†]

There exist non-abelian groups. One example of a non-abelian group is the group of symmetries of a triangle (see example 2, where $ac = f$ but $ca = d$, and so $ac \neq ca$).

22. Determine whether the following groups are commutative (see problems 2 and 4–7): (a) the group of rotations of a triangle; (b) the group of rotations of a square; (c) the group of symmetries of a square; (d) the group of symmetries of a rhombus, and (e) the group of symmetries of a rectangle.

23. Prove that in any group: (a) $(ab)^{-1} = b^{-1}a^{-1}$, (b) $(a_1 \cdot \ldots \cdot a_n)^{-1} = a_n^{-1} \cdot \ldots \cdot a_1^{-1}$. (An aide-mémoire: A jacket is put on after a shirt and is taken off before.)

If we have an equality $a = b$ in an arbitrary group G (which means that the left hand side and the right hand side denote the same element), then from it we can deduce new equalities by multiplying both sides of the original equality by some element c of the group G. However, because the product may depend on the order of elements, we must either multiply both sides of the equality on the right to get $ac = bc$, or multiply both sides on the left to give $ca = cb$.

24. Let a and b be any elements of a group G. Prove that the equations $ax = b$ and $ya = b$ each have a unique solution in G.

The uniqueness of problem 24 can be expressed the following way: if $ab_1 = ab_2$ or $b_1a = b_2a$, then $b_1 = b_2$.

25. Let $a \cdot a = e$ for every element a in a group G. Prove that the group G is commutative.

By a^m, where m is an arbitrary positive integer and a an arbitrary element of a group G, we will mean the product $a \cdot a \cdot \ldots \cdot a$, where the number of terms is m.

26. Prove that $(a^{-1})^m = (a^m)^{-1}$, where m is a positive integer.

Thus $(a^{-1})^m$ and $(a^m)^{-1}$, where m is a positive integer, are the same element, which we will denote by a^{-m}. We also set $a^0 = e$ for every element a of the group.

27. Prove that $a^m \cdot a^n = a^{m+n}$ for any two integers m and n.

28. Prove that $(a^m)^n = a^{mn}$ for any two integers m and n.

§4. Cyclic groups

The simplest (yet very important) groups are the cyclic groups, which we will now study.

[†]Named after the Norwegian mathematician Niels Henrik Abel (1802–1829).

Definition 7. Let a be an element of a group G. The smallest positive integer n such that $a^n = e$ is called the *order of the element* a. If such an n does not exist, then we say that a is an *element of infinite order*.

29. Find the orders of all elements in the groups of symmetries of an equilateral triangle, square and rhombus (see problems 3, 5 and 6).

30. Suppose an element a has order n. Prove that: (a) the elements e, a, a^2, $\ldots$, a^{n-1} are all distinct, and (b) for any integer m, the element a^m equals one of the elements listed above.

Definition 8. If an element a has order n and all the elements of a group G are equal to one of e, a, a^2, $\ldots$, a^{n-1}, then the group G is called a *cyclic group of order* n, *generated by* a, and the element a is called a *generator* of the group.

Example 8. Suppose we have a regular n-gon in the plane. Let us consider all rotations of the plane that send the regular n-gon to itself.

31. Prove that these rotations form a cyclic group of order n.

32. Find all the generators of the group of rotations of a triangle and of a square (examples 1 and 3).

33. Suppose the element a has order n. Prove that $a^m = e$ if and only if $m = nd$, where d is an arbitrary integer.

34. Suppose that a has prime order p and let m be an arbitrary integer. Prove that either $a^m = e$ or a^m has order p.

35. Let the greatest common divisor (highest common factor) of two positive integers m and n be d, and suppose that a has order n. Prove that a^m has order n/d.

36. Find all generators of the group of rotations of a regular 12-gon.

37. Let a be an element of infinite order. Prove that the elements $\ldots$, a^{-2}, a^{-1}, $a^0 = e$, a, a^2, $\ldots$ are all distinct.

Definition 9. If a is an element of infinite order in the group G, and all of the elements of G are equal to one of $\ldots$, a^{-2}, a^{-1}, e, a, a^2, $\ldots$, then G is called an *infinite cyclic group, generated by* a.

38. Prove that the group of integers under addition (example 7) is an infinite cyclic group. Find its generators.

Example 9. Let n be a positive integer greater than 1. Let us consider all possible remainders that we can get when dividing integers by n, i.e., the numbers 0, 1, 2, $\ldots$, $n-1$. We define a binary operation on the set of these remainders as follows. We add remainders as usual, then find the remainder of the sum when divided by n, and define the result to be that remainder. We call this operation *addition modulo* n. For example, modulo 4, we get $1 + 2 = 3$ and $3 + 3 = 2$.

39. Make the tables of addition modulo: (a) 2, (b) 3, (c) 4.

40. Prove that the remainders with the operation of addition modulo n form a cyclic group of order n.

Let us look again at an arbitrary cyclic group of order n with elements e, a, a^2, $\ldots$, a^{n-1}.

41. Prove that $a^m \cdot a^r = a^k$, where $0 \leqslant m < n$, $0 \leqslant r < n$ and $0 \leqslant k < n$, if and only if the equality $m + r = k$ holds modulo n.

It follows from the result of the last problem that, in some sense, the multiplication of elements in an arbitrary cyclic group of order n corresponds to the addition of remainders modulo n. In the same way, multiplication of elements in an infinite cyclic group corresponds to the addition of integers (see problem 27). Here we come to an important concept in group theory, that of an isomorphism.

§5. Isomorphisms

Definition 10. Suppose we are given two groups G_1 and G_2, and suppose there exists a bijective map φ from the elements of the group G_1 to the elements of the group G_2 (see section 2), such that multiplication in G_1 corresponds to multiplication in G_2, i.e., if $\varphi(a) = a'$, $\varphi(b) = b'$, $\varphi(c) = c'$ and $ab = c$ in the group G_1, then in the group G_2, $a'b' = c'$. Then φ is called an *isomorphism* from the group G_1 to the group G_2, and groups between which we can define an isomorphism are called *isomorphic*. The condition that the bijective map φ is an isomorphism can be written the following way: $\varphi(ab) = \varphi(a) \cdot \varphi(b)$ for any elements a and b of the group G_1; here the multiplication ab is performed in G_1 and the multiplication $\varphi(a) \cdot \varphi(b)$ is performed in G_2.

42. Which of the following groups are isomorphic to each other? (1) The group of rotations of a square; (2) the group of symmetries of a rhombus; (3) the group of symmetries of a rectangle; (4) the group of remainders under addition modulo 4.

43. Let $\varphi : G_1 \to G_2$ be an isomorphism. Prove that the inverse map $\varphi^{-1} : G_2 \to G_1$ is also an isomorphism.

44. Let $\varphi_1 : G_1 \to G_2$ and $\varphi_2 : G_2 \to G_3$ be isomorphisms. Prove that $\varphi_2\varphi_1 : G_1 \to G_3$ is also an isomorphism.

It follows from the last two problems that if two groups are both isomorphic to a third, then they are isomorphic to each other. (Why?)

45. Prove that any cyclic group of order n is isomorphic to the group of remainders under addition modulo n.

46. Prove that any infinite cyclic group is isomorphic to the group of integers under addition.

47. Let $\varphi : G \to F$ be an isomorphism. Prove that $\varphi(e_G) = e_F$, where e_G and e_F are the identity elements of G and F, respectively.

48. Let $\varphi : G \to F$ be an isomorphism. Prove that $\varphi(g^{-1}) = (\varphi(g))^{-1}$ for every element g of the group G.

49. Let $\varphi : G \to F$ be an isomorphism and $\varphi(g) = h$. Prove that the orders of g and h are the same.

When we are studying the group operation itself, and the nature of the elements of the group does not play any role, then isomorphic groups may be considered to be the same. For example, we say that there exists only one cyclic group of order n up to isomorphism (see problem 45), which we will denote by C_n, and that there is only one infinite cyclic group up to isomorphism (see problem 46), which we will denote by C_∞.

If a group G_1 is isomorphic to a group G_2, then we will write $G_1 \cong G_2$.

50. Find, up to isomorphism, all groups that have: (a) 2 elements, (b) 3 elements.

51. Give an example of two groups that have the same number of elements and are non-isomorphic.

52. Prove that the group of all real numbers under addition is isomorphic to the group of all positive real numbers under multiplication.

53. Let a be an arbitrary element of the group G. Consider the map φ_a from the set of elements of the group G to itself, defined by $\varphi_a(x) = ax$ for any element x in G. Prove that φ_a is a transformation of the set of elements of the group G (i.e., it is a bijective map from the set of elements of G to itself).

54. We can define the transformation φ_a (see previous problem) for every element a of the group G. Prove that the set of these transformations φ_a forms a group with the usual operation of composition of transformations.

55. Prove that the group G is isomorphic to the group of transformations defined in the previous problem. (This result is known as Cayley's Theorem.)

§6. Subgroups

Let us consider a subset H of the group G. It may turn out that H itself is a group under the same binary operation as G.

In that case, H is called a *subgroup* of the group G. For example, the group of rotations of a regular n-gon is a subgroup of the group of all symmetries of a regular n-gon.

If a is an element of the group G, then the set of all elements of the form a^m is a subgroup of the group G. (This subgroup is a cyclic group, like those we studied in section 4.)

56. Let H be a subgroup of the group G. Prove that: (a) the identity elements of G and H are the same; (b) if a is an element of the subgroup H, then the inverse of a is the same in both G and H.

57. For H to be a subgroup of G (under the same binary operation) it is necessary and sufficient for following conditions to be satisfied: (1) if a and b are both in H, then the element ab (the multiplication performed in G) is also in H; (2) e (the identity element of G) is in H; (3) if a is in H, then a^{-1} (inversion performed in G) is also in H. Prove this.

Note. Condition (2) follows from conditions (1) and (3) if H is assumed to be non-empty.

58. Find all subgroups of the group of: (a) symmetries of an equilateral triangle; (b) symmetries of a square.

59. Find all subgroups of the cyclic groups: (a) C_5, (b) C_8, (c) C_{15}.

60. Prove that all subgroups of C_n are of the form $\{e, a^d, a^{2d}, \dots, a^{(\frac{n}{d}-1)d}\}$, where d is a factor of n and a is a generator of C_n.

61. Prove that all subgroups of the infinite cyclic group C_∞ are of the form $\{\dots, a^{-2r}, a^{-r}, e, a^r, a^{2r}, \dots\}$, where a is a generator and r is an arbitrary non-negative integer.

62. Prove that in any infinite group there are an infinite number of subgroups.

63. Prove that the intersection of any number of subgroups[†] of a group G is also a subgroup of the group G.

Example 10. Let us look at a regular tetrahedron with vertices labelled A, B, C and D. If we look from the side of the vertex D at the triangle ABC, then the points A, B, C can go either clockwise or anticlockwise (see fig. 5). This is the way we will distinguish the two orientations of the tetrahedron.

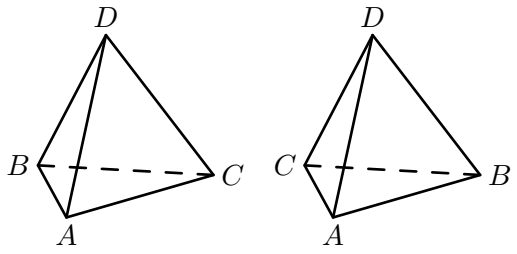


FIGURE 5.

[†]The intersection of several sets is the set of all elements that belong to all of the given sets at the same time.

64. Do the following transformations preserve orientation of the tetrahedron: $a = \left(\begin{smallmatrix} A & B & C & D \\ B & C & A & D \end{smallmatrix}\right)$, rotation by 120° about the altitude; $b = \left(\begin{smallmatrix} A & B & C & D \\ D & C & B & A \end{smallmatrix}\right)$, rotation by 180° about the axis that goes through the centres of the edges AD and BC; $c = \left(\begin{smallmatrix} A & B & C & D \\ A & C & B & D \end{smallmatrix}\right)$, reflection about the plane that goes through the edge AD and the centre of the edge BC; $d = \left(\begin{smallmatrix} A & B & C & D \\ B & C & D & A \end{smallmatrix}\right)$, the transformation that permutes the vertices cyclically?

All symmetries of the regular tetrahedron clearly form a group, which is called the *group of symmetries of a tetrahedron*.

65. How many elements are there in the group of symmetries of a tetrahedron?

66. In the group of symmetries of a tetrahedron, find subgroups isomorphic to: (a) the group of symmetries of a triangle; (b) the cyclic group C_4.

67. Prove that all symmetries of a tetrahedron which preserve orientation form a group. How many elements are in that group?

The group of symmetries of a tetrahedron which preserve orientation is called the *group of rotations of a tetrahedron*.

68. In the group of rotations of a tetrahedron, find subgroups isomorphic to the cyclic groups: (a) C_2, (b) C_3.

§7. Direct products

Given two groups, we can form a new group in the following way.

Definition 11. The *direct product* of two groups G and H, denoted by $G \times H$ (and usually pronounced "G cross H"), is the set of all ordered pairs (g, h), where g is an arbitrary element of G and h is an arbitrary element of H, with the following binary operation: $(g_1, h_1) \cdot (g_2, h_2) = (g_1 g_2, h_1 h_2)$, where the multiplication $g_1 g_2$ is performed in the group G and $h_1 h_2$ in H.

69. Prove that $G \times H$ is a group.

70. Suppose that the group G has n elements and the group H has k elements. How many elements are in the group $G \times H$?

71. Prove that the groups $G \times H$ and $H \times G$ are isomorphic.

72. Find subgroups of $G \times H$ which are isomorphic to G and to H respectively.

73. Suppose that G and H are commutative. Prove that $G \times H$ is also commutative.

74. Let G_1 be a subgroup of G and H_1 be a subgroup of H. Prove that $G_1 \times H_1$ is a subgroup of $G \times H$.

75. Let G and H be arbitrary groups. Is it necessarily true that every subgroup of the group $G \times H$ can be written as $G_1 \times H_1$ where G_1 is a subgroup of G and H_1 is a subgroup of H?

76. Prove that the group of symmetries of a rhombus is isomorphic to the group $C_2 \times C_2$.

77. Are the following pairs of groups isomorphic? (a) $C_2 \times C_3$ and C_6; (b) $C_2 \times C_4$ and C_8.

78. Prove that $C_m \times C_n \cong C_{mn}$ if and only if the numbers m and n are relatively prime.[†] (The "if" part of this result is often known as the *Chinese Remainder Theorem*, and can be also be stated in terms of modular arithmetic as follows (see problem 45): If m and n are relatively prime, then the simultaneous equations $x \equiv a \pmod{m}$ and $x \equiv b \pmod{n}$ have a unique solution modulo mn.)

§8. Cosets and Lagrange's Theorem

Each subgroup H of a group G gives rise to a partition of the set of elements of G, as we now show.

For any element x in G, consider the set of all elements of the form xh, where h is an element of the subgroup H. This set, denoted by xH, is called the *left coset of H generated by the element x.*

79. Find all left cosets of the following subgroups of the group of symmetries of a triangle: (a) rotations of a triangle; (b) reflections with respect to one axis $\{e, c\}$ (see examples 1 and 2).

80. Let H be a subgroup of a group G. Prove that every element of G belongs to some left coset of H.

81. Suppose that the element y belongs to the left coset of H generated by the element x. Prove that the left cosets of H generated by the elements x and y coincide. (In symbols: if y is an element of xH, then $xH = yH$.)

82. Suppose that the two left cosets of H generated by elements x and y have an element in common. Prove that these cosets coincide.

Therefore the left cosets generated by any two elements either do not intersect or they coincide, giving us a partition of all elements of the group G into non-intersecting classes. This partition is called *the left coset decomposition of the group G by the subgroup H.*

Assume G to be finite. The number of elements in a subgroup of G is called the *order* of that subgroup. Let m be the order of the subgroup H. If $h_1 \neq h_2$, then $xh_1 \neq xh_2$, so each left coset also has m elements. Therefore, if n is the order of the group G and r is the number of left cosets in the coset decomposition of G by H, we have $mr = n$, which proves

[†]Two numbers are called *relatively prime* or *coprime* if their greatest common divisor (highest common factor) is 1.

Theorem 1 (Lagrange's[†] Theorem)**.** *The order of a subgroup divides the order of the group.*

83. Prove that the order of any element (see page 9) divides the order of the group.

84. Prove that any group of prime order is cyclic and any element in it except e is a generator.

85. The group G contains 31 elements. How many subgroups can the group G have?

86. Prove that all groups of prime order p are isomorphic to each other.

87. Suppose that m divides n. Construct a group of order n which contains a subgroup isomorphic to a given group G of order m.

88. Suppose that m divides n. Must a group of order n have a subgroup of order m?

We can also construct the *right cosets* Hx and the *right coset decomposition* of the group G by the subgroup H. If the order of the subgroup H is equal to m, then as before, all right cosets also have m elements and their number is equal to the integer n/m, where n is the order of the group. So the number of right cosets is equal to the number of left cosets.

Note. In practice, to find the coset decomposition of a finite group, it is not necessary to construct the coset of each element, because this way we will have repeating classes. Rather, we should take elements that are not in cosets already constructed. Because $eH = He = H$, the subgroup itself is always a right and a left coset.

89. Construct the left and right coset decompositions of the group of symmetries of a triangle by the subgroup of: (a) rotations $\{e, a, b\}$, (b) reflections with respect to one axis $\{e, c\}$.

90. Construct the left and right coset decompositions of the group of symmetries of a square by the subgroup of: (a) reflections with respect to the central reflections $\{e, a\}$, (b) reflections with respect to a diagonal $\{e, d\}$.

91. Construct the coset decomposition of the group of integers under addition by the subgroup of numbers divisible by 3.[‡]

92. Find, up to isomorphism, all groups of order: (a) 4, (b) 6, (c) 8.

§9. Inner automorphisms

Let us begin with an example. We will study the group of symmetries of an equilateral triangle. If we denote the vertices by A, B, and C, then every element of

[†]Langrange, Joseph Louis (1736–1813) was a French mathematician and mechanical physicist.

[‡]We do not specify here which decomposition to construct, left or right, because in a commutative group these two decompositions obviously coincide.

the group can be defined by a permutation of the three letters A, B and C. For example, the reflection of the triangle with respect to its altitude from the vertex A to the side BC will be written as $\left(\begin{smallmatrix} A & B & C \\ A & C & B \end{smallmatrix}\right)$. To multiply two elements of the group of symmetries of the triangle, it is sufficient to perform the two permutations one after another. This way, we get an isomorphism of the group of symmetries of a triangle to the group of permutations of three letters, A, B, and C. Note, though, that the isomorphism is not unique: it depends on which vertex we denote by A, which by B and which by C. Relabelling the vertices can itself be viewed as a permutation of the three letters A, B and C. For example, $g = \left(\begin{smallmatrix} A & B & C \\ B & C & A \end{smallmatrix}\right)$ corresponds to the following relabelling of the vertices:

old name of vertex	A	B	C
new name of vertex	B	C	A

With these new names, every element of the group of symmetries of the triangle will get a new name as a permutation of the letters A, B, and C. For example, the reflection of the triangle with respect to the vertical altitude (fig. 6) is denoted in the following way:

$$\text{old name of permutation: } \left(\begin{smallmatrix} A & B & C \\ A & C & B \end{smallmatrix}\right)$$
$$\text{new name of permutation: } \left(\begin{smallmatrix} A & B & C \\ C & B & A \end{smallmatrix}\right)$$

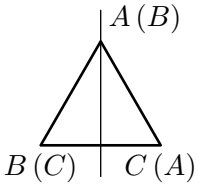


FIGURE 6.

93. Consider an element of the group of symmetries of a triangle, which, under some naming of the vertices, corresponds to the permutation h. Which permutation will correspond to the same element of the group of symmetries of a triangle when the vertices are renamed by g?

Note that this relabelling using g turns the element h of the group of transformations into ghg^{-1}. This is the case not only in this example of the group of symmetries of a triangle, but also in general. An element of the form ghg^{-1} is called a *conjugate* of h and the process is called *conjugation*. Doing this relabelling to an entire group, though, is so important that it warrants the following definition.

Definition 12. Let G be a group and g an element of G. Define a map φ_g from the group G to itself by the formula $\varphi_g(h) = ghg^{-1}$ (where h is any group element). This map is called the *inner automorphism* of the group G generated by the element g.

94. Prove that an inner automorphism of a group is an isomorphism from the group to itself. (The word *automorphism* means an isomorphism from the object to itself.)

95. To what does the reflection with respect to an altitude get sent by all possible inner automorphisms of the group of symmetries of a triangle?

96. To what does the rotation of the triangle by 120° get sent by all possible inner automorphisms of the group of symmetries of a triangle?

97. Which pairs of elements of the group of symmetries of a tetrahedron can be sent to each other by inner automorphisms and which cannot? Answer the same question for the group of rotations of a tetrahedron.

98. Prove that the orders of elements ab and ba are equal in any group.

We note that under any inner automorphism of a group (as well as with any other isomorphism of the group to itself), a subgroup is generally sent to a different subgroup. For example, reflections with respect to one altitude are sent to reflections with respect to another altitude. However, some highly symmetric subgroups stay in place under all inner automorphisms, for example, the subgroup of rotations of the group of symmetries of a triangle. We will now look at such subgroups.

§10. Normal subgroups

Definition 13. A subgroup of a group is called a *normal subgroup* if it is sent to itself by every inner automorphism of the group. In other words, a subgroup N of the group G is called a normal subgroup of G if for any element a in N and any element g in G, the element gag^{-1} is in N.

Thus, the subgroup of rotations is a normal subgroup of the group of symmetries of a triangle, while the subgroup of reflections with respect to the altitude from A to BC (consisting of two elements) is not a normal subgroup of the group of symmetries of a triangle.

99. Prove that in a commutative group, every subgroup is normal.

100. Is the subgroup of central symmetries of the group of symmetries of a square consisting of elements $\{e, a\}$ normal (examples 3 and 4)?

Theorem 2. *A subgroup N of a group G is normal if and only if the left and right coset decompositions (see section 8) of the group G by the subgroup N coincide.*[†]

101. Prove the theorem stated above.

102. Let G be a group of order n and H a subgroup with order $m = n/2$. Prove that H is a normal subgroup of G.

[†]In this case, the decompositions will simply be called decompositions by a normal subgroup.

103. Prove that an intersection (see the footnote on page 12) of any number of normal subgroups of a group G is a normal subgroup of the group G.

104. The set of elements of the group G that commute with all elements of the group is called the *centre* of the group G. Prove that the centre is a subgroup, in fact, a normal subgroup, of the group G.

105. Let N_1 and N_2 be normal subgroups of the groups G_1 and G_2 respectively. Prove that $N_1 \times N_2$ is a normal subgroup of the group $G_1 \times G_2$.

The following example shows that a normal subgroup of a normal subgroup of a group G does not have to be a normal subgroup of the group G itself.

Example 11. We examine the subgroup of the group of symmetries of a square which consists of reflections with respect to the diagonals and the centre (see examples 3 and 4), that is, $\{e, a, d, f\}$. This subgroup contains half of the elements of the group and hence is a normal subgroup (see problem 102 above). The subgroup $\{e, d\}$, which consists of reflections with respect to one of the diagonals, contains half of the elements of the subgroup $\{e, a, d, f\}$ and is, therefore, a normal subgroup of it. On the other hand, the subgroup $\{e, d\}$ is not a normal subgroup of the entire group of symmetries of a square, because d goes to $bdb^{-1} = f$ under the inner automorphism of the group generated by reflection with respect to another diagonal.

§11. Quotient groups

We begin with an example. We will look at the decomposition of the group of symmetries of a square by the normal subgroup consisting of the central symmetries e and a (see examples 3 and 4). It is an easy result that the decomposition will have four cosets and be as described in table 2. Let us denote each coset by a letter;

e	b	d	g
a	c	f	h
E	A	B	C

Table 2.

we have used E, A, B, and C in our table. If we multiply any element from the class A by any element in the class B, then the result turns out to be in the class C, independently of which elements were chosen from A and B. From the solution of the following problem, it follows that this isn't by accident.

106. Suppose we have a decomposition of the group G by a normal subgroup N, and suppose elements x_1 and x_2 lie in the same coset and elements y_1 and y_2 also lie in the same coset. Prove that the elements x_1y_1 and x_2y_2 both lie in the same coset.

So by taking a representative from each of the two cosets and multiplying them in a given order, we arrive at a coset which does not depend on which representatives we have chosen. Therefore, when we decompose a group by a normal subgroup N into a set of cosets, we can define a binary operation in the following way: if $A = xN$ and $B = yN$, then let $A \cdot B = (xy)N$. The result of problem 106 shows that this operation is well defined and does not depend on the choice of elements x and y, which generate the classes A and B. So, in the example above, $A \cdot B = C$.

Problems 107–109 deal with decompositions of a group G by a normal subgroup N.

107. Let T_1, T_2 and T_3 be cosets of N. Prove that $(T_1T_2)T_3 = T_1(T_2T_3)$.

108. Prove that $NT = TN = T$ for every coset T of N.

109. Prove that for any coset T of N, there exists a coset T^{-1} such that $TT^{-1} = T^{-1}T = N$.

From the statements of problems 107–109, it follows that the set of cosets with the binary operation described above is a group. This group is called the *quotient group of G by the normal subgroup N* and is denoted G/N. It is also sometimes called the *factor group* of G by N.

Clearly, $G/\{e\} \cong G$ and $G/G \cong \{e\}$. It is also clear that in the finite case, the order of the quotient group is equal to n/m, where n is the order of the group G and m is the order of the normal subgroup N. For example, the quotient group of the group of symmetries of a square by the subgroup of central symmetries has four elements.

110. Determine whether the quotient group of the group of symmetries of a square by the subgroup of central symmetries is isomorphic to the group of rotations of a square or to the group of symmetries of a rhombus.

111. Find all normal subgroups and the corresponding quotient groups[†] by them in the following groups: (a) the group of symmetries of a triangle, (b) $C_2 \times C_3$, (c) the group of symmetries of a square, (d) the group of quaternions (see the solution of problem 92).

112. Find all normal subgroups and quotient groups of the groups: (a) C_n, (b) C_∞.

113. Find all normal subgroups and quotient groups of the group of rotations of a tetrahedron.

114. In the direct product of groups $G_1 \times G_2$, consider the subgroup $G_1 \times \{e\}$. Prove this is a normal subgroup and that the quotient group by it is isomorphic to the group G_2.

[†]Subsequently, to find a quotient group means to point out some group, studied earlier, to which the sought quotient group is isomorphic.

§12. The commutator subgroup

Recall that two elements a and b of a group G are said to *commute* if $ab = ba$. The extent to which two elements of a group do not commute can be measured using the product $aba^{-1}b^{-1}$, which equals the identity if and only if a and b commute (prove this).

Definition 14. The element $aba^{-1}b^{-1}$ is called the *commutator* of the elements a and b. The *commutator subgroup* $K(G)$ of the group G is the set of all products of a finite number of commutators in the group G.

115. Prove that the commutator subgroup is actually a subgroup.

116. Prove that the commutator subgroup is a normal subgroup.

117. Prove that the commutator subgroup equals the identity subgroup $\{e\}$ if and only if the group is commutative.

118. Find the commutator subgroup of the groups: (a) of symmetries of a triangle, (b) of symmetries of a square, (c) of quaternions (see the solution of problem 92).

119. Prove that the commutator subgroup of the group of symmetries of a regular n-gon is isomorphic to the group C_n for n odd and the group $C_{n/2}$ for n even.

120. Find the commutator subgroup of the group of rotations of a tetrahedron.

121. Prove that if a normal subgroup of the group of rotations or the group of symmetries of a tetrahedron contains at least one rotation around an axis that goes through a vertex, then it contains the entire group of rotations of a tetrahedron.

122. Find the commutator subgroup of the group of symmetries of a tetrahedron.

Let us examine two more groups, the group of rotations of a cube and the group of rotations of a regular octahedron (fig. 7).

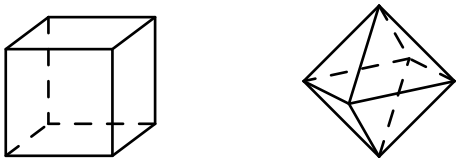


Figure 7.

123. How many elements are there in each of the above groups? List the elements of the group of rotations of a cube.

124. Prove that the groups of rotations of a cube and an octahedron are isomorphic.

125. How many different ways can we paint the faces of a cube with six colours (each face a different colour), if two paintings are considered different only if they cannot be superimposed with one another by a rotation of a cube? Answer the same question for a matchbox.

126. To which of the groups we have already studied is the group of rotations of a matchbox isomorphic?

The computation of the commutator subgroup of the group of rotations of a cube can be done conveniently by inscribing a tetrahedron into the cube as shown in the figure (fig. 8). If we connect the remaining vertices B, D, A_1 and C_1, then we will

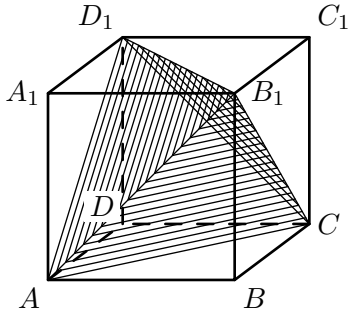


FIGURE 8.

get a second tetrahedron. Any rotation of the cube either sends each tetrahedron to itself, or swaps them.

127. Prove that the set of all rotations of the cube that send each tetrahedron to itself forms: (a) a subgroup, (b) a normal subgroup of the group of rotations of a cube.

128. Prove that the commutator subgroup of the group of rotations of a cube is isomorphic to the group of rotations of a tetrahedron.

We now prove the following three properties of the commutator subgroup, which we will need subsequently.

129. Prove that the quotient group of an arbitrary group G by its commutator subgroup is commutative.

130. Suppose N is a normal subgroup of the group G and with the quotient group G/N commutative. Prove that N contains the commutator subgroup of the group G.

131. Suppose N is a normal subgroup of the group G, and let $K(N)$ be the commutator subgroup of the group N. Prove that $K(N)$ is a normal subgroup of G. (Compare this with the example on page 18.)

§13. Homomorphisms

A *homomorphism* from the group G to the group F is a map $\varphi : G \to F$ such that $\varphi(ab) = \varphi(a) \cdot \varphi(b)$ for any elements a and b of the group G. (Here the multiplication ab is performed in the group G and the multiplication $\varphi(a) \cdot \varphi(b)$ is performed

in F.) A homomorphism is different from an isomorphism because a homomorphism does not have to be bijective.

Example 12. Let G be the group of rotations of a cube and C_2 be the group of permutations of the two tetrahedra inscribed into it (see page 21). Every rotation of the cube corresponds to a permutation of the tetrahedra; when two rotations are performed successively, the corresponding permutations of the tetrahedra are multiplied. This way, the map from the group of rotations of the cube to the group of permutations of the tetrahedra is a homomorphism.

132. Suppose $\varphi : G \to F$ is a surjective homomorphism from the group G to the group F. If the group G is commutative, then F is also commutative. Prove this. Is the converse true?

133. Prove that a homomorphism from the group G to the group F sends the identity of the group G to the identity of the group F.

134. Prove that $\varphi(a^{-1}) = (\varphi(a))^{-1}$, where $\varphi : G \to F$ is a homomorphism, on the left hand side the inverse is taken in the group G and on the right hand side the inverse is taken in the group F.

135. Suppose that $\varphi_1 : G \to F$ and $\varphi_2 : F \to H$ are homomorphisms. Prove that $\varphi_2\varphi_1 : G \to H$ is a homomorphism.

We can obtain some important examples of homomorphisms by the following construction of the *natural homomorphism*.

Let N be a normal subgroup of the group G. Consider the following map φ from the group G to the quotient group G/N: let every element g in the group G be mapped by φ to the coset T of N which contains g.

136. Prove that $\varphi : G \to G/N$ is a surjective homomorphism.

The map φ is called the *natural homomorphism* from the group G to the quotient group G/N, and is surjective. We have thus shown that there is a homomorphism associated to every normal subgroup. We will also show the converse, that every surjective homomorphism from a group G to a group F can be viewed as a natural homomorphism from G to the quotient group G/N for a suitable normal subgroup N.

If $\varphi : G \to F$ is a homomorphism, then the set of elements g such that $\varphi(g) = e_F$ is called the *kernel of the homomorphism* φ, and is denoted by $\ker \varphi$.

137. Prove that $\ker \varphi$ is a subgroup of the group G.

138. Prove that $\ker \varphi$ is a normal subgroup of the group G.

Let's look at the decomposition of the group G with respect to the kernel $\ker \varphi$.

139. Prove that g_1 and g_2 lie in the same coset if and only if $\varphi(g_1) = \varphi(g_2)$.

Theorem 3 (The First Isomorphism Theorem)**.** *Suppose $\varphi : G \to F$ is a surjective homomorphism from the group G to the group F. Then the map $\psi : G/\ker\varphi \to F$ which sends each coset of $\ker\varphi$ to the value of $\varphi(g)$ for any element g in the coset is an isomorphism.* (*This is well-defined by problem 139.*)

The proof of this theorem is contained in the solutions to the following problems.

140. Prove that ψ is a surjective map.

141. Prove that ψ is a bijective map.

142. Prove that ψ is an isomorphism.

Let us give an example of an application of the above theorem.

Example 13. In problem 110, we had to determine whether the quotient group of the symmetries of a square by the normal subgroup of central reflections was isomorphic to the group of rotations of a square or the group of symmetries of a rhombus. Every element of the group corresponds to a certain permutation of the axes l_1, l_2, l_3, l_4 (see figure 9). Note that the diagonals l_1 and l_3 can only be sent to

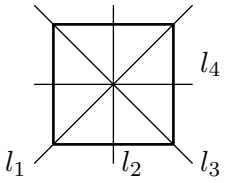


FIGURE 9.

each other, and the axes l_2 and l_4 can also only be sent to each other. In this way, we get a map from the group of symmetries of a square to a group of permutations of four elements: l_1, l_2, l_3 and l_4. This map is a surjective homomorphism to the group of permutations that sends l_1 and l_3 to l_1 and l_3 and sends l_2 and l_4 to l_2 and l_4 (check this). This group consists of four permutations and is isomorphic to the group of symmetries of the rhombus $L_1L_2L_3L_4$ (figure 10). The kernel of the constructed

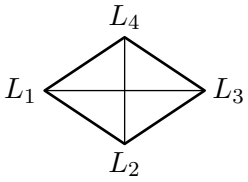


FIGURE 10.

homomorphism consists of all symmetries of the square that send each one the four axes to itself. It is easy to check that the only such transformations are e and the central symmetry a. Therefore, by Theorem 3, the subgroup of reflections $\{e, a\}$ with respect to the centre is a normal subgroup in the group of symmetries of a square and its quotient group is isomorphic to the group of symmetries of a rhombus.

The following problems can be solved in a similar way.

143. Prove that the rotations of a tetrahedron by 180° around the axes that go through the centres of opposite edges, together with the identity transformation, form a normal subgroup of the group of symmetries of a tetrahedron, and find the quotient group.

144. Prove that the rotations of a cube by 180° around axes going through the centres of opposite faces, together with the identity transformation, form a normal subgroup of the group of rotations of a cube, and find the quotient group.

145. Suppose we have a regular n-gon centred at O in the plane. Let R be the group of all rotations of the plane around the point O. Let us look at the subgroup C_n of all rotations of the plane that send the regular n-gon to itself. Prove that this is a normal subgroup of the group R and that $R/C_n \cong R$.

146. Let N_1 and N_2 be normal subgroups of the groups G_1 and G_2 respectively. Prove that $N_1 \times N_2$ is a normal subgroup of $G_1 \times G_2$, and that $G_1 \times G_2/N_1 \times N_2 \cong G_1/N_1 \times G_2/N_2$.

147. Can two non-isomorphic groups have isomorphic normal subgroups with isomorphic quotient groups?

148. Can a group have two isomorphic normal subgroups with non-isomorphic quotient groups?

149. Can a group have non-isomorphic normal subgroups with isomorphic quotient groups?

Let us now consider what happens to subgroups, normal subgroups and commutator subgroups under homomorphisms. If $\varphi : G \to F$ is a homomorphism and we choose a subset M of G, then we call the set of all elements of F that have at least one preimage in M the *image* of the set M under the homomorphism φ (denoted by $\varphi(M)$). Conversely, if P is a subset of F, then we call the set of all elements of G whose image is in P the *preimage* of the set P (denoted by $\varphi^{-1}(P)$). Note that φ^{-1} by itself, without specifying P, does not have a meaning: in general, there does not exist an inverse map for a homomorphism. Note also that if $\varphi(M) = P$, then $\varphi^{-1}(P)$ contains M but does not necessarily coincide with it (see figure 11).

150. Prove that under the homomorphism $\varphi : G \to F$, the image of a subgroup H of the group G is a subgroup of the group F.

151. Suppose H is a subgroup of F and $\varphi : G \to F$ is a homomorphism. Prove that $\varphi^{-1}(H)$ is a subgroup of G.

152. Suppose that N is a normal subgroup of F and $\varphi : G \to F$ is a homomorphism. Prove that $\varphi^{-1}(N)$ is a normal subgroup of G.

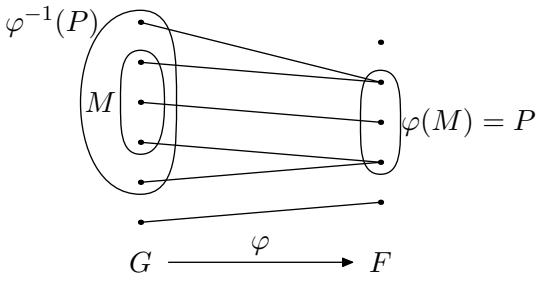


FIGURE 11.

153. Suppose that $\varphi : G \to F$ is a homomorphism and let K_1 and K_2 be the commutator subgroups of the groups G and F respectively. Prove that $\varphi(K_1)$ is a subset of K_2 and that K_1 is a subset of $\varphi^{-1}(K_2)$.

154. Suppose that N is a normal subgroup of the group G and that φ is a surjective homomorphism from the group G to the group F. Prove that $\varphi(N)$ is a normal subgroup of the group F.

155. Let K_1 and K_2 be the commutator subgroups of the groups G and F respectively and suppose that φ is a surjective homomorphism from the group G to the group F. Prove that $\varphi(K_1) = K_2$. Is it necessarily true that $K_1 = \varphi^{-1}(K_2)$?

§14. Solvable groups

There exists an important class of groups, related to commutative groups: the so-called solvable groups. They are called solvable, because, as we will see, the ability to solve a polynomial equation in radicals depends on whether a certain group is solvable.

Let G be a group and $K(G)$ its commutator subgroup. The commutator subgroup $K(G)$ is itself a group and so it itself has a commutator subgroup $K(K(G))$. That commutator subgroup also has a commutator subgroup, and so on. The group $K(K(\dots(K(G))\dots))$ with r Ks will be denoted by $K_r(G)$, so $K_{r+1}(G) = K(K_r(G))$.

Definition 15. A group G is called *solvable* if the chain of groups G, $K(G)$, $K_2(G)$, $K_3(G)$, ... ends for some finite n with the identity group, i.e., if for some n we have $K_n(G) = \{e\}$.

For example, any commutative group is solvable, because if G is a commutative group, then at the first step we get $K(G) = \{e\}$ (see problem 117). Similarly, a group G is solvable if its commutator subgroup is commutative, because then $K_2(G) = \{e\}$.

156. Determine whether or not the following groups are solvable: (a) the cyclic group C_n, (b) the group of symmetries of a triangle, (c) the group of quaternions (see the solution to problem 92), (d) the group of rotations of a tetrahedron, (e) the group of symmetries of a tetrahedron, (f) the group of rotations of a cube.

We will be exploring the solvability of other groups later. In particular, we need to show that *the group of rotations of a regular dodecahedron* (figure 12) is non-solvable.

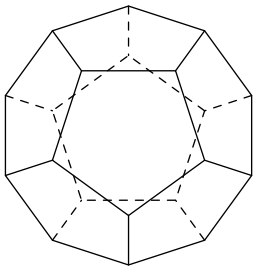


FIGURE 12.

157. How many elements are there in the group of rotations of a dodecahedron?

The rotations of a dodecahedron can be broken down into four classes: (1) the identity transformation; (2) rotations around axes going through the centres of opposite faces; (3) rotations around axes going through opposite vertices; (4) rotations around axes going through the centres of opposite edges.

158. How many elements are there in each class? (The identity transformation does not belong to classes (2)–(4).)

159. Suppose that N is a normal subgroup in the group of rotations of a dodecahedron, and suppose that N has some element of one of the four classes (1)–(4). Prove that N then contains that entire class.

So each of the classes (1)–(4) either is entirely in N or entirely not in N.

160. Prove that the group of rotations of a dodecahedron has no normal subgroups besides $\{e\}$ and the entire group.

161. Let a group G be non-commutative and have no normal subgroups besides $\{e\}$ and G. Prove that the group G is non-solvable.

From problems 160 and 161, it follows that the group of rotations of a dodecahedron is non-solvable.

Let us look at several more problems whose results we will need later.

162. Prove that any subgroup of a solvable group is solvable.

163. Let $\varphi : G \to F$ be a surjective homomorphism from the group G to the group F and suppose G is solvable; prove that the group F is also solvable.

164. In the setup of the previous problem, give an example where the group F is solvable, while the group G is non-solvable.

165. Suppose that the group G is solvable and that N is a normal subgroup of G. Prove that the quotient group G/N is solvable.

166. Suppose that N is a normal subgroup of the group G. Prove that if both N and G/N are solvable, then the group G is also solvable.

167. Suppose that the groups G and F are solvable. Prove that the group $G \times F$ is also solvable.

168. Suppose that the group G is solvable. Prove that there exists a chain of subgroups G_0, G_1, ..., G_n of G such that: (1) $G_0 = G$; (2) each group G_i with $1 \leqslant i \leqslant n$ is a normal subgroup of the group G_{i-1}, and each of the quotient groups G_{i-1}/G_i is abelian; (3) the group G_n is abelian.

169. Suppose that for a group G there exists a chain of subgroups with the properties described in the previous problem. Prove that the group G is solvable.

The results of problems 168 and 169 show that, for a group G, the existence of a chain of subgroups as described in problem 168 is equivalent to the solvability of G and so could be taken as the definition of solvability. Another equivalent definition can be derived from the results of the following two problems.

170. Suppose that the group G is solvable. Prove that there exists a chain of groups $G_0, G_1, \ldots, G_n$ such that: (1) $G_0 = G$; (2) each group G_i with $1 \leqslant i \leqslant n-1$ contains a commutative normal subgroup N_i such that $G_i/N_i \cong G_{i+1}$; (3) the group G_n is commutative.

171. Suppose that for a group G there exists a chain of groups with the properties described in the previous problem. Prove that the group G is solvable.

§15. Permutations

Let us look now in more detail at *permutations* (i.e., transformations) on the set of the first n positive integers 1, 2, ..., n; we will call these permutations *permutations of degree n*. Note that permutations of an arbitrary set with n elements can be reduced to the permutations described above by just numbering the elements of that set by positive integers 1, 2, ..., n. An arbitrary permutation of degree n can be written as $\left(\begin{smallmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{smallmatrix}\right)$, where i_m is the image of the element m under this permutation. Recall that a permutation is a bijective map, so all the elements in the second row are distinct.

172. How many different permutations of degree n are there?

Definition 16. The group of all permutations of degree n with the usual operation of multiplication (i.e., composition)[†] is called the *symmetric group of degree n* and is denoted by S_n.

173. Prove that for $n \geqslant 3$, the group S_n is non-commutative.

A permutation can move some elements and leave others fixed; it can turn out that the moving elements are moving as if in a circle. For example, the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 2 & 6 & 3 & 5 & 1 & 7 \end{pmatrix}$$

leaves the elements 2, 5 and 7 fixed, and moves the rest of the elements in a circle: $1 \to 4$, $4 \to 3$, $3 \to 6$, $6 \to 1$. Permutations of this sort are called *cyclic permutations* or simply *cycles*. For cyclic permutations, we will use also a different notation: the expression $(1\,4\,3\,6)$ will mean the permutation that sends $1 \to 4$, $4 \to 3$, $3 \to 6$, $6 \to 1$ and leaves the rest of the elements fixed. So, if this is a permutation of degree 7, then it coincides with the permutation described above.

Not all permutations are cyclic. For example, the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 4 & 1 & 2 & 6 \end{pmatrix}$$

is not cyclic, but it can be expressed as a product of two cycles:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 4 & 1 & 2 & 6 \end{pmatrix} = (1\,3\,4) \cdot (2\,5).$$

The two cycles move different elements, and such cycles are called *independent*. It is easy to see that the product of independent cycles does not depend on the order in which they are performed. If we do not distinguish products of independent cycles differing only in the order in which they are performed, then the following statement will be true.

174. Any permutation splits uniquely (up to the order of the terms) into several independent cycles. Prove this.

Cycles of the form $(i\,j)$ that move only two elements are called *transpositions*.

175. Prove that any cycle can be split into a product of transpositions (not necessarily independent).

The transpositions $(1\,2)$, $(2\,3)$, $\ldots$, $(n-1, n)$ are called *elementary transpositions*.

176. Prove that an arbitrary transposition can be expressed as a product of elementary transpositions.

[†]With our definitions, we will view products of transformations (see page 5) as being performed from right to left. Sometimes products of transformations are considered as being performed from left to right. The groups produced in the two cases are isomorphic.

From the results of problems 174–176, it follows that an arbitrary permutation of degree n can be expressed as a product of elementary transpositions. In other words, the following theorem is true.

Theorem 4. *If a subgroup of the symmetric group S_n contains all elementary transpositions, then this subgroup coincides with the entire group S_n.*

Let the numbers 1, 2, ..., n be written in a line in an arbitrary order. We say that a pair of numbers i, j forms an *inversion* in this line if $i < j$ but j is earlier in the line than i. The number of inversions characterises the disorder of the line with respect to the usual order of numbers 1, 2, ..., n.

177. Find the number of inversions in the line 3, 2, 5, 4, 1.

Subsequently we will be interested not in the actual number of inversions, but in its parity.

178. Prove that the parity of the number of inversions in a line changes if two arbitrary numbers switch places.

Definition 17. A permutation $\left(\begin{smallmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{smallmatrix}\right)$ is called *even* or *odd* according to whether the number of inversions in the second line is even or odd.

For example, the identity permutation $\left(\begin{smallmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{smallmatrix}\right)$ is an even permutation, because the number of inversions in the second line is equal to zero.

179. Find the parity of the permutation $\left(\begin{smallmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 4 & 1 & 3 \end{smallmatrix}\right)$.

180. Prove that when an even permutation is multiplied by a transposition, either on the left or the right, we get an odd permutation, and, vice versa, when an odd permutation is multiplied by a transposition we get an even permutation.

181. Prove that an even permutation can only be split into a product of an even number of transpositions, and an odd permutation can only be split into a product of an odd number of transpositions.

182. Find the parity of an arbitrary cycle of length: (a) 3, (b) 4, (c) m.

183. Prove that when two permutations of the same parity are multiplied, we get an even permutation, and when two permutations of different parities are multiplied, we get an odd permutation.

184. Prove that the permutations a and a^{-1} have the same parity, where a is an arbitrary permutation.

From the results of problems 183 and 184, it follows that the set of all even permutations forms a subgroup of the group S_n.

Definition 18. The group of all even permutations of degree n is called the *alternating group of degree n* and is denoted by A_n.

185. Prove that for $n \geqslant 4$, the group A_n is non-commutative. (Compare problem 173.)

186. Prove that the alternating group A_n is a normal subgroup of the group S_n, and construct the coset decomposition of the group S_n with respect to the subgroup A_n.

187. Find the number of elements in the group A_n.

188. Prove that the groups S_2, S_3 and S_4 are solvable.

Let us prove now that the alternating group A_5 is non-solvable. One of the methods for proving this consists of the following. Inscribe five tetrahedra numbered 1, 2, 3, 4 and 5 into a dodecahedron. Then we can check that every rotation of the dodecahedron corresponds to an even permutation of the tetrahedra and that different rotations correspond to different permutations. In this way, we establish an isomorphism between the group of rotations of a dodecahedron and the group A_5 of even permutations of degree five. Then the non-solvability of the group A_5 will follow from the non-solvability of the group of rotations of a dodecahedron.

189. Inscribe five tetrahedra into a dodecahedron using the method described above.

Another method of proving the non-solvability of the group A_5 amounts to a repetition of the idea in the proof of the non-solvability of the group of rotations of a dodecahedron. For that, the following problems need to be solved.

190. Prove that any non-identity even permutation of degree five can be split into independent cycles in precisely one of the following three ways: (1) $(i_1\, i_2\, i_3\, i_4\, i_5)$, (2) $(i_1\, i_2\, i_3)$, (3) $(i_1\, i_2)(i_3\, i_4)$.

191. Suppose that N is a normal subgroup of the group A_5. Prove that if N contains at least one permutation that splits into independent cycles in one of the ways described in the previous problem, then N contains all permutations that split in the same way into independent cycles.

192. Prove that the group A_5 has no normal subgroups except for the identity subgroup and the entire group.

From the statements of problems 192 and 161, and from the fact that the group A_5 is non-commutative, it follows that the group A_5 is non-solvable.

193. Prove that the symmetric group S_n contains a subgroup isomorphic to the group A_5 for $n \geqslant 5$.

From the statements of problems 193 and 162 we get the following theorem.

Theorem 5. *For $n \geqslant 5$, the symmetric group S_n is non-solvable.*

The above theorem, as well as other results from this chapter, will be needed in the next chapter to prove the non-solvability in radicals of general polynomial equations of degree greater than four.

CHAPTER 2

Complex numbers

As we progress through secondary school, we gradually expand the set of numbers that we study. In doing this, the fact most stressed is that such expansion allows us to operate on the numbers more freely. So in moving from the natural numbers to the integers, it becomes possible to subtract any numbers; in moving to the rationals, it becomes possible to divide any numbers, and so on. A more significant feature of such expansions turns out to be the fact that the properties of the expanded systems often allow us to get new results about the original systems. For example, many difficult problems of number theory relating only to positive integers were solved using real and complex numbers.

Historically, complex numbers appeared specifically to solve certain problems about the real numbers. For example, the Italian mathematician Cardano (1501–1576), when solving cubic equations, found the real roots of the equation using intermediate "unreal" square roots of negative numbers.

Over time, complex numbers began taking on a more prominent position in mathematics and its applications. First of all, they deeply penetrated into the theory of polynomial equations, because over the complex numbers, the study of such equations turned out to be much more convenient. For example, any polynomial equation of degree n ($n \geqslant 1$) with real or complex coefficients has at least one complex root (see "the fundamental theorem of algebra" below, page 55). On the other hand, not every polynomial equation with real coefficients has at least one real root.

After an interpretation of the complex numbers appeared using points in a plane and vectors in a plane, it became possible to apply geometric concepts such as continuity and geometric transformations to the study of the complex numbers. The connection between complex numbers and vectors allowed the reduction of many problems in mechanics, especially in hydro- and aerodynamics, as well as the theory of electricity, the theory of heat, etc., to problems about complex numbers and equations in them.

The study of complex numbers has now developed into a large and important branch of modern mathematics known as the theory of functions of a complex variable.

The reader can expect to gain a reasonably deep understanding of complex numbers and functions of a complex variable from this chapter.

§1. Fields and polynomials

Real numbers can be summed and multiplied, and it is also possible to perform the inverse operations, subtraction and division. In sums, we can arbitrarily switch summands and arbitrarily distribute parentheses. We can treat the terms of a product in the same way. All of these properties, as well as the connection between addition and multiplication, can be expressed concisely in the following way.

The real numbers have the following three properties:

(1) They form an abelian group (see chapter 1, section 3) under addition; the identity element of this group is denoted by 0 and is called the zero.
(2) If we ignore 0, then the rest of the numbers form an abelian group under multiplication.
(3) Addition and multiplication are connected through the distributive law: for any numbers a, b and c,

$$a(b+c) = ab + ac.$$

Having these three properties is very important, because they allow us to simplify arithmetic and algebraic expressions and to solve many equations. The set of real numbers is not the only set that has these three properties. We introduce a special term to describe such sets.

Definition 19. If on a set we have two binary operations defined, called addition and multiplication, satisfying the three properties described above, then such a set is called a *field*.

194. Determine whether the following subsets of the real numbers under the usual operations of addition and multiplication are fields: (a) all positive integers; (b) all integers; (c) all rational numbers; (d) all numbers of the form $r_1 + r_2\sqrt{2}$, where r_1 and r_2 are arbitrary rational numbers.

195. Prove that in any field we have $a \cdot 0 = 0 \cdot a = 0$ for any element a in the field.

196. Prove that in any field we have $(-a)\cdot b = a\cdot(-b) = -(a\cdot b)$ and $(-a)\cdot(-b) = ab$ for any elements a and b in the field.

197. Suppose that a and b are elements of an arbitrary field, with $a \cdot b = 0$. Prove that $a = 0$ or $b = 0$.

Example 14. Consider the set $\{0, 1, \ldots, n-1\}$ of remainders modulo n, with the operations of addition modulo n (see example 9 on page 9) and *multiplication modulo* n, where the result of the multiplication of two numbers is the remainder when their product is divided by n.

198. Construct the multiplication tables for multiplication modulo 2, 3 and 4.

199. Prove that this set of remainders under the operations of addition and multiplication modulo n forms a field if and only if n is a prime.

Definition 20. The *difference* of elements b and a in an arbitrary field (denoted by $b - a$) is the solution to the equation $x + a = b$ (or to $a + x = b$). If $a \neq 0$, then the *quotient* of the element b by a (denoted by b/a) is the solution to the equation $ya = b$ (or to $ay = b$).

From the result of problem 24 and the fact that in a field addition and multiplication are commutative, it follows that the elements $b - a$ and b/a (with $a \neq 0$) are defined uniquely in any field.

Because a field is a group under addition and, without the zero, is a group under multiplication, the equality $x + a = b$ is equivalent to the equality $x = b + (-a)$, and the equality $ya = b$ with $a \neq 0$ is equivalent to the equality $y = ba^{-1}$. Thus, $b - a = b + (-a)$ and $b/a = ba^{-1}$.

The reader can easily prove that the operations of addition, subtraction, multiplication and division have all of the basic properties of these operations in the field of real numbers. In particular, in any field both sides of any equality can be multiplied or divided by any non-zero element, any term can be moved from one side of an equality to the other with the opposite sign, etc. As an example, let us consider the distributive property which connects subtraction and multiplication.

200. Prove that in any field, $(a - b)c = ac - bc$ for any elements a, b, and c in the field.

If K is a field, then we can consider polynomials with coefficients from the field K, also known as polynomials over the field K, in the same way as we consider polynomials over the field of real numbers.

Definition 21. Let n be a positive integer. A *polynomial of degree n in one variable x over the field K* is any expression of the form

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0, \tag{1.1}$$

where a_n, a_{n-1}, $\ldots$, a_0 are elements of the field K, and $a_n \neq 0$. If a is an element of the field K, then the expression a is also considered a polynomial over the field K; if $a \neq 0$, then this is a polynomial of degree 0, while if $a = 0$, then the degree of the polynomial is undefined.

The elements a_n, a_{n-1}, $\ldots$, a_0 are called the *coefficients* of the polynomial (1.1), and a_n is called the *leading coefficient*.

Two polynomials in the variable x are considered equal if and only if all of their corresponding coefficients are pairwise equal.

Let

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0.$$

If we plug in some element a of the field K in the right hand side of this equality in place of x and perform the necessary operations, taking the operations of addition and multiplication as operations in the field K, then as a result we will get some element b of the field K. Then we write $P(a) = b$. If $P(a) = 0$, where 0 is the zero

element of the field K, then we say that a is a *root of the equation* $P(x) = 0$; in this case we also say that a is a *root of the polynomial* $P(x)$.

Polynomials over an arbitrary field K can be added, subtracted and multiplied.

The *sum* of the polynomials $P(x)$ and $Q(x)$ is the polynomial $R(x)$ where the coefficient of x^k in $R(x)$ ($k = 0, 1, 2, \dots$) is equal to the sum (in the field K) of the coefficients of x^k in the polynomials $P(x)$ and $Q(x)$. The *difference* of two polynomials is defined in an analogous way. Obviously, the degree of the sum or the difference is no larger than the greater of the degrees of the given polynomials.

To calculate the *product* of the polynomials $P(x)$ and $Q(x)$, we need multiply each term ax^k of the polynomial $P(x)$ by each term bx^l of the polynomial $Q(x)$ using the rule: $ax^k \cdot bx^l = abx^{k+l}$, where ab is a product in the field K and $k + l$ is the usual sum of non-negative integers. After this, all of the resulting expressions have to be added and like terms collected, that is, all terms involving the same degree r of the variable x are collected, and the sum $d_1x^r + d_2x^r + \cdots + d_sx^r$ is changed to the expression $(d_1 + d_2 + \cdots + d_s)x^r$.

If

$$P(x) = a_nx^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_0$$

and

$$Q(x) = b_mx^m + b_{m-1}x^{m-1} + b_{m-2}x^{m-2} + \cdots + b_0,$$

then[†]

$$\begin{aligned} P(x) \cdot Q(x) = a_nb_mx^{n+m} + (a_nb_{m-1} + a_{n-1}b_m)x^{n+m-1} + \\ (a_nb_{m-2} + a_{n-1}b_{m-1} + a_{n-2}b_m)x^{n+m-2} + \cdots + \\ (a_1b_0 + a_0b_1)x + a_0b_0. \end{aligned}$$

Because $a_0 \neq 0$ and $b_0 \neq 0$, we have $a_0b_0 \neq 0$ (see problem 197), so the degree of the polynomial $P(x) \cdot Q(x)$ is equal to $n + m$, i.e., the degree of the product of two non-zero polynomials is equal to the sum of the degrees of the given polynomials.

Using the fact that the operations of addition and multiplication of elements in a field K have the properties of commutativity, associativity and distributivity, it is easy to show that the operations we have introduced, of addition and multiplication of polynomials over the field K, also have the properties of commutativity, associativity and distributivity.

If

$$P(x) + Q(x) = R_1(x), \quad P(x) - Q(x) = R_2(x), \quad P(x) \cdot Q(x) = R_3(x),$$

and a is an arbitrary element of the field K, then it is also easy to show that

$$P(a) + Q(a) = R_1(a), \quad P(a) - Q(a) = R_2(a), \quad P(a) \cdot Q(a) = R_3(a).$$

[†]The coefficient of x^k in the product $P(x) \cdot Q(x)$ is equal to $a_0b_k + a_1b_{k-1} + \cdots + a_{k-1}b_1 + a_kb_0$, where we take $a_i = 0$ for $i > n$ and $b_j = 0$ for $j > m$.

Polynomials over an arbitrary field K can be divided by each other with a remainder. To divide a polynomial $P(x)$ by the polynomial $Q(x)$ with a remainder means to find polynomials $S(x)$ (the quotient) and $R(x)$ (the remainder) such that

$$P(x) = S(x) \cdot Q(x) + R(x),$$

and either the degree of the polynomial $R(x)$ is less than the degree of the polynomial $Q(x)$, or $R(x) = 0$.

Let $P(x)$ and $Q(x)$ be arbitrary polynomials over the field K, with $Q(x) \neq 0$. We show that we can divide the polynomial $P(x)$ by the polynomial $Q(x)$ with a remainder.

Suppose that

$$P(x) = a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_0$$

and

$$Q(x) = b_m x^m + b_{m-1}x^{m-1} + b_{n-2}x^{m-2} + \cdots + b_0.$$

If $n < m$, then letting $S(x) = 0$ and $R(x) = P(x)$ we get the required quotient and remainder. If $n \geqslant m$, then let us look at the polynomial

$$P(x) - \frac{a_n}{b_m}x^{n-m}Q(x) = R_1(x).$$

The polynomial $R_1(x)$ does not contain an x^n term, so either its degree is no more than $n-1$ or $R_1(x) = 0$. If

$$R_1(x) = c_k x^k + c_{k-1}^{k-1} + \cdots + c_0$$

and $k \geqslant m$, then we look at the polynomial

$$R_1(x) - \frac{c_k}{b_m}x^{k-m}Q(x) = R_2(x),$$

and so on. Because the degree of the resulting polynomial is strictly less than the degree of the previous polynomial, this process will have to end, i.e., at some step we will get

$$R_{s-1}(x) - \frac{d_l}{b_m}x^{l-m}Q(x) = R_s(x),$$

where the degree of the polynomial $R_s(x)$ is less than m, or $R_s(x) = 0$. Then we get

$$\begin{aligned}
P(x) &= \frac{a_n}{b_m}x^{n-m}Q(x) + R_1(x) \\
&= \frac{a_n}{b_m}x^{n-m}Q(x) + \frac{c_k}{b_m}x^{k-m}Q(x) + R_2(x) \\
&= \cdots \\
&= \frac{a_n}{b_m}x^{n-m}Q(x) + \frac{c_k}{b_m}x^{k-m}Q(x) + \cdots + \frac{d_l}{b_m}x^{l-m}Q(x) + R_s(x) \\
&= \left(\frac{a_n}{b_m}x^{n-m} + \frac{c_k}{b_m}x^{k-m} + \cdots + \frac{d_l}{b_m}x^{l-m}\right) \cdot Q(x) + R_s(x).
\end{aligned}$$

Here, the expression in the parentheses is the quotient of the division of the polynomial $P(x)$ by $Q(x)$ and $R_s(x)$ is the remainder. The process of division of a polynomial by another polynomial described here is basically the process of long division.

The following problem shows that if $P(x)$ and $Q(x)$ are two polynomials with $Q(x) \neq 0$, then, no matter what method we use for dividing $P(x)$ by $Q(x)$ with a remainder, the quotient and the remainder are defined uniquely.

201. Suppose that

$$P(x) = S_1(x) \cdot Q(x) + R_1(x),$$

and also that

$$P(x) = S_2(x) \cdot Q(x) + R_2(x),$$

where the degrees of polynomials $R_1(x)$ and $R_2(x)$ are less than the degree of the polynomial $Q(x)$ (allowing $R_1(x) = 0$ or $R_2(x) = 0$). Prove that

$$S_1(x) = S_2(x), \quad R_1(x) = R_2(x).$$

§2. The field of complex numbers

From the solution to problem 194, it follows that there exist fields smaller than the field of real numbers, for example the field of rational numbers. We will now construct a field larger than the field of real numbers, namely the field of complex numbers.

Let us look at all ordered pairs of real numbers, i.e., pairs of the form (a, b), where a and b are arbitrary real numbers. We say that $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$. On the set of all such pairs we define two operations, addition and multiplication, in the following way:

$$(a, b) + (c, d) = (a + c, b + d), \tag{2.1}$$

$$(a, b) \cdot (c, d) = (ac - bd, ad + bc). \tag{2.2}$$

(In the terms on the right hand sides of the equalities, the operations are the usual ones on real numbers.) For example, we have

$$(\sqrt{2}, 3) + (\sqrt{2}, -1) = (2\sqrt{2}, 2),$$
$$(0, 1) \cdot (0, 1) = (-1, 0).$$

Definition 22. The set of all possible ordered pairs of real numbers, with the operations of addition and multiplication defined by (2.1) and (2.2), is called the *set of complex numbers*.

From this definition it is clear that in the set of complex numbers there isn't anything "supernatural": complex numbers are just pairs of real numbers. However, one may question whether it is right to call such objects numbers. We will discuss

this question at the end of this section. Another question that the reader may have is why we defined the operations of addition and multiplication of complex numbers, especially the odd-looking multiplication, in this way and not in some other way. We will answer this question in section 3.

We will now explore useful properties possessed by the complex numbers.

202. Prove that the complex numbers form an abelian group under addition. Which complex number is the identity element (the zero) of this group?

Subsequently, it will be convenient to denote complex numbers with a single letter, for example z or w.

203. Prove that the operation of multiplication of complex numbers is commutative and associative, i.e., $z_1 \cdot z_2 = z_2 \cdot z_1$ and $(z_1 \cdot z_2) \cdot z_3 = z_1 \cdot (z_2 \cdot z_3)$ for any complex numbers z_1, z_2 and z_3.

It is easy to check that

$$(a, b) \cdot (1, 0) = (1, 0) \cdot (a, b) = (a, b)$$

for any complex number (a, b). Therefore, the complex number $(1, 0)$ is an identity element in the set of complex numbers under multiplication.

204. Let z be an arbitrary complex number with $z \neq (0, 0)$. Prove that there exists a complex number z^{-1} such that

$$z \cdot z^{-1} = z^{-1} \cdot z = (1, 0).$$

The results of problems 203 and 204 show that the non-zero complex numbers form an abelian group under multiplication.

205. Prove that the operations of addition and multiplication satisfy the distributive law, i.e., $(z_1 + z_2) \cdot z_3 = z_1 \cdot z_3 + z_2 \cdot z_3$ for any complex numbers z_1, z_2 and z_3.

It follows from the results of problems 202–205 that the complex numbers together with the operations of addition and multiplication defined in (2.1) and (2.2) form a field. This is called the *field of complex numbers*.

For complex numbers of the form $(a, 0)$, where a is an arbitrary real number, equations (2.1) and (2.2) give

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

Thus, if we make a correspondence between each complex number of the form $(a, 0)$ and the real number a, then the operations on numbers of the form $(a, 0)$ will correspond to the usual operations on real numbers. Therefore, we will simply identify the complex number $(a, 0)$ with the real number a,[†] and we will say that the field of complex numbers contains the field of real numbers.

[†]In the same way, for example, that the rational number $n/1$ is identified with the integer n.

The complex number $(0, 1)$ is not real (with our identification), and we will denote it by i, i.e., $i = (0, 1)$. Because the field of complex numbers contains all real numbers and the number i, it will also contain all numbers of the form $b \cdot i$ and $a + b \cdot i$, where a and b are arbitrary real numbers, and the operations of addition and multiplication are understood as operations on complex numbers.

206. Let (a, b) be a complex number. Prove that

$$(a, b) = a + b \cdot i.$$

From the result of problem 206, it is clear that we have $a + bi = c + di$ if and only if $a = c$ and $b = d$.

In this way, any complex number can be expressed in a unique way in the form $a + bi$, where a and b are real numbers. If $z = a + bi$, then, following historical tradition, we call a the *real part* of the complex number z, we call bi the *imaginary part*, and b is called the *coefficient of the imaginary part*.

We call the expression of a complex number z in the form $z = a+bi$ the *algebraic form* of the complex number z.

For complex numbers in algebraic form, formulæ (2.1) and (2.2) can be written in the following way:

$$(a + bi) + (c + di) = (a + c) + (b + d)i, \tag{2.3}$$

$$(a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i. \tag{2.4}$$

207. Solve the equation (find the formula for a difference):

$$(a + bi) + (x + yi) = (c + di).$$

208. Solve the equation (find the formula for a quotient):

$$(a + bi) \cdot (x + yi) = (c + di), \quad \text{where } a + bi \neq 0.$$

It is easy to check that

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

i.e., $i^2 = -1$. Hence, in the field of complex numbers, we can take square roots of some negative real numbers.

209. Calculate: (a) i^3, (b) i^4, (c) i^n.

210. Find all complex numbers $z = x + yi$ such that: (a) $z^2 = 1$, (b) $z^2 = -1$, (c) $z^2 = a^2$, (d) $z^2 = -a^2$ (where a is some real number).

Definition 23. The complex number $a - bi$ is called the *complex conjugate* of the complex number $z = a + bi$, and is denoted by $\bar{z}$. It is easy to check that

$$z + \bar{z} = 2a, \; z \cdot \bar{z} = a^2 + b^2.$$

211. Let z_1 and z_2 be arbitrary complex numbers. Prove the following identities: (a) $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$, (b) $\overline{z_1 - z_2} = \bar{z}_1 - \bar{z}_2$, (c) $\overline{z_1 \cdot z_2} = \bar{z}_1 \cdot \bar{z}_2$, (d) $\overline{z_1/z_2} = \bar{z}_1/\bar{z}_2$ if z_2 is non-zero.

212. Let

$$P(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0,$$

with z a complex number and all the a_i real numbers. Prove that $\overline{P(z)} = P(\bar{z})$.

Moving to the complex numbers is the next step in the sequence: positive integers–integers–rational numbers–real numbers. However, the reader may be of the opinion that real numbers are in fact numbers, while complex numbers are not numbers but objects of more a complicated nature. Of course, any terminology can be accepted, but in reality, complex numbers do deserve to be called numbers.

The first objection may be that they are not numbers, but pairs of numbers. Remember, however, that this is the same way that the rational numbers are formally introduced. A rational number is a class of equivalent fractions, and fractions are pairs of integers written in the form m/n (where $n \neq 0$), where the operations on these rational numbers are simply operations on pairs of integers. Therefore this first objection is not well-founded. Another objection may be that numbers are something with which we can measure. If we take this to mean that numbers are what we can use to measure anything we want, then we should forbid, for example, negative numbers, because there are no segments of length -3 cm, and a train cannot travel for -4 days. If we consider numbers as something that can be used (or are convenient) for measuring at least something, then complex numbers turn out to be no worse than other numbers; they can be used to conveniently write down, for example, the current and resistance in electrical circuits of alternating current, and this technique is widely used in electronics.

Thus, in moving from real numbers to complex numbers is as natural as, for example, the transition from integers to rationals.

§3. The uniqueness of the field of complex numbers

We now move on to consider the question of why we defined the complex numbers in exactly the way we did, and not in another way. The answer to the question is that we wanted to get a field that was an extension of the field of real numbers. Is it possible to construct a different field which is also an extension of the field of real numbers? We will answer this very question in this section.

Definition 24. An *isomorphic map* (or simply an *isomorphism*) from one field to another is a bijective map φ which is an isomorphism with respect to both addition and multiplication, i.e., $\varphi(a + b) = \varphi(a) + \varphi(b)$ and $\varphi(ab) = \varphi(a)\varphi(b)$ for all a, b in the first field. Fields between which we can define an isomorphism are called *isomorphic*.

If we only study the operations of addition and multiplication in the field, then in isomorphic fields, all essential properties turn out to be the same. Therefore, in the same way as with groups, we will not distinguish between isomorphic fields.

As we saw in the previous section, in the field of complex numbers there is an element i such that $i^2 = -1$. The following problem shows that adjoining such an element to the field of real numbers necessarily gives the field of complex numbers.

213. Suppose that M is a field containing both the field of real numbers and an element i_0 satisfying $i_0^2 = -1$. Prove that M contains a field M' that is isomorphic to the field of complex numbers.

We will say that a field is a *minimal field with given properties* if it has the given properties and does not contain any smaller field with the same properties.

For example, the result of problem 213 can be stated the following way: the minimal field that contains the field of real numbers and an element i_0 such that $i_0^2 = -1$ is the field of complex numbers. This proves, in some sense, the uniqueness of the field of complex numbers. However, there exists a stronger result. Namely, let's remove the requirement that M needs to contain an element i_0 such that $i_0^2 = -1$ and pose the problem of finding all fields that are minimal non-trivial extensions of the field of real numbers. It turns out that there are only two such extensions (up to isomorphism), and one of them is the field of complex numbers. We now show this.

Suppose that a field M contains the field of real numbers, i.e., M contains all real numbers and operations on them in the field M coincide with the usual operations on real numbers. Let the field M also contain an element j, distinct from all real numbers. Then for any real numbers a_{n-1}, a_{n-2}, $\dots$, a_0 there exists an element in M equal to

$$j^n + a_{n-1}j^{n-1} + a_{n-2}j^{n-2} + \cdots + a_0. \tag{3.1}$$

We will call n the *degree* of the expression (3.1).

There are two possible situations:

(1) there exists an expression of the form (3.1) with $n \geqslant 1$ that gives an element equal to 0;
(2) there does not exist an expression of the form (3.1) with $n \geqslant 1$ that gives an element equal to 0.

Suppose first that we have situation (1).

Definition 25. A polynomial with coefficients from a field K is called *reducible over the field K* if it can be expressed as a product of two polynomials of smaller degree with coefficients in K. In the other case it is called *irreducible over the field K*.[†]

For example, the polynomials $x^3 - 1$ and $x^2 - x - 1$ are reducible over the field of real numbers, because $x^3 - 1 = (x-1)(x^2+x+1)$ and $x^2 - x - 1 = \left(x - \frac{1+\sqrt{5}}{2}\right)\left(x - \frac{1-\sqrt{5}}{2}\right)$, whereas the polynomials $x^2 + 1$ and $x^2 + x + 1$ are irreducible over the field of real numbers. Clearly, polynomials of degree one over any field are irreducible.

[†]Irreducible polynomials over the field K are analogous to prime numbers in the set of integers.

214. We can select, from among all expressions of the form (3.1) which are equal to 0, an expression with the smallest degree n ($n \geqslant 1$). Suppose this expression is

$$j^n + a_{n-1}j^{n-1} + a_{n-2}j^{n-2} + \cdots + a_0 = 0.$$

Prove that the polynomial

$$x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_0$$

is irreducible over the field of real numbers.

Later, we will show that any polynomial of degree greater than two with real coefficients is reducible over the field of real numbers (see problem 272). Therefore the smallest degree n in problem 214 has to be no greater than 2. And since $n \neq 1$ (or we would have $j + a = 0$, so that j is equal to the real number $-a$, contrary to our assumption), we must have $n = 2$.

Thus, in situation (1), the following equality has to be true for some real numbers p and q in the field M:

$$j^2 + pj + q = 0,$$

where the polynomial $x^2 + px + q$ is irreducible over the field of real numbers.

215. Prove that in situation (1), the field M contains an element i_0 such that $i_0^2 = -1$.

From the results of problems 215 and 213, it follows that in situation (1), the field M contains a field M', isomorphic to the field of complex numbers. Therefore, if the field M is a minimal non-trivial extension of the field of real numbers, then the field M must coincide with the field M'. Thus, in situation (1), any minimal field which is a non-trivial extension of the field of real numbers coincides with (i.e., is isomorphic to) the field of complex numbers. So, in situation (1), there exists a unique field, up to isomorphism, which is a minimal non-trivial extension of the field of real numbers, namely the field of complex numbers.

216. The *field of rational functions* in the variable x over the reals consists of all elements of the form $P(x)/Q(x)$, where $P(x)$ and $Q(x)$ are polynomials over the reals, with $Q(x) \neq 0$. Prove that if a field M is an extension of the field of real numbers, and M is in situation (2) above, then M contains a subfield isomorphic to the field of rational functions in the variable x over the reals.

§4. Geometric representation of complex numbers

We introduce a rectangular coordinate system XOY, and associate with each complex number $a + bi$ the point (a, b) in the coordinate plane. We thus get a bijective map between all complex numbers and all points of the plane. This gives us the *first geometric representation* of the complex numbers.

217. Which complex numbers correspond to the points in figure 13?

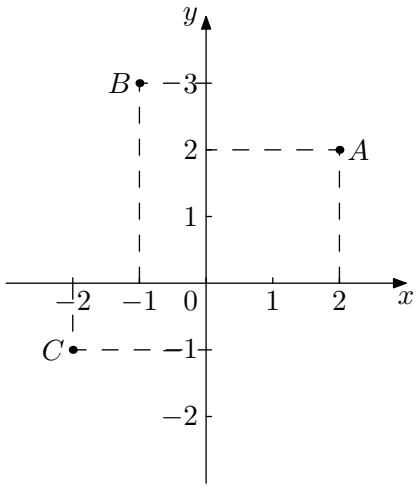


FIGURE 13.

218. Let complex numbers be represented by points in the plane. What is the geometric meaning of the map φ, if for any complex number z: (a) $\varphi(z) = -z$, (b) $\varphi(z) = 2z$, (c) $\varphi(z) = \bar{z}$ (where $\bar{z}$ is the complex conjugate of z).

Let $A = (x_A, y_A)$ and $B = (x_B, y_B)$ be two points in the plane (see figure 14). The segment AB with the noted direction from A to B is called the *vector* $\overrightarrow{AB}$. The

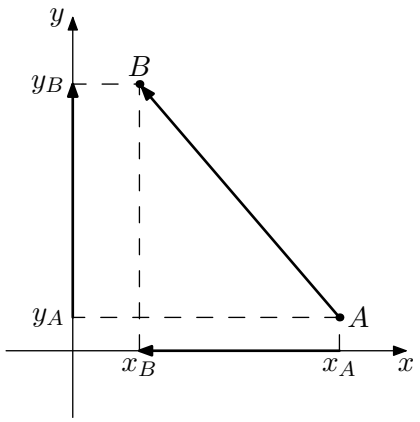


FIGURE 14.

coordinates of the vector $\overrightarrow{AB}$ are, by definition, computed in the following way:

$$x_{\overrightarrow{AB}} = x_B - x_A, \quad y_{\overrightarrow{AB}} = y_B - y_A.$$

Two vectors are considered equal if they are parallel, pointing in the same direction and equal in length.

219. Prove that two vectors are equal if and only if their coordinates are equal.

Equal vectors are usually considered to be the same one vector, which is characterised only by its coordinates—this is the so called *free vector*. If we create a correspondence such that to each complex number $a + bi$ we associate a free vector with coordinates (a, b), we get the *second geometric representation* of the complex numbers.

220. Suppose that the complex numbers z_1, z_2 and z_3 correspond to the free vectors u, v and w. Prove that $z_1 + z_2 = z_3$ if and only if $u + v = w$, where the sum of vectors is calculated using the parallelogram rule.

221. Prove the following relationship between the two geometric representations of the complex numbers: if z_A, z_B, and $z_{\overrightarrow{AB}}$ are the complex numbers corresponding to the points A and B and the vector $\overrightarrow{AB}$, then $z_{\overrightarrow{AB}} = z_B - z_A$.

From the definition of equal vectors, we see that equal vectors have the same length. This length is also taken to be the length of the corresponding free vector.

Definition 26. The *modulus* of a complex number z (denoted by $|z|$) is defined to be the length of the corresponding free vector.[†]

222. Suppose that $z = a + bi$. Prove that

$$|z|^2 = a^2 + b^2 = z \cdot \bar{z}.$$

223. Prove the following inequalities:

$$\text{(a) } |z_1 + z_2| \leqslant |z_1| + |z_2|,$$
$$\text{(b) } |z_1 - z_2| \geqslant \big||z_1| - |z_2|\big|,$$

where z_1 and z_2 are arbitrary complex numbers. In what cases does equality hold?

224. Prove, using the complex numbers, that in an arbitrary parallelogram, the sum of the squares of the lengths of the diagonals is equal to the sum of the squares of the lengths of all of the sides.

§5. Trigonometric representation of complex numbers

Recall that the angle between the rays OA and OB is the angle by which we need to rotate the ray OA around the point O anticlockwise to get the ray OB. (If the rotation is clockwise, it is denoted by a "minus" sign.) Note that the angle is not defined uniquely, but only up to addition of $2k\pi$, where k is any integer.

[†]For real numbers (as a special case of complex numbers), the concept introduced here coincides with the concept of the *absolute value*. In fact, the real number $a + 0i$ corresponds to the vector with coordinates $(a, 0)$, parallel to the x-axis, and its length is equal to $|a|$, the absolute value of the number a.

Definition 27. Let the point O be the origin and the vector OA with coordinates (a, b) correspond to the complex number $z = a + bi$ (see figure 15). The *argument* of the complex number z (denoted by $\operatorname{Arg} z$) is the angle between the positive direction of the axis OX and the ray OA (angle φ in figure 15). If $z = 0$, then $\operatorname{Arg} z$ is undefined.

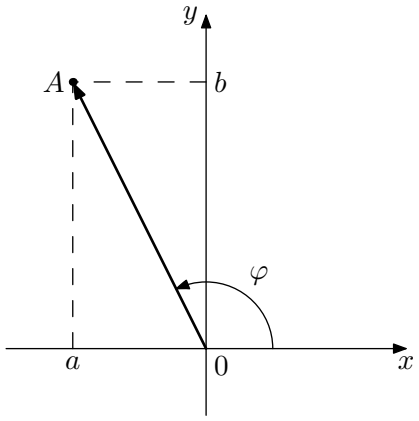


FIGURE 15.

Since, for a given number $z \neq 0$, the angle is not defined uniquely, when we write $\operatorname{Arg} z$, we will mean a multi-valued function that for each $z \neq 0$ takes an infinite set of values, the difference between each of which is a multiple of 2π. By writing $\operatorname{Arg} z = \varphi$, we will mean that one of the values of the argument of z is equal to φ.

Suppose that $z = a + bi \neq 0$ and let $r = |z|$. The vector $\overrightarrow{OA}$ with coordinates (a, b) corresponds to the complex number $a + bi$, and therefore its length is equal to r. Suppose too that $\operatorname{Arg} z = \varphi$. Then, by definition of the trigonometric functions (see figure 15), we have

$$\cos\varphi = \frac{a}{r}, \quad \sin\varphi = \frac{b}{r}.$$

From this we get

$$z = a + bi = r\cos\varphi + ir\sin\varphi = r(\cos\varphi + i\sin\varphi),$$

where $r = |z|$ and $\varphi = \operatorname{Arg} z$. This is called the *trigonometric representation* of the complex number z.

For example, if $z = -1 + \sqrt{3}\,i$, then $|z| = \sqrt{1+3} = 2$ (see problem 222) and $\cos\varphi = -\frac{1}{2}$, $\sin\varphi = \frac{\sqrt{3}}{2}$. We can take $\varphi = \frac{2\pi}{3}$, so $z = -1 + \sqrt{3}\,i = 2(\cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3})$.

225. Represent the following complex numbers using their trigonometric form: (a) $1+i$, (b) $-\sqrt{3} - i$, (c) $3i$, (d) -5, (e) $1 + 2i$.

226. Let $z_1 = r_1(\cos\varphi_1 + i\sin\varphi_1)$ and $z_2 = r_2(\cos\varphi_2 + i\sin\varphi_2)$. Prove that

$$z_1 z_2 = r_1 r_2(\cos(\varphi_1+\varphi_2) + i\sin(\varphi_1+\varphi_2)),$$
$$\frac{z_1}{z_2} = \frac{r_1}{r_2}(\cos(\varphi_1-\varphi_2) + i\sin(\varphi_1-\varphi_2)) \quad \text{if } z_2 \neq 0.$$

So, when multiplying complex numbers, their moduli are multiplied and their arguments are added; when dividing, the moduli are divided and the arguments subtracted, that is, $|z_1 z_2| = |z_1|\cdot|z_2|$, $\operatorname{Arg}(z_1 z_2) = \operatorname{Arg} z_1 + \operatorname{Arg} z_2$, and so on.

227. Prove *De Moivre's formula*[†]:

$$(r(\cos\varphi + i\sin\varphi))^n = r^n(\cos n\varphi + i\sin n\varphi)$$

for any positive integer n.

228. Calculate $(1-\sqrt{3}\,i)^{100}/2^{100}$.

229. Let $z = r(\cos\varphi + i\sin\varphi)$ be a fixed non-zero complex number and n a positive integer. Find all complex numbers w satisfying the equality

$$w^n = z. \tag{5.1}$$

Definition 28. By $\sqrt[n]{z}$ (the nth degree root of z), we mean a multi-valued function that puts each complex number $z \neq 0$ in correspondence with all n solutions to equation (5.1). When $z = 0$, we let $\sqrt[n]{0} = 0$.

230. Find all values of the following: (a) $\sqrt{-1}$, (b) $\sqrt[3]{8}$,(c) $\sqrt[4]{\cos 100^\circ + i\sin 100^\circ}$, (d) $\sqrt[3]{1+i}$.

It is convenient to introduce the following notation for subsequent use:

$$\varepsilon_n = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}.$$

231. Prove that all values of $\sqrt[n]{1}$ are 1, ε_n, ε_n^2, $\ldots$, ε_n^{n-1}.

Note. Because $\varepsilon_n^n = 1$, the set of elements 1, ε_n, ε_n^2, $\ldots$, ε_n^{n-1} forms a cyclic group under multiplication.

232. Let z_1 be one of the values of $\sqrt[n]{z_0}$. Find all values of $\sqrt[n]{z_0}$.

Henceforth, we will usually use the representation of complex numbers as points in the plane, i.e., the complex number $z = a + bi$ will correspond to the point with coordinates (a, b). Also, instead of saying "the point which corresponds to the complex number z", we will simply say "the point z".

233. Let the complex numbers be represented by points in the plane. What are the geometric meanings of the expressions: (a) $|z|$, (b) $\operatorname{Arg} z$, (c) $|z_1 - z_2|$, (d) $\operatorname{Arg}(z_1/z_2)$?

[†]De Moivre (1667–1754), an English mathematician.

234. Find the locus of points z satisfying the following conditions, where z_0, z_1 and z_2 are fixed complex numbers, and R is a fixed positive real number:
(a) $|z| = 1$, (b) $|z| = R$, (c) $|z - z_0| = R$, (d) $|z - z_0| \leqslant R$, (e) $|z - z_1| = |z - z_2|$, (f) $\operatorname{Arg} z = \pi$, (g) $\operatorname{Arg} z = \frac{9\pi}{4}$, (h) $\operatorname{Arg} z = \varphi$.

235. How are the values of $\sqrt[n]{z}$ positioned in the plane, where z is a fixed complex number?

§6. Continuity

In what follows, the concept of continuity, and in particular the concept of a continuous curve, will play an important role for us. If the reader does not know the formal definitions of these concepts, then he or she will probably still intuitively understand what a continuous curve is, as well as what a continuous function of a real variable is. (On an intuitive level, we can say that a continuous function is a function whose graph is a continuous curve.) However, if a function of a real variable is sufficiently complicated (for example, consider $f(x) = (x^3 - 2x)/(x^2 - \sin x + 1)$), it can be rather difficult to determine whether it is continuous or not simply by using the intuitive concept of continuity. Therefore we will give a rigorous definition of continuity, and use it to prove several basic results about continuous functions. We will give the definition of continuity not only for functions of a real variable, but also for functions of a complex variable.

If we look at the graph of a function of a real variable, then this graph can be continuous at some points and yet have discontinuities at some other points. Therefore, it is natural to first introduce a definition, not of continuity of a function, but of continuity at a point.

If we try to describe more precisely our intuitive understanding of the continuity of a function $f(x)$ at a given point x_0, we will see that continuity means the following: with only small changes to the variable x near the point x_0, the function also changes by only a relatively small amount around the value $f(x_0)$. Note that we can make the change in the function as small as we want by choosing a sufficiently small interval of changing the variable x around x_0. More formally, this can be stated in the following way.

Definition 29. Let $f(z)$ be a function of a real or complex variable z. We say that the function $f(z)$ is *continuous at the point* z_0 if for any real number $\varepsilon > 0$, we can find a real number $\delta > 0$ (depending on z_0 and ε) such that, for all numbers z satisfying the condition $|z - z_0| < \delta$, the inequality $|f(z) - f(z_0)| < \varepsilon$ holds.[†]

Example 15. We shall prove that the function of a complex variable $f(z) = 2z$ is continuous at any point z_0. Suppose we are given a point z_0 and an arbitrary real

[†]For the geometric meaning of the inequalities $|z - z_0| < \delta$ and $|f(z) - f(z_0)| < \varepsilon$, see problems 233(c) and 234(d).

number $\varepsilon > 0$. We need to pick a real number $\delta > 0$ such that, for all numbers z satisfying the condition $|z - z_0| < \delta$, the inequality $|f(z) - f(z_0)| = |2z - 2z_0| < \varepsilon$ holds. It is not difficult to see that we can pick $\delta = \varepsilon/2$ (independently of the point z_0). Indeed, from the condition $|z - z_0| < \delta$, it follows that:

$$|2z - 2z_0| = |2(z - z_0)| = |2| \cdot |z - z_0| < 2\delta = \varepsilon,$$

using 226, so $|2z-2z_0| < \varepsilon$. Thus the function $f(z) = 2z$ is continuous at any point z_0. In particular, it is continuous for all real values of the variable z. Therefore, if we restrict ourselves to real values of the variable, we see that the function of a real variable $f(x) = 2x$ is continuous at all real values of x.

236. Let a be a fixed complex (or, as a special case, real) number. Prove that the constant function of a complex (or real) variable, $f(z) = a$, is continuous for all values of the variable z.

237. Prove that the function of a complex variable $f(z) = z$ and the function of a real variable $f(x) = x$ are continuous for all values of the variable.

238. Prove that the function of a complex variable $f(z) = z^2$ is continuous for all values of the variable.

Definition 30. Let $f(z)$ and $g(z)$ be two functions of a complex (or real) variable. The function of complex (or real) variable $h(z)$ is the *sum* of the functions $f(z)$ and $g(z)$ if for every z_0, the equality $h(z_0) = f(z_0)+g(z_0)$ holds. Also, if the value of $f(z_0)$ or of $g(z_0)$ is undefined, then the value of $h(z_0)$ is also undefined. In the same way, we can define the *difference*, *product* and *quotient* of two functions.

239. Suppose that the functions $f(z)$ and $g(z)$ of a complex or real variable are continuous at the point z_0. Prove that at the point z_0, the following functions are also continuous: (a) $h(z) = f(z) + g(z)$, (b) $h(z) = f(z) - g(z)$, (c) $h(z) = f(z)g(z)$.

From the result of problem 239(c) we see, in particular, that if the function $f(z)$ is continuous at a point z_0, and n is a positive integer, then the function $(f(z))^n$ is also continuous at the point z_0.

240. Suppose that the functions $f(z)$ and $g(z)$ of a complex or real variable are continuous at the point z_0, and that $g(z_0) \neq 0$. Prove that at the point z_0, the following functions are also continuous: (a) $h(z) = 1/g(z)$, (b) $h(z) = f(z)/g(z)$.

Definition 31. Let $f(z)$ and $g(z)$ be two functions of a complex or real variable. The function $h(z)$ is the *composition* of the functions $f(z)$ and $g(z)$ if at every point z_0, the equality $h(z_0) = f(g(z_0))$ holds. Also, if $g(z_0)$ is undefined or if the function $f(z)$ is not defined at the point $g(z_0)$, then $h(z_0)$ is also not defined.

241. Let $f(z)$ and $g(z)$ be functions of a complex or real variable. Suppose that $g(z_0) = z_1$, that the function $g(z)$ is continuous at the point z_0, and that the function $f(z)$ is continuous at the point z_1. Prove that the function $h(z) = f(g(z))$ is continuous at the point z_0.

From the results of problems 239–241, it follows in particular that if we construct an expression from several functions of a complex (or real) variable, each of which is continuous for all values of the variable, using the operations of addition, subtraction, multiplication, division, raising to a positive integer power and composition, then the resulting expression will be a function which is continuous everywhere that none of the denominators is zero.

For example, considering the results of problems 236 and 237, we see that the functions $f(z) = z^n$, $f(z) = az^n$, and, more generally, $f(z) = a_n z^n + a_{n-1} z_{n-1} + \cdots + a_0$, are continuous functions of the variable z for any complex numbers a, a_n, a_{n-1}, …, a_0.

242. Prove that the functions of real variables $f(x) = \sin x$ and $f(x) = \cos x$ are continuous for all values of the variable x.

243. Consider the function $f(x) = \sqrt[n]{x}$ for all real values $x \geqslant 0$, where n is some positive integer, and $\sqrt[n]{x}$ is taken to be the non-negative nth degree root. Prove that this function is continuous for all $x > 0$.

When studying continuity, we often encounter statements that intuitively seem completely obvious, but for which constructing a formal proof presents great technical difficulties and requires a more formal definition of the real numbers than seen in high school, as well as a study of the basics of set theory and topology. One example of such a statement is this: if a function of a real variable $f(x)$ is continuous on some interval and on that interval it only takes integer values, then on that interval it takes only one value. Intuitively, it seems obvious that when the point x is moving along the interval, the value of the function $f(x)$ must change continuously and cannot "jump" from one integral value to another. However, it is rather difficult to prove this statement formally.

For this reason, in the sequel, we will often rely on the intuition of the reader, and assume without proof several "intuitively obvious" statements related to continuity. In particular, we will assume without proof the statement above that was used as an example. A formal proof of this statement can be found in standard analysis textbooks.

§7. Continuous curves

Let the variable t take real values on the interval $0 \leqslant t \leqslant 1$, and let every such value of t correspond to some complex number

$$z(t) = x(t) + iy(t).$$

The plane in which we illustrate the values of z will subsequently be called simply "the z-plane". If the functions $x(t)$ and $y(t)$ are continuous for $0 \leqslant t \leqslant 1$, then when t varies between 0 and 1, the point $z(t)$ will describe some *continuous curve* in the z-plane. We will consider this curve together with its direction, taking the point $z(0)$

to be the beginning of the curve and the point $z(1)$ to be its end. We call the function $z(t)$ a *parametric equation* of this curve.

Example 16. Let $z(t) = t + it^2$. Then $x(t) = t$ and $y(t) = t^2$. Therefore, $y(t) = (x(t))^2$ for every t, i.e., the point $z(t)$ lies on the parabola $y = x^2$ for every t. When t varies from 0 to 1, then $x(t)$ also varies from 0 to 1, and the point $z(t)$ runs along the parabola $y = x^2$ from the point $z(0) = 0$ to the point $z(1) = 1 + i$ (see figure 16).

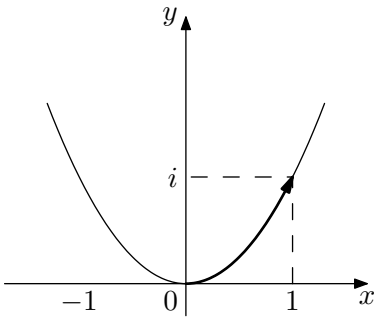


FIGURE 16.

244. Construct the curves that are given by the following parametric equations in the z-plane: (a) $z(t) = 2t$; (b) $z(t) = it$; (c) $z(t) = it^2$; (d) $z(t) = t - it$; (e) $z(t) = t^2 + it$; (f) $z(t) = R(\cos 2\pi t + i \sin 2\pi t)$; (g) $z(t) = R(\cos 4\pi t + i \sin 4\pi t)$; (h) $z(t) = R(\cos \pi t + i \sin \pi t)$;

$$\text{(i) } z(t) = \begin{cases} \cos 2\pi t + i \sin 2\pi t & \text{when } 0 \leqslant t \leqslant \frac{1}{2}, \\ 4t - 3 & \text{when } \frac{1}{2} < t \leqslant 1. \end{cases}$$

245. Write a parametric equation for the line segment connecting $z_0 = a_0 + b_0 i$ and $z_1 = a_1 + b_1 i$.

Note. In subsequent problems, some of the parametric equations have indices. These indices simply denote the number of the curve, but we consider all of these curves to be in the same z-plane.

246. What geometric transformations need to be performed on the curve C_1 satisfying the equation $z_1(t)$ to get the curve C_2 satisfying the equation $z_2(t)$ if:

(a) $z_2(t) = z_1(t) + z_0$, where z_0 is a fixed complex number;
(b) $z_2(t) = a \cdot z_1(t)$, where a is a positive real number;
(c) $z_2(t) = z_0 \cdot z_1(t)$, where $|z_0| = 1$;
(d) $z_2(t) = z_0 \cdot z_1(t)$, where z_0 is an arbitrary complex number?

247. Suppose that $z_1(t)$ is a parametric equation of the curve C. What curve is described by the equation $z_2(t)$, if $z_2(t) = z_1(1 - t)$?

248. Suppose that $z_1(t)$ and $z_2(t)$ are the parametric equations of the curves C_1 and C_2, and that $z_1(1) = z_2(0)$. What curve is described by the equation $z_3(t)$, if

$$z_3(t) = \begin{cases} z_1(2t) & \text{when } 0 \leqslant t \leqslant \frac{1}{2}, \\ z_2(2t-1) & \text{when } \frac{1}{2} < t \leqslant 1? \end{cases}$$

249. Let $z(t) = \cos \pi t + i \sin \pi t$ (see figure 17). For each t, find all values of $\operatorname{Arg} z(t)$ in terms of t.

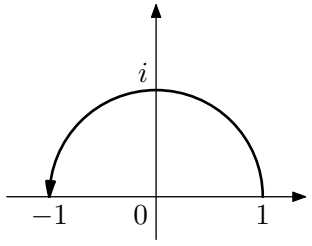


FIGURE 17.

250. Let $z(t) = \cos \pi t + i \sin \pi t$. How should we pick one value for $\operatorname{Arg} z(t)$ for each t so that the selected values vary continuously with t varying from 0 to 1, given that $\operatorname{Arg} z(0)$ is picked to be: (a) 0, (b) 2π, (c) -4π, (d) $2\pi k$, where k is an integer?

The following statement seems intuitive enough that we will give it without proof.

Theorem 6. *Suppose that a continuous curve C with parametric equation $z(t)$ does not pass through the origin (i.e., $z(t) \neq 0$ for $0 \leqslant t \leqslant 1$), and let the argument of the starting point of the curve C (i.e., $\operatorname{Arg} z(0)$) be chosen to equal φ_0. Then we can pick one value for the argument for all points of the curve C so that while moving along the curve the argument changes continuously, starting from the value φ_0.*

In other words, there is a function $\varphi(t)$ so that $\varphi(t)$ is one of the values of $\operatorname{Arg} z(t)$ for every t, the function $\varphi(t)$ is continuous for $0 \leqslant t \leqslant 1$, and $\varphi(0) = \varphi_0$.[†]

251. Let $\varphi_1(t)$ and $\varphi_2(t)$ be two continuous functions describing the varying of $\operatorname{Arg} z(t)$ along the curve C. Prove that $\varphi_1(t) - \varphi_2(t) = 2\pi k$, where k is some fixed integer not depending on t.

252. Prove that if we pick some value $\varphi(0) = \varphi_0$, then the function $\varphi(t)$ describing a continuous varying of $\operatorname{Arg} z(t)$ along the curve C is determined uniquely.

253. Suppose that the function $\varphi(t)$ describes a continuous varying of $\operatorname{Arg} z(t)$. Prove that the function $\psi(t) = \varphi(t) - \varphi(0)$ is uniquely defined by $z(t)$ and does not depend on the choice of $\varphi(0)$.

[†]It is possible to give a formal definition of an angle swept out by a given curve. Using this concept, it is easy to obtain the statement of Theorem 6: we need only take $\varphi(t) = \varphi_0 + \varphi_1(t)$, where $\varphi_1(t)$ is the angle swept out by the part of the curve from $z(0)$ to $z(t)$.

From the statement of problem 253 it follows by taking $t = 1$ that on a given continuous curve C, not passing through the point $z = 0$, the value of $\varphi(1) - \varphi(0)$ is uniquely determined by the condition of that $\varphi(t)$ is continuous.

Definition 32. The value $\varphi(1) - \varphi(0)$ will be called the *change in argument along the curve* C.

254. What is the value of the change in argument along the curves with the following parametric equations:

(a) $z(t) = \cos \pi t + i \sin \pi t$,
(b) $z(t) = \cos 2\pi t + i \sin 2\pi t$,
(c) $z(t) = \cos 4\pi t + i \sin 4\pi t$,
(d) $z(t) = (1 - t) + it$?

255. What is the value of the change in argument along the curves in figure 18?

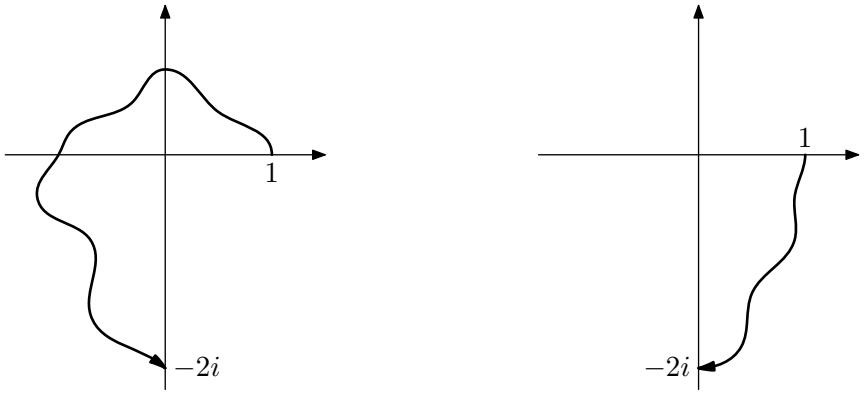


FIGURE 18.

If a continuous curve C is closed, i.e., $z(1) = z(0)$, then the value $\varphi(1) - \varphi(0)$ has the form $2\pi k$, where k is an integer, said to be the number of times the curve C goes around the point $z = 0$.

256. How many times do the following curves go around the point $z = 0$:

(a) $z(t) = 2 \cos 2\pi t + 2i \sin 2\pi t$ (see figure 19(a)),
(b) $z(t) = \frac{1}{2} \cos 4\pi t - \frac{1}{2} i \sin 4\pi t$ (see figure 19(b)),
(c) the curve in figure 19(c),
(d) the curve in figure 19(d)?

257. Prove that the number of times a closed curve goes around the point $z = 0$ does not depend on the choice of starting point, but only on the direction of the curve.

258. Suppose that the closed curve C with equation $z_1(t)$ goes k times around the point $z = 0$. How many times does the curve with equation $z_2(t)$ go around the

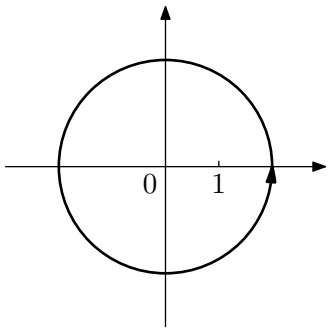


Figure (a)

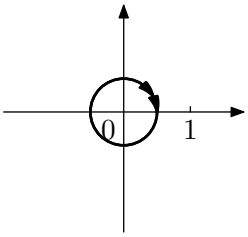


Figure (b)

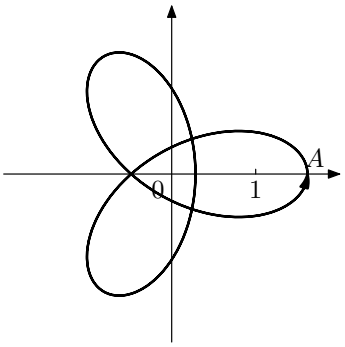


Figure (c)

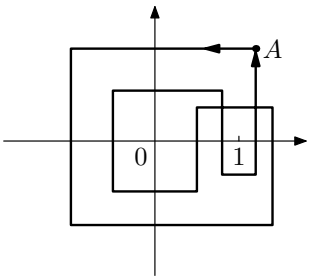


Figure (d)

FIGURE 19.

point $z = 0$ if: (a) $z_2(t) = 2z_1(t)$; (b) $z_2(t) = -z_1(t)$; (c) $z_2(t) = z_0 \cdot z_1(t)$, where $z_0 \neq 0$ is a fixed complex number; (d) $z_2(t) = \overline{z_1(t)}$?

Definition 33. Suppose that a continuous closed curve C with equation $z_1(t)$ does not pass through the point $z = z_0$. Then we say that the curve C goes k times around the point z_0 if the curve with equation $z_2(t) = z_1(t) - z_0$ goes k times around the point $z = 0$ (see figure 20).

Thus, to determine the number of times the curve goes around the point $z = z_0$, we simply have to observe the rotation of the vector $z_1(t) - z_0$, which can be viewed as the vector connecting the points z_0 and $z_1(t)$ (see problem 221).

259. How many times do the curves in problem 256 go around the point $z = 1$?

260. Let $z_1(t)$ and $z_2(t)$ be the equations of curves C_1 and C_2, which do not pass through the point $z = 0$. Let φ_1 and φ_2 be the respective changes in argument along

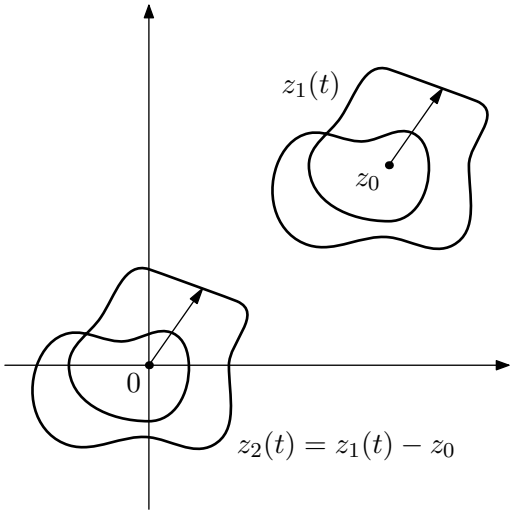


FIGURE 20.

these curves. What is the value of the change in argument along the curve C with equation $z(t)$ if (a) $z(t) = z_1(t)z_2(t)$, (b) $z(t) = z_1(t)/z_2(t)$?

§8. Maps of curves and the fundamental theorem of algebra

Suppose we are given two planes of complex numbers: a z-plane and a w-plane, and that we are given a function $w = f(z)$ which associates to each value of z a uniquely determined value w. If we have a continuous curve C in the z-plane with equation $z(t)$, then every point of this curve is sent to some point in the w-plane by the function $w = f(z)$. If this function $f(z)$ is continuous, then in the w-plane we will also get a continuous curve with equation $w_0(t) = f(z(t))$. This curve will be denoted by $f(C)$.

261. What does the curve $f(C)$ look like if $w = f(z) = z^2$ and the curve C is:

(a) a quarter circle:

$$z(t) = R\left(\cos\frac{\pi t}{2} + i\sin\frac{\pi t}{2}\right),$$

(b) a semicircle: $z(t) = R(\cos\pi t + i\sin\pi t)$,

(c) a circle: $z(t) = R(\cos 2\pi t + i\sin 2\pi t)$?

262. Suppose that the change in argument along the curve C is φ. What is the change in argument along the curve $f(C)$ if: (a) $f(z) = z^2$, (b) $f(z) = z^3$, (c) $f(z) = z^n$, where n is an arbitrary positive integer?

263. Suppose that the closed curve C goes k times around the point $z = z_0$. How many times does the curve $f(C)$ go around the point $w = 0$ if $f(z) = (z - z_0)^n$?

264. Suppose that the number of times that the closed curve C goes around the points $z = 0$, $z = 1$, $z = i$ and $z = -i$ is k_1, k_2, k_3 and k_4, respectively. How many times does the curve $f(C)$ go around the point $w = 0$ if: (a) $f(z) = z^2 - z$, (b) $f(z) = z^2 + 1$, (c) $f(z) = (z^2 + iz)^4$, (d) $f(z) = z^3 - z^2 + z - 1$?

Let us consider the equation

$$a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 = 0$$

with $n \geqslant 1$, where the a_i are arbitrary complex numbers and $a_n \neq 0$. Our next goal is to show that this equation has at least one complex root. If $a_0 = 0$, then the equation has the root $z = 0$. We will therefore assume that $a_0 \neq 0$.

Let us denote the greatest of the numbers $|a_0|$, $|a_1|$, $\ldots$, $|a_n|$ by A. Because $a_n \neq 0$, we have $A > 0$. Pick positive real numbers R_1 and R_2 with R_1 small enough for the two inequalities $R_1 \leqslant 1$ and $R_1 < |a_0|/10An$ to hold, and R_2 large enough for the two inequalities $R_2 \geqslant 1$ and $R_2 > 10An/|a_n|$ to hold.

265. Let $|z| = R_1$. Prove that

$$|a_n z^n + \cdots + a_1 z| < \frac{|a_0|}{10}.$$

266. Let $|z| = R_2$. Prove that

$$\left|\frac{a_{n-1}}{z} + \cdots + \frac{a_0}{z^n}\right| < \frac{|a_n|}{10}.$$

We will denote the curve with equation $z(t) = R(\cos 2\pi t + i \sin 2\pi t)$, (i.e., the circle with radius R drawn anticlockwise) by C_R. Because the curve C_R is closed ($z(1) = z(0)$), the curve $f(C_R)$, where $f(z) = a_n z^n + \cdots + a_0$, is also closed ($f(z(1)) = f(z(0))$). Let $\nu(R)$ be the number of times that the curve $f(C_R)$ goes around the point $w = 0$ if $f(C_R)$ does not pass through the point $w = 0$.

267. What are the values of $\nu(R_1)$ and $\nu(R_2)$?

Let us now change the radius R continuously from R_1 to R_2. The curve $f(C_R)$ will be continuously deformed from the curve $f(C_{R_1})$ to the curve $f(C_{R_2})$. If for some value of R the curve $f(C_R)$ does not pass through the point $w = 0$, then for sufficiently small changes of R, the curve $f(C_R)$ will be deformed so little that the number of times it goes around the point $w = 0$ will not change, i.e., for this value of R, the function $\nu(R)$ is continuous. If none of the curves $f(C_R)$ with $R_1 \leqslant R \leqslant R_2$ pass through the point $w = 0$, then $\nu(R)$ will be continuous for all $R_1 \leqslant R \leqslant R_2$. Because the function $\nu(R)$ only takes on integer values, it can only be continuous if for all values of R in the interval $R_1 \leqslant R \leqslant R_2$ it takes the same value; in particular, it follows that $\nu(R_1) = \nu(R_2)$. But from the solution to problem 267 it follows that $\nu(R_1) = 0$ while $\nu(R_2) = n$. Therefore the assumption that none of the curves $f(C_R)$

pass through the point $w = 0$ for $R_1 \leqslant R \leqslant R_2$ is false, so for some z it has to be true that $f(z) = 0$. So, we have deduced the following theorem[†]

Theorem 7 (The fundamental theorem of algebra[‡]). *Any equation*

$$a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 = 0,$$

where $n \geqslant 1$ and the a_i are arbitrary complex numbers with $a_0 \neq 0$, has at least one complex root.

268. Prove *Bézout's theorem*[§] that if z_0 is a root of the equation $a_n z^n + \cdots + a_1 z + a_0 = 0$, then the polynomial $a_n z^n + \cdots + a_1 z + a_0$ is divisible by $z - z_0$ without a remainder.

269. Prove that the polynomial $a_n z^n + \cdots + a_1 z + a_0$, where $a_n \neq 0$, can be expressed in the form $a_n z^n + \cdots + a_1 z + a_0 = a_n(z - z_1)(z - z_2) \cdots (z - z_n)$.

Note. Suppose that the polynomial $P(z)$ is factored into factors:

$$P(z) = a_n(z - z_1)(z - z_2) \cdots (z - z_n).$$

The right part is equal to 0 if and only if at least one of the factors is equal to 0 (see problems 195 and 197). Therefore the roots of the equation $P(z) = 0$ are precisely $z_1, z_2, \ldots, z_n$.

270. Let z_0 be a root of the equation

$$a_n z^n + \cdots + a_1 z + a_0 = 0,$$

where all of the a_i are real numbers. Prove that $\bar{z}_0$ is also a root of this equation.

271. Suppose that the equation with real coefficients

$$a_n z^n + \cdots + a_1 z + a_0 = 0$$

has a complex root z_0 which is not a real number. Prove that the polynomial $a_n z^n + \cdots + a_1 z + a_0$ is divisible by some polynomial of degree two with real coefficients.

272. Prove that any polynomial with real coefficients can be expressed as a product of polynomials of degree one and two with real coefficients.

Note. From the result of problem 272, it follows that the irreducible polynomials (see page 40) over the field of real numbers are precisely the polynomials of degree one and the polynomials of degree two which do not have real roots. We used this result earlier in section 3 of this chapter. Over the field of complex numbers, it follows from problem 269 that the only irreducible polynomials are those of degree one.

Let us return now to polynomials with arbitrary complex coefficients.

[†]Our discussion is somewhat informal and should be viewed only as an outline of a proof. This discussion can (although not easily) be made absolutely rigorous.

[‡]This theorem was proven in 1799 by the German mathematician Gauss (1777–1855).

[§]Bézout (1730–1783) was a French mathematician

Definition 34. Suppose that z_0 is a root of the equation

$$a_n z^n + \cdots + a_1 z + a_0 = 0.$$

We say that z_0 is a *root of multiplicity* k if the polynomial $a_n z^n + \cdots + a_1 z + a_0$ is divisible by $(z - z_0)^k$ but not divisible by $(z - z_0)^{k+1}$.

273. What are the multiplicities of the roots $z = 1$ and $z = -1$ in the equation

$$z^5 - z^4 - 2z^3 + 2z^2 + z - 1 = 0?$$

Definition 35. The *derivative* of the polynomial

$$P(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_k z^k + \cdots + a_1 z + a_0$$

is the polynomial

$$P'(z) = a_n n z^{n-1} + a_{n-1}(n-1) z^{n-2} + \cdots + a_k k z^{k-1} + \cdots + a_1.$$

The derivative is usually denoted with a prime.

274. Let $P(z)$ and $Q(z)$ be two polynomials. Prove the following equalities:

(a) $(P(z) + Q(z))' = P'(z) + Q'(z)$
(b) $(cP(Z))' = cP'(z)$, where c is an arbitrary complex number;
(c) $(P(z)Q(z))' = P'(z)Q(z) + P(z)Q'(z)$.

275. Let $P(z) = (z - z_0)^n$ with $n \geqslant 1$ an integer. Prove that $P'(z) = n(z - z_0)^{n-1}$.

276. Prove that if the equation $P(z) = 0$ has a root z_0 of multiplicity $k > 1$, then the equation $P'(z) = 0$ has the root z_0 with multiplicity $k - 1$, and if the equation $P(z) = 0$ has a root z_0 of multiplicity 1, then $P'(z_0) \neq 0$.

§9. The Riemann surface for $w = \sqrt{z}$

Earlier we looked at single-valued functions for which every value of the variable corresponded to only one value of the function. Henceforth, however, we will be especially interested in multi-valued functions for which some values of the variable correspond to several values of the function.[†] Our interest in these functions is easily explained. The final goal of this book is to prove Abel's theorem, that the function that describes the roots of the general fifth degree polynomial in terms of its coefficients cannot be expressed in terms of radicals. But this function is multi-valued, because a fifth degree polynomial with fixed coefficients has, in general, five roots. Functions that can be expressed in terms of radicals are also multi-valued.

The general idea behind our proof of Abel's theorem consists of the following. We will associate to a multi-valued function of a complex variable some group—the so-called *Galois*[‡] *group* of the function. It will turn out that Galois group of the

[†]In cases where it does not cause confusion, we will often omit the word "multi-valued".

[‡]Évariste Galois (1811–1832), a French mathematician who determined general conditions of solvability of equations in radicals and who established the fundamentals of group theory.

function that expresses the roots of the general fifth degree polynomial cannot be the Galois group of a function expressed in radicals, and therefore the function itself cannot be expressed in radicals.

To introduce the concept of a Galois group, we will first introduce another important concept from the theory of functions of a complex variable—the concept of a *Riemann*† *surface* for a multi-valued function. We will begin with the construction of a Riemann surface for one of the simplest examples of a multi-valued function, namely the function $w = \sqrt{z}$.

We know that the function $w = \sqrt{z}$ takes on only one value $w = 0$ at $z = 0$ and two values for all $z \neq 0$ (see problem 229). Note that if w_0 is one of the values of $\sqrt{z_0}$, then the other value of $\sqrt{z_0}$ is $-w_0$.

277. Find all values of: (a) $\sqrt{1}$, (b) $\sqrt{-1}$, (c) $\sqrt{i}$, (d) $\sqrt{1 + i\sqrt{3}}$ (where by $\sqrt{3}$ we mean the positive value of the root).

Let us make a cut in the z-plane along the negative part of the real axis from 0 to $-\infty$. For all z not on the cut, we will pick the value $w = \sqrt{z}$ that lies on the right half of the w-plane. In this way, we get a function that is single-valued and continuous on the entire z-plane excluding the cut. We will denote this function by ${}_1\sqrt{z}$. This function gives a single-valued and continuous map of the z-plane, excluding the cut, to the right half of the w-plane (see figure 21).

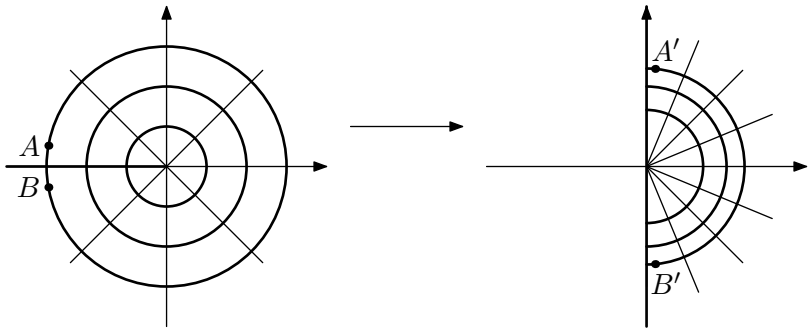


FIGURE 21.

Note. If we pick $\operatorname{Arg} z$ such that $-\pi < \operatorname{Arg} z < \pi$, then for the function ${}_1\sqrt{z}$ we will get $\operatorname{Arg} {}_1\sqrt{z} = \frac{1}{2} \operatorname{Arg} z$ (see problem 229). Therefore under the map $w = {}_1\sqrt{z}$, the z-plane is folded back as a fan towards the positive side of the real axis by decreasing the angle of the "fan" by a factor of two and by changing lengths along the rays of the "fan".

If we now pick the value of $w = \sqrt{z}$ that lies in the left half of the w-plane for all z that do not lie on the cut, then we will get another function, also continuous and single-valued on the entire z-plane excluding the cut. This function, which we will

†Named after B. Riemann (1826–1866), a German mathematician.

denote by ${}_2\sqrt{z}$, gives a single-valued and continuous map of the z-plane, excluding the cut, to the left half of the w-plane (see figure 22). Here ${}_2\sqrt{z} = -{}_1\sqrt{z}$.

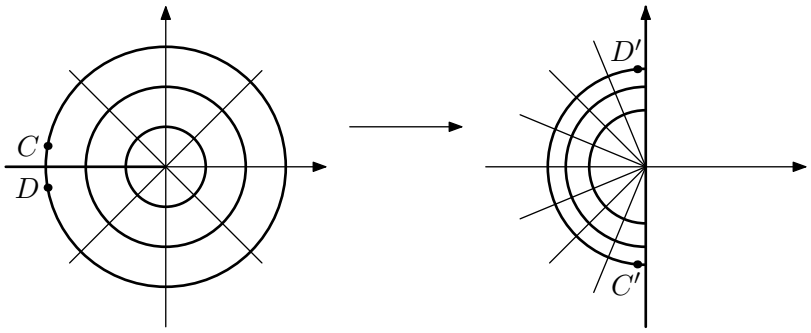


FIGURE 22.

The constructed functions ${}_1\sqrt{z}$ and ${}_2\sqrt{z}$ are called the *single-valued continuous branches* of the function $w = \sqrt{z}$ (with the given cut).

We now take two copies of the z-plane, which we will also call *sheets*, and on each we make a cut along the negative part of the real axis from 0 to $-\infty$ (see figure 23). On the first sheet we define the function ${}_1\sqrt{z}$ and on the second sheet the

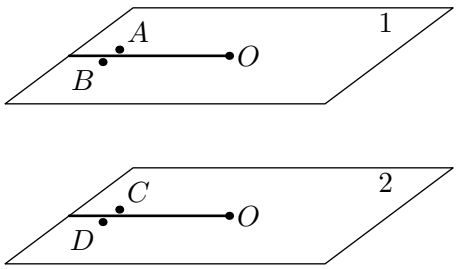


FIGURE 23.

function ${}_2\sqrt{z}$. Then we can view the functions ${}_1\sqrt{z}$ and ${}_2\sqrt{z}$ together as one single-valued function, defined not on the z-plane, but on a more complicated surface consisting of two separate sheets.

If a point z is moving continuously on the first sheet (or on the second sheet), not crossing the cut, then the single-valued function we constructed changes continuously. If, however, the point z moving on the first sheet goes across the cut, continuity is broken. This can be seen, for example, from the fact that the points A and B, which are close together in the z-plane, are sent under the map $w = {}_1\sqrt{z}$ to the two points A' and B', which are far apart from each other (see figure 21).

On the other hand, from figures 21 and 22, it is easy to see that the image A' of the point A under the map $w = {}_1\sqrt{z}$ turns out to be close to the image D' of the point D under the map $w = {}_2\sqrt{z}$.

Thus, if when crossing the cut, the point z goes from the upper side of the cut on one sheet to the lower side of the cut on the other sheet, then the single-valued function we constructed will change continuously. To guarantee the necessary movement of the point z, we will consider the upper side of the cut on the first sheet to be glued to the lower side of the cut on the second sheet, and the upper side of the cut on the second sheet to be glued to the lower side of the cut on the first sheet (see figure 24). Furthermore, during the gluing, between the edges of the cuts that

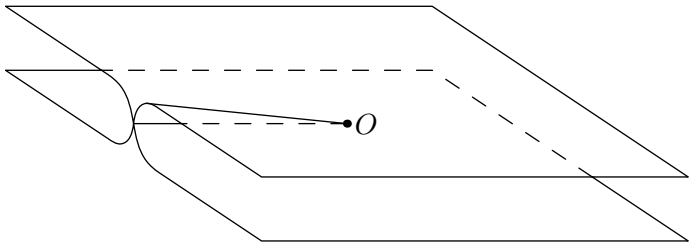


FIGURE 24.

are being glued we will add a ray of points from 0 to $-\infty$. During the first gluing, we will pick values of $w = \sqrt{z}$ that lie on the positive part of the imaginary axis, and during the second gluing, we pick values of $w = \sqrt{z}$ that lie on the negative part of the imaginary axis.

After performing the described gluing, what we find is that the double-valued function $w = \sqrt{z}$ is changed to another function which is single-valued and continuous, not on the z-plane, but on some other new, more complicated surface. We will call this surface the *Riemann surface* of the function $w = \sqrt{z}$.

Attempts to glue this surface in the way described without intersections (and without turning one plane over) are unsuccessful. Despite this, we consider figure 24 to be an illustration of the Riemann surface of the function $w = \sqrt{z}$, assuming an additional condition, namely that the intersection along the negative part of the real axis is not actually present. For comparison, consider the following example. In figure 7 (page 20), we have a drawing of the frame of a cube. Although some of the segments in the figure intersect, we can easily agree that these intersections are only present in the illustration, and this allows us to avoid mistakes.

The Riemann surface of an arbitrary multi-valued function $w(z)$ can be constructed similarly to the way we constructed the Riemann surface of the function $w = \sqrt{z}$. To do this, we first need to separate the single-valued continuous branches of the function $w(z)$, with some points z (the cuts) being excluded from consideration. After that, the branches are glued, restoring the cuts in the process, so that we get a single-valued continuous function on the constructed surface. We call the resulting surface the Riemann surface of the multi-valued function $w(z)$.[†]

[†]Such constructions cannot be performed for every multi-valued function, but for those functions which we will be considering in the sequel, such constructions really can be performed.

So it remains for us to determine how to separate the continuous single-valued branches of an arbitrary function $w(z)$ and how to glue them together afterwards. To obtain answers to these questions, let us look again in more detail at the function $w = \sqrt{z}$.

Let $w(z)$ be a multi-valued function, and suppose we fix one of the values w_0 of the function $w(z)$ at some point z_0. Let $w^*(z)$ be a continuous and single-valued branch of the function $w(z)$ defined on some region of the z-plane (for example, on the entire plane excluding some cuts),[†] and such that $w^*(z_0) = w_0$. Also, let C be a continuous curve, lying entirely within this region, going from the point z_0 to some other point z_1. Then when the point z moves along the curve C, the function $w^*(z)$ will change continuously from $w^*(z_0)$ to $w^*(z_1)$.

This property can also be used in reverse, namely to determine the function $w^*(z)$. For suppose that at some point z_0 we fix one of the values w_0 of $w(z)$, and we let C be a continuous curve going from the point z_0 to some point z_1. As we move along the curve C, we can pick a value of $w(z)$ for each point z on C such that the values change continuously as we move along the curve C, beginning with the value w_0. When we reach the point z_1, we will have a uniquely determined value $w_1 = w(z_1)$. We will say that w_1 is the value of $w(z_1)$ *determined by continuity along the curve* C given the condition that $w(z_0) = w_0$. If the chosen values of the function $w(z)$ along the curve C are plotted in the w-plane, then we will get a continuous curve which begins at the point w_0 and ends at the point w_1. This curve is one of the continuous images of the curve C under the map $w = w(z)$.

278. Suppose we pick the value $w(1) = \sqrt{1} = 1$ for the function $w(z) = \sqrt{z}$. Determine $w(-1) = \sqrt{-1}$ by continuity along: (a) the upper semicircle of radius 1 centred at the origin, (b) the lower semicircle (see figure 25).

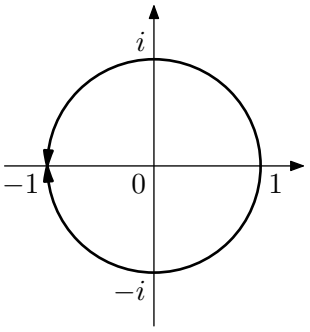


FIGURE 25.

It turns out that when defining functions by continuity along a curve, we can encounter some difficulties. Consider the following example.

[†]We also require this region to be *path-connected*, that is, given any two points in the region, there is a continuous curve between them which lies entirely in the region. This will always be true in the cases we will be considering.

279. Find all continuous images $w_0(t)$ of the curve C with parametric equation $z(t) = 2t - 1$ (figure 26) under the map $w = \sqrt{z}$, beginning: (a) at the point i, (b) at the point $-i$.

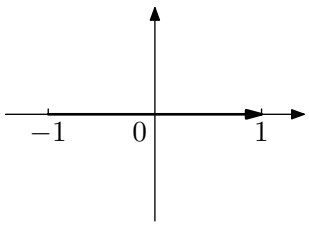


FIGURE 26.

From the solution to problem 279, we see that even with a fixed image of the beginning point of the curve C, the continuous image of the curve C under the map $w = \sqrt{z}$ may not be uniquely defined. The uniqueness is broken where the curve C passes through the point $z = 0$. It turns out that the only place where the uniqueness of the image of the function $w = \sqrt{z}$ is broken is where two images of the point $z(t)$ come close to each other, blending into one point.

To avoid the problem of non-uniqueness of continuous images of curves under the map $w = \sqrt{z}$, we can remove the point $z = 0$ and forbid curves from passing through this point. But even this restriction does not give us the ability to separate single-valued continuous branches of the function $w = \sqrt{z}$. For if we fix one of the values $w_0 = w(z_0)$ at some point z_0 and determine $w(z_1)$ at some other point z_1 by continuity of $w(z)$ along various curves from z_0 to z_1, we may still get different values for $w(z_1)$ (see, for example, problem 278). Let us see how we can avoid such non-uniqueness issues.

280. Suppose that the change in argument of $z(t)$ along the curve C is φ. Find the change in argument along any continuous image $w_0(t)$ of the curve C under the map $w(z) = \sqrt{z}$.

281. Let $w(z) = \sqrt{z}$ and suppose we pick $w(1) = \sqrt{1} = -1$. Find the value of $w(i) = \sqrt{i}$ by continuity along: (a) the line segment connecting $z = 1$ and $z = i$; (b) the curve with parametric equation $z(t) = \cos\frac{3}{2}\pi t - i\sin\frac{3}{2}\pi t$; (c) the curve with parametric equation $z(t) = \cos\frac{5}{2}\pi t + i\sin\frac{5}{2}\pi t$.

282. Let $w(z) = \sqrt{z}$ and let C be a closed curve beginning at $z = 1$. Suppose that we pick $w(1) = \sqrt{1} = 1$. Determine by continuity along the curve C the value of $w(1) = \sqrt{1}$ at the ending point of C, if the curve C has the equation: (a) $z(t) = \cos 2\pi t + i\sin 2\pi t$, (b) $z(t) = \cos 4\pi t - i\sin 4\pi t$, (c) $z(t) = 2 - \cos 2\pi t - i\sin 2\pi t$.

283. Let C be a closed curve in the z-plane which does not pass through the point $z = 0$. Prove that the value of the function $\sqrt{z}$ at the ending point of the curve C,

determined by continuity, will coincide with the value at the beginning point if and only if the curve C goes around the point $z = 0$ an even number of times.

For later use, it is convenient to introduce the following notation.

Definition 36. Let C be a continuous curve with parametric equation $z(t)$. We will denote the curve that geometrically coincides with C but goes in the opposite direction by C^{-1}, which has parametric equation $z_1(t) = z(1-t)$ (see problem 247).

Definition 37. Let the beginning point of the curve C_2 coincide with the ending point of the curve C_1. Then by C_1C_2 we will mean the curve we get if we first walk along C_1 and then along C_2 (see problem 248).

284. Suppose that C_1 and C_2 are two curves connecting the point z_0 to the point z_1, neither one passing through the point $z = 0$, and suppose we pick one of the values $w_0 = \sqrt{z_0}$. Prove that the values of $\sqrt{z_1}$ determined by continuity along the curves C_1 and C_2 will be the same if and only if the curve $C_1^{-1}C_2$ (see figure 27) goes around the point $z = 0$ an even number of times.

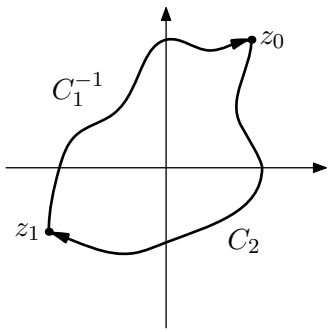


FIGURE 27.

From the statement of the last problem, it follows in particular that if the curve $C_1^{-1}C_2$ goes 0 times around the point $z = 0$, then the value of the function $\sqrt{z}$ at the ending points of the curves C_1 and C_2 determined by continuity will be the same if their beginning point values are the same.

Thus to separate single-valued continuous branches of the function $w = \sqrt{z}$, it is sufficient to ensure that the curve $C_1^{-1}C_2$ cannot go around the point $z = 0$ at all. For that, it is sufficient to make a cut from the point $z = 0$ to infinity and to forbid curves from intersecting the cut. This is exactly what we did above, making the cut from the point $z = 0$ to $-\infty$ along the negative part of the real axis.

If, after making the cut, we fix at some point z_0 not on the cut one of the values $w_0' = \sqrt{z_0}$, then the value at any other point z_1 will be determined by continuity along any curve C going from z_0 to z_1 and not passing through the cut. This determines a single-valued continuous branch ${}_1\sqrt{z}$ of the function $w = \sqrt{z}$. If at the point z_0 we now fix the other value $w_0'' = \sqrt{z_0}$, then from that we can similarly determine the other branch ${}_2\sqrt{z}$ of the function $w = \sqrt{z}$.

285. Prove that for any point z not on the cut, we have ${}_1\sqrt{z} \neq {}_2\sqrt{z}$

286. Let us fix at some point z' not on the cut the value $w' = {}_1\sqrt{z'}$ and determine the value of the function $w = \sqrt{z}$ at all other points of the z-plane (excluding the cut) by continuity of $\sqrt{z}$ along curves from the point z' and not going across the cut. Prove that we will get a single-valued continuous branch which coincides with the function ${}_1\sqrt{z}$ as defined above from the point z_0.

When we are trying to separate the single-valued continuous branches, it follows from the result of problem 286 that choosing as a beginning point different points of the z-plane will result in the same collection of single-valued continuous branches, depending only on where we put the cut.

287. Suppose that the points z_0 and z_1 do not lie on the cut, and suppose that a curve C connecting z_0 to z_1 crosses the cut exactly once (see figure 28). Suppose we pick a value $w_0 = \sqrt{z_0}$, and by continuity along the curve C we determine the value $w_1 = \sqrt{z_1}$. Prove that the values w_0 and w_1 correspond to different branches of the function $w = \sqrt{z}$.

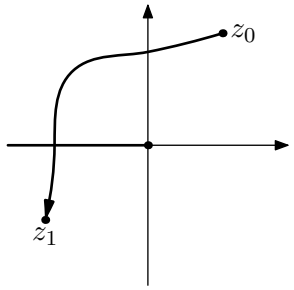


FIGURE 28.

Thus when crossing the cut, we go from one branch of the function $w = \sqrt{z}$ to the other branch, i.e., the branches connect in exactly the same way we connected them earlier (see figure 24). Hence we get the Riemann surface of the function $w = \sqrt{z}$ as before.

We will say that some property holds when walking around the point z_0 if it holds when we walk once anticlockwise around any circle centred at z_0 of small enough radius.[†]

288. Prove that when walking around the point z_0 we stay on the same sheet of the Riemann surface of the function $w = \sqrt{z}$ if $z_0 \neq 0$, and we move to the other sheet when $z_0 = 0$.

The following concept is very important in what follows.

[†]More formally, this means the following: we can find a real number $\delta > 0$ such that the mentioned property holds when we walk once around any circle centred at z_0 whose radius is less than δ.

Definition 38. If walking around a point moves us from one sheet to another (i.e., it changes the value of the function), then this point is called a *branch point* of the given multi-valued function.

The Riemann surface of the function $w = \sqrt{z}$ can be drawn as a schematic diagram, as in figure 29. The diagram shows that this Riemann surface has two

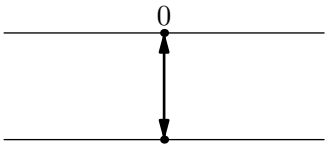


FIGURE 29.

sheets, that the point $z = 0$ is a branch point of the function $w = \sqrt{z}$, and that walking around the point $z = 0$ we go from the first sheet to the second sheet, and also walking around the same point takes us from the second sheet to the first. The arrows at the point $z = 0$ are used to indicate that we move from the first sheet to the second and back again not only when walking around the point $z = 0$, but also at any crossing of the cut going from the point $z = 0$ to infinity. We will see below that the connection between branch points and the cuts made starting at them is not coincidental. (Note that if there were three sheets and moving around $z = 0$ were to take us from the first to the second, from the second to the third and from the third back to the first, then we would draw single-headed arrows from the first to the second, from the second to the third and from the third back to the first.)

Henceforth, we will generally draw not the Riemann surface of a multi-valued function itself but its schematic diagram instead.

§10. Riemann surfaces for more complicated functions

Let us now look at at the multi-valued function $w = \sqrt[3]{z}$.

289. Assume that the curve C with parametric equation $z(t)$ does not pass through $z = 0$. Suppose that the change in the argument along the curve C is φ, and let $w_0(t)$ be a continuous image of the curve $z(t)$ under the map $w = \sqrt[3]{z}$. Find the change in argument along the curve $w_0(t)$.

290. Find the branch points of the function $w = \sqrt[3]{z}$.

291. Suppose a cut is made from the point $z = 0$ to $-\infty$ along the negative part of the real axis, and let the single-valued branches of the function $w = \sqrt[3]{z}$ be given by

the following conditions: $f_1(1) = 1$,

$$f_2(1) = \cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i,$$
$$f_3(1) = \cos\frac{4\pi}{3} + i\sin\frac{4\pi}{3} = -\frac{1}{2} - \frac{\sqrt{3}}{2}i.$$

Find: (a) $f_1(i)$, (b) $f_2(i)$, (c) $f_1(8)$, (d) $f_3(8)$, (e) $f_3(-i)$.

292. Construct the Riemann surface and its schematic diagram for the function $w = \sqrt[3]{z}$.

293. Let C be a continuous curve with parametric equation $z(t)$ and let w_0 be one of the values of $\sqrt[n]{z(0)}$. Prove that there is at least one continuous image of the curve C under the map $w(z) = \sqrt[n]{z}$ starting at the point w_0.

294. Assume that the curve $z(t)$ does not pass through the point $z = 0$, and let $w_0(t)$ be a continuous image of the curve $z(t)$ under the map $w = \sqrt[n]{z}$. Letting φ be the change in argument along the curve C, find the change in argument along the curve $w_0(t)$.

295. Find the branch points of the function $\sqrt[n]{z}$.

Earlier (see page 45), we introduced the notation

$$\varepsilon_n = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}.$$

There we also looked at some properties of this complex number.

296. Assume that the curve $z(t)$ does not pass through the point $z = 0$, and let $w_0(t)$ be one of the continuous images of the curve $z(t)$ under the map $w = \sqrt[n]{z}$. Find all continuous images of the curve $z(t)$ under the map $w = \sqrt[n]{z}$.

Suppose that two continuous curves C_1 and C_2 go from some point z_0 to another point z_1. Just as for the function $w = \sqrt{z}$ (see problem 284), we can prove that if the curve $C_1^{-1}C_2$ does not go around the point $z = 0$, then the value of the function $w = \sqrt[n]{z}$ at z_1 determined by continuity along the curves C_1 and C_2 will be the same if the values at z_0 are the same. Therefore, in the same way as for the function $w = \sqrt{z}$, if we make a cut from the point $z = 0$ to infinity, then the function $w = \sqrt[n]{z}$ will separate into continuous single-valued branches.

297. Let us make a cut from the point $z = 0$ to infinity not passing through the point $z = 1$, and define continuous single-valued branches of the function $\sqrt[n]{z}$ by the conditions $f_i(1) = \varepsilon_n^i$, where i runs from 0 to $n - 1$. Express $f_i(z)$ in terms of $f_0(z)$.

298. Construct the schematic diagram for the Riemann surface of the function $\sqrt[n]{z}$.

299. For the function $\sqrt{z-1}$, find the branch points and construct the schematic diagram of the Riemann surface.

300. Find the branch points and construct the schematic diagram of the Riemann surface for the function $\sqrt[n]{z+i}$.

In those situations where the multi-valued function has several branch points, to separate the continuous single-valued branches we will make cuts from each branch point to infinity using non-intersecting lines. In such a case, the schematic diagram of the Riemann surface of the given function may depend on the actual lines along which the cuts were made. (An example of this will be considered below in problems 327 and 328.) When this happens, we will make a note of how the cuts are made, otherwise we will not.

Schematic diagrams of Riemann surfaces constructed by the reader while solving the problems below may differ from the schematic diagrams in the solutions by the numbering of the sheets. If the sheets are re-numbered accordingly, the schematic diagrams should coincide.

301. Let $f(z)$ be a single-valued continuous function and C a continuous curve in the z-plane starting at the point z_0. Let w_0 be one of the values of $\sqrt[n]{f(z_0)}$. Prove that there exists at least one continuous image of the curve C under the map $w = \sqrt[n]{f(z)}$ starting at the point w_0.

From the solution to the problem 301, we have the possibility of determining the function $w = \sqrt[n]{f(z)}$ by continuity along any curve not passing through points at which the single-valuedness of continuous images of curves is violated (that is, where $f(z) = 0$).

302. Let $f(z)$ be a single-valued continuous function and $w_0(z)$ be one of the continuous single-valued branches (with corresponding cuts) of the function $w(z) = \sqrt[n]{f(z)}$. Find all single-valued continuous branches (with the same cuts) of the function $w(z)$.

303. Find all branch points and construct the schematic diagrams of the Riemann surfaces of the functions: (a) $\sqrt{z(z-i)}$, (b) $\sqrt{z^2+1}$.

304. Construct schematic diagrams of the Riemann surfaces of the following functions: (a) $\sqrt[3]{z^2-1}$, (b) $\sqrt[3]{(z-1)^2 z}$, (c) $\sqrt[3]{(z^2+1)^2}$.

305. Find the continuous single-valued branches and construct a schematic diagram of the Riemann surface of the function $\sqrt{z^2}$.

Note. From the solution to problem 305 we see that the point $z = 0$ is not a branch point of the function $\sqrt{z^2}$. At the same time, the images of curves passing through the point $z = 0$ are not determined uniquely. For example, the continuous images of the piecewise-linear curve AOB in figure 30 under the map $w = \sqrt{z^2}$ are the piecewise-linear curves COD, COF, EOD, and EOF. When moving through the point $z = 0$, we can stay on the same sheet (as in curves COD and EOF) or move to

the other sheet (as in curves COF and EOD). The Riemann surface of the function $w(z) = \sqrt{z^2}$ has the form shown in figure 31.†

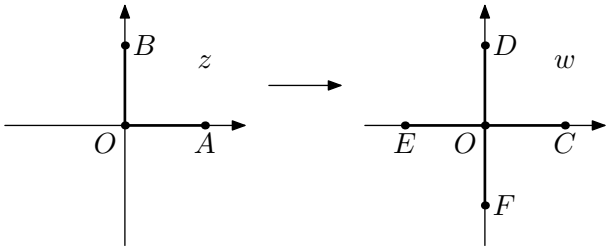


FIGURE 30.

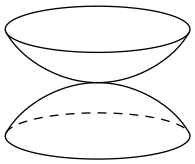


FIGURE 31.

Definition 39. Points at which the single-valuedness of continuous images of curves is violated are called *singular points* of the given function.

When constructing a schematic diagram of such a Riemann surface, we do not need to make cuts to infinity, but simply to remove these points, that is to not allow curves to pass through them.

306. Construct schematic diagrams of the Riemann surfaces of the following functions: (a) $\sqrt[4]{z^2+2}$, (b) $\sqrt[4]{z^2}$, (c) $\sqrt[4]{(z-1)^2(z+1)^3}$, (d) $\sqrt[4]{z(z^3-1)}$.

Below we will also consider functions that are not defined at some points. It may turn out that those points are branch points.

307. Construct a schematic diagram of the Riemann surface of the function $\sqrt{1/z}$.

308. Construct schematic diagrams of the Riemann surfaces of the following functions:

$$\text{(a) } \sqrt{\frac{1}{z-i}}, \qquad \text{(b) } \sqrt[3]{\frac{z-1}{z+1}}, \qquad \text{(c) } \sqrt[4]{\frac{(z+i)^2}{z(z-1)^3}}.$$

†Editor's footnote: In modern mathematics, Riemann surfaces are not permitted to have singular points such as these in them, so the function $w = \sqrt{z^2}$ would not be considered. This is similar to the problems exhibited by the derivative of the real function $y = |x| = \sqrt{x^2}$ at $x = 0$.

When solving the problems in this section, we always discover that after making non-intersecting cuts from all branch points to infinity, the function considered separates into single-valued continuous branches which are then connected in a certain way along the cuts. It turns out that a rather large class of multi-valued functions has this property. In particular, all of the functions considered below have this property, namely functions that can be expressed in radicals (section 11) and the algebraic functions (section 14).[†]

The proof of this statement is beyond the scope of this book. We could easily refer the reader to existing literature on this matter[‡] and accept the statement formulated above without proof. (The reader may wish to do this and to skip ahead to section 11.) However, the reader may then be left with some feeling of dissatisfaction. Although we cannot free the reader completely from that feeling, we can still show that the statement formulated above follows from another property, the so-called *monodromy property*, which seems more intuitively obvious.

We know that to separate single-valued continuous branches of a multi-valued function $w(z)$ in some region of the z-plane, we need the function $w(z)$ to be defined by continuity in the same way along any two curves C_1 and C_2 lying in that region and going from some point z_0 to another point z_1. The monodromy property is connected to this condition.

Definition 40. Suppose that $w(z)$ is a multi-valued function such that if any of its values w_0 is fixed at an arbitrary point z_0, then the function can be determined (perhaps non-uniquely) by continuity along any continuous curve which begins at the point z_0 and does not pass through any of the points at which $w(z)$ is not defined. We say that $w(z)$ has the *monodromy property* if it satisfies the following condition.

Let C_1 and C_2 be continuous curves in the z-plane which both start at some point z_0, end at some point z_1, and do not pass through any of the branch points or singular points of the function $w(z)$. Suppose that the curve C_1 can be continuously deformed into the curve C_2 in such a way that the curves we get during the deformation process do not pass through any of the branch points of the function $w(z)$ and such that the endpoints of the curve remain fixed. (See figure 32 as an example, where a and b are branch points.) The intermediate curves may, however, pass through non-branch singular points. Then if we fix the value of $w(z_0)$, the value of $w(z_1)$ determined by continuity along the curve C_1 will be the same as that determined by continuity along the curve C_2.

Let us determine what follows from the property of monodromy.

309. Suppose that the function $w(z)$ has the monodromy property. Let us make non-intersecting cuts in the z-plane from all branch points of the function $w(z)$ to

[†]Both types of functions are special cases of a larger class of functions, the so-called analytic functions, which also have these properties.

[‡]See, for example, George Springer, Introduction to Riemann Surfaces, chapter 3.

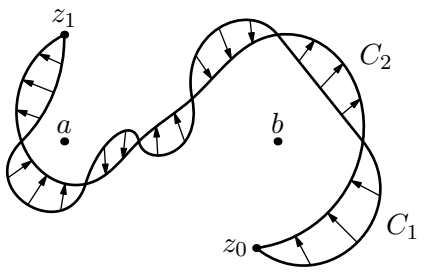


FIGURE 32.

infinity and also remove any singular points of the function. Prove that then the function $w(z)$ separates into single-valued continuous branches.

310. Suppose that, under the conditions of the previous problem, the cuts do not pass through the singular points of the function $w(z)$ and that $w(z)$ has a finite number of branch points. Prove that when crossing a cut (in a given direction), we move from any fixed branch of the function $w(z)$ to exactly the same other branch independently of where we cross the cut.

Note 1. When walking around a branch point we cross the cut going from that point to infinity precisely once. Therefore, because of problem 310, the movements from one branch to another when crossing a cut in an arbitrary place coincide with the movements that we get when walking around the branch point from which that cut is made (in the corresponding direction), which therefore coincide with movements noted at that point in the schematic diagram of the Riemann surface.

Note 2. From the results of problems 309 and 310, it follows that if a multi-valued function $w(z)$ has the monodromy property, then we can construct a Riemann surface for $w(z)$. Also, to find out the structure of this surface, it suffices to find the branch points of the function $w(z)$ and to determine the movements between the branches of the function $w(z)$ when walking around these points.

All functions which we will consider below have the monodromy property. We cannot prove this statement formally here, because for that we would need the notion of an analytic function. However, we can give an outline of the proof that a multi-valued function $w(z)$ has the property of monodromy, assuming that the function is "sufficiently good". What that means will be clear from the idea of the proof.

So suppose that we have satisfied the conditions which follow from the monodromy property. Let C_1' and C_2' be continuous images of the curves C_1 and C_2 under the map $w(z)$, starting at the same point $w_0 = w(z_0)$. We need to prove that the curves C_1' and C_2' end at the same point.

Suppose first that curves that we get by deforming C_1 to C_2 not only do not pass through any branch points, but also do not pass through any singular points of the function $w(z)$ (see page 67). Let C be any of these curves. Then there exists a unique

continuous image C' of the curve C under the map $w(z)$ which begins at $w_0 = w(z_0)$. If the function $w(z)$ is "sufficiently good",[†] then under continuous deformation of the curve C from position C_1 to position C_2, the curves C' are continuously deformed from position C_1' to position C_2'. Thus the ending point of the curve C' is also deformed continuously. But the curve C ends at the point z_1, so the ending point of the curve C' must coincide with one of the images $w(z_1)$ of the point z_1. If the function $w(z)$ takes on only a finite number of values at every point z, and in particular at z_1 (and we are only considering such functions), then the ending point of the curve C' cannot jump from one image of the point z_1 to another, because then the deformation would not be continuous. Therefore the ending points of all of the curves C', and in particular of the curves C_1' and C_2', coincide.

We can now look at what happens when the curve C passes through a singular point of the function $w(z)$ which is not a branch point. We need only consider the special case where the curve is changing only near the singular point a (see figure 33). If at the point z_0 we fix the value $w_0 = w(z_0)$, then by continuity we can determine

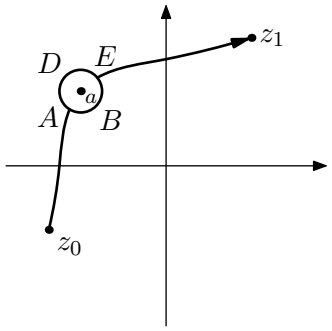


FIGURE 33.

the value of $w(z)$ at the point A. After that, we can determine the value of $w(z)$ at the point E by continuity along either of the curves ADE or ABE. These yield the same value, because otherwise when walking around the curve $EDABE$, the value of the function $w(z)$ would change, and the point a would be a branch point of the function $w(z)$. As the value of $w(z)$ at the point E is then determined, the value of $w(z)$ at the point z_1 can now be determined by continuity along the curve Ez_1. Thus the two curves z_0ABEz_1 and z_0ADEz_1 determine the same value of $w(z_1)$.

So the missing piece in our exposition is a justification of the fact that the functions studied below are "sufficiently good". This the reader has to take on trust or turn to a deeper study of analytic functions.[‡]

[†]The monodromy theorem is usually proved for arbitrary analytic functions. See, for example, George Springer, Introduction to Riemann Surfaces, chapter 4.

[‡]See, for example, George Springer, Introduction to Riemann Surfaces, chapters 4 and 5.

§11. Functions expressible in radicals

Definition 41. Let $f(z)$ and $g(z)$ be two multi-valued functions. By $f(z)+g(z)$, we mean the multi-valued function, all of whose values at a point z_0 are given by adding each value of $f(z_0)$ to each value of $g(z_0)$. In a similar way, we define the functions $f(z)-g(z)$, $f(z)g(z)$ and $f(z)/g(z)$.

By $(f(z))^n$, where n is a positive integer, we mean the function, all of whose values at the point z_0 are given by raising each value of $f(z_0)$ to the nth power.

By $\sqrt[n]{f(z)}$, where n is a positive integer, we mean the function, all of whose values at the point z_0 are given by calculating all possible values of $\sqrt[n]{f(z_0)}$ for every value of $f(z_0)$.

311. Find all values of: (a) $\sqrt[3]{-8}+\sqrt{2i}$, (b) $(1-\sqrt{-2i})/(\sqrt{-4})$, (c) $\sqrt{i+\sqrt{-1}}$, (d) $(\sqrt[4]{(1+i)^2})^2$, (e) $(\sqrt{i}+\sqrt{i})^2$.

Definition 42. We say that a multi-valued function $h(z)$ can be *expressed in radicals* if we can produce it from the function $f(z)=z$ and the constant functions $f_a(z)=a$ (with a an arbitrary fixed complex number) using the operations of addition, subtraction, multiplication, division, raising to a positive integer power and taking a root of a positive integer degree.

For example, the function $h(z)=\left(\sqrt[3]{\sqrt{z}+3z^2}-i/\sqrt{z}\right)^4$ is expressed in radicals. Some functions that can be expressed in radicals have already been studied above.

312. Suppose that the function $h(z)$ can be expressed in radicals and that C is a continuous curve in the z-plane, starting at the point z_0 and not passing through points z where $h(z)$ is undefined. Prove that if w_0 is one of the values of $h(z_0)$, then there exists at least one continuous image of the curve C under the map $w=h(z)$ starting at the point w_0.

From the result of problem 312 we deduce that an arbitrary function $h(z)$ that can be expressed in radicals can be defined by continuity along any continuous curve C that does not go through points where the function $h(z)$ is undefined. If the curve C does not pass through any branch points or singular points (see page 67) of the function $h(z)$, then the function $h(z)$ is defined uniquely by continuity along the curve C.

We already noted in previous section that functions that can be expressed in radicals are "sufficiently good",† i.e., they have the monodromy property. Therefore, for any function that can be expressed in radicals, we can construct a Riemann surface (see 309 and 310).‡ We will now determine the structure of these Riemann surfaces.

From now on, we will assume that all functions being considered can be expressed in radicals.

†Functions that can be expressed in radicals are analytic.

‡Any function that can be expressed in radicals only has a finite number of branch points.

313. Let $h(z) = f(z) + g(z)$. Remove from the plane all singular points of the function $h(z)$ and make non-intersecting cuts to infinity from all points that are branch points of at least one of the functions $f(z)$ and $g(z)$. Let $f_1(z)$, ..., $f_n(z)$ and $g_1(z)$, ..., $g_m(z)$ be the continuous single-valued branches of the functions $f(z)$ and $g(z)$ on the surface with cuts. Find the continuous single-valued branches of the function $h(z)$.

If when going around the point z_0 we go from the branch $f_{i_1}(z)$ to the branch $f_{i_2}(z)$ and from the branch $g_{j_1}(z)$ to the branch $g_{j_2}(z)$, then it is clear that from the branch $h_{i_1 j_i}(z) = f_{i_1}(z) + g_{j_1}(z)$ we will go to the branch $h_{i_2 j_2}(z) = f_{i_2}(z) + g_{j_2}(z)$. This is suggested by the following formal method of constructing the schematic diagram of the Riemann surface of the function $h(z) = f(z) + g(z)$ given that the schematic diagrams of $f(z)$ and $g(z)$ have already been constructed (with the same cuts). For every pair of branches $f_i(z)$ and $g_j(z)$ we put into correspondence a sheet where we define the branch $h_{i,j}(z) = f_i(z) + g_j(z)$. If in the schematic diagrams of the Riemann surfaces of the functions $f(z)$ and $g(z)$ we have denoted movements at the point z_0 from the branch $f_{i_1}(z)$ to the branch $f_{i_2}(z)$ and from the branch $g_{j_1}(z)$ to the branch $g_{j_2}(z)$, then in the schematic diagram of the Riemann surface of the function $h(z)$ we will denote at the point z_0 a movement from the branch $h_{i_1,j_1}(z)$ to the branch $h_{i_2,j_2}(z)$.

314. Construct the schematic diagrams of the Riemann surfaces of the functions: (a) $\sqrt{z} + \sqrt{z-1}$, (b) $\sqrt[3]{z^2-1} + \sqrt{1/z}$, (c) $\sqrt{z} + \sqrt[3]{z}$, (d) $\sqrt{z^2-1} + \sqrt[4]{z-1}$.

The formal method of constructing a schematic diagram of the Riemann surface of the function $h(z) = f(z) + g(z)$ described above does not always give the right result, because it does not account for the fact that some of the branches $h_{i,j}(z)$ may turn out to be equal. For simplicity we will suppose that the cuts do not go through the singular points of the function $h(z)$. In that case, crossing any cut we will go from sheets that correspond to equal branches of the function $h(z)$ to other sheets also corresponding to equal branches, because of uniqueness. Therefore, if we identify sheets that correspond to equal branches of the function $h(z)$, i.e., replace each such set of sheets with just one sheet, then we will have uniquely defined movements between the resultant sheets when going around any branch point z_0.

315. Find all values of $f(1)$ if: (a) $f(z) = \sqrt{z} + \sqrt{z}$, (b) $f(z) = \sqrt{z} + \sqrt[4]{z^2}$, (c) $\sqrt[3]{z} + \sqrt[3]{z}$.

316. For the following functions, construct both the schematic diagram of the Riemann surface using the formal method and the correct schematic diagram of the Riemann surface: (a) $f(z) = \sqrt{z} + \sqrt{z}$, (b) $f(z) = \sqrt{z} + \sqrt[4]{z^2}$, (c) $\sqrt[3]{z} + \sqrt[3]{z}$.

Finally we see that to construct a schematic diagram of the Riemann surface of a function $h(z) = f(z) + g(z)$ using the schematic diagrams of the functions $f(z)$ and $g(z)$ (constructed with the same cuts), it is sufficient to construct a schematic

diagram using the formal method described above and then perform some identifications.

It is easy to see how this algorithm can also be used for constructing the schematic diagrams of the Riemann surfaces of the functions $h(z) = f(z) - g(z)$, $h(z) = f(z)g(z)$ and $h(z) = f(z)/g(z)$.

317. Construct the schematic diagrams of the Riemann surfaces of the following functions: (a) $i\sqrt{z} - \sqrt[4]{z^2}$, (b) $\sqrt{z-1} \cdot \sqrt[4]{z}$, (c) $\sqrt{z^2-1}/\sqrt[4]{z+1}$, (d) $\left(\sqrt{z} + \sqrt{z}\right) \Big/ \sqrt[3]{z(z-1)}$.

318. Let $f_1(z)$, $f_2(z)$, ..., $f_m(z)$ be all of the continuous single-valued branches of the function $f(z)$. Find, with the same cuts, all continuous single-valued branches of the function $h(z) = (f(z))^n$, where n is some positive integer.

From the result of the last problem, it follows that the schematic diagram of the Riemann surface of the function $h(z) = (f(z))^n$ would coincide with the schematic diagram of the Riemann surface of the function $f(z)$ if all of the branches $h_i(z) = (f_i(z))^n$ were different. However, this is not always the case. If we have equal branches in $h(z)$, then because of uniqueness we will always go from equal branches to equal branches, as before.

Thus to construct the schematic diagram of the Riemann surface of the function $h(z) = (f(z))^n$, it is sufficient to look at the schematic diagram of the Riemann surface of the function $f(z)$, from the branches $f_i(z)$ getting branches $h_i(z) = (f_i(z))^n$. If some branches turn out to be equal, then we simply identify the corresponding sheets.

319. Construct the schematic diagrams of the Riemann surfaces of the following functions: (a) $(\sqrt[4]{z})^2$, (b) $(\sqrt{z} + \sqrt{z})^2$, (c) $(\sqrt{z} \cdot \sqrt[3]{z-1})^3$.

Let us now study how the schematic diagram of the function $\sqrt[n]{f(z)}$ is connected to the schematic diagram of the Riemann surface of the function $f(z)$.

320. What are the possible branch points of the function $\sqrt[n]{f(z)}$?

We introduce cuts in the z-plane from the branch points of the function $f(z)$ to infinity such that they do not pass through the points where one of the values of $f(z)$ is 0. Choose continuous single-valued branches of the function $f(z)$; suppose these are the (single-valued) functions $f_1(z)$, $f_2(z)$, ..., $f_m(z)$. Let us make additional cuts to infinity from the points where one of the values of $f(z)$ is 0. Let $g(z)$ be one of the continuous single-valued branches of the function $\sqrt[n]{f(z)}$ under these cuts.

321. Prove that the function $(g(z))^n$ coincides with one of the functions $f_i(z)$ everywhere except for the cuts.

The result of the previous problem implies that every branch of the function $\sqrt[n]{f(z)}$ corresponds to some branch of the function $f(z)$.

322. Let $g(z)$ be a continuous single-valued branch of the function $\sqrt[n]{f(z)}$ corresponding to the branch $f_i(z)$ of the function $f(z)$. Find all continuous single-valued branches of the function $\sqrt[n]{f(z)}$ corresponding to the branch $f_i(z)$.

It follows from the result of the preceding problem that for every branch $f_i(z)$ of the function $f(z)$, there is a corresponding "packet" of n branches of the function $\sqrt[n]{f(z)}$. We can enumerate the branches in this packet $f_{i,0}(z)$, $f_{i,1}(z)$, $\ldots$, $f_{i,n-1}(z)$ so that for every k we have $f_{i,k}(z) = \varepsilon_n^k \cdot f_{i,0}(z)$.

Let z_0 be a branch point of the function $f(z)$, and suppose that walking around z_0 takes us from the branch $f_i(z)$ to the branch $f_j(z)$. Then clearly for the function $\sqrt[n]{f(z)}$, walking around z_0 starting on a sheet in the packet for the branch $f_i(z)$ will move us to a sheet in the packet for the branch $f_j(z)$.

323. Let C be a curve in the z-plane given parametrically by the function $z(t)$. Suppose that the curve in the w-plane given parametrically by $w_0(t)$ is a continuous image of the curve C under the map $w = \sqrt[n]{f(z)}$. Prove that the curve given by $w_k(t) = \varepsilon_n^k \cdot w_0(t)$ also is a continuous image of the curve C under the map $w = \sqrt[n]{f(z)}$.

324. Suppose that the curve C in the z-plane passes through neither branch points nor singular points of the function $\sqrt[n]{f(z)}$. Prove that if moving along the curve C takes us from the branch $f_{i,s}(z)$ to the branch $f_{j,r}(z)$, then it will also take us from the branch $f_{i,s+k}(z)$ to the branch $f_{j,r+k}(z)$. Here the sums $s + k$ and $r + k$ are evaluated modulo n (see example 9 on page 9).

Therefore to determine where a walk around a given branch point of $\sqrt[n]{f(z)}$ will take us from any sheet of a given packet, it is sufficient to determine where such a walk would take us from a single sheet in this packet, for walks from other sheets in the same packet are then completely determined by the result of problem 324.

325. Construct a schematic Riemann surface diagram for the function $\sqrt{\sqrt{z} - 1}$.

326. Construct schematic Riemann surface diagrams for the following functions: (a) $\sqrt[3]{\sqrt{z} - 2}$, (b) $\sqrt{\sqrt[3]{z} - 1}$.

The following two problems consider cases where the Riemann surface diagram of a function depends on the cuts made.

327. Construct the Riemann surface diagrams for the function $f(z) = \sqrt{z^2 + 1} - 2$ with the cuts pictured in figure 34 (a) and (b). In each case, determine whether the points z with $f(z) = 0$ lie on the same or different sheets.

328. Using the previous problem, construct the Riemann surface diagrams for the function $h(z) = \sqrt{\sqrt{z^2 + 1} - 2}$ with the cuts pictured in figure 35 (a) and (b).

Let us restate the results of this section which we will need later on.

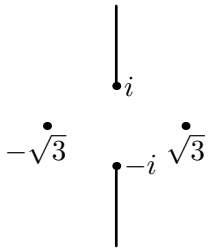


Figure (a)

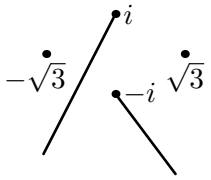


Figure (b)

FIGURE 34.

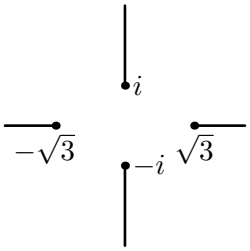


Figure (a)

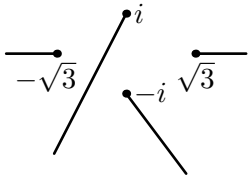


Figure (b)

FIGURE 35.

Theorem 8. *Given Riemann surface diagrams for the functions $f(z)$ and $g(z)$, the following construction will yield Riemann surface diagrams for the functions $h(z) = f(z)+g(z)$, $h(z) = f(z)-g(z)$, $h(z) = f(z)g(z)$, and $h(z) = f(z)/g(z)$ under the same cuts as the original diagrams:*

(a) *for every pair of branches $f_i(z)$ and $g_j(z)$, define a sheet with the value of the branch $h_{i,j}(z)$ on it equal to $f_i(z) + g_j(z)$, $f_i(z) - g_j(z)$, $f_i(z)g_j(z)$ and $f_i(z)/g_j(z)$ respectively;*

(b) *if the walk around z_0 takes us from the branch $f_{i_1}(z)$ to the branch $f_{i_2}(z)$ and from the branch $g_{i_1}(z)$ to the branch $g_{i_2}(z)$, then, for the function $h(z)$, let the same walk take us from the branch $h_{i_1,j_1}(z)$ to the branch $h_{i_2,j_2}(z)$;*

(c) *identify the sheets where the branches $h_{i,j}(z)$ are equal.*

Theorem 9. *Given a Riemann surface diagram for the function $f(z)$, the following construction will yield a Riemann surface diagram for the function $h(z) = (f(z))^n$ with the same cuts as the original diagram:*

(a) *for every branch $f_i(z)$, define a sheet with the value of the branch $h_i(z)$ on it equal to $(f_i(z))^n$.*
(b) *identify the sheets where the branches $h_i(z)$ are equal.*

Theorem 10. *Given a Riemann surface diagram for the function $f(z)$, the following hold for the Riemann surface diagram for the function $h(z) = \sqrt[n]{f(z)}$ constructed with the same cuts as the original diagram and additional cuts introduced as necessary from the points where 0 is a value of $f(z)$:*

(a) *every sheet of the Riemann surface diagram for $f(z)$ is replaced by a packet of n sheets;*
(b) *every walk around a branch point of the function $h(z)$ takes us from packets to packets consistently, that is, the packet we end on depends only upon the starting packet and not on the particular sheet in it;*
(c) *these transitions from one packet to another coincide with the transitions between the corresponding sheets of the Riemann surface diagram for $f(z)$;*
(d) *if the sheets in the packets are numbered so that $f_{i,k}(z) = \varepsilon_n^k \cdot f_{i,0}(z)$, then the transitions from one packet to another preserve the order of the sheets up to a cyclic shift (see problem 324).*

§12. Galois groups of multi-valued functions

We will now associate a certain permutation group with every Riemann surface diagram.

329. Suppose that the curve C in the z-plane does not pass through any branch point or singular point of the function $w(z)$. Prove that if we start on different sheets of the Riemann surface diagram and move along the continuous image of the curve C under the function $w(z)$, we will move to different sheets of the diagram.

Because of the result of this problem, it follows that to each anticlockwise walk around a branch point of the function $w(z)$, there is a corresponding permutation of the sheets of the Riemann surface diagram for $w(z)$, specifying for each sheet the new sheet we walk to.

330. Consider the Riemann surface diagrams for the functions in problem 314. Suppose that the sheets of these diagrams are numbered 1, 2, ... from top to bottom. For each of these functions, write down the permutation corresponding to a walk about each branch point.

331. Let $g_1, \ldots, g_s$ be elements of an arbitrary group G. Consider all elements of G which may be obtained from $g_1, \ldots, g_s$ by repeated applications of multiplication and taking the inverse. Prove that the resulting collection of elements is a subgroup in the group G.

Definition 43. The subgroup constructed in problem 331 is called the *subgroup generated by the elements $g_1, \ldots, g_s$*.

Definition 44. Let $g_1, \ldots, g_s$ be the sheet permutations of a Riemann surface diagram corresponding to anticlockwise walks around each of the branch points. The subgroup of all possible sheet permutations generated by $g_1, \ldots, g_s$ will be called the *sheet permutation group* of the given Riemann surface diagram.

Note 1. If the number of sheets in a diagram is finite (and we only consider such diagrams in this book), then for the construction of the sheet permutation group for this diagram it is sufficient to use multiplication alone, without resorting to inverses. Indeed in this case every sheet permutation has some finite order k: $g^k = e$, thus $g^{-1} = g^{k-1} = g \cdot g \cdot \ldots \cdot g$ ($k-1$ times).

Note 2. We will only consider the sheet permutation groups constructed below up to isomorphism. Hence the sheet numbering will be unimportant since with different numberings, the resulting subgroups of S_n are still isomorphic, even if different.

332. Which familiar groups are isomorphic to the sheet permutation groups of the Riemann surface diagrams of the following functions: (a) $\sqrt{z}$, (b) $\sqrt[3]{z}$, (c) $\sqrt[n]{z}$, (d) $\sqrt[3]{z^2-1}$ (see problem 304),(e) $\sqrt[4]{(z-1)^2(z+1)^3}$ (see problem 306)?

333. Which familiar groups are isomorphic to the sheet permutation groups of the Riemann surface diagrams of the functions appearing in problems: (a) 314, (b) 317, (c) 319?

334. Describe the sheet permutation groups for the two Riemann surface diagrams of the function $\sqrt{\sqrt{z^2+1}-2}$ constructed in problem 328.

Suppose that the point z_0 is neither a branch point nor a singular point of the multi-valued function $w(z)$. Let $w_1, w_2, \ldots, w_n$ be the values of $w(z)$ at z_0. Consider a continuous curve C starting and ending at the point z_0 and passing through neither branch points nor singular points of the function $w(z)$. If we choose a value $w_i = w(z_0)$ and determine the value of $w(z_0)$ by continuity along the curve C, we will obtain a new value $w_j = w(z_0)$. Thus if we begin with different values w_i, we will obtain different values w_j (otherwise the uniqueness of $w(z)$ determined by continuity along the curve C^{-1} would be violated). Consequently the curve C is associated with a certain permutation of the values $w_1, w_2, \ldots, w_n$. Furthermore, if a curve C is associated with a permutation g, then the curve C^{-1} is associated with the permutation g^{-1}, and if the curves C_1 and C_2 (both with endpoints at z_0) are associated with the permutations g_1 and g_2 respectively, then the curve C_1C_2 is associated with g_2g_1 (recalling that we multiply permutations from right to left).

Therefore the associated permutations of all possible curves starting and ending at z_0 form a permutation group of the values of $w(z_0)$.

335. Let G_1 be the permutation group of the values of $w(z_0)$ and let G_2 be the sheet permutation group of a Riemann surface diagram for the function $w(z)$. Prove that the groups G_1 and G_2 are isomorphic.

Note that to define the permutation group of the values of $w(z_0)$, we did not use any Riemann surface diagrams for the function $w(z)$. Therefore, by problem 335, it follows that the permutation group of the values of $w(z_0)$ for an *arbitrary* point z_0 and the sheet permutation group for an *arbitrary* Riemann surface diagram of the function $w(z)$ are isomorphic. Consequently, the permutation groups of the values of $w(z_0)$ for *all* points z_0 and the sheet permutation groups for *all* Riemann surface diagrams of the function $w(z)$ are isomorphic, and we may view them as the same group. We call this group the *Galois group* of the multi-valued function $w(z)$.[†]

§13. Galois groups of functions expressible in radicals

We have now come to one of the main results of this book.

Theorem 11. *If a multi-valued function $h(z)$ can be expressed in radicals, then its Galois group is solvable* (*see section 14 of chapter 1*).

The proof of Theorem 11 is contained in the solutions to the following problems.

336. Suppose that $f(z)$ and $g(z)$ are multi-valued functions, and that either $h(z) = f(z) + g(z)$ or $h(z) = f(z) - g(z)$ or $h(z) = f(z)g(z)$ or $h(z) = f(z)/g(z)$. Consider the formally constructed Riemann surface diagram of the function $h(z)$, as in Theorem 8 (a) and (b), page 75. Prove that if F and G are the sheet permutation groups of the Riemann surface diagrams of the functions $f(z)$ and $g(z)$, then the sheet permutation group of the formally constructed diagram is isomorphic to some subgroup of the direct product $F \times G$ (see section 7 of chapter 1).

337. Under the conditions of the previous problem, let H_1 be the sheet permutation group of the formally constructed diagram and let H_2 be the sheet permutation group of the true Riemann surface diagram for the function $h(z)$. Prove that there is a surjective homomorphism (see section 13 of chapter 1) from the group H_1 onto the group H_2.

338. Suppose that the Galois groups of the functions $f(z)$ and $g(z)$ are solvable. Prove that the Galois groups of the following functions are also solvable: $h(z) = f(z) + g(z)$, $h(z) = f(z) - g(z)$, $h(z) = f(z) \cdot g(z)$ and $h(z) = f(z)/g(z)$.

339. If the Galois group of a function $f(z)$ is solvable, prove that the Galois group of $h(z) = (f(z))^n$ is also solvable.

340. Let H be the sheet permutation group of the Riemann surface diagram for the function $h(z) = \sqrt[n]{f(z)}$. Let F be the sheet permutation group of the Riemann surface diagram for the function $f(z)$, constructed with the same cuts. Construct a *surjective* homomorphism from the group H onto the group F.

[†]This group is also called the *monodromy group*.

341. Prove that the kernel of the homomorphism (see section 13 of chapter 1) constructed in the previous problem is commutative.

342. Suppose the Galois group of the function $f(z)$ is solvable. Prove that the Galois group of the function $h(z) = \sqrt[n]{f(z)}$ is also solvable.

The constant functions $f_a(z) = a$ and the function $f(z) = z$ are single-valued and continuous in the entire z-plane, therefore their Riemann surfaces have only one sheet, and consequently the corresponding Galois groups are trivial (just $\{e\}$) and hence solvable. From this, recalling the definition of functions expressible in radicals (page 71), and the results of problems 338, 339, and 342, we have proven Theorem 11.

Note. Let us point out the following fact to a reader familiar with the theory of analytic functions. If the Galois group is defined as the permutation group of the values of $h(z)$ at some point z_0 (see page 78), then Theorem 11 will hold for a wider class of functions. For example, in constructing the function $h(z)$, in addition to constants, the identity function, the arithmetic operations and radicals, we may use any single-valued analytic functions (e.g., e^z, $\sin z$, etc.), the multi-valued function $\log z$, and certain other functions. The Galois group of $h(z)$ will still be solvable, though not necessarily finite.

§14. Abel's theorem

Consider the equation

$$3w^5 - 25w^3 + 60w - z = 0. \tag{14.1}$$

We will consider z to be a parameter, and for every complex z, will look for all complex roots w of this equation. By problem 269, this equation has 5 roots (counting multiplicities) for every z.

343. Which values of w may be multiple roots (roots of multiplicity greater than 1, see page 56) of the equation $3w^5 - 25w^3 + 60w - z = 0$? For which values of z are they multiple roots?

From the solution of the above problem, we know that equation (14.1) has 4 different roots when $z = \pm 38$ and $z = \pm 16$, and it has 5 different roots for other values of z. Thus the function $w(z)$ expressing the roots of (14.1) in terms of the parameter z takes 4 different values at $z = \pm 38$ and $z = \pm 16$, and takes 5 different values at other z. Let us study this function $w(z)$. First, we will prove that small variations of z result in only small variations of the roots of (14.1). This property is more precisely expressed in the following problem.

344. Let z_0 be an arbitrary complex number and let w_0 be one of the roots of (14.1) at $z = z_0$. Consider a disc of a small radius r centred at w_0. Prove that there is a real number $\rho > 0$ such that if $|z_0' - z_0| < \rho$, then there is at least one root of (14.1) at $z = z_0'$ lying inside the disc.

Let $w(z)$ express the roots of equation (14.1) in terms of the parameter z, and let w_0 be one of the values $w(z_0)$. Then the result of problem 344 implies that if z moves continuously along a curve from z_0, then it is possible to keep choosing one of the values of $w(z)$ so that the point w also moves continuously along a curve from w_0. In other words, $w(z)$ may be defined via continuity along any curve C. If, in addition, the curve C does not pass through any branch points or singular points of the function $w(z)$ (see page 67), then $w(z)$ will be uniquely defined along C.

345. Prove that if $z \neq \pm 38$ and $z \neq \pm 16$, then z is neither a branch point nor a singular point for the function $w(z)$.

The function $w(z)$ is algebraic[†] and therefore "well-behaved" (see section 10, page 69), i.e., it has the monodromy property. Therefore it is possible to construct a Riemann surface for $w(z)$ (see problems 309 and 310). This surface will clearly have 5 sheets.

By problem 345, $z = \pm 38$ and $z = \pm 16$ are the only possible candidates for branch points or singular points of the function $w(z)$. Nevertheless, it is not yet clear whether these points are indeed branch points or singular points.

346. Suppose it is known that $z_0 = +38$ (or $z_0 = -38$, or $z_0 = \pm 16$) is a branch point for the function $w(z)$. How are the sheets of the Riemann surface of $w(z)$ glued at z_0? (To be more precise, how are the sheets glued along the cut from z_0 to infinity? See note 2 on page 69.)

347. Suppose that z_0 and z_1 are arbitrary points different from $z = \pm 38$ and $z = \pm 16$, and let w_0 and w_1 be any of their respective images under the map $w(z)$. Prove that there is a continuous curve from z_0 to z_1 which does not pass through $z = \pm 38$ and $z = \pm 16$ and whose continuous image, beginning at w_0, ends at w_1.

348. Prove that all four points $z = \pm 38$ and $z = \pm 16$ are branch points for the function $w(z)$. What can a Riemann surface diagram for $w(z)$ look like? Sketch all possible cases. (We consider two diagrams different if one cannot be obtained from the other by a permutation of sheets and of branch points.)

349. Find the Galois group for the function $w(z)$.

350. Prove that the function $w(z)$ cannot be expressed in radicals.

351. Prove that the general polynomial equation of degree five

$$a_5 w^5 + a_4 w^4 + a_3 w^3 + a_2 w^2 + a_1 w + a_0 = 0$$

(a_0, a_1, a_2, a_3, a_4, a_5 are complex parameters, $a_5 \neq 0$) is not solvable in radicals, i.e., there is no formula which expresses the roots of the above equation in terms of the

[†]A multi-valued function $w(z)$ is called *algebraic* if it expresses, in terms of the parameter z, all roots of some equation

$$a_n(z)w^n + a_{n-1}(z)w^{n-1} + \cdots + a_0(z) = 0,$$

where each $a_i(z)$ is a polynomial in z. All algebraic functions are analytic.

coefficients using the operations of addition, subtraction, multiplication, division, raising to the power of a natural number and taking radicals of the degree of a natural number.

352. Considering the equation

$$(3w^5 - 25w^3 + 60w - z)w^{n-5} = 0,$$

prove that for $n > 5$, the general polynomial equation of degree n is not solvable in radicals.

The results of problems 351 and 352 comprise the central statement of this book. We have proven the following theorem.

Abel's Theorem. *When $n \geqslant 5$, the general polynomial equation of degree n,*

$$a_n w^n + a_{n-1} w^{n-1} + \cdots + a_1 w + a_0 = 0,$$

is not solvable in radicals.

Note 1. In the introduction, we obtained Cardano's formula for solving the general cubic equation. Recall that among the values given by the formula, not all were roots of the equation, but only those for which some additional condition held. It is therefore reasonable to ask whether, for the general polynomial equation of degree n, there is a formula in radicals whose values include the roots of the equation as a subset. Let us show that no such formula exists even for the equation (14.1).

As before, let $w(z)$ express the roots of equation (14.1) in terms of the parameter z. If the values of $w(z)$ form a subset of values of some function $w_1(z)$ expressible in radicals, then the Riemann surface for the function $w(z)$ is a separate piece of the Riemann surface for $w_1(z)$. If G is the Galois group of the function $w_1(z)$, then each permutation in G yields a permutation of the five sheets corresponding to $w(z)$. This map is a homomorphism from the group G *onto* the group S_5. Since the group S_5 is not solvable, the group G is also not solvable (see problem 163). On the other hand, G must be solvable as the Galois group of a function expressible in radicals, and thus we have a contradiction.

Note 2. The note on page 79 implies that Abel's theorem remains true if, in addition to radicals, we allow the use of certain other functions, such as any single-valued analytic functions (e^z, $\sin z$, etc.), the function $\log z$ and some others.

Note 3. Consider equation (14.1) over the real numbers only. Suppose the function $y(x)$ expresses the real roots of equation

$$3y^5 - 25y^3 + 60y - x = 0$$

in terms of the real parameter x. Perhaps it is possible to express the function $y(x)$ in radicals? The answer again turns out to be no. For the reader familiar with the theory of analytic functions, we note that this fact follows from analytic continuation. The function $w(z)$ expressing the roots of equation (14.1) in terms of the parameter z is an analytic function. Therefore, had the function $y(x)$ been expressible in

radicals, then the corresponding formula, viewed over the complex numbers, would have defined the function $w(z)$ by the uniqueness of analytic continuation. But this would mean that $w(z)$ is expressible in radicals.

Therefore Abel's theorem remains true even if we only consider the real roots of the general nth degree equation ($n \geqslant 5$) with real coefficients. In addition, as in note 2, the theorem remains true even if we allow the use of some other functions, for example, any functions which admit a single-valued analytic continuation (e^x, $\sin x$, etc.), $\log x$ and others.

Note 4. The class of algebraic functions (see the footnote on page 80) is a very rich and interesting class. In particular, one can show that all functions expressible in radicals are algebraic. We have proved that any function expressible in radicals has a solvable Galois group (Theorem 11, page 78). It turns out that if we limit our attention to algebraic functions, the converse is also true: if a Galois group of an algebraic function is solvable, then the function is expressible in radicals. Therefore an algebraic function is expressible in radicals if and only if its Galois group is solvable. This statement is one of the results of general Galois Theory.

Index