Little Mathematics Library

L.A.KALUZHNIN

THE FUNDAMENTAL THEOREM OF ARITHMETIC

Mir Publishers · Moscow

ИЗДАТЕЛЬСТВО · МИР ·

ПОПУЛЯРНЫЕ ЛЕКЦИИ ПО МАТЕМАТИКЕ

Л. А. Калужнин

ОСНОВНАЯ ТЕОРЕМА АРИФМЕТИКИ

Издательство «Наука» Москва

LITTLE MATHEMATICS LIBRARY

L.A.Kaluzhnin

THE FUNDAMENTAL THEOREM OF ARITHMETIC

Translated from the Russian
by
Ram S. Wadhwa

MIR PUBLISHERS
MOSCOW

First published 1979

На английском языке.

CONTENTS

INTRODUCTION

It is customary to think that arithmetic precedes algebra, and that it is a more elementary part of mathematics. At school, arithmetic is taught from the first form while algebra only from the fifth. Since a vast majority of people know about mathematics mainly from what they have learnt at school, the idea about the elementariness of arithmetic has taken deep roots. Meanwhile arithmetic, if considered as a study of properties of integers, and of operations upon them, is a difficult and far from elementary section of mathematics. True, in such a generalized sense, this section is rather known as "higher arithmetic" or "theory of numbers" so as to distinguish it from school arithmetic. But these designations do not alter facts. And the fact is that both school arithmetic and higher arithmetic belong to one and the same sphere of knowledge. In my view, it would be very useful if schoolboys from higher classes, having interest in mathematics, enriched the knowledge that they have acquired in lower classes. Actually, such an enrichment is also essential in order to get acquainted with higher arithmetic in future.

This brochure is intended to be of help in this direction.

As a starting point, we shall consider the so-called *fundamental theorem of arithmetic*. This somewhat scientific designation need not be frightening: everybody knows this theorem well and often use it for arithmetical calculations (e. g. while finding the common denominator of fractions), not realizing at the same time that this is an important theorem requiring a careful and detailed proof. We shall explain what it is all about.

Every integer can be expressed as a product of prime numbers. For example,

$$420 = 2 \times 2 \times 3 \times 5 \times 7 \tag{1}$$

Now, if the number is sufficiently large, then for finding the corresponding factorization, it is necessary sometimes to spend a long time. Nevertheless, we can accomplish this factorization in all cases if we like. But may be, we have been just lucky so far? Are we sure that any arbitrary whole number can be represented as a product of prime numbers? It is actually so, but this fact requires a proof. The first part of the fundamental theorem in fact comprises the statement:

Every whole number can be represented as a product of prime numbers.

The proof of this statement is carried out in this brochure. In fact it is very simple and it would be useful for the reader to work it out independently. The proof of the second part of the theorem is more difficult (it is, however, considered self-evident at school).

Before going on to its formulation, let us once again consider the above example of factorization of the number 420 into prime multipliers. The procedure, well-known from school, also represented schematically thus:

$$\begin{array}{r|l} 420 & 2 \\ 210 & 2 \\ 105 & 3 \\ 35 & 5 \\ 7 & 7 \\ 1 & \end{array}$$

actually gives the factorization (1). But may be, there are also other methods of factorization? How to know whether they will also give the same result? Of course, for example, we can try to expand the given number as a product of two smaller numbers (not necessarily prime numbers) and then each of them as a product of smaller numbers and so on until we arrive at numbers which cannot be factorized further (i. e. at prime numbers). However, from the very first step it is clear that such a process is not unique. In fact, for example, for the same number 420, we have

$$420 = 20 \times 21, \qquad 420 = 15 \times 28$$

Thus it is quite natural to ask: are there whole numbers which can be expressed in different ways as products of prime numbers? It turns out that such whole numbers do not exist, and the corresponding statement about the *uniqueness* of factorization of numbers as product of prime multipliers does, actually, constitute the second part of the fundamental theorem:

If some whole number n is expanded in two ways as a product of prime multipliers

$$n = p_1 \cdot p_2 \ldots p_k = q_1 \cdot q_2 \ldots q_l$$

then these factorizations exactly coincide except for the order of multipliers: both of them contain one and the same number of *multipliers, $k = l$, and every multiplier occurring in the first factorization is repeated the same number of times in the second*[1].

We shall give quite a detailed proof of this statement. It is, as we pointed out earlier, much more complicated than the proof of the first statement. This complication is not accidental but is connected with the fundamental properties of the arithmetic of whole numbers. It turns out that apart from this primary arithmetic, there are in existence, and of great use, many other 'arithmetics'. In some of the arithmetics, the statements of the fundamental theorem are valid, in others – not, more so, the statement about uniqueness of expansion is not fulfilled. We shall give examples of arithmetics of the first as well as second kind.

We shall consider in greater detail one arithmetic of the first kind – the arithmetic of complex whole numbers, or as they are often called, Gaussian whole numbers. We may mention, by the way, that we shall sometimes call the ordinary whole numbers as *rational* whole numbers (so as not to confuse them with Gaussian whole numbers). However, at places where they don't lead to confusion, we shall be speaking of just whole numbers, meaning thereby rational whole numbers. In the arithmetic of Gaussian whole numbers, the theorem is likewise applicable and this applicability carries along with it a whole lot of interesting and far from obvious properties of rational whole numbers.

At the end of this brochure, we shall give an example of the arithmetic in which the fundamental theorem is not applicable: true, the numbers being considered there may be expressed as a product of prime multipliers, but it may turn out that the prime numbers occurring in the two expansions are different. We shall not investigate this arithmetic in greater detail: this would require the introduction of a number of new concepts and a study of their properties, which is possible only in the framework of a serious university course.

For an understanding of our exposition, the reader is not required to possess more knowledge than is imparted by a school curriculum in mathematics, but for one important exception. While proving the theorems, we shall be making extensive use of the method of mathematical induction[2]. This method in mathematics is

[1] If we consider any artitrary whole number (positive or negative), then by the uniqueness of factorization into prime multipliers it should be understood that two factorizations $n = p_1 \cdot p_2 \ldots p_k$ and $n = q_1 \cdot q_2 \ldots q_l$ may differ not only in the order of·the multipliers, but also in signs of corresponding multipliers; see § 1 – formulation of the fundamental theorem.

[2] Sometimes it is also called "complete induction".

unfortunately rarely taught at school. A detailed substantiation and elucidation of this theory would lead us too far from our topic. To the readers, who would like to get acquainted with the method of proof by induction, we may recommend the brochure by I. S. Sominsky "The Method of Mathematical Induction" (Mir Publishers, 1975) which was published in the series "Little Mathematics Library", or the book "On Mathematical Induction" in the series "Popular Lectures on Mathematics" (Nauka, 1967) by I. S. Sominsky, L. I. Golovina and I. M. Yaglom.

At the end of this brochure, we shall mention some books which explain in a comprehensible form the theoretical-numerical facts that are more or less closely linked with the question being investigated here.

§ 1. The Fundamental Theorem of Arithmetic. Proof of the First Part

We shall give a single formulation for the statements given in the introduction, i. e. formulate completely the fundamental theorem of arithmetic.

Any non-zero whole number may be represented in the form of a product of prime numbers; moreover, such a representation is unique except for the order of the multipliers and their signs.

As has already been stated, the above-mentioned theorem contains two statements: first, a statement about the existence of a representation for any number as a product of prime numbers, and second, a statement about the uniqueness of such a representation. We shall prove both these statements. In this paragraph, we shall prove only the first of these. To begin with, we shall make two simple observations:

1. One (1) is, for many reasons, not considered as a prime number in spite of the fact that it cannot be expressed as a product of smaller numbers. Then the question arises: in what way is the above-mentioned theorem valid for integer 1? Or, in other words, in what way is integer 1 represented in the form of a product of prime numbers? Mathematics, in contrast with, say, grammar, does not like exceptions. We shall consider that

$$1 = 1$$

actually is the expansion of integer 1 into a product of prime numbers. Moreover, the number of prime multipliers in the right hand side is equal to zero. This relation reminds the definition of zero order

$a^0 = 1$ (number of multipliers a is equal to zero) and is convenient in many ways. We make a similar agreement for integer -1 also.

2. As a second remark, we shall simply give an example to explain the concept of uniqueness of expansion of a whole number into prime multipliers. Two expansions for number 18

$$18 = 2 \times 3 \times 3$$

and

$$18 = (-3) \times (-2) \times 3$$

are considered indistinguishable.

Proof for the existence of expansion of a rational whole number into a product of prime multipliers. We shall first confine ourselves to the case of positive whole numbers. The possibility of their expansion into prime multipliers is proved by the method of mathematical induction:

(a) For $n = 1$ $1 = 1$ is the required representation: 1 is a product of merely a large number of prime numbers.

(b) Let us suppose that for all positive numbers m, that are less than n, expansion into a product of prime numbers is already established. We shall then go on to show that for number n also such an expansion will occur. If n is a *prime* number, then

$$n = n$$

is the required expansion (one prime multiplier).

Let n be a *complex* number. Then it is a product $n = n_1 \cdot n_2$ of two whole numbers n_1 and n_2 each of which is different from 1 or from n. Consequently, $n_1 < n$ and $n_2 < n$. But then, by the principles of induction, the expansion of numbers n_1 and n_2 as products of prime numbers is already established:

$$n_1 = p_1 \cdot p_2 \dots p_r$$
$$n_2 = q_1 \cdot q_2 \dots q_s$$

where p_j and q_i are prime numbers. We have $n = p_1 \cdot p_2 \dots \dots p_r \cdot q_1 \cdot q_2 \dots q_s$, i. e. we have got the required expansion of the number n.

If n is a negative whole number, then $-n$ is a positive number. As has already been proved, $-n$ is expandable into a product of prime numbers. Let

$$-n = p_1 \cdot p_2 \dots p_k$$

Then

$$n = (-1)p_1 \cdot p_2 \ldots p_k$$

or, for example, $n = (-p_1) \cdot p_2 \ldots p_k$ is the required expansion of the number n. This also proves the first part of the theorem. There are many proofs for uniqueness of expansion. The one which we shall deduce is neither the shortest nor the simplest. However, our proof has the advantage that it can be directly generalized into a number of other cases, for example, to the case of polynomials of one variable, and to the case of complex whole numbers. Apart from this, during the course of the proof, we shall obtain a number of important theorems of arithmetic as a sort of by-product.

§2. Division with Remainder and Greatest Common Divisor (GCD) of Two Numbers. Proof of the Second Part of the Fundamental Theorem

The statement about the possibility of "division with a remainder" in the case of whole number is the starting point for our consideration. This statement can be precisely formulated as under:

THEOREM 1. *Let a and b be whole numbers and $b \neq 0$. Then there exist whole numbers q and r*[1]*, where* $0 \leqslant |r| < b$, such that

$$a = q \cdot b + r \tag{1}$$

The equality $r = 0$ in the equation (1) is equivalent to the fact that number a is divisible by b[2]. We shall denote such a fact in future by $b|a$ – this is an accepted notation in the number theory.

We shall prove the possibility of such a representation. For this, we observe that for every rational number τ a whole number t

[1] The remainder r can be any whole number – positive, negative, or zero.

[2] For two whole numbers a and b, the statements "number a is divisible by number b", "number a is a multiple of number b", "number b is a divisor of number a", or, finally, "number b divides number a" mean one and the same thing; we shall use each one of them.

can be found so that $|\tau - t| < 1$[1]. Let $\tau = \frac{a}{b}$; a and b being whole numbers. We select a whole number q so that $\left|\frac{a}{b} - q\right| < 1$ and express

$$r = b\left(\frac{a}{b} - q\right) = a - bq$$

Thus r is a whole number $|r| = |b| \left|\frac{a}{b} - q\right| < |b| \times 1 = |b|$ and

$$a = q \cdot b + r,$$

q.e.d.[2]

Theorem (1) allows us to deduce the idea of GCD of two numbers and prove many of its properties.

DEFINITION 1. If a and b are two non-zero whole numbers and if c is a number such that $c|a$ and $c|b$, then c is called a *common divisor* of numbers a and b. We shall note that any two numbers always have common divisors. These are numbers 1 and -1. If no other divisors exist, then numbers a and b are called *mutually prime* numbers. We shall talk about the mutually prime numbers later.

DEFINITION 2. Number d is called the *greatest common divisor* of numbers a and b (GCD), if: (1) d is a common divisor of a and b and (2) d is divisible by any other common divisor of numbers a and b. (Thus, for example, 6 is GCD of numbers 18 and 30, since $6|18$ and $6|30$, and on the other hand, 6 is divisible by all common divisors of these numbers: 1, -1, 2, -2, 3, -3, 6, -6.)

The reader must be aware even from school that a GCD exists for any pair of whole numbers and must also be conversant with the method of its determination. But if we recall and carefully

[1] As a matter of fact the nearest whole number to τ differs from it by not more than $\frac{1}{2}$ but we shall not require the precision.

[2] We note that in representation (1), the whole numbers q and r are not determined uniquely. For example for $a = 13$ and $b = 3$ we have $13 = 4 \cdot 3 + 1 \, (q = 4, r = 1)$ or $13 = 5 \cdot 3 + (-2)\,(q = 5, r = -2)$. This is also seen from our proof. In fact if a is not divisible by b, then $\frac{a}{b}$ is a fractional number but then $n < \frac{a}{b} < n + 1$ where n is a whole number. For number q we can choose $q = n$ or $q = n + 1$ which gives two representations of the form (1). Only in case where $b|a$, is the number q singularly represented, $a = q \cdot b$; in this case $r = 0$.

analyze this method we can easily deduce that it makes use of the factorization of numbers a and b into prime multipliers and of the uniqueness of such a factorization. This method is still forbidden to us since we are just going to prove the corresponding theorem.

From our definition (Definition 2) it does not directly follow that, for any two numbers a and b, a GCD always exists. We shall now prove that it is actually so; moreover, this proof shall not make use of factorization of numbers a and b into prime multipliers.

THEOREM 2. *For any pair of whole numbers $a \neq 0$ and $b \neq 0$, there exists a GCD.*

PROOF. In addition to numbers a and b, we shall consider all the possible numbers of the type $xa + yb$ where x and y are any integers. Numbers of such kind,

$$v = xa + yb \tag{2}$$

are called *linear combinations* of numbers a and b. For example, for $a = 6$, $b = 22$, the linear combination will be numbers 28 $(28 = 1 \cdot 6 + 1 \cdot 22)$, $10(10 = (-2) \cdot 6 + 1 \cdot 22)$, $-92(-92 = 3 \cdot 6 + + (-5) \cdot 22)$, etc. Generally, for any given numbers a and b there exists an infinitely large number of their linear combinations. We shall denote the set of such numbers through M. We observe that this set contains, in particular, also the numbers a (for $y = 0$, $x = 1$) and b (for $x = 0$, $y = 1$) as well as number 0 ($x = 0$, $y = 0$). All numbers v from the set M are obviously whole numbers. If v belongs to M then $-v$ also belongs to M (if $v = xa + yb$, then $-v = (-x)a + (-y) \cdot b$). We also notice one more property of numbers v belonging to M, that we shall need at once: all such numbers are divisible by all common divisors of numbers a and b. In fact, if $c|a$ and $c|b$ and say $a = a'c$ and $b = b'c$, then $v = xa + yb = xa'c + yb'c = (xa' + yb')c$, i. e. $c|v$. Now let $d \neq 0$ – the minimum number by taking modulus out of all non-zero numbers in M[1]. We shall prove that d is the GCD for numbers a and b. It satisfies the property of GCD as per definition (2), since all numbers from M possess this

[1] Such number in the set M actually does exist. We notice that in the set M are contained numbers which are not equal to zero (for example a or b) and their moduli are positive integers, i. e. natural numbers. But one of the fundamental properties of natural numbers, usually applied as an axiom (see I.S. Sominsky, "The Method of Mathematical Induction") is that any non-void collection of natural numbers always contains a minimum number.

property. All that is now required is to establish that it also possesses the property (1), i. e. d is a common divisor of numbers a and b. We shall show that $d|a$. Since d belongs to M, it can be expressed in the form $d = sa + tb$ where s and t are suitable integers. We shall divide a by d with remainder, i. e. we shall find such numbers q and r, $r < |d|$, so that

$$a = qd + r$$

But then the remainder r also must belong to set M. Actually,

$$r = a - qd = a - q(sa + tb) = (1 - qs)a + tb$$

We now recall that d by modulus is the minimum number among non-zero numbers of set M and $r < d$. It follows that $r = 0$ and $d|a$. In exactly similar way the divisibility $d|b$ is proved. The theorem is thus proved.

We have established the existence of GCD of two non-zero whole numbers. Apart from that we shall deduce from the proof the following fact which we shall soon require:

THEOREM 3. *GCD of numbers a and b is represented in the form of a linear combination of these numbers.*

The question arises: has the GCD of numbers a and b been singularly determined? The answer is, of course, in the negative: if number d possesses the properties (1) and (2) of the definition of GCD, then $-d$ also possesses these properties. But this exhausts the non-singularity. Actually, let d and d' be two GCDs of numbers a and b. Since d possesses the property (2) and d' – property (1), $d'|d$. But analogously $d|d'$. Thus $\alpha = \frac{d}{d'}$, and $\frac{d'}{d} = \frac{1}{d/d'} = \frac{1}{\alpha}$ are integers. But the only integers whose reciprocals are also integers are numbers 1 and -1. Thus $\alpha = 1$ or $\alpha = -1$, whence $d' = d$ or $d' = -d$. If in the definition of GCD, we require that this number were positive – it sometimes (but not always) is convenient – then it could be said that GCD of *two non-zero integers exists and is singularly determined.*

In future we shall express GCD of numbers a and b through (a, b) as is usually the practice in the literature on number theory.

Let's go over to the question of pairs of mutually prime numbers. We have already come across this concept. Now we shall repeat its definition.

DEFINITION 3. Integers $a \neq 0$ and $b \neq 0$ are called *mutually prime* if their GCD is equal to 1.

In other words, it may be said that mutually prime numbers are such numbers for which the only common divisors are numbers 1 ând -1.

From the aforesaid (Theorem 3), it follows that if $(a, b) = 1$, then 1 can be expressed in the form

$$1 = sa + tb \tag{3}$$

with suitable integers s and t. Conversely, if the equality (3) holds for suitable s and t, then a and b are mutually prime. Really (see proof of Theorem 1), $d = (a, b)$ – this is the lowest number by modulus among non-zero numbers of the type $xa + yb$. Consequently, if (3) holds, then $|d| \leqslant 1$ and $d \neq 0$, so $d = \pm 1$.

From this directly follows the most important property of mutually prime numbers.

THEOREM 4. *If* $a|bc$ *and* $(a, b) = 1$, *then* $a|c$ (this property reads: *if number a divides the product of two numbers and is mutually prime to one of them, then it is a divisor of the other*).

PROOF. Since $(a, b) = 1$, we can find such numbers s and t so that

$$1 = sa + tb \tag{4}$$

Multiplying both sides by c we have

$$c = (sc)a + t(bc)$$

Both items on the right-hand side are divisible by a, consequently c is divisible by a.

The following statement is also useful.

THEOREM 5. *If number a is mutually prime with numbers b and c, then it is mutually prime with the product bc.*

PROOF. Since $(a, b) = 1$, we can find whole numbers s and t satisfying the equality

$$1 = sa + tb$$

Analogously, since $(a, c) = 1$, then

$$1 = ua + vc$$

for suitable u and v. Multiplying these two equations we get

$$1 = (sa + tb)(ua + vc) = sua^2 + savc + tbua + tbvc = (sua + svc + \\ + tbu)a + (tv)\cdot(bc)$$

If $m = sua + svc + tbu$ and $n = tv$, then m and n are integers and

$$1 = ma + n(bc)$$

This shows that a and bc are mutually prime.

The statement of the last theorem can be easily extended for an indefinite number of factors.

THEOREM 6. *If a is mutually prime with numbers b_1, b_2, ..., b_k, then a is mutually prime with the product $b_1 \cdot b_2 \ldots b_k$.*

The proof of this theorem is carried out by the method of mathematical induction for k factors.

PROOF *of uniqueness* of factorization of an integer as a product of prime multipliers.

Now, at last, we can prove the second part of the fundamental theorem of arithmetic. For this, we observe that by definition of a prime number, two different prime numbers are mutually prime. The proof of uniqueness of factorization shall be carried out by induction for absolute value of number n.

(a) If $|n| = 1$, then $n = \pm 1$ and

$$1 = 1, \quad -1 = -1$$

i. e. the factorization is unique for numbers 1 and -1.

(b) Let us suppose that the property to be proved is already true for all numbers m for which $|m| < |n|$. Let

$$n = p_1 \cdot p_2 \ldots p_k = q_1 \cdot q_2 \ldots q_l$$

be two factorizations for the number n as products of prime numbers p_1, p_2 ..., p_k and q_1, q_2, ..., q_l respectively. We state that prime number p_k occurs among prime numbers q_1, q_2, ..., q_l (or, may be, is opposite in sign to some one of them). Really, if it is not so, i. e. if $p_k \neq \pm q_i$, $i = 1, 2, \ldots, l$, then p_k would be mutually prime with all the numbers q_i and, consequently, according to Theorem 6, also with their product, i. e. with the number n. But this is impossible since $p_k | n$, i. e. $(p_k, n) = p_k$. Thus p_k is equal to some one of the prime numbers $\pm q_i$. We may assume that $p_k = q_l$ because if it is not so we can obtain such an equality by rearranging the multipliers q_i and then, if at all $p_k = -q_l$, by changing the sign of q_l by changing it in some other q_i also.

Thus we get

$$n = p_1 \cdot p_2 \ldots p_{k-1} \cdot p_k = q_1 \cdot q_2 \ldots q_{l-1} \cdot p_k$$

whence

$$m = \frac{n}{p_k} = p_1 \cdot p_2 \ldots p_{k-1} = q_1 \cdot q_2 \ldots q_{l-1}$$

But $|m| < |n|$ and by assumption of induction, the statement of theorem for m has already been proved, i. e. $k - 1 = l - 1$, the sequences in $p_1, p_2, \ldots, p_{k-1}$ and $q_1, q_2 \ldots q_{l-1}$ contain, except for the accuracy in signs, the same prime numbers and corresponding prime numbers occur in both the factorizations the same number of times, and since $p_k = q_l$, then it is also valid for sequences $p_1, p_2, \ldots, p_{k-1}, p_k$ and $q_1, q_2, \ldots, q_{l-1}, q_l$, q.e.d.

§ 3. Algorithm of Euclid and Solution of Linear Diophantine Equations with Two Unknowns

According to Theorem 2 two integers a and b have a GCD. We shall now describe a single procedure for determining GCD which was indicated even in the 'Elements of Euclid' and is called "Euclidean Algorithm".

For this we shall assume that

$$|a| \geqslant |b|$$

First step. Let us divide a by b with remainder:

$$a = q_1 \cdot b + r_1, \quad |r_1| < |b| \qquad (1)$$

If $r_1 = 0$, then $b|a$ and $(a, b) = b$. If $r_1 \neq 0$, then we take the

Second step. Let us divide b by r_1:

$$b = q_2 \cdot r_1 + r_2, \quad |r_2| < |r_1| \qquad (2)$$

If $r_2 \neq 0$, then we take the

Third step.

$$r_1 = q_3 \cdot r_2 + r_3, \quad |r_3| < |r_2| \qquad (3)$$

and so on. At every step the new remainder is less than the remainder in the previous step

$$|b| > |r_1| > |r_2| > \ldots$$

and at some kth step ($k < |b|$) the remainder becomes equal to zero.

kth *step.*

$$r_{k-2} = q_k \cdot r_{k-1} \qquad (k)$$

We shall show that the last non-zero remainder r_{k-1} is the required (a, b). Really, we get a chain of equalities:

$$
\begin{array}{ll}
(1) & a - q_1 \cdot b + r_1 \\
(2) & b = q_2 \cdot r_1 + r_2 \\
(3) & r_1 = q_3 \cdot r_2 + r_3 \\
 & \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\
(k-1) & r_{k-3} = q_{k-1} \cdot r_{k-2} + r_{k-1} \\
(k) & r_{k-2} = q_k \cdot r_{k-1}
\end{array}
$$

From the last equality we get $r_{k-1}|r_{k-2}$, from the last but one – $r_{k-1}|r_{k-1}$ and $r_{k-1}|r_{k-2}$ and, consequently, $r_{k-1}|r_{k-3}$. From the previous equality we can analogously conclude that $r_{k-1}|r_{k-4}$ and thus going step by step to earlier equations, we conclude that $\ldots, r_{k-1}|r_2, r_{k-1}|r_1, r_{k-1}|b, r_{k-1}|a$. We see that r_{k-1} is the common divisor of numbers a and b.

Now let $c|a$ and $c|b$. Then from (1), (2), ..., $(k-1)$ successively, we get $c|r_1$, $c|r_2$, $\ldots$, $c|r_{k-1}$. Thus r_{k-1} is really the GCD for numbers a and b.

Let us take a numerical example: $a = 858$, $b = 253$. We have

$$\begin{aligned}
&(1) & 858 &= 3\cdot 253 + 99\\
&(2) & 253 &= 2\cdot 99 + 55\\
&(3) & 99 &= 1\cdot 55 + 44\\
&(4) & 55 &= 1\cdot 44 + 11\\
&(5) & 44 &= 4\cdot 11
\end{aligned}$$

whence (858, 253) = 11. Thus, with the help of Euclidean algorithm, GCD of two numbers is determined without factorizing them into prime multipliers.

In Theorem 3 we established that $(a, b) = d$ can be expressed in the form

$$d = s\cdot a + t\cdot b$$

but in the proof there was no indication as to how the corresponding s and t can be found. With the help of Euclidean algorithm, this problem is very easily solved. We won't describe the procedure for a general case, but shall explain it for the already solved numerical example.

So, we have to find whole numbers s and t such that

$$11 = s\cdot 858 + t\cdot 253$$

From (4), (3), (2), (1) successively, we get

$$\begin{aligned}
11 &= 55 + (-1)\cdot 44\\
44 &= 99 + (-1)\cdot 55\\
55 &= 253 + (-2)\cdot 99\\
99 &= 858 + (-3)\cdot 253
\end{aligned}$$

Now substituting in the first of these equalities the expression for 44 from the second, then for 55 the expression from the next equality and so on, we get

$$\begin{aligned}11 &= 55 + (-1)\cdot(99 + (-1)\cdot 55)\\ &= 2\cdot 55 + (-1)\cdot 99\\ &= 2\cdot(253 + (-2)\cdot 99) + (-1)\cdot 99\\ &= 2\cdot 253 + (-5)\cdot 99\\ &= 2\cdot 253 + (-5)\cdot(858 + (-3)\cdot 253)\\ &= (-5)\cdot 858 + 17\cdot 253\end{aligned}$$

Finally: $s = -5$, $t = 17$.

The reader can easily make out how this algorithm can be used in a general case. The equalities occurring in the Euclidean algorithm while finding the GCD of numbers a and b allow us to solve equations of the type

$$d = xa + yb$$

(where $d = (a, b)$).

In general, the equation of the type

$$xa + yb = c$$

where a, b, c are the given integers for which one seeks solution x, y in integers, is called a *linear diophantine equation* with two unknowns. It is called linear since the unknowns x and y occur in it in the *first* order. The term "diophantine"[1] indicates that the coefficients of the equation are *integers* and the required solution are also *integers*.

We observe that we have really learnt how to solve the linear diophantine equations of the type

$$xa + yb = c \tag{I}$$

But we must discuss the question about all the solutions of the equation (I) in greater detail. We shall notice first that not every equation of this type has a solution. Actually, if equation (I) does have a solution in integers, say $x = x_0$ and $y = y_0$: $c = x_0a + y_0b$, and if $d = (a, b)$, then, since $d|a$, $d|b$, d divides both terms on the right-hand side and, consequently, also divides c. From this we draw the following conclusion:

In order that a solution in terms of integers of equation (I) *may exist, it is necessary that the right-hand side of the equation is divisible by the greatest common divisor of the numbers a and b.*

[1] Named after the ancient Greek mathematician Diophantos (around 250 B.C.) who investigated equations for integers in his book "Arithmetica". At the end of our exposition we shall stop for a while on the quadratic *diophantine equations.*

For example, the equation

$$9x + 15y = 7$$

does not have a solution, since 7 is not divisible by $3 = (9, 15)$. On the contrary, if $d|c$, then the equation (I) does have a solution in terms of integers and we even know how to find such a solution. Actually let $c = c'd$, and let s and t are such integers (they can be found out with the help of Euclidean algorithm) that

$$d = as + bt$$

Then

$$c = c'd = a(sc') + b(tc')$$

i. e. $x_0 = sc'$, $y_0 = tc'$ are the solutions of the equation (I).

Let us solve, for example, the diophantine equation

$$33 = 858x + 253y \tag{II}$$

We have already shown that

$$11 = 858 \cdot (-5) + 253 \cdot 17$$

Multiplying this equality by 3, we get

$$33 = 858 \cdot (-15) + 253 \cdot 51$$

Thus $x = -15$, $y = 51$ are the solutions of the equation (II). It should not be thought that the desired solution is unique. Generally, it turns out that *if a diophantine equation of the type* (I) *does have a solution, then it has an infinite number of solutions.* We shall now study this question in greater detail: we shall prove the formulated statement and find a general form for all possible solutions of the equation (I). Let us begin with the elucidation of the general form. Let us suppose that we already know that, in addition to the solution in terms of integers x_0, y_0, the equation (I) also has the solution x_1, y_1, we have

$$c = ax_0 + by_0$$
$$c = ax_1 + by_1$$

Subtracting the second equality from the first, we get

$$a(x_0 - x_1) + b(y_0 - y_1) = 0$$

or

$$a(x_0 - x_1) = b(y_1 - y_0) \tag{III}$$

If $d = (a, b)$, then we put $a' = a/d$, $b' = b/d$, i. e.

$$a = a'd$$
$$b = b'd$$

where a' and b' are mutually prime numbers. Dividing the equality III by d, we arrive at the equality

$$a'(x_0 - x_1) = b'(y_1 - y_0)$$

But then since a', b' are mutually prime, $a' \mid (y_1 - y_0)$ and, analogously, $b' \mid (x_0 - x_1)$. Substituting

$$y_1 - y_0 = a'k_1$$
$$x_0 - x_1 = b'k_2$$

we get $a'b'k_1 = a'b'k_2$, whence $k_1 = k_2 = k$. Thus, finally

$$y_1 = y_0 + a'k = y_0 + \frac{a}{d}k \qquad \text{(IV)}$$

$$x_1 = x_0 - b'k = x_0 \neg \frac{b}{d}k \qquad \text{(V)}$$

where k is some integer. Conversely, it is easy to check that if x_0, y_0 is the solution of equation (I), then all pairs of numbers IV, V for any integer k give solution to the equation (I). Actually,

$$\begin{aligned} ax_1 + by_1 &= a\left(x_0 - \frac{b}{d}k\right) + b\left(y_0 + \frac{a}{d}k\right) \\ &= ax_0 + by_0 + \left(-\frac{ab}{d}k + \frac{ab}{d}k\right) \\ &= c + 0 = c \end{aligned}$$

Thus, *if* x_0, y_0 *are solutions of equation* (I), *then all numbers of the type* $x_0 - \frac{b}{d}k$, $y_0 + \frac{a}{d}k$ *are also solutions* (it means that for every case, there are infinite solutions – one for every k) *and there are no other solutions.**

§ 4. Gaussian Numbers and Gaussian Whole Numbers

The natural generalization of rational whole numbers is the complex whole numbers or, as they are usually called, "Gaussian whole numbers", after the great German mathematician K.F. Gauss, who first studied them in detail.

DEFINITION 4. A complex number is called "Gaussian whole number" if its real and imaginary parts are essentially rational whole numbers. In other words, they are complex numbers of the form α

$$\alpha = a + bi \tag{1}$$

where a and b are whole (rational) numbers. In addition to the Gaussian whole numbers, we shall also need (simple) *Gaussian numbers*, i. e. complex numbers, whose real and imaginary parts are rational numbers.

The relation between the field of Gaussian numbers and Gaussian whole numbers is analogous to the relation between rational numbers and rational whole numbers. More precisely we mean the following statement which we shall frequently use without special reservation, and which the reader can easily verify directly.

I. *Sum, difference and product of two whole Gaussian numbers are also Gaussian whole numbers* (this property is expressed in short by saying that Gaussian whole numbers form a *ring*).

II. *Sum, difference, product and quotient* (in case the divisor is not equal to zero) *of two Gaussian numbers are also Gaussian numbers.* (This property is expressed shortly as: Gaussian numbers form a *field*.)

III. *Quotient of two Gaussian whole numbers is a Gaussian number and, conversely, every Gaussian number can be represented as a quotient of two Gaussian whole numbers.*

The last statement requires a little explanation. Let $\alpha = a + bi$ and $\beta = c + di$ are Gaussian whole numbers (i. e. a, b, c, d are whole rational numbers) and let $\beta \neq 0$. We shall show that $\gamma = \alpha/\beta$ – Gaussian number. Actually

$$\gamma = \frac{a + bi}{c + di} = \frac{(a + bi)(c - di)}{(c + di)(c - di)}$$

$$= \frac{ac + dd - adi + bci}{c^2 + d^2}$$

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

Numbers $\dfrac{ac + bd}{c^2 + d^2}$ and $\dfrac{bc - ad}{c^2 + d^2}$ – real and imaginary parts of the number γ are rational and, consequently, γ is a Gaussian number.

We observe finally that obviously any rational number is Gaussian

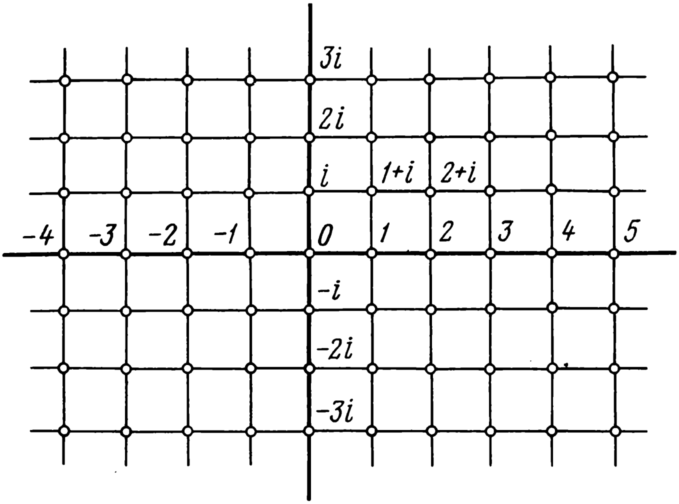

Fig. 1

(imaginary part is equal to zero) and that every rational whole number is a Gaussian whole number.

For future, it will be useful to have an idea about the arrangement of Gaussian whole numbers on a complex plane. By definition itself, the Gaussian whole numbers are represented by points with integral coordinates (Fig. 1). They are located on top of the mesh of squares with sides equal to unity, covering the complex surface.

From the theory of complex numbers we shall need the ideas of norm and modulus of a complex number. We remind that *norm* of a complex number $\alpha = x + yi$ is the non-negative real number $N(\alpha) = x^2 + y^2$, the *modulus* of the complex number α denoted by $|\alpha|$ is the real number $\sqrt{x^2 + y^2}$. Geometrically, modulus of a complex number is the distance of the corresponding point on the complex surface from the origin of coordinates. Norm $N(\alpha)$ of a number α is represented as the product $N(\alpha) = \alpha \cdot \bar{\alpha}$, where $\bar{\alpha}$ is the complex conjugate $x - iy$ of number α. *The property of multiplicability* of norm is also supposed to be a familiar property, i. e.

$$N(\alpha \cdot \beta) = N(\alpha) \cdot N(\beta) \qquad (2)$$

We shall at once note that if α is a Gaussian number then $N(\alpha)$

is a non-negative rational number and even if α is a Gaussian whole number, then $N(\alpha)$ is a non-negative whole number[1].

However, not every positive rational whole number is a norm of a Gaussian whole number. In fact we shall now prove the following theorem.

THEOREM 7. *A positive rational whole number c is norm of some Gaussian whole number if and only if the number c, can be represented in the form of sum of squares of two integers.*

PROOF. If $\alpha = a + bi$ is a Gaussian whole number, then $N(\alpha) = a^2 + b^2$ is the sum of squares of whole numbers a and b. Conversely, if $c = x^2 + y^2$ where x and y are rational whole numbers, then $c = N(x + yi)$ where $x + yi$ is a Gaussian whole number. The theorem is thus proved.

It is not difficult to show that not every positive whole number can be represented as a sum of two squares. We shall show, for example, that a positive odd integer t, which can be represented as a sum of two squares of integers, gives a remainder equal to 1 upon division by 4, i. e. is a number of the type $t = 4k + 1$. Actually let $t = x^2 + y^2$, then one of the numbers, say, x, must be even, the other y – odd. Let $x = 2m$ and $y = 2n + 1$. Then $x^2 = 4m^2$ and $y^2 = 4(n^2 + n) + 1$ and, finally, $t = 4(m^2 + n^2 + n) + 1$, which proves our statement. In this way, numbers 7, 11, 15 and others which cannot be represented in the form of a sum of two squares are, consequently, not norms of Gaussian numbers.

We shall explain the question, precisely which whole numbers can be represented in the form of sum of two squares or, in other words, which numbers are the norms of Gaussian whole numbers, after studying the arithmetic of Gaussian whole numbers. We shall now go on to a study of this arithmetic.

As in the domain (ring) of rational whole numbers, so also in the domain of Gaussian whole numbers, the question of divisibility is of main interest.

We shall say that a Gaussian whole number α divides a Gaussian whole number β and denote this fact as $\alpha|\beta$ – if for some Gaussian whole number γ, the equation

$$\beta = \alpha \cdot \gamma \tag{3}$$

holds. Since from (3) follows $N(\beta) = N(\alpha) \cdot N(\gamma)$, the necessary condition for $\alpha|\beta$ is the divisibility $N(\alpha)|N(\beta)$ where $N(\alpha)$ and $N(\beta)$ are rational whole numbers.

[1] Modulus $|\alpha|$ of a Gaussian number is not necessarily a rational number; therefore, in future we shall mainly use norm instead of modulus.

In case of rational whole numbers, there are only two numbers which divide all integers: $+1$ and -1. In case of Gaussian whole numbers, there are four such numbers: $+1$, -1, $+i$, $-i$. It is easily seen that these four numbers satisfy this property. Actually,

$$\alpha = \alpha \cdot 1$$
$$\alpha = (-\alpha)\cdot(-1)$$
$$\alpha = (-\alpha i)\cdot i$$
$$\alpha = (\alpha i)\cdot(-i)$$

There are no other numbers among Gaussian whole numbers with the given properties. In fact, if some Gaussian whole number ξ divides all Gaussian whole numbers, then this must, in particular, divide number 1 (therefore such numbers are called *unitary divisors*). From $N(\xi)|1$ it follows that $N(\xi) = 1$. If $\xi = x + yi$, then $x^2 + y^2 = 1$. It is obvious that this equation has precisely four solutions among rational whole umbers: $x = 1$, $y = 0$; $x = -1$, $y = 0$; $x = 0$, $y = 1$; $x = 0$, $y = -1$. These four solutions exactly correspond to Gaussian whole numbers $+1$, -1, i, $-i$.

For Gaussian whole numbers, in a way analogous to rational whole numbers, we develop the concept of *common divisor*, *greatest common divisor*, *mutually prime numbers* and *prime numbers*. The first three concepts are determined exactly in the same way as in the case of rational whole numbers. However, we must deal with the definition of simple Gaussian integers at a little greater length.

DEFINITION 5. A Gaussian whole number π is called *prime* if in all its factorizations $\pi = \tau\cdot\sigma$ as product of two Gaussian whole numbers, one of the factors (τ or σ) is a unitary divisor. (Here the unitary divisors are not considered simple numbers.)

This property may be expressed in other words as follows: a simple Gaussian number π is a nonzero whole Gaussian number whose norm is greater than unity and which cannot be expanded as a product of two Gaussian whole numbers whose norms are less than the norm of number π.

According to this definition, simple Gaussian numbers are, for example, numbers $\pi_1 = 2 + i(N(\pi_1) = 5)$, $\pi_2 = 3 + 2i(N(\pi_2) = 13)$. In general all numbers, whose norms are simple rational numbers, are simple numbers. Below we shall see that simple Gaussian whole numbers are not exhausted by these examples. We shall describe all simple Gaussian numbers. As for now, we go on to the formulation and proof of the fundamental theorem of arithmetic for Gaussian whole numbers.

FUNDAMENTAL THEOREM. *Any Gaussian whole number* $\alpha \neq 0$ *can be expressed as a product of simple Gaussian numbers*

$$\alpha = \pi_1 \cdot \pi_2 \, \dots \, \pi_k \tag{4}$$

(π_i are simple Gaussian numbers not necessarily different from one another). *Such an expansion is unique in the following sense: if*

$$\alpha = \sigma_1 \cdot \sigma_2 \, \dots \, \sigma_l \tag{5}$$

is another expansion of number α *into a product of simple Gaussian numbers* σ_j, *then both these expansions have one and the same number of multipliers,* $k = l$, *and factors* (4) *and* (5) *may differ from each other only by the order of factors and by multipliers which are unitary divisors.*

As regards the part of formulation concerning uniqueness of expansion, we shall make yet another observation. If, say,

$$\alpha = \pi_1 \cdot \pi_2 \cdot \pi_3$$

is a product of simple numbers π_1, π_2, π_3, then, for example,

$$\alpha = (-\pi_3) \cdot (i\pi_2) \cdot (i\pi_1) = (\pi_1 \cdot \pi_2 \cdot \pi_3)$$

is the other representation of number α as a product of simple numbers $-\pi_3$, $i\pi_2$, $i\pi_1$ as differing from simple numbers π_1, π_2, π_3. However, it is easy to notice that any of the numbers $-\pi_3$, $i\pi_2$, $i\pi_1$ is obtained by multiplying one of the numbers π_1, π_2, π_3 by some unitary divisor; moreover, the initial order of numbers is also changed. Such differences in the expansion of one and the same number are allowed. The second part of the formulation of the theorem actually states that non-uniqueness of such kind in different expansions vanishes. This case is not different from the case of rational whole numbers in arithmetic. It is simply complicated by the fact that in case of arithmetic of Gaussian whole numbers, we are provided with a large number of unitary divisors [1]. The statement about the uniqueness of the expansion may be formulated more briefly by introducing the idea of associability of Gaussian whole numbers.

DEFINITION 6. Two Gaussian whole numbers are called *associative* if they differ from each other by a factor equal to a unitary divisor, i. e. β, $-\beta$, $i\beta$, $-i\beta$ are associative Gaussian whole numbers if β is an arbitrary Gaussian whole number.

[1] We note that uniqueness of expansion, except for the signs of the multipliers about which we talked in the case of whole *rational* numbers, also means uniqueness except for multipliers which are unitary divisors, since $+1$ and -1 are the only unitary divisors in this case.

By using this definition, the statement about uniqueness in the fundamental theorem is formulated as under:

If $\alpha = \pi_1 \cdot \pi_2 \ldots \pi_k$ *and* $\alpha = \sigma_1 \cdot \sigma_2 \ldots \sigma_l$, *where* $\pi_i (i = 1, 2, \ldots, k)$ *and* $\sigma_j (j = 1, 2, \ldots, l)$ *are prime numbers, then* $l = k$ *and the multipliers* σ_j *may be expressed so that every* σ_j *will be associative with the corresponding prime number* π_j.

We shall outline the *proof of the fundamental theorem.* It is done in the same way as the proof of the corresponding statements for rational whole numbers. Therefore we shall not do it exhaustively but strongly recommend the reader to do it himself.

The first statement of the theorem – about the existence of an expansion – may be done by induction for norm of the number:

(a) If $N(\alpha) = 1$, then $\alpha = 1, -1, i, -i$, i. e. α can be expanded into a product of an empty set of prime numbers[1].

(b) Let $N(\alpha) = n$, and for all Gaussian whole numbers with minimum norm, the statement has already been proved. Then either α is a prime number and everything is proved, or $\alpha = \rho \cdot \tau$ where $N(\rho) < n$ and $N(\tau) < n$. According to assumptions of induction, factorization for ρ and τ do exist: $\rho = \pi_1 \cdot \pi_2 \ldots \ldots \pi_k$ and $\tau = \sigma_1 \cdot \sigma_2 \ldots \sigma_l$. Then $\alpha = \pi_1 \cdot \pi_2 \ldots \pi_k \cdot \sigma_1 \cdot \sigma_2 \ldots \ldots \sigma_l$ is the factorization for α.

The proof of the statement about the uniqueness may be carried out by means of establishing the properties of GCD and properties of mutually prime numbers in the case of Gaussian whole numbers. The statement about the possibility of division with a remainder for the case of Gaussian whole numbers provides the clue to the whole proof. Here it is formulated as under:

Let $\alpha, \beta, (\beta \neq 0)$ *be two Gaussian whole numbers, then there exist Gaussian whole numbers* γ *and* ρ, *where* $N(\rho) < N(\beta)$, *so that*

$$\alpha = \gamma \cdot \beta + \rho$$

The proof is based on a very simple geometrical fact: if P is a point lying in a square with side a, or on one of its sides, then the distance of this point P from the nearest corner is less than a. Really, the centre of the square is the point farthest from all corners. But its distance from any corner is equal to $\frac{1}{\sqrt{2}} a < a$. Any other point in the square is situated even closer to the nearest corner.

[1] About 'factorizability' of unitary divisors into products of prime multipliers, we accept the same terms as for ± 1 in case of rational whole numbers, see p. 11.

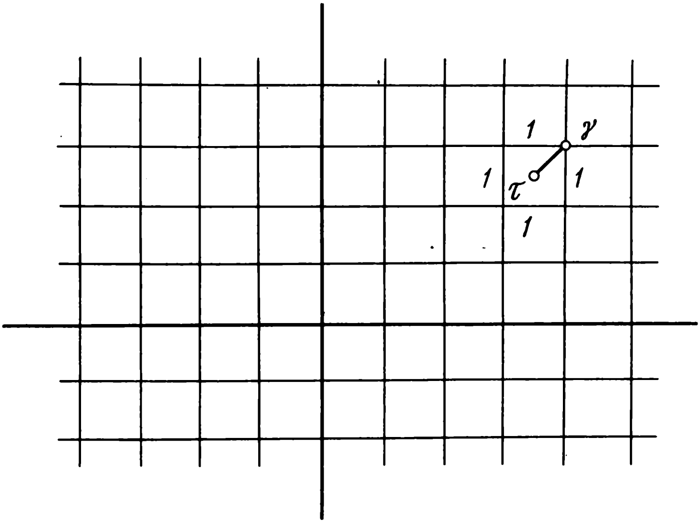

Fig. 2

It is now clearly visible from this simple proof that for any point τ on a complex surface, we can find a point γ with integral coordinates – representing a Gaussian whole number – and removed from τ by a distance less than 1 (Fig. 2). In other words, for any complex number τ there exists a Gaussian whole number γ so that $N(\tau - \gamma) < 1$. Let us find such a γ for number $\tau = \alpha/\beta$ and put $\rho = \alpha - \gamma\beta$. Then ρ is Gaussian whole number

$$N(\rho) = N(\beta) \cdot N\left(\frac{\alpha}{\beta} - \gamma\right) < N(\beta)$$

and

$$\alpha = \gamma\beta + \rho$$

The statement is proved.

Having already got the theorem on division with remainder, we can prove all other properties in the same way as we did above in the case of rational numbers: (1) we prove the existence of a GCD for two Gaussian whole numbers α and β in the form of numbers $\delta \neq 0$ with minimum norm from a set of numbers that can be represented in the form $\alpha\xi + \beta\eta$ (ξ and η are Gaussian whole numbers), (2) the concept of Gaussian whole numbers that are prime to each other is introduced and the fundamental lemma is proved: if α is mutually prime to β_1 and if α is mutually prime to β_2, then α is mutually prime to $\beta_1 \cdot \beta_2$. After this it is very easy to prove by induction of norm, the uniqueness of factorization into prime multipliers.

§ 5. Gaussian Prime Numbers and Representation of Rational Whole Numbers as Sum of Two Squares

We now go on to a description of all Gaussian prime numbers. We shall first prove some auxiliary statements – lemmae.

LEMMA 1. *Every Gaussian prime number is a divisor of a prime rational number*[1].

Actually, since $N(\alpha) = \alpha \cdot \bar{\alpha}$, *any Gaussian whole number divides its norm*: $\alpha | N(\alpha)$. Now let π be a prime Gaussian number, then $\pi \,|\, N(\pi)$, and let $N(\pi) = p_1 \cdot p_2 \, \ldots \, p_r$ is the factorization of number $N(\pi)$ as a product of prime rational numbers. We have $\pi \,|\, p_1 \cdot p_2 \, \ldots \, p_r$, hence π divides one of the prime numbers p_i. In fact, if the prime Gaussian whole number π did not divide any of the numbers p_i then it would be mutually prime to each of them and consequently to their product $N(\pi)$. But this is impossible since $\pi \,|\, N(\pi)$. So π is a divisor of one of the prime rational whole numbers p_i and the lemma is proved.

LEMMA 2. *Norm $N(\pi)$ of a prime Gaussian number π is either a prime rational number or the square of a prime rational number.*

Really, as we already know, π divides some prime rational number p. Let $p = \pi \cdot \gamma$. Taking the norms $N(\pi) \cdot N(\gamma) = p^2$, only two possibilities exist: (1) $N(\pi) = N(\gamma) = p$,and (2) $N(\pi) = p^2 =$ $= N(p)$ and $N(\gamma) = 1$. The lemma is thus proved.

The second case means that γ is a unitary divisor and one of the equalities $\pi = p$, $\pi = -p$, $\pi = ip$, $\pi = -ip$ is valid. Consequently, p is such a prime rational number that it is also a prime

[1] We observe that a prime rational number is always a whole Gaussian number also; however, as a Gaussian number it is not necessarily prime, but may be divided into Gaussian whole numbers with a lower norm. Thus, for example, 2 is a prime number if it is considered as a rational whole number. But it is not prime if we consider it as a Gaussian whole number. Actually, in the domain of Gaussian whole numbers, 2 can be factorized as $2 = (1 + i)(1 - i)$ and neither of the factors $1 + i$ and $1 - i$ is a unitary divisor. It is obvious that 5 is also not prime in the domain of Gaussian numbers, since $5 = (2 + i)(2 - i)$.

Gaussian number. In case (1) γ is a prime Gaussian number since $N(\gamma) = p$. It may be stated that $\gamma = \bar{\pi}$. Actually, $N(\pi) = p = \pi \cdot \bar{\pi}$ and $\bar{\pi}$ is a prime number. But we also have $p = \pi \cdot \gamma$ so that $\bar{\pi} = \gamma$.

On the other hand, let p be some prime rational number. Then if it is not prime Gaussian number, it is divisible by some prime Gaussian number other than p and, in addition, as we have seen, $p = \pi \cdot \bar{\pi}$. So p is a product of two prime Gaussian complex conjugate numbers. In this case p is the norm of a Gaussian whole number and can therefore be represented as a sum of two squares. Such a prime number if it is odd (i. e. $p \neq 2$) is a number of the form $4n + 1$, representable as a sum of two squares. It can be shown that *all prime numbers of the form* $4n + 1$ *can be represented as a sum of two squares*, i. e. they are the norms of some Gaussian whole numbers and, consequently, belong to the class of such prime rational numbers which can be factorized into products of two complex conjugate prime Gaussian numbers. We shall not carry out proof of this statement [1]. It is all prime rational numbers other than numbers of the type $4n + 1$ or 2, i. e. numbers of the type $4n + 3$ that form the set of prime rational numbers which are both prime and are in the domain of Gaussian numbers.

Two (2) holds a special position. It is easy to see that

$$2 = i \cdot (1 - i)^2, \quad N(1 - i) = 2$$

Thus 2 is divisible by the square of a prime Gaussian number $(1 - i)$.

Assuming that all prime numbers of the type $4n + 1$ can be represented as a sum of two squares, we can now determine all the rational whole numbers which can be represented as a sum of two squares. As we already know, the only necessary and sufficient condition for any such number t is that it should be norm of some Gaussian whole number α: $t = N(\alpha)$. Number α is expanded as a product of prime Gaussian numbers

$$\alpha = \pi_1 \cdot \pi_2 \ \ldots \ \pi_r \tag{6}$$

We divide all prime numbers $\pi_i (i = 1, 2, \ldots, r)$ into two classes. The first class contains such numbers π_i whose norms are prime, and correspondingly the second class contains numbers whose

[1] The proof of this fact based on theory of comparisons and given by L. Euler may be found in any textbook on number theory. It is dealt with in great detail in ref. [3] in the list of literature at the end of the brochure.

norms are squares of prime integers [1]. We denote the various numbers of first class as $\sigma_j (j = 1, 2, \ldots, l)$ and those of second class as $\rho_k (k = 1, 2, \ldots, s)$. We have: $N(\sigma_j) = p_j$, $N(\rho_k) = q_k^2$, where p_j is a prime number of the type $4n + 1$ or 2, and q_k is a prime number of the type $4n + 3$. Combining the equal prime numbers in the right-hand side of (6) we can write the product in the form of powers of prime numbers σ_j and ρ_k

$$\alpha = \sigma_1^{a_1} \cdots \sigma_l^{a_l} \cdot \rho_1^{b_1} \cdots \rho_s^{b_s} \tag{7}$$

Changing to norms, we get

$$N(\alpha) = t = N(\sigma_1^{a_1}) \ldots N(\sigma_l^{a_l}) \cdot N(\rho_1^{b_1}) \ldots N(\rho_s^{b_s})$$

$$t = p_1^{a_1} \ldots p_l^{a_l} q_1^{2b_1} \ldots q_s^{2b_s} \tag{8}$$

We see that prime numbers q_k enter the factorization of number t in *even powers.* Conversely, suppose number t can be represented in the form (8) where each p_j is a prime number of the type $4n + 1$ or number 2, q_k are prime numbers of the type $4n + 3$ and $a_1, \ldots, a_l, b_1, \ldots, b_s$ are non-negative whole numbers. Then, since every p_j is a sum of two squares, we may select σ_j so that $N(\sigma_j) = p_j$. Further, putting $\rho_k = q_k$ and $\alpha = \sigma_1^{a_1} \ldots \sigma_l^{a_l} \cdot \rho_1^{b_1} \ldots \rho_s^{b_s}$, we get $t = N(\alpha)$, i. e. t can be represented as a sum of two squares. Finally, we have the following theorem:

THEOREM 8. *The necessary and sufficient condition that a rational whole number could be represented as a sum of two squares is that prime numbers of the type $4n + 3$ in the factorization of this number should occur in even orders* [2].

We observe that this theorem gives a criterion for a diophantic equation of 2nd order, $x^2 + y^2 = t$, to have a solution (whole number). We shall not stop here to explain how such a solution is actually found out.

[1] Of course, it may turn out that one of the classes is empty. This, however, does not substantially allect the course of our discussions. We shall only have to consider that all numbeŗs a_j or all numbers b_k (in factorization (7) or (8)) may be zeros.

[2] Such a formulation also covers the case when the factorization of the number under consideration does not include any prime numbers of the type $4n + 3$, because 0 is also even!

In general, a study of diophantic equations of the type

$$ax^2 + 2bxy + cy^2 = t$$

is closely linked with arithmetics whose number domains are analogous to the domains of Gaussian whole numbers.

In such investigations, the following astonishing fact, which mathematicians encountered in the middle of the last century, is important. *The theorem about uniqueness of factorization of numbers into prime numbers does not hold in all like arithmetics.* Without going into the details of the situation arising here, we shall cite an example of one "arithmetic" in which the fundamental theorem is not valid.

§ 6. Yet Another "Arithmetic"

We shall consider complex numbers of the type

$$\alpha = x + y\sqrt{-5} \tag{1}$$

where x and y are rational whole numbers. It is easily seen that sum, difference and product of numbers of type (1) are also numbers of the same type. We denote the set of all such numbers by Γ. Obviously, Γ contains all rational whole numbers (for $y = 0$). Just as in the case of rational and Gaussian whole numbers, we can talk about divisibility of Γ: α divides $\beta(\alpha|\beta)$, if $\beta|\alpha$ is again a number from Γ, i. e. representable in form (1). As also in the case of Gaussian whole numbers, the norms of numbers from Γ play an important role in the question of divisibility

$$N(\alpha) = N(x + y\sqrt{-5}) = (x + y\sqrt{-5})(x - y\sqrt{-5})$$
$$= x^2 + 5y^2$$

In this way, the norm of any number from Γ is a rational whole number and since $N(\xi \cdot \eta) = N(\xi) \cdot N(\eta)$, the condition $N(\alpha) | N(\beta)$ is necessary (though generally not sufficient) so that $\alpha | \delta$.

Just like the case of Gaussian whole numbers, the idea of unitary divisors and prime numbers is introduced. As regards unitary divisors, things are even simpler here than for Gaussian whole numbers. That is, only numbers ± 1 are unitary divisors. Actually, for unitary divisors $\xi = u + v\sqrt{-5}$, the condition $N(\xi) = u^2 + 5v^2 = 1$ must hold. But this diophantic equation obviously cannot have any other solution except $u = \pm 1$ and $v = 0$.

The fact that each number from Γ can be expressed as a product of prime numbers from Γ is proved by induction for norms in exactly the same way as for Gaussian whole numbers. But the *statement about uniqueness* of such a factorization *is not valid* here, and we shall prove it by a simple example.

We shall first show that numbers $2 = 2 + 0 \cdot \sqrt{-5}$, $3 = 3 + 0 \times \sqrt{-5}$, $1 + \sqrt{-5}$, $1 - \sqrt{-5}$ are prime numbers in Γ. Actually, $N(2) = 4$, $N(3) = 9$, $N(1 + \sqrt{-5}) = N(1 - \sqrt{-5}) = 6$. If any of these numbers were not prime in Γ, then it could be divisible only by some number $\alpha = x + y\sqrt{-5}$ for which $N(\alpha) = x^2 + 5y^2 = 2$ or $N(\alpha) = x^2 + 5y^2 = 3$. But there are no such numbers in Γ since, obviously, the equations

$$x^2 + 5y^2 = 2 \tag{2}$$

and

$$x^2 + 5y^2 = 3 \tag{3}$$

do not have whole number solutions.

Thus, the given 4 numbers are prime numbers in Γ. We now consider an easily verifiable equality

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}) \tag{4}$$

It shows that number 6 from Γ has two different factorizations into prime numbers.

The German mathematician E. Kummer (1810-1893) encountered this effect while trying to solve the so-called great Fermi theorem. The difficulties that arose later in connection with the non-validity of the fundamental theorem of arithmetic in some important domains of numbers were successfully overcome by Kummer himself as well as by other mathematicians – R. Dedekind, E. Zolotarev, L. Kronecker, etc. Thus arose a vast new branch of mathematics, called the theory of algebraic numbers, which is being successfully developed right till the present time.

LITERATURE

We shall indicate here a number of works through which the reader may get additional information about the number theory, and in particular about the subject of our booklet. The proposed list does not claim to be exhaustive. Some of these works, in turn, carry references to literature on number theory. We shall list these books in the order of increasing difficulty. Thus, while the first few works do not require any special background beyond the framework of school mathematics, the latter are textbooks for students in universities and teachers-training institutes. Experience has shown that these can also be used for self-study.

1. Khinchin, A.Ya., Elements of Number Theory: Encyclopaedia of Elementary Mathematics, Book 1, *Arithmetic*, Gostekhizdat, 1951.

   This article by the eminent Soviet mathematician A.Ya. Khinchin is written with great skill and may be recommended both for teachers as well as senior schoolboys. Chapter I "Divisibility and Prime Numbers" contains, among other things, a detailed proof of the fundamental theorem of arithmetic. Chapter II "Method of Comparisons" is, in our view, one of the best introductions to the theory of comparisons – a section of number theory having a very large number of applications in arithmetic as well as modern algebra.

2. Markushevich, A.I., Division with Remainder in Arithmetic and Algebra (Series *Pedagogical Teachers' Library*), pub. Academy of Pedagogical Sciences of RSFSR, 1949.

   The book contains almost all the material presented in this booklet, as also many other sections of algebra and arithmetic directly connected with the theory of divisibility.

3. Davenport, H., Higher Arithmetic, Nauka, 1965.

   The subtitle of the book "An Introduction to Number Theory" itself indicates that it contains a systematic treatment of the elements of this field. Written by an eminent English specialist on number theory, Davenport's book does not require a background beyond the framework of school mathematics. It may be specially recommended for junior mathematics students, but may also be used by non-mathematicians. The latter, of course, must possess the skill to properly understand descriptions. The readers of our booklet will certainly find Chapter II on "Comparisons" and Chapter V on "Sums of Squares" (giving a complete proof of representability of prime numbers of the type $4n + 3$ as a sum of two squares) of special interest.

The following books may be recommended for mathematics students for self-study in the number theory.

4. Vinogradov, I. N., Principles of Number Theory, Nauka, 1965.
5. Arnold, I. V., Theoretical Arithmetic, Uchpedgiz, 1939.
6. Bukhshtab, A. A., Number Theory, Prosveshchenie, 1966.
7. Hasse, G., Lectures on Number Theory, For. Lang. Publ. House, 1953.

TO THE READER

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

Our address is:
USSR, 129820, Moscow, I-110, GSP
Pervy Rizhsky Pereulok, 2
Mir Publishers

Printed in the Union of Soviet Socialist Republics

N. VILENKIN, D.SC.

Method of Successive Approximations

This book explains in a popular form the methods of approximation, solutions of algebraic, trigonometric, model and other equations.

Intended for senior schoolchildren, polytechnic students, mathematics teachers and for those who encounter solutions of equations in their practical work. In the course of exposition some elementary ideas about higher mathematics are introduced in the book. About 20 solved exercises are included in the appendix.

The book has already been translated into Spanish and French.

Contents. Successive Approximations. Achilles and the Tortoise. Division by Computers. Determination of Square Roots by the Method of Successive Approximations. Determination of Roots with a Natural Index by Means of Successive Approximations. Iterational Method. Geometrical Meaning of Iterational Method. Compressible Representations. Compressible Representations and Iterational Method. Method of Chords. Perfected Method of Chords. Derivative of a Polynomial. Newton's Method of Approximate Solution of Algebraic Equations. Geometrical Meaning of Derivative. Geometrical Meaning of Newton's Method. Derivatives of Any Function. Calculation of Derivatives. Determination of First Approximations. Combined Method for Solution of Equations. Criterion for Convergence of Iterational Process. Speed of Convergence of Iterational Process. Solution of a System of Linear Equations by the Method of Successive Approximations.

E 0081

Educated at the Paris University, Lev Arkadievich Kaluzhnin D.Sc., is a Professor at the Kiev State University. He is the author of some 100 published works, which include a number of textbooks for university and secondary school students.

In this booklet, Prof. Kaluzhnin deals with one of the fundamental propositions of arithmetic of rational whole numbers - the uniqueness of their expansion into prime multipliers. Having establisehd a conncetion between arithmetic and Gaussian numbers and the question of representing integers as sum of squares, Prof. Kaluzhnin has shown the uniqueness of expansion also holds in the arithmetic of complex (Gaussian) whole numbers.

The author hopes that the booklet will not only be of interest to senior schoolboys but will also be useful for teachers.

Mir Publishers
Moscow

ИЗДАТЕЛЬСТВО • МИР •