DOCUMENT RESUME

ED 035 551 SE 007 732

AUTHOR	Barnett, I. A.
TITLE	Some Ideas About Number Theory.
INSTITUTION	National Council of Teachers of Mathematics, Inc., Washington, D.C.
PUB DATE	61
NOTE	77p.
AVAILABLE FROM	National Council of Teachers of Mathematics, 1201 Sixteenth Street, N.W., Washington, D. C. 20036
EDRS PRICE	EDRS Price MF-$0.50 HC Not Available from EDRS.
DESCRIPTORS	Algebra, Arithmetic, College Mathematics, *Mathematical Enrichment, Mathematics, *Number Concepts, *Number Systems, Resource Materials, Secondary School Mathematics

ABSTRACT

The material in this booklet is designed for non-professional mathematicians who have an interest in the theory of numbers. The author presents some elementary results of number theory without involving detailed proofs. Much of the material has direct application for secondary school mathematics teachers. A brief account of the nature of number theory is given in order to acquaint the reader with the subject. Topics discussed include: prime and composite numbers, the Euler-phi-function, some proven facts about number theory, conjectures and unsolved problems, congruences, Diophantine equations, some generalizations of number theory, and many more. Proofs of general theorems are avoided; instead the author verifies general theorems for special cases. A bibliography is provided for the reader who wishes to investigate the proofs of these theorems. (FL)

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS

SOME IDEAS ABOUT number theory

I. A. BARNETT

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS
1201 Sixteenth Street, N. W., Washington, D. C. 20036

SOME IDEAS ABOUT number theory

I. A. BARNETT
Ohio University
Athens, Ohio

Second printing 1963
Third printing 1965
Fourth printing 1967

Library of Congress Catalog Card Number: 61–12377

Printed in the United States of America

Preface

The author has no intention of adding to the already long list of textbooks on the theory of numbers. This booklet gives an informal account of some of the more elementary results of the subject without going into detailed proofs. The theory of numbers lends itself naturally to such treatment. Its appeal lies in the fact that its results may be readily understood by those who are not professional mathematicians, and the truth of the general theorems may be verified easily for special cases.

Some of the material will have direct application for teachers of arithmetic, algebra, and geometry. Furthermore, it is hoped that the general reader, as well as the teacher, will be stimulated to go more deeply into the subject. A few of the books from which the author himself has benefited are listed in the bibliography for those who wish to go into the proofs and learn aspects of the subject that cannot be included in a brief treatment.

Table of Contents

Chapter I *The Natural Numbers*

Chapter II *Primes and Composites*

Chapter III *The Divisors of a Number*

Chapter IV *Some Proven Facts of Number Theory*

Chapter V *Some Conjectures and Unsolved Problems*

Chapter VI ***Fermat Numbers and Regular Polygons***

Chapter VII ***Congruences***

Chapter VIII ***Diophantine Equations***

Chapter IX ***Generalizations of Number Theory***

CHAPTER

I

The Natural Numbers

The positive integers 1, 2, 3, $\cdots$ are called the *natural numbers*. The doʋs represent numbers that have been omitted, and the sequence continues indefinitely; in other words, there is no largest positive integer. The natural numbers form the basis for the study of the theory of numbers, or higher arithmetic as it is sometimes called. The mathematician studies the behavior of these numbers just as the chemist classifies the elements according to their atomic weights and their reactions with each other.

1. What is number theory?

From earliest times man has shown curiosity about numbers. This was particularly true of the ancient Greeks and Chinese, whose interest was mainly in the study of relationships among numbers. It was not until the seventeenth century that the first serious study of the subject was made. This work was done by the famous French mathematician Pierre Fermat (1601–1665), who is considered the founder of the theory of numbers.

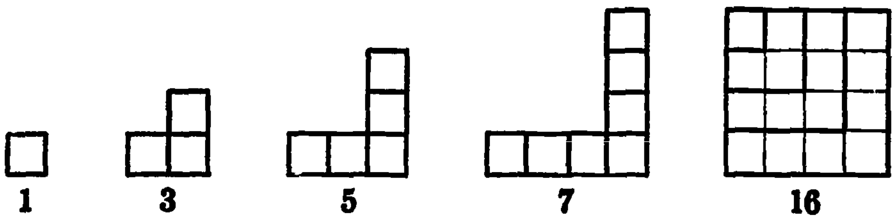
The theory of numbers is regarded as the purest branch of pure mathematics because it has very few applications to other sciences. Many of its general results were discovered and suggested by special cases which were observed not only by mathematicians but also by amateurs. The world-famous German mathematician, Carl Friedrich Gauss (1777–1855), referred to mathematics as the "Queen of the Sciences" and to higher arithmetic as the "Queen of Mathematics."

Gauss made many original contributions to the subject; he also systematized all the materials available and put them in the form we have today. In this respect his work in number theory may be compared with that of Euclid (about 300 B.C.) in geometry.

2. The sequence of consecutive odd numbers

We begin with some very simple observations regarding the natural numbers. We note first that they divide themselves into two groups, the odd and the even numbers, which resemble the parent group in that neither of the groups has a largest number.

Adding consecutive odd numbers, one obtains a striking result: $1 + 3 = 4 = 2^2$; $1 + 3 + 5 = 9 = 3^2$; $1 + 3 + 5 + 7 = 16 = 4^2$; and so on. It is true without exception that *the sum of the first n odd numbers is equal to n^2*. Thus the sum of the first 1000 odd numbers is equal to $(1000)^2$, or 1,000,000. This result, as well as many others of a similar character, was proved geometrically by the ancient Greeks. The proof of the statement under consideration is given below (12: 1–11):*



3. The Fibonacci sequence

Another interesting sequence of natural numbers arises from a problem suggested by the thirteenth century mathematician Leonardo of Pisa (about 1170–1250), who is called *Fibonacci* since he was the son (*figlio*) of Bonaccio. In his famous book *Liber Abaci* the following problem appears (7: 27–31; 6: 76–78): How many pairs of rabbits can be produced from a single pair, if it is supposed that every month each pair begets a new pair, which from the second month on becomes productive?

We are led to the following sequence of natural numbers:

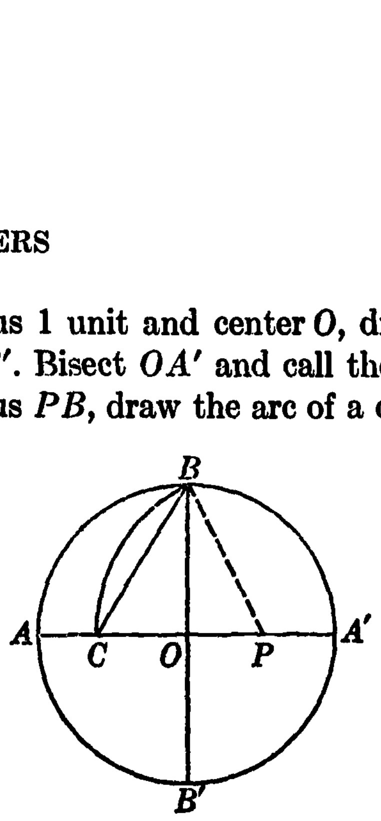
$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \cdots,$$

where each new term is formed by adding the last term to its predecessor. Thus $2 = 1 + 1, 3 = 2 + 1, 5 = 3 + 2, 8 = 5 + 3, 13 = 8 + 5, \cdots$. Fibonacci sequences occur in plant growth and in art, as well as in geometry. We shall not attempt to explain the occurrence of these numbers in plant growth; the reader is referred to the topic "Phyllotaxis" in an encyclopedia.

We now turn to a geometric application of Fibonacci numbers. Let us first recall how a regular decagon (a ten-sided polygon, all of whose sides and angles are equal) may be inscribed in a given circle. We shall describe the construction, which is to be made with ruler and compasses only, without going into the proof (8: 213).

* The symbol $(x:y)$ will be used to refer to page y of reference x in the numbered list at the end of this pamphlet.

In a circle with radius 1 unit and center O, draw two perpendicular diameters AA' and BB'. Bisect OA' and call the midpoint P. With P as center and with radius PB, draw the arc of a circle cutting AO at C.



Then BC is the length of the side of the inscribed regular pentagon (the five-sided figure all of whose sides are chords of the given circle and are of equal length); and OC is the length of the side of the regular decagon. In the proof of the construction the ratio AC to CO (the ratio of the segments into which C divides the radius) turns out to be

$$\frac{2}{1+\sqrt{5}} \quad \text{or} \quad \frac{\sqrt{5}-1}{2},$$

and this ratio is approximately .618.

Let us now return to the Fibonacci sequence of numbers and find, starting with 2, the ratios of pairs of consecutive terms: $2/3 = .667$; $3/5 = .600$; $5/8 = .625$; $8/13 = .615$; $13/21 = .619$; $21/34 = .618$; $34/55 = .618$; $55/89 = .618$; $89/144 = .618$, and so on. We note that the ratios of these pairs of numbers seem to be getting closer to the ratio $2/(1 + \sqrt{5})$. In fact, it has been shown (12:44–45) that *the limiting value of the ratio of two successive terms of the Fibonacci numbers, as we go out indefinitely, is precisely the ratio of the side of the regular inscribed decagon to the radius of the circumscribing circle.*

The expressions *golden section* and *divine proportion* refer to the ratio $2/(1 + \sqrt{5})$ or .618. To the Greeks the most pleasing rectangular figures were those whose sides were in the ratio 3/5 (approximately the divine proportion .618).

CHAPTER

II

Primes and Composites

4. What are prime and composite numbers?

Both Aristotle (384–322 B.C.) and Euclid distinguished between such numbers as 2, 3, 5, 7, 11, 13, 17, which they called *primes*, and 4, 6, 8, 9, 10, 12, 14, which they called *composites*. A prime has no exact divisor other than itself and unity. It is obvious that the only even prime is 2. The number 1 will not be considered a prime for reasons that will be apparent later.

Anyone can tell at a glance whether a number is odd or even; to determine whether a number is prime or composite is much more difficult and may not even be feasible. If we want to find out whether 30,031 is prime or composite, we have the tedious task of testing the primes in succession as divisors, until we find one prime which divides exactly into 30,031. In this case the first prime divisor is 59, and the other factor, also a prime, is 509. Thus we find that 30,031 is composite. Tables of primes have been constructed up to 10,000,000, and certain other large primes have been found by ingenious methods. At the present time the largest verified prime is $2^{3217} - 1$, which, when written in the usual form, contains 969 digits.

5. The distribution of primes

If we start with a small prime, we may readily find the next consecutive prime by trial. But nobody yet knows the prime that immediately follows $2^{3217} - 1$. Such a prime exists, since, as we shall prove later, there is an endless sequence of primes.

The smallest interval between two consecutive primes starting with 3 is, obviously, 2, but we can show that there are consecutive primes as far apart as we choose, though we cannot give their exact values. Let us illustrate. First, we may write as many consecutive composite numbers as we please. If we wish to have 1000 consecutive composite numbers, all we need do is write the sequence

$$1001! + 2, 1001! + 3, 1001! + 4, \cdots, 1001! + 1001,$$

where the notation 1001! has the usual meaning (2) (3) (4) (5) $\cdots$ (1001). Since 1001! contains the factor 2, the first term, 1001! + 2, is exactly divisible by 2. Likewise the term 1001! + 3 contains the factor 3, and so on, until we arrive at the thousandth member of the sequence, 1001! + 1001, which is exactly divisible by 1001. Thus each of the 1000 members is a composite number. The largest prime preceding 1001! + 2, and the first prime following 1001! + 1001, are two consecutive primes whose difference is at least 1000. In this way we can prove that there exist two primes that are arbitrarily far apart, even though we cannot give the values of these primes.

The primes are very irregularly distributed. There are 25 primes among the first 100 natural numbers. The following table shows the distribution of primes among the natural numbers up to 1000 (10: 75–77).

Numbers from	1	100	200	300	400	500	600	700	800	900
to	100	200	300	400	500	600	700	800	900	1000
Number of Primes	25	21	16	16	17	14	16	14	15	14

The primes seem to occur less frequently as we go farther out in the sequence of natural numbers. It has been shown by ingenious calculations that in the interval from 10^{12} (one trillion) to 10^{12} + 1000 there is the following distribution of primes.

Numbers from	10^{12} +	0	100	200	300	400	500	600	700	800	900
to	10^{12} +	100	200	300	400	500	600	700	800	900	1000
Number of Primes		4	6	2	4	2	4	3	5	1	6

6. A formula for the number of primes

In spite of the fact that the primes are irregularly spaced, mathematicians have found a formula giving the approximate number of primes between 1 and any natural number N. Although the formula gives approximate results, the approximation becomes more and more exact as N increases in size.

This formula is most easily described in terms of the number e, which is known as the natural or Naperian base of logarithms after the Scottish mathematician John Napier (1550–1617).

The meaning of this number e may be explained in connection with a problem of compound interest. If \$1 is invested at 100 percent interest, the investment will amount to \$2 in a year. If, however, the money is compounded twice a year, the investment will amount to \$2.25; if compounded quarterly, the amount will be \$2.44. If we now figure the compound interest continuously—not annually, semi-annually, or quarterly, but at each instant—the original \$1 grows to \$2.72 (to the nearest cent) at the end of one year. This limiting value 2.72 is the approximate value of the number e. The reader acquainted with the concept of limit will recognize that the number e just described is really

$$\lim_{t\to\infty}\left(1+\frac{1}{t}\right)^t.$$

We are now ready to give the formula for the number of primes from 1 to N. This number, which must of course depend on N, is usually denoted by $\pi(N)$. The Greek letter π corresponds to p in the English alphabet and refers to the first letter of the word *prime*. The π used in connection with the circle has no relation to the π used here as a notation for this particular function of N.

The first estimate concerning the magnitude of $\pi(N)$ seems to have been made independently (about 1800) by the French mathematician Adrien Legendre (1752–1833) and by Gauss. The formula is

$$\pi(N) \doteq \frac{N}{\log_e N},$$

where $\doteq$ signifies approximate value.

The following table gives values of $\pi(N)$ and the corresponding values of $N/\log_e N$. It also shows that the ratio of $\pi(N)$ to $N/\log_e N$ seems to approach 1. The values of $\pi(100{,}000{,}000)$ and $\pi(1{,}000{,}000{,}000)$ were obtained (about 1870) by special methods.

N	$\pi(N)$	$N/\log_e N$	$\pi(N) - N/\log_e N$	$\pi(N)\Big/\frac{N}{\log_e N}$
1,000	168	148	20	1.135
10,000	1,229	1,086	143	1.123
100,000	9,592	8,686	906	1.104
1,000,000	78,498	72,380	6,118	1.085
10,000,000	664,579	620,440	44,139	1.071
100,000,000	5,761,455	5,428,610	332,845	1.061
1,000,000,000	50,847,478	48,255,600	2,591,878	1.054

It was actually proved that the ratio

$$\frac{\pi(N)}{\dfrac{N}{\log_e N}}$$

tends to 1 as a limit as N approaches infinity. This remarkable result is one of the mathematical highlights of the nineteenth century and is known as the *Prime Number Theorem.* It was proved independently in 1896 by the French mathematician Jacques Hadamard (1865–) and by the Belgian mathematician C. J. de la Vallee Poussin (1866–).

7. The fundamental theorem of arithmetic

Before discussing the question of the existence of an infinite number of primes, we note that every composite number, no matter how large, may be decomposed into a product of primes (1: 12–21). For example, $15 = (3)(5)$; $36 = (2)(2)(3)(3) = (2^2)(3^2)$; $150 = (2)(3)(5^2)$. It is also true, although more difficult to prove, that this decomposition can be accomplished in only one way, apart from the order in which the factors are written. We find that $1665 = (3^2)(5)(37)$, and this number cannot be expressed in any other way as a product of primes.

These two results together constitute what is known as the *fundamental theorem of arithmetic or of number theory*. The theorem tells us that every natural number can be obtained by the single operation of the multiplication of specific primes, and the same number can never be the product of a different set of primes. We see now why the number 1 is not called a prime; it may be introduced into a factorization as frequently as we wish without changing the value of the number, and the uniqueness of factorization does not apply.

In our daily use of arithmetic, we unconsciously assume the fundamental theorem without realizing it. If, for example, the product of two integers a and b is exactly divisible by c, and if a and c have no factors in common, then b is divisible by c. Thus 144 is divisible by 8, so that if we write $144 = (9)(16)$, it follows that the 16 absorbs all the factors of 8, since 9 does not have a factor in common with 8. This is a result of the unique factorization of 144, and of 8, since we are really assuming that there is no other way of using up all the prime factors of 8 except through those of 16. Without going into further instances, we may say that all the divisibility properties of the natural numbers depend on the uniqueness of the factorization of a natural number into its prime factors (1: 28).

8. The number of primes is infinite

We shall now show by very simple reasoning that there is an infinite number of primes. Euclid was the first to do this, and his proof is still a model of beauty and simplicity. His method consists in showing that if we assume there is a largest prime, we arrive at a contradiction. Saying that there is no largest prime is tantamount to saying that there must be an infinite number of primes. To illustrate the method let us suppose that the number 89 is the largest prime. We shall then show that there exists a larger prime. The same method will work whether the largest assumed prime is 89 or $2^{3217} - 1$.

Form the product of all the primes from 2 through 89, and add 1 to the product. Let N be this result so that

$$N = (2)(3)(5) \cdots (83)(89) + 1.$$

This number N is very large, obviously greater than 89, and it must be either prime or composite. If it is prime, 89 is not the largest prime number, as we had originally supposed, and we have arrived at a contradiction. If it is composite, it may be factored into a product of primes. (We are now using only the first part of the fundamental theorem of arithmetic.) The number 2 is obviously not an exact divisor of N, since we see that, if we divide N by 2, we obtain a remainder of 1. Indeed, every prime in the series $3, 5, \cdots, 89$, when used as a divisor of N, leaves a remainder of 1. As every composite number must have prime divisors, we conclude that even the smallest prime divisor of N must be greater than 89, which contradicts our assumption that 89 is the largest prime. We may apply this method to any prime p which we assume to be the largest prime, and our proof is complete.

9. Formulas that yield only primes

Attempts have been made to give formulas that yield only primes (10: 80–81). For example, we may easily verify that the formula

$$x^2 - x + 41$$

will give a prime for every value of x that is a whole number, from $x = 0$ to $x = 40$. Notice however that the primes thus obtained will not be consecutive. Thus when $x = 0$, the value of the expression above is 41; when $x = 1$, its value is again 41; when $x = 2$, the value is 43; when $x = 4$, the value is 53. When $x = 40$, the value is 1601, which may be verified to be a prime. However when $x = 41$, the expression obviously reduces to 41^2, which is composite. This is one example showing that a property may hold in many instances and yet not be true in all cases. Another such formula is $x^2 + x + 17$, which yields a prime for each of the 17 integral values from $x = 0$ to $x = 16$. But we see that for $x = 17$, we obtain the composite number (17)(19).

Other such expressions could be given. The two cited are examples of polynomials in the single variable x. A polynomial is a sum of terms each of which has a numerical factor multiplied by a non-negative integral power of x. Explicitly, a polynomial has the form

$$a + bx + cx^2 + \cdots + lx^n$$

where $a, b, c, \cdots, l$ are numerical coefficients. The number n, the highest power of x, is called the *degree* of the polynomial. In the two examples given above, the degree of each of the polynomials is 2, and the numerical coefficients are integral; for the case $x^2 - x + 41, a = 41, b = -1, c = 1$.

The reader may wonder whether there is a polynomial in x that will yield only primes when integral values are substituted for x. The answer is *no*. It has been proved that *no polynomial with integral coefficients, irrespective of its degree, can yield only prime values when all possible natural numbers are substituted for x in the polynomial* (10: 80–81).

While no polynomial can yield only primes, an exponential expression has been found that always yields primes. In 1947 the American mathematician W. H. Mills proved the following theorem: *There exists some real number R for which the greatest integral value in R^{3^n} gives only primes as n assumes the infinite set of values* $1, 2, 3, \cdots$.

A real number is one such as $2\frac{1}{2}$ or $\sqrt{7}$. If in the formula R^{3^n}, we let $R = 2\frac{1}{2}$ and $n = 1$, we obtain $15\frac{5}{8}$; the largest integer in $15\frac{5}{8}$ is 15. Obviously this value of R does not give a prime even for $n = 1$, let alone for all values of n. In fact, the actual value of this real number R cannot be determined; we know only that there is such a number.

We have proved (Section 8) that there always exists a prime larger than any given prime. For example, $2^{11213} - 1$ is the largest prime known today, discovered in 1964 by two American mathematicians, Alexander Hurwitz and John L. Selfridge. This number, when written in the ordinary number system, contains nearly 3400 digits.

10. Dirichlet's theorem

Euclid's result on the infinitude of primes stood for more than 2000 years without essential modification. In the nineteenth century the German mathematician Lejeune Dirichlet (1805–1859) looked for sequences of natural numbers, other than the odd numbers, among which there would be an infinite number of primes. He stated and proved a far-reaching generalization of Euclid's result.

To express the fact that all integers are even or odd, we need only say that they are of the form $2n$ or $2n + 1$, where n is any natural number. We may equally well classify integers on the basis of their divisibility or non-divisibility by 3 instead of 2. Thus we may say that every natural number has one of the forms $3n$, $3n + 1$, $3n + 2$, where n assumes any integral value. For example, 16 is of the form $(3)(5) + 1$ while 23 is of the form $(3)(7) + 2$.

Let us now give some numbers of the form $3n + 1$:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$3n + 1$	4	7	10	13	16	19	22	25	28	31	34	37	40	43	46

We see that among these 15 numbers there are six primes: 7, 13, 19, 31, 37, 43. Dirichlet proved that in this infinite sequence of natural numbers of the special form $3n + 1$, there is an infinite number of primes. The same is true for numbers of the form $3n + 2$. In fact, if we form the arithmetical sequence or progression $an + b$, where a and b are relatively prime (have no common factor), we obtain a sequence containing an infinite number of primes.

The general theorem stated and proved by Dirichlet is that *in every arithmetical progression of the form* $an + b$ *(as* n *takes on the values* 1, 2, 3, $\cdots$) *there is an infinitude of primes.* His proof (2: 269–305) used very advanced methods of the calculus. Only a few years ago the American mathematician A. Selberg proved the same result by using elementary methods dealing with the natural numbers only.

CHAPTER

III

The Divisors of a Number

11. The number of divisors of a number

In considering the decomposition of a natural number into its prime factors, we have already noticed that each prime factor may occur any number of times. We might think of each natural number as a compound that is decomposed into its elements (prime factors) and each element may occur once, twice, or more often. Thus, $15 = (3)(5)$; $60 = (2^2)(3)(5)$; $144 = (2^4)(3^2)$. If we now consider all the prime and composite divisors, including 1 and the number itself, we find that these divisors play an important role in the development of number theory.

As an example, we see that while 15 has only two distinct prime factors, 3 and 5, it has four divisors, 1, 3, 5, 15. The number 60 has the distinct prime factors 2, 3, and 5, and it has the 12 divisors:

$$1, 2, 3, 4, 5, 6, 10, 12, 15, 20, 30, 60.$$

These divisors may be counted systematically as follows. Since the prime decomposition of 60 is $(2^2)(3)(5)$, we note first that the two factors 2 give rise to three divisors, 1, 2, and 2^2; the factor 3 gives rise to two divisors 1 and 3; the factor 5 gives rise to two divisors 1 and 5. Now each of the three divisors 1, 2, and 4 may be combined with each of the two divisors 1 and 3; this gives rise to the six divisors:

$$1, 2, 4;\ 3, 6, 12.$$

Each of these six divisors may be combined with each of the divisors 1 and 5. In this way we obtain the 12 divisors:

$$1, 2, 4;\ 3, 6, 12;\ 5, 10, 20;\ 15, 30, 60.$$

The number of divisors is found from the following rule (1: 22–23). Let the natural number N have the factorization

$$N = p^a q^b r^c \cdots,$$

where $p, q, r, \cdots$ are the prime factors raised to the powers $a, b, c, \cdots$, respectively. The number of divisors of N, denoted by $d(N)$, is found by the formula

$$d(N) = (a + 1)(b + 1)(c + 1) \cdots,$$

where the dots mean that we continue until we have exhausted all the exponents.

When $N = 60 = (2^2)(3)(5)$, we see that $p = 2, q = 3, r = 5; a = 2, b = 1, c = 1$, so that $d(60) = (2 + 1)(1 + 1)(1 + 1) = 12$.

Note that the number of divisors depends only on the exponents of the prime factors and not on the prime factors themselves. If $N = (7^2)(3)(5)$, the number of divisors of N is still 12.

12. The sum of the divisors

Let us next consider the sum of the divisors. For $N = 15$, this sum is equal to $1 + 3 + 5 + 15 = 24$. For $N = 60$, the sum is $1 + 2 + 3 + 4 + 5 + 6 + 10 + 12 + 15 + 20 + 30 + 60 = 168$. Using the systematic method for finding all the divisors of 60, we obtain the sum of the divisors in the following form:

$$1 + 2 + 2^2 + 3(1) + 3(2) + 3(2^2) + 5(1) + 5(2) \\ + 5(2^2) + 15(1) + 15(2) + 15(2^2),$$

and this sum may be put in the more abbreviated form

$$(1 + 2 + 2^2)(1 + 3)(1 + 5).$$

In general, if $N = p^a q^b r^c \cdots$, the sum of the divisors denoted by $\sigma(N)$ is given by the formula (1: 23):

$$\sigma(N) = (1 + p + p^2 + \cdots + p^a)(1 + q + q^2 + \cdots + q^b) \\ \cdot(1 + r + r^2 + \cdots + r^c) \cdots,$$

where again we take all the products until the prime factors of N are exhausted. The sum of the divisors is denoted by $\sigma(N)$ because σ is the Greek letter corresponding to our *s*, the first letter in *sum*. The following examples illustrate the fact that the formula gives the same sums we found above by direct addition:

$$N = 15 = (3)(5);\ p = 3, q = 5;\ a = 1, b = 1; \\ \sigma(15) = (1 + 3)(1 + 5) = 24;$$

$$N = 60 = (2^2)(3)(5);\ \sigma(60) = (1 + 2 + 2^2)(1 + 3)(1 + 5) = 168.$$

We note that the formula for $\sigma(N)$ involves not only the exponents but also the prime factors themselves.

13. The Euler phi-function or totient

The two expressions $d(N)$ and $\sigma(N)$ are examples of number-theoretic functions. The term *function* is used because the values of $d(N)$ and $\sigma(N)$ depend upon N. They are called number-theoretic because the functions always yield integers when N is itself an integer. Another such function is $\pi(N)$, the number of primes from 1 to N, which we discussed in Section 6. On the other hand, the function $\sqrt{N}$ is not of this type, since, for example, $\sqrt{2}$ is not an integer.

We now consider another function, called the totient of N or the phi-function of N, denoted by $\phi(N)$. This was first introduced about 1760 by the Swiss mathematician Leonard Euler (1707–1783). By $\phi(N)$ we mean the number of natural numbers less than N that have no factor in common with N. Thus $\phi(6) = 2$, since there are only two numbers, 1 and 5, less than 6 having no factor in common with 6. In contrast to this, $\pi(6) = 3$ since there are three prime numbers, 2, 3, and 5, less than 6 (1 is not a prime). The following values of $\phi(N)$ for $N = 1, 2, \cdots, 15$ —assigning the value 1 to $\phi(1)$—may be easily verified:

$$\begin{array}{lll}
\phi(1) = 1 & \phi(6) = 2 & \phi(11) = 10 \\
\phi(2) = 1 & \phi(7) = 6 & \phi(12) = 4 \\
\phi(3) = 2 & \phi(8) = 4 & \phi(13) = 12 \\
\phi(4) = 2 & \phi(9) = 6 & \phi(14) = 6 \\
\phi(5) = 4 & \phi(10) = 4 & \phi(15) = 8
\end{array}$$

We observe that the values of $\phi(N)$ are irregular. In spite of this, Euler gave an exact formula for $\phi(N)$:

If $N = p^a q^b r^c \cdots$, then

$$\phi(N) = N(1 - 1/p)(1 - 1/q)(1 - 1/r) \cdots,$$

where the dots mean that we continue until all the prime factors of N are exhausted (10: 110–11). The following examples demonstrate the use of the formula:

$$15 = (5)(3); \phi(15) = 15(1 - 1/3)(1 - 1/5) = (15)(2/3)(4/5) = 8;$$

$$144 = (2^4)(3^2); \phi(144) = 144(1 - 1/2)(1 - 1/3) = 48.$$

The second example tells us that there are 48 natural numbers less than 144 that have no factor in common with 144.

Of course it is quite obvious that if N is a prime p, then $\phi(p) = p - 1$, since each one of the $p - 1$ numbers less than p can have no factor in common with p. In particular, $\phi(2) = 1, \phi(3) = 2, \phi(5) = 4, \phi(7) = 6, \phi(11) = 10, \cdots, \phi(p) = p - 1$.

The following interesting property of the Euler phi-function is worth mentioning (10: 119). If we denote the divisors of N, including N itself, by the letters $u, v, w, \cdots, N$, then $\phi(u) + \phi(v) + \phi(w) + \cdots + \phi(N)$ is equal to N. Note the following examples:

If $N = 9$, the divisors of N are 1, 3, and 9, and $\phi(1) + \phi(3) + \phi(9) = 1 + 2 + 6 = 9$;
if $N = 30$, the divisors are 1, 2, 3, 5, 6, 10, 15, 30, and $\phi(1) + \phi(2) + \phi(3) + \phi(5) + \phi(6) + \phi(10) + \phi(15) + \phi(30) = 1 + 1 + 2 + 4 + 2 + 4 + 8 + 8 = 30$.

A property common to all the three functions $d(N)$, $\sigma(N)$, and $\phi(N)$ is the following: if N is the product of two natural numbers m and n which have no factor in common, then

$$d(mn) = d(m)d(n), \quad \sigma(mn) = \sigma(m)\sigma(n), \quad \phi(mn) = \phi(m)\phi(n).$$

Thus in each case the function of the product of the two factors is equal to the product of the functions of the separate factors. The following examples illustrate this common property:

$$\begin{aligned} d(15) &= d(3)d(5), \text{ since } d(15) = 4, d(3) = 2, d(5) = 2; \\ \sigma(15) &= \sigma(3)\sigma(5), \text{ since } \sigma(15) = 24, \sigma(3) = 4, \sigma(5) = 6; \\ \phi(15) &= \phi(3)\phi(5), \text{ since } \phi(15) = 8, \phi(3) = 2, \phi(5) = 4. \end{aligned}$$

This property is usually referred to as the *multiplicative* property (1: 48) and is not possessed by the function $\pi(N)$, as we see from an example: $\pi(12)$ is not $\pi(3)\pi(4)$, since $\pi(12) = 5$, $\pi(3) = 1$, $\pi(4) = 2$.

14. Perfect numbers

The formula for the sum of the divisors of a number N has many interesting consequences. One of these is connected with perfect numbers. *A perfect number N is a number for which the sum of its divisors including* 1 *but excluding N is equal to N itself.*

A simple example of a perfect number is 6, since the divisors are 1, 2, and 3, and their sum is 6. The next perfect number is 28. Its divisors are 1, 2, 4, 7, and 14, and $1 + 2 + 4 + 7 + 14 = 28$. Numerologists from time immemorial have attributed special significance to the numbers 6 and 28, because God created the world in 6 days, and 28 is the number of days required for the moon to circle the earth.

Seventeen perfect numbers were known in 1953. The twelfth one, containing 37 digits, is

2,658,455,991,569,831,744,654,692,615,953,842,176.

The Greeks knew the first five: 6, 28, 496, 8,128, and 33,550,336. In fact, Euclid proved (12: 80–81) that *if p is a natural number that* *makes* $(2^p - 1)$ *a prime, then* $2^{p-1}(2^p - 1)$ *is a perfect number*. The following examples illustrate this relationship:

When $p = 2$, $2^p - 1 = 3$, which is a prime number.
Then $2^{p-1} = 2$, and $(2)(3) = 6$, which is a perfect number.
When $p = 3$, $2^p - 1 = 7$, which is a prime number.
Then $2^{p-1} = 2^2 = 4$, and $(4)(7) = 28$, which is a perfect number.
When $p = 5$, $2^p - 1 = 31$, which is a prime number.
Then $2^{p-1} = 2^4 = 16$, and $(16)(31) = 496$, which is a perfect number.

If we let $p = 4$, we find that $2^p - 1 = 15$, which is not prime. Then $2^{p-1} = 8$, and $(8)(15) = 120$. Euclid's formula does not tell us that 120 is a perfect number. Neither does it tell us that it is not perfect. Actually every even value of p except 2 will make $2^p - 1$ a composite number. Furthermore when p itself is a prime, $2^p - 1$ is not necessarily a prime. For example, when $p = 11$, $2^p - 1 = 2047$, which is $(23)(89)$. In these cases the formula fails.

Although Euclid's formula gives us a way of discovering even perfect numbers, the question may be asked: Are there other formulas that would yield even perfect numbers? The answer is *no*. Euler proved (10: 91–93) that *if a number is to be an even perfect number, it must be expressible in the form* $2^{p-1}(2^p - 1)$, *where the value of* p *makes* $(2^p - 1)$ *a prime*. Thus the entire question of discovering even perfect numbers depends on finding the values of p for which $(2^p - 1)$ is prime. The search for such values of p is still going on.

Numbers of the form $2^p - 1$ with p a prime are called Mersenne numbers after the Franciscan Father M. Mersenne (1588–1648). It is clear that once we know a Mersenne number which is prime, then $2^{p-1}(2^p - 1)$ would be a perfect number by Euclid's result. With the advent of computing machines, large Mersenne primes have been found, and the known perfect numbers are much more numerous than the Greeks anticipated. In 1952 the Mersenne prime $(2^{3217} - 1)$ was discovered, and this gave rise to the perfect number $2^{3216}(2^{3217} - 1)$, which contains 1,937 digits.

Euclid's formula, as already indicated, gives only even perfect numbers. No formula has been devised that will give odd perfect numbers In fact, not a single odd perfect number has as yet been discovered; and calculations indicate (10: 359*a*) that none exists less than $(1.4)(10^{14})$.

We shall now indicate the proof of a well known property of even perfect numbers, that they always end in 28 or in 6 preceded by an odd digit. This property is true not only for the perfect numbers but also for every member of the sequence a_n where $a_n = 2^{2n}(2^{2n+1} - 1)$. Obviously, the even perfect numbers $2^{p-1}(2^p - 1)$ are included among these if we confine ourselves to the odd primes p, so that $p - 1$ will be an even number $2n$ and $p = 2n + 1$. (The only even perfect number not included in the sequence a_n is 6, which results when $p = 2$.)

Let us first consider the odd values of n in the sequence a_n, namely, $a_1, a_3, a_5, \cdots a_{2k+1}$. We may verify the following by direct computation.

$$\begin{aligned} a_1 &= 28 \\ a_3 &= 256a_1 + (60)(16) \\ a_5 &= 256a_3 + (60)(16^2) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{2k+1} &= 256a_{2k-1} + (60)(16^k) \end{aligned}$$

Since $a_1 = 28$, we see that $a_3 = (256)(28) + (60)(16)$. When the product (256)(28) is divided by 100, the remainder is 68; when the product (60)(16) is divided by 100, the remainder is 60. When the sum of $68 + 60$ is divided by 100, the remainder is 28. Hence a_3 has the remainder 28. In a similar manner we find that $a_5, a_7, \cdots, a_{2k+1}$ all leave remainders of 28 when divided by 100.

Proceeding to the terms of a_n with even subscripts, namely, $a_2, a_4, \cdots, a_{2k+2}$, we may verify the following:

$$\begin{aligned} a_2 &= (16)(31) = 496 \\ a_4 &= 256a_2 + (240)(16) \\ a_6 &= 256a_4 + (240)(16^2) \\ a_8 &= 256a_6 + (240)(16^3) \\ a_{10} &= 256a_8 + (240)(16^4) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{2k+2} &= 256a_{2k} + (240)(16^k). \end{aligned}$$

The first term a_2 leaves a remainder of 96 when divided by 100. Similarly, a_4 leaves a remainder of 16; a_6 leaves a remainder of 36; a_8 leaves a remainder of 56; a_{10} leaves a remainder of 76. Finally, a_{12} leaves the same remainder as a_2.

Thus we have verified that in the sequence $a_1, a_2, a_3, \cdots$ the a's with odd subscripts end in 28 while those with even subscripts end in 6, preceded by an odd digit.

15. Amicable numbers

Two numbers are said to be amicable if the sum of the divisors of the first number is equal to the second number itself and if the sum of the divisors of the second number is equal to the first number. Here the divisors of a number include all the divisors except the number itself (see Section 11). The smallest pair of amicable numbers is 220 and 284. To verify that they are amicable, let us first compute the sum of the divisors of $220 = (2^2)(5)(11)$. By the formula in Section 12, $\sigma(220) = (1 + 2 + 2^2)(1 + 5)(1 + 11) = 504$. Since this sum includes the number itself, we subtract 220 from 504 to obtain 284. Similarly, $284 = (2^2)(71)$, so that $\sigma(284) = (1 + 2 + 2^2)(1 + 71) = 504$. When we subtract the number itself, 284, we obtain 220, the first number.

The ancients believed that two people wearing talismans bearing these numbers would be friendly and in harmony, hence amicable. Such numbers were known to the Greeks about A.D. 320, more than 500 years after the discovery of perfect numbers.

During the middle ages the Arabs were attracted to the study of amicable numbers, and the search for additional pairs has continued to the present day. In 1747 Euler published a list of 60 pairs; in 1866 an amateur mathematician found a remarkably small pair, 1184 and 1210 (10: 96–100).

CHAPTER

IV

Some Proven Facts of Number Theory

In this chapter we shall state some results for which we shall give numerical illustrations without formal proofs.

16. Chebyshev's theorem

The Russian mathematician Chebyshev (1821–1894) proved that *between every integer greater than* 1 *and its double, there is at least one prime.* Between 2 and 4, for example, there is the prime 3. Between 6 and 12 there are two primes, 7 and 11.

The proof of this result, while not lengthy, requires some mathematical maturity for its understanding (12: 86).

17. Fermat's little theorem

In the year 1640 Fermat mentioned the following result in a letter to another French mathematician (1: 46): *If p is a prime that is not a divisor of the natural number a, then the expression $a^{p-1} - 1$ is exactly divisible by p.* Following are two examples:

If $a = 2$ and $p = 7$, then $2^{7-1} - 1 = 63$, which is divisible by 7;
if $a = 8$ and $p = 5$, then $8^{5-1} - 1 = 4095$, which is divisible by 5.

The reason for the restriction on a is obvious. If a is divisible by p, then a^{p-1} is also divisible by p, and hence $a^{p-1} - 1$ cannot be divisible by p.

There is no doubt that Fermat himself had a proof for this theorem, but no record of it exists. Euler was the first to publish a proof; many others have since been given.

A converse of Fermat's little theorem has been proved, but it is rather involved and will not even be stated (10: 327, 339).

Fermat's little theorem is so called to distinguish it from another result stated by Fermat and known as Fermat's last theorem, for which no proof has yet been found (Section 28).

From the statement of the little theorem, it is clear that when p is not a prime, the result is in doubt. For example, if $p = 4$ and $a = 3$, $3^{4-1} - 1 = 26$, which is not divisible by 4. However, when $p = 4$ and $a = 5$, $5^{4-1} - 1 = 124$, which is divisible by 4.

18. Euler's generalization of Fermat's little theorem

In 1760 Euler gave the following extension of Fermat's little theorem, which includes the case when p is not prime: *Let m and a be any two natural numbers with no factor in common. Then the expression $a^{\phi(m)} - 1$ is divisible by m.* The $\phi(m)$ is the Euler phi-function (Section 13); that is, the number of natural numbers less than m having no factor in common with m (1: 47). For example, let $m = 10$, so that $\phi(10) = 4$. If $a = 3$, $3^4 - 1 = 80$, and 80 is divisible by 10.

That Euler's result is a true generalization is seen by substituting p for m. Then $\phi(m) = \phi(p) = p - 1$, and the Euler result becomes $a^{p-1} - 1$, which is Fermat's little theorem.

19. Wilson's theorem

In 1770 Edward Waring published the following theorem, which he ascribed to his student John Wilson (1741–1793): *If p is a prime, then the product of all the natural numbers up to and including $p - 1$, that is,* $(1)(2)(3)(4) \cdots (p - 1)$, *increased by* 1, *is divisible by p.* In other words,

$$\frac{(1)(2)(3) \cdots (p-1) + 1}{p}$$

is a natural number (10: 259).

This result may also be stated in another form: *The expression* $(p - 1)! + 1$ *is divisible by p.* For example, if $p = 7$, then $6! + 1 = 721$, which is divisible by 7.

The first proof of this theorem was given in 1770 by the French mathematician J. L. Lagrange (1736–1813).

The converse to Wilson's theorem is: *If the product of all the numbers from* 1 *to* $(n - 1)$ *increased by* 1 *is exactly divisible by n, then n must be a prime* (10: 261).

To prove the converse, suppose n is not a prime so that $n = ab$, where a lies between 1 and n. Then a must occur as a factor in $(1)(2)(3) \cdots (n - 1) = (n - 1)!$, and hence $(n - 1)! + 1$ could not be divisible by a, much less divisible by n. Thus we have reached a contradiction, and n must be a prime.

However, this is not a practical test for primeness since for any sizable n, the value of $(n - 1)! + 1$ would be a very large number, and it would involve much computation to determine whether n is an exact divisor.

20. Fermat's two-square theorem

If we ignore the even prime 2, the other primes may be arranged in two classes:

A: 5, 13, 17, 29, 37, 41, $\cdots$;
B: 3, 7, 11, 19, 23, 31, $\cdots$.

All the primes in A leave a remainder of 1 when divided by 4; those in B leave a remainder of 3 when divided by 4. We may now state the following result discovered by Fermat: *Any prime of the class A can be represented as the sum of the squares of two natural numbers.* For example, $5 = 1^2 + 2^2$; $13 = 2^2 + 3^2$; $593 = 23^2 + 8^2$. *No prime of the class B can be represented as a sum of two squares.*

The proof of these results is not simple and can be understood only by an experienced mathematician (1: 115–20).

In the examples given for the class of primes A, each prime is represented as the sum of the squares of two natural numbers, and cannot be the sum of two other squares. If now we consider a number like 65, which is not itself a prime but is the product of the two primes 5 and 13, each of the class A, there are exactly two ways of expressing 65 as a sum of two squares:

$$65 = 4^2 + 7^2 \quad \text{and} \quad 65 = 1^2 + 8^2.$$

We shall now state a more general result for the number N which is the product of any number of prime factors, each of the form $4h + 1$. This is the same as saying that each prime factor when divided by 4 leaves a remainder of 1 and thus belongs to class A.

When N has r distinct prime factors each of the form $4h + 1$, there are exactly 2^{r-1} ways of expressing N as a sum of the squares of two natural numbers which are relatively prime to each other.

In particular, when N has only one such prime factor (when N is itself prime), as in the case $N = 13$, already cited, $r = 1$ and $2^{1-1} = 2^0 = 1$. Hence, there is only one way of writing a prime of the form $4h + 1$ as a sum of two squares, apart from the obvious possibilities of interchanging the squares and changing their signs.

It is apparent that the number $125 = 5^3$ may be written as a sum of two squares in two ways: $125 = 11^2 + 2^2$, and $125 = 10^2 + 5^2$. In this case $r = 1$, since 5 is the only distinct prime factor; hence $2^{r-1} = 2^{1-1} = 2^0 = 1$. However, in the representation $125 = 10^2 + 5^2$ the numbers 10 and 5 are not relatively prime, and there is no contradiction.

On the other hand, when $N = 65 = (5)(13)$, $r = 2$ and $2^{r-1} = 2^1 = 2$, and there are two ways of expressing 65 as a sum of two squares.

If the number $N = (5)(13)(17) = 1105$, we find the four ways of writing N as a sum of two squares: $4^2 + 33^2$, $9^2 + 32^2$, $12^2 + 31^2$, and $23^2 + 24^2$. Here $r = 3$ and $2^{3-1} = 4$.

The first recorded proof of the number of representations was given by Euler in 1746, although more than 100 years earlier Fermat indicated in a letter to another mathematician that he had a proof. There is no reason to doubt Fermat's claim.

21. Lagrange's theorem—sum of four squares

The following is another result first announced by Fermat: *Every natural number N, no matter how large, may be written as a sum of four or fewer squares* (1: 124–26).

Euler made many efforts to find a proof, but without success. The first proof was given in 1770 by Lagrange. It is not simple and required considerable insight, so that Lagrange is entitled to the honor of having this theorem named for him.

The following examples illustrate Lagrange's theorem:

$$600 = 24^2 + 4^2 + 2^2 + 2^2 = 576 + 16 + 4 + 4;$$
$$600 = 20^2 + 10^2 + 10^2;$$
$$102 = 10^2 + 1^2 + 1^2;$$
$$102 = 8^2 + 5^2 + 3^2 + 2^2.$$

Fermat's two-square theorem and Lagrange's theorem together tell us that some numbers may be written as the sum of two squares, but *every* number may be written as a sum of, at most, four squares.

Mathematicians have successfully investigated the number of representations of a number as a sum of four squares, but the formula will not be stated here (1: 128; 11: 175–78).

22. Sum of three squares

A more difficult problem is the representation of N as the sum of three squares. Certain numbers such as 15 and 23 cannot be represented as the sum of three squares.

In general, *all numbers are representable as a sum of three squares, except those of the form $4^r(8k + 7)$, where r and k may be any natural numbers or zero.* A few examples will illustrate this theorem. When $r = 0$ and $k = 1$, the expression equals 15, which cannot be represented as the sum of three squares. Similarly, when $r = 0$ and $k = 2$, or when $r = 1$ and $k = 5$, we obtain 23 or 188, respectively, neither of which can be represented as the sum of three squares. On the other hand, 29 is not of the form $4^r(8k + 7)$, and it can be represented as the sum of squares as follows: $29 = 2^2 + 3^2 + 4^2 = 2^2 + 5^2$.

23. Waring's theorem

It seems natural that once we have disposed of the problem of expressing a number as a sum of squares, we may proceed to think of expressing a number as a sum of cubes or as a sum of higher powers.

In 1770, the same year in which Lagrange proved his famous theorem that every number may be expressed as a sum of four squares, the English mathematician Edward Waring (1734–1798) announced his belief that every number may be expressed as the sum of a limited number of cubes, fourth powers, or higher powers. Waring offered no proof, but he arrived at this conclusion from much numerical evidence.

In 1909 David Hilbert (1862–1943), one of the most famous mathematicians of modern times, proved that *every number*, no matter what its size, *may be written as the sum of a finite number of specified powers.* For example, if we specify that the sum should be of kth powers, Hilbert's result tells us that every number N may be written as the sum of a finite number of kth powers, and this finite number depends only on the power specified and not on the size of N. Lagrange's theorem is a special case of Waring's theorem when $k = 2$ and the limited number of kth powers is actually four.

Note that $N = 1^k + 1^k + 1^k \cdots$, to N terms, will not lead to a limited number of kth powers as N increases in size.

Hilbert used intricate methods of the calculus, and he proved only that the number of kth powers is limited; he could not determine in general how many terms it would take to express the number N. Later, other proofs were given by mathematicians in England, Russia, and America, and their work included some special results for particular values of k. Thus it is known that every natural number is expressible as a sum of 9 or fewer cubes, as a sum of 19 or fewer fourth powers. Some numbers take the maximum of 9 when written as a sum of cubes. For example, 23 may be written in only one way as the sum of cubes: $23 = 2^3 + 2^3 + 1^3 + 1^3 + 1^3 + 1^3 + 1^3 + 1^3 + 1^3$, and no fewer than 9 cubes will do. As a matter of fact, it has been shown that 23 is the only number that takes the full complement of 9 cubes.

For fourth powers we find that 79 takes 19 of them: $79 = 2^4 + 2^4 + 2^4 + 2^4 + 1^4 + 1^4 + 1^4 + 1^4 + 1^4 + 1^4 + 1^4 + 1^4 + 1^4 + 1^4 + 1^4 + 1^4 + 1^4 + 1^4 + 1^4$.

CHAPTER

V

Some Conjectures and Unsolved Problems

A fascinating aspect of number theory is the great variety of problems that have been proposed and that often are simple enough for a layman to understand, though their solutions have eluded the efforts of mathematicians. The nature of some of the unsolved problems will be indicated briefly.

24. Twin prime problem

The name *twin prime* is given to the pairs of primes which differ by 2. Thus 3 and 5 are twin primes. A short list of twin primes is: 3 and 5; 5 and 7; 11 and 13; 17 and 19; 29 and 31; 41 and 43; 59 and 61; 10,006,427 and 10,006,429.

Twin primes may be found as far out as we are able to go in the table of primes, although they become scarcer as we move out. A striking result concerning twin primes occurs if we form the infinite sum of the reciprocals of the primes that form the twins, as follows:

$$1/3 + 1/5 + 1/7 + 1/11 + 1/13 + 1/17 + 1/19 + 1/29 + 1/31 + \cdots.$$

This sum will actually be a finite number, or, as the mathematician says, this sum will *converge*. From this fact two conclusions are possible: either the twin primes come to an end, or they are infinite in number but so scarce that their sum still has a limit. The conjecture about the twin prime problem is that the number of pairs of twin primes is unlimited. No proof of this conjecture is known.

25. A conjecture on a quadratic progression

The problem about to be stated appears so elementary that it seems almost unbelievable that its solution has not been found. We have already seen that among the arithmetical progressions of the form $an + b$, an infinite number of primes occurs as n takes on the values 1, 2, 3, $\cdots$. However, we do not know whether the progression $n^2 + 1$, for $n = 1, 2, 3, \cdots$, contains an infinite number of primes (4: 18–19).

The following table shows the values of $n^2 + 1$ for $n = 1, 2, \cdots, 16$:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$n^2 + 1$	2	5	10	17	26	37	50	65	82	101	122	145	170	197	226	257

We see from the table that among the first 16 numbers of the form $n^2 + 1$ there are only seven primes. The conjecture is that there is an infinite number of primes in the infinite sequence $n^2 + 1$.

26. Is there a prime between consecutive squares?

Chebyshev proved that between any natural number and its double there is at least one prime, but it is not known whether there is a prime between every two consecutive squares.

When the numbers studied are small, one may actually find the primes. Thus between 6^2 and 7^2 are 37, 41, and 47. Between 11^2 and 12^2 we find 127, 131, 137, 139. However, no mathematician has as yet found an answer to the question in general.

27. The Goldbach conjecture

Christian Goldbach (1690–1764), an obscure eighteenth century mathematician, carried on an extensive correspondence with Euler on problems of number theory. In this correspondence Goldbach considered the following two questions:

A. Is every even number the sum of two odd primes? (Of course the numbers 2 and 4 would be ruled out.)

B. Is every odd number the sum of three odd primes? (Now 3, 5, and 7 must be ruled out.)

It is easy to see that if we could answer A in the affirmative, B would also be true. For if N is an odd number and if p is any odd prime smaller than N, then $N - p$ is even. If now $N - p$ could be written as the sum of two odd primes, q and r, then N itself could be expressed as the sum of three odd primes, namely, $N = p + q + r$.

For small values we may verify A. Thus:

$$
\begin{array}{rl}
6 = 3 + 3 & 14 = 3 + 11 = 7 + 7 \\
8 = 3 + 5 & 16 = 3 + 13 = 5 + 11 \\
10 = 3 + 7 & 18 = 5 + 13 = 7 + 11 \\
12 = 5 + 7 & 20 = 3 + 17 = 7 + 13.
\end{array}
$$

These verifications have been made for all even numbers up to 100,000.

For small values, B may also be verified. Thus:

$$\begin{array}{ll} 9 = 3 + 3 + 3 & 25 = 3 + 5 + 17 = 7 + 7 + 11 \\ 11 = 3 + 3 + 5 & 39 = 7 + 13 + 19 = 13 + 13 + 13 \\ 19 = 3 + 5 + 11 & 57 = 3 + 7 + 47 = 19 + 19 + 19. \end{array}$$

While no proof of the Goldbach conjectures has been given, certain advances have been made, especially in the attempts to prove conjecture B. One of these was due to the Russian mathematician J. Vinogradoff, who in 1937 showed that every sufficiently large odd number is the sum of three odd primes. However, the proof is an existence proof (see Section 9), and no one knows how large the numbers are that can be written as the sum of three odd primes. The methods Vinogradoff used are intricate methods of the calculus and not those involving the natural numbers.

The difficulties involved in the Goldbach conjecture arise from the fact that we are trying to decompose a natural number into a sum of primes. But the primes and their properties all depend on the operation of multiplication. All questions relating to such problems are included in additive number theory.

28. Fermat's last theorem

Perhaps the most famous conjecture in all number theory is Fermat's last theorem, so called to distinguish it from Fermat's little theorem.

Fermat seems to have had very little interest in publishing any of his results. He corresponded extensively with other mathematicians of his day, and these correspondents recognized his great originality. Had he published his ideas as they occurred to him, the world would have had to credit him with more than the discovery of numerous original results in number theory. He might have been regarded, along with René Descartes (1596–1650), as the co-inventor of analytic geometry. He would have shared the honor of discovering the differential calculus with Sir Isaac Newton (1643–1716) and with Gottfried Leibniz (1646–1716). Finally, he would have been recognized, along with Blaise Pascal (1623–1662), as having laid the foundations of the theory of probability.

Everyone who has studied plane geometry knows the Pythagorean Theorem (about 570 B.C.): *In a right triangle the square of the hypotenuse is equal to the sum of the squares of the other two sides.* If, therefore, numerical values are assigned to any two of the sides, the value of the third side is found by a simple calculation. As a problem in number theory, however, our interest lies in the study of those integral values of the hypotenuse for which the sides are also integers. Thus when the hypotenuse is 5, the other two sides are 3 and 4, since $3^2 + 4^2 = 5^2$; when the hypotenuse is 13, the other two sides are 5 and 12, since $5^2 + 12^2 = 13^2$. But this relationship does not hold for arbitrary integral values of the hypotenuse such as 2, 3, or 4.

Fermat encountered the problem of right triangles with integral sides in a work on number theory by the Greek mathematician Diophantus (about A.D. 250). Diophantus considered the problem of separating certain integral squares into a sum of two other integral squares. In the margin of his copy of Diophantus' book Fermat left the following note: "However, it is impossible to write a cube as the sum of two cubes, a fourth power as the sum of two fourth powers; and, in general, any power beyond the second as the sum of two similar powers. For this I have discovered a truly wonderful proof but the margin is too small to contain it."

This is the celebrated Fermat's last theorem, which still remains unsolved although some of the outstanding mathematicians of the last three centuries have tried to prove or to disprove it.

Algebraically stated, this conjecture says that *it is impossible to find three natural numbers a, b, c, which will satisfy the equation* $a^n + b^n = c^n$, *if n is an integer greater than* 2. If any of the quantities a, b, or c is zero, we obtain the "trivial" solutions, $a = 0$, $b = c$, etc. Such solutions are not to be considered.

The simplest case of Fermat's theorem occurs when the exponent is 4, so that $a^4 + b^4 = c^4$. Fermat proved this case to be impossible (1: 161–63), and other mathematicians have proved the impossibility for many other exponents.

Mathematicians believe that when x, y, or z does not have a factor in common with n, then the equation $x^n + y^n = z^n$ is impossible in natural numbers when $n > 2$. But even this has not yet been proved or disproved.

It has been shown that if $x^p + y^p = z^p$ has a solution when p is an odd prime that is not a factor of x, y, or z, and if $(2^{p-1} - 1)$ is not divisible by p^2, then Fermat's conjecture is correct.

For example, we may conclude from this last remark that there are no natural numbers x, y, and z, which are not multiples of 7 and for which $x^7 + y^7 = z^7$. When $p = 7$, $2^{7-1} - 1 = 2^6 - 1 = 63$, which is not divisible by 7^2.

The same procedure proves that $x^p + y^p = z^p$ is impossible for all x, y, and z that have no factors in common with p for all primes p up to 1093. Since it has been shown, however, that $2^{1092} - 1$ is divisible by $(1093)^2$, the truth or falsity of the conjecture is left in doubt (4: 73). Using similar criteria, mathematicians have proved that Fermat's conjecture is true for all exponents up to 250,000,000, provided no one of the x, y, and z has a factor in common with p.

When x, or y, or z has a factor in common with n, some mathematicians doubt that the conjecture is true at all.

Whether or not Fermat actually had a proof is a matter of speculation. It seems most unlikely that we shall ever find out. In all Fermat's statements on his discoveries, where he claimed to have a proof, everything he said has been substantiated. His guess that a certain sequence of natural numbers was composed entirely of primes (see Section 30) was later shown to be incorrect, but Fermat had never claimed that he had a proof of this.

The interest aroused by Fermat's last theorem has been so great that it is not surprising to learn that an award was offered for its solution. In 1908 a German mathematician who had worked on the problem offered a prize of 100,000 marks (about $25,000 at the time) for the first complete proof. This problem continues to attract amateur mathematicians, and numerous incorrect "proofs" have been submitted.

The reader may wonder what effect a proof of Fermat's last theorem would have on the progress of mathematics. Would its solution have any bearing on initiating new mathematics? In seeking a solution of this problem, mathematicians have been stimulated to create new types of number theory and have contributed greatly to the development of other aspects of mathematics. Once this problem is solved its value as a stimulant for research will cease. It is very likely, however, that the method itself may be applied to other problems still unsolved.

29. The method of infinite descent

The method of infinite descent, a modification of mathematical induction, is a powerful tool used in proving Fermat's conjecture for the particular exponents for which the conjecture has been verified. Fermat used this technique to show that certain equations do not have solutions in natural numbers. For example, Fermat proved in this way that when all three sides of a right triangle can be expressed in natural numbers, the area can never be the square of a natural number.

Briefly, the method consists in expressing the problem in the form of an equation whose solution we are seeking in natural numbers. We assume that the equation has a solution, and then we show that there is a contradiction. We accomplish this by using certain facts of number theory and of algebra to obtain a new equation of the same form as the original, but one whose solution is in smaller natural numbers than those assumed. By repeating this process we eventually come to the stage where one of the numbers in the solution is zero. The contradiction lies in the fact that zero is not a natural number. Thus a solution of the equation in natural numbers is not possible. (See Section 46.)

For example, the proof of the impossibility of $x^4 + y^4 = z^4$ in natural numbers is made by assuming that there is a solution $x = a, y = b, z = c$, where a, b, and c are natural numbers. From this assumed solution it can be shown that there is another set of natural numbers a', b', and c' for which $(a')^4 + (b')^4 = (c')^4$, but with c' less than c. By repeating this process, we obtain a set of solutions with a succession of z's, each smaller than its predecessor. Eventually we arrive at a solution where the value of z is zero, and hence not a natural number, as was assumed. This is the contradiction (1: 162).

While Fermat himself had proved the impossibility of satisfying the equation $x^4 + y^4 = z^4$ with natural numbers, many years elapsed before the proof of the impossibility of satisfying $x^3 + y^3 = z^3$ was given by Euler in 1753. But Euler's proof lacked rigor at one point. In 1798 the French mathematician A. M. Legendre (1752–1833) gave a complete proof.

For some time we have known that Fermat's conjecture is correct for n up to about 700. More recently, with the aid of electronic computers it has been shown that the conjecture is correct for exponents to about 4000. However, no method seems to give promise of answering the conjecture in general, either in the affirmative or in the negative.

CHAPTER

VI

Fermat Numbers and Regular Polygons

In this chapter we shall discuss the constructibility of regular polygons and their relation to Fermat numbers. By a geometrical construction the Greeks meant a construction performed using only the straight edge and compasses. With these tools they attempted to solve the three famous problems of antiquity. They wanted (a) to trisect the general angle, (b) to duplicate the cube (construct the side of a cube having twice the volume of a given cube), and (c) to square the circle (construct the side of a square whose area is equal to that of a given circle). In the past century it has been proved that these problems cannot be solved with straight edge and compasses.

30. Fermat numbers

The Fermat numbers can be represented in the form

$$2^{2^t} + 1,$$

where t assumes the values 0, 1, 2, 3, 4, 5, $\cdots$.

The following table gives the values of $2^{2^t} + 1$ for the first six values of t:

t	0	1	2	3	4	5
$2^{2^t} + 1$	3	5	17	257	65,537	4,294,967,297

In numerous letters to his contemporaries, Fermat expressed the belief that the numbers obtained from $2^{2^t} + 1$ would be primes; this is verified in the above table for $t = 0, 1, 2, 3, 4$. It was not until almost 100 years later that Euler disproved this conjecture by actually exhibiting the factors of $2^{2^5} + 1$:

$$2^{2^5} + 1 = 4{,}294{,}967{,}297 = (641)(6{,}700{,}417).$$

Since Euler's time many other Fermat numbers have been shown to be composites. In fact no Fermat prime beyond the case for $t = 4$ has yet been found. Not many have been investigated, since they become so stupendous in size. When $t = 10$, for example, the number $2^{2^{10}} + 1$ contains 155 digits. Contrary to Fermat's guess, it is now thought that no Fermat number beyond the fifth is prime, but this is still a conjecture.

31. Constructible quantities

Certain concepts and results in mathematics have lain dormant for many years, even for centuries, until they have been connected with other parts of mathematics or with the physical sciences. This was the case, for example, with the conic sections such as the ellipse. The ancient Greeks studied them for their own beauty, but centuries later Johannes Kepler (1571–1630) showed that the planets move about the sun in orbits that are elliptical.

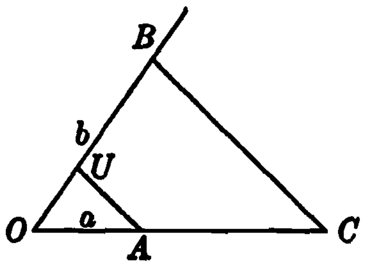
For more than 150 years after Fermat introduced the Fermat numbers, no application was found to connect them with other parts of mathematics. About 1800 Gauss took up the problem of the construction of regular polygons with ruler and compasses, and he showed that this problem is related to Fermat numbers.

In the discussion of the Fibonacci numbers we indicated the construction with ruler and compasses of a regular pentagon and of a regular decagon. Nearly everyone knows from plane geometry how to inscribe an equilateral triangle, a square, and a regular hexagon in a circle. Gauss became interested in finding out which regular polygons in addition to the triangle, square, pentagon, hexagon, and decagon can be inscribed in a circle, and he gave the complete answer to this question.

It should be understood at the outset that the word *ruler* means a straight-edge without any graduations, enabling us to draw a straight line through two given points. The compasses permit the drawing of arcs of circles; they are also used to lay off a given length along a line.

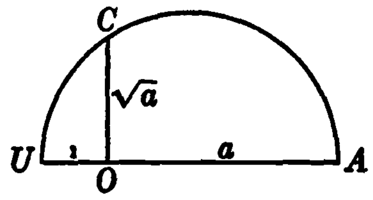
Given any two line segments of lengths a and b measured in terms of the same unit, we may make the obvious construction of a new line segment equal in length to the sum or difference of a and b, using only the straight-edge and compasses. While the constructions of the product and quotient of two line segments of lengths a and b are not as familiar, they too can be performed. We shall give only the construction of the product of two given line segments.

Draw two intersecting straight lines, as in the figure. Starting at the intersection O, lay off a on one of the lines, and b on the other, thus determining the points A and B, respectively. On OB lay off the unit



OU, in terms of which the lengths of a and b are measured. Join U to A. Through B construct BC parallel to UA. Then the segment OC represents the product ab. The proof depends only on the similarity of the triangles OAU and OCB.

More familiar is the construction of a line segment whose length is the square root of a. Along any line, lay off $OA = a$. Then, as in the figure, lay off OU, the unit in terms of which a is measured. Now con-



struct a circle on UA as diameter. At O construct a perpendicular to UA and let it intersect the circle at C. Then $OC = \sqrt{a}$. Again we have used only the straight-edge and compasses.

By the expression *constructibility of a figure* we mean that the construction can be made using only the operations of addition, subtraction, multiplication, division, or extracting a square root, and any combinations of these operations. The mathematician is then able to transform the geometric criteria into algebraic ones. In fact, he is able to prove that if a geometric quantity is constructible, this quantity must be the root of a special algebraic equation. One requirement is that the degree of this equation must be a power of 2. Gauss was able to use these algebraic criteria for determining which regular polygons can be constructed (10: 340–58).

32. Gauss's criterion for the constructibility of regular polygons

Gauss proved that *a regular polygon of p sides, where p is an odd prime, can be inscribed in a circle with the aid of ruler and compasses only, provided p is a Fermat prime; that is, p has the form* $2^{2^t} + 1$. *Furthermore, no regular polygon with a prime number of sides other than those just described is so constructible.*

Gauss's theorem tells us that it is possible to construct polygons of 3, 5, 17, 257, and 65,537 sides, since these numbers are Fermat primes. It also tells us that it is impossible, by means of ruler and compasses, to construct regular polygons of 7, 11, 13, 19, and 23 sides, since these numbers, although primes, cannot be written in the form $2^{2^t} + 1$. Thus the problem of finding other constructible polygons depends only on finding additional Fermat numbers that are prime.

Gauss's theorem does not tell us how to draw the constructible polygons. Euclid knew how to inscribe an equilateral triangle and a regular pentagon. Almost 2000 years later (1796) Gauss gave the construction of the 17-sided regular polygon (13: 372–73). The method of constructing the 257-sided polygon was given in 1887 (13: 378).

Gauss requested, it is said, that a regular polygon of 17 sides be inscribed on his grave. This was not done on his simple grave in Göttingen, but the polygon does appear on the monument in his native town of Brunswick (10: 358).

The question still remains: Are polygons with a composite number of sides constructible? First we note that since an angle may be bisected by the use of ruler and compasses, it follows that if a regular polygon is constructible, so is the regular polygon having twice as many sides.

Gauss proved the following more general rule of constructibility: *A regular polygon of n sides can be inscribed by ruler and compasses if and only if the number n is representable as the product of a power of 2 and distinct Fermat primes.*

Thus, since regular polygons of 3 and 5 sides are inscribable, so also are $(2)(3) = 6$ and $(2^2)(3) = 12$ and $(2)(3)(5) = 30$, and $(2^2)(3)(5) = 60$. However, a regular polygon of $9 = (3)(3)$ sides is not inscribable, since (3)(3) is not a product of distinct primes. Nor would a 21-sided figure be inscribable, since n is a product of two distinct primes only one of which is a Fermat prime. Below is a table listing all the regular polygons of n sides, where n lies between 3 and 25, showing, by the Gauss criterion, which are inscribable and which are not.

Regular polygons of *n* sides

Inscribable	3	4	5	6	8	10	12	15	16	17	20	24
Non-in-scribable	7	9	11	13	14	18	19	21	22	23	25	

CHAPTER

VII

Congruences

It often happens that development of a subject is retarded because suitable means of recording its results are lacking. This was true in music and also in mathematics. Had the ancient Greeks known our present system of writing numbers, there is little doubt that they could have contributed much more to the development of mathematics. Even such giants as Fermat and Euler had difficulties with statements and proofs because of the lack of a suitable notation.

It was Gauss who first saw that, by extending the concept of equality, he could express some of the known facts of number theory more simply. With the notation discussed below many new results were found.

33. Meaning of congruence

Gauss introduced the idea of congruence, which he expressed in writing as

$$a \equiv b \pmod{m}.$$

This is read, "a is congruent to b (modulo m)," and means that $a - b$ is exactly divisible by the number m.

Here a and b stand for any positive or negative integers or zero, while m is a natural number. If $a = b$, so that $a - b$ is zero, the difference is divisible by *every* modulus and there would be no point in writing $a \equiv a \pmod{m}$. If $m = 1$, it will always be true that $a \equiv b \pmod{1}$ no matter what the integers a and b are, since $a - b$ is always divisible by 1.

Examples of congruences are $7 \equiv 3 \pmod{2}$, since $7 - 3$ is divisible by 2; and $7 \equiv 3 \pmod{4}$, since $7 - 3$ is divisible by 4. However 7 is not congruent to 3 (mod 3), since $7 - 3$ is *not* divisible by 3.

An equally useful way of defining congruence, entirely equivalent to the preceding, is to say that $a \equiv b \pmod{m}$ means that when a is divided by m, we obtain the same remainder as when b is divided by m. Thus $93 \equiv 18 \pmod{5}$, since 93 divided by 5 and 18 divided by 5 both leave the remainder 3. Also $93 \equiv 18 \pmod{25}$ since 93 and 18 both leave the remainder 18 when divided by 25.

34. Operations with congruences

Congruences may be applied to calendar problems, card tricks, magic squares, and games of all sorts (12: 159, 206, 244). We shall content ourselves here with a few simple applications to arithmetic and algebra. Before explaining these applications, we shall assume the validity of certain operations with congruences. These are easily proved just as they are for equations (1:41–42). Two numbers, each congruent to a third, are congruent to each other provided each of the congruences has the same modulus. Two congruences may be added, subtracted, and multiplied, and we still obtain true congruences provided the same modulus is used throughout. Thus, if $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then $a + c \equiv b + d \pmod{m}$ and $ac \equiv bd \pmod{m}$. In particular, $a^2 \equiv b^2 \pmod{m}$; and $a^e \equiv b^e \equiv$ (mod m).

In division we must be a little more careful. Although $76 \equiv 28 \pmod{8}$, it is not permissible to divide each member of the congruence by 4, since this gives $19 \equiv 7 \pmod{8}$, which is false. The reason is clear. Although $76 - 28$ is divisible by 8, $^{76}\!/\!_4 - {}^{28}\!/\!_4$ is not divisible by 8. However, dividing both members of a congruence by the same integer is valid if the divisor has no factor in common with the modulus. For example, $50 \equiv 5 \pmod{9}$ implies that $10 \equiv 1 \pmod{9}$; but $76 \equiv 28 \pmod{8}$ does not imply that 19 is congruent to 7, modulo 8, since the divisor has a factor in common with the modulus.

We call attention to another difference between an algebraic equation and a congruence. In ordinary algebra the product of two numbers equals zero only when at least one of the factors is zero. Although $(2)(4) \equiv 0 \pmod{8}$, neither 2 nor 4 is divisible by 8. However, if the modulus is a prime p, then $ab \equiv 0 \pmod{p}$ implies that at least one of the congruences $a \equiv 0 \pmod{p}$ or $b \equiv 0 \pmod{p}$ must hold.

As an application of how congruences may be used in simplifying computations, let us verify the fact, already mentioned, that $2^{2^5} + 1 \equiv 0 \pmod{641}$. First we observe that $2^2 = 4$, $2^4 = 16$, $2^8 = 246$, $2^{16} = 65{,}536 \equiv 154 \pmod{641}$, since it may be verified that 154 is the remainder obtained when 65,536 is divided by 641. Furthermore, two congruences having the same modulus may be multiplied, so that $(2^{16})^2 = 2^{32} \equiv (154)^2 \pmod{641}$. But $(154)^2 = 23{,}716 \equiv 640 \equiv -1 \pmod{641}$. Hence, $2^{32} + 1$ is divisible by 641, which is another way of saying that $2^{2^5} + 1 \equiv 0 \pmod{641}$.

35. Some old and new results stated in congruential form

If we use the congruence notation, Fermat's little theorem, which tells us that $a^{p-1} - 1$ is exactly divisible by the odd prime p, may be written

$$a^{p-1} \equiv 1 \pmod{p},$$

provided a does not contain p as a factor. If we do not insist on this last restriction, we may write the result in the form $a^p \equiv a \pmod{p}$, for if a contains the factor p, $a^p - a$ will surely be divisible by p.

We shall indicate a proof of Fermat's little theorem by the use of a new property of congruences in addition to those given in Section 34. By the hypothesis of Fermat's theorem, p is a prime, and a is not a multiple of p. Then each member of the set

$$a, 2a, 3a, \cdots, (p-1)a$$

is congruent (mod p) to one and only one of the integers

$$1, 2, 3, \cdots, p-1,$$

but not necessarily in the order written (1:43). We shall not give a proof of this elementary result, but shall merely illustrate it by an example. Let $p = 7$ and $a = 4$. Then the integers 4, 8, 12, 16, 20, 24 are congruent (mod 7) to 4, 1, 5, 2, 6, 3, respectively; and these are precisely the numbers 1, 2, 3, 4, 5, 6, when rearranged.

Since congruences with the same modulus may be multiplied together, we obtain from the preceding remark that

$$(a)(2a)(3a) \cdots (p-1)a \equiv (1)(2)(3) \cdots (p-1)(\text{mod } p).$$

But this congruence may also be written

$$a^{p-1}(p-1)! \equiv (p-1)! \pmod{p}.$$

We may divide both members by $(p-1)!$, since $(p-1)!$ has no factoı in common with the modulus p, and thus we obtain $a^{p-1} \equiv 1 \pmod{p}$.

Euler's generalization of Fermat's theorem in congruence form is

$$a^{\phi(m)} \equiv 1 \pmod{m},$$

provided a and m have no factors in common.

Wilson's theorem takes the form

$$(p-1)! \equiv -1 \pmod{p}.$$

A series of interesting congruences results from considering the sums of powers of integers. If p is an odd prime, it may be shown that (1:51)

$$1 + 2 + 3 + \cdots + (p-1) \equiv 0 \pmod{p};$$

$$1^2 + 2^2 + 3^2 + \cdots + (p-1)^2 \equiv 0 \pmod{p};$$

$$1^3 + 2^3 + 3^3 + \cdots + (p-1)^3 \equiv 0 \pmod{p};$$

....

$$1^k + 2^k + 3^k + \cdots + (p-1)^k \equiv 0 \pmod{p};$$

provided k is not a multiple of $p - 1$. If $k = s(p - 1)$, then

$$1^{p-1} + 2^{p-1} + 3^{p-1} + \cdots + (p - 1)^{p-1} \equiv -1 \pmod{p};$$

$$1^{2(p-1)} + 2^{2(p-1)} + 3^{2(p-1)} + \cdots + (p - 1)^{2(p-1)} \equiv -1 \pmod{p};$$

. . . .

Only the simplest of the first group of congruences will be proved; namely,

$$1 + 2 + 3 + \cdots + (p - 1) \equiv 0 \pmod{p}.$$

As we know from algebra,

$$1 + 2 + 3 + \cdots + (p - 1) = \frac{(p - 1)\,p}{2};$$

since p is odd, $\frac{(p - 1)p}{2}$ is an integer divisible by p, and the congruence is verified.

Next we show that

$$1^{p-1} + 2^{p-1} + 3^{p-1} + \cdots + (p - 1)^{p-1} \equiv -1 \pmod{p}.$$

Since p is an odd prime, the numbers 1, 2, 3, $\cdots$, $(p - 1)$ have no factor in common with p, so that we obtain by Fermat's little theorem

$$1^{p-1} \equiv 1 \pmod{p}; \quad 2^{p-1} \equiv 1 \pmod{p};$$
$$3^{p-1} \equiv 1 \pmod{p}; \cdots; (p - 1)^{p-1} \equiv 1 \pmod{p}.$$

We may add the $p - 1$ congruences, since they all have the same modulus (1: 41–42) and obtain

$$\begin{aligned} 1^{p-1} + 2^{p-1} + 3^{p-1} + \cdots + (p - 1)^{p-1} &\equiv \underbrace{1 + 1 + 1 \cdots 1}_{\text{p−1 terms}} \pmod{p} \\ &\equiv p - 1 \pmod{p} \\ &\equiv -1 \pmod{p}. \end{aligned}$$

The other congruences of this group are proved in a similar fashion.

Some of the familiar facts of arithmetic, algebra, and trigonometry may be written as congruences. For example, consider the statement, and its converse, that numbers which have the same last digit differ by a multiple of 10. Thus 12 and 102 differ by 90; 102 and 1002 differ by 900; 37 and 497 differ by 460. These relationships may be expressed as follows: If $M \equiv N \pmod{10}$, then the last digit of M equals the last digit of N, and conversely.

The congruence notation may also be used in connection with the well-known fact that the value of -1 to any power, written $(-1)^n$, depends only on the evenness or oddness of n. In congruence form, this says: If $m \equiv n \pmod{2}$, then $(-1)^m = (-1)^n$, and conversely.

In trigonometry, angles which differ by multiples of 360° have the same trigonometric ratios. Thus if $A \equiv B \pmod{360°}$, then $\sin A = \sin B$, $\cos A = \cos B$, and so on.

36. Divisibility by 9, 3, and 11

We are now ready for the application of congruences. Our everyday system of writing numbers is based on powers of 10, so that 257 means $(2)(10^2) + (5)(10^1) + (7)(10^0)$, and the number 501.23 means $(5)(10^2) + (0)(10^1) + (1)(10^0) + (2)(10^{-1}) + (3)(10^{-2})$.

More generally, if u, t, and h, represent the units, tens, and hundreds digits of a number N, the number itself is

$$N = u + 10t + 100h + \cdots .$$

Since $10 \equiv 1 \pmod 9$, it follows that $10t \equiv \mathrm{t} \pmod 9$, because we are multiplying both members of the congruence by the same number t. Also, $100 \equiv 1 \pmod 9$, so that $100h \equiv h \pmod 9$. Hence we may write N as

$$N \equiv u + t + h + \cdots \pmod 9.$$

By the very meaning of congruence this says that *any number differs from the sum of its digits by a multiple of* 9. For example, $72381 - (7 + 2 + 3 + 8 + 1) = 72{,}381 - 21 = 72{,}360 = (9)(8{,}040)$.

A consequence of the previous statement is that *a number is exactly divisible by* 9 *if the sum of its digits is divisible by* 9, *and only then.* Of course it is easier to test the sum of the digits for divisibility by 9 than to test the number itself.

For example, if the number is 30,100,002, the sum of its digits is $3 + 1 + 2 = 6$, which is not divisible by 9; hence, the number itself is not. On the other hand, 511,101,000,009 is divisible by 9 because the sum of its digits $5 + 1 + 1 + 1 + 1 + 9 = 18$, a multiple of 9.

A similar rule applies to divisibility by 3; that is, *a number is divisible by* 3 *if the sum of its digits is divisible by* 3, *and only then.* Thus 30,100,002 has 6 for the sum of its digits so that the number itself is divisible by 3.

The rule for the divisibility of a number by 11 is based on the fact that

$$10 \equiv -1 \pmod{11}, \qquad 10^2 \equiv 1 \pmod{11}, \qquad 10^3 \equiv -1 \pmod{11}, \cdots,$$

so that the equality

$$N = u + 10t + 100h + \cdots$$

may be replaced by the congruence

$$N \equiv u - t + h - \cdots \pmod{11}.$$

Stated in words, this last congruence tells us that *a number is divisible by* 11 *if the units digit diminished by the tens digit increased by the hundreds digit, etc., is divisible by* 11. The converse is also true. For example,

since $9{,}581 \equiv 1 - 8 + 5 - 9 \equiv -11 \pmod{11}$,

9,581 is divisible by 11;

but $9{,}234 \equiv 4 - 3 + 2 - 9 \equiv -6 \pmod{11}$, so that

9,234 is not divisible by 11.

37. Casting out nines as a check in arithmetic

If we apply the principle that *after all the* 9*'s are divided out of a number, the remainder is the same as that obtained when all the* 9*'s are divided out of the sum of the digits*, we obtain simple checks for the operations of addition, subtraction, multiplication, and division.

We shall give here the check for multiplication only. We shall not go into the details of how the operations with congruences are applied, since they may readily be supplied by the reader. Consider the product of 3602 by 978.

$$\begin{array}{r l l}
3602 & \rightarrow 11 \rightarrow & 2 \\
\underline{978} & \rightarrow 24 \rightarrow & \underline{6} \\
28816 & & 12 \rightarrow 3 \\
25214 & & \\
\underline{32418} & & \\
3522756 & \rightarrow 30 \rightarrow & 3
\end{array}$$

We have indicated by the arrows to the right of 3602 that we have cast out the nines by adding the digits, first obtaining 11 and then 2; for 978 we obtain 24 and then 6. We then multiply 2 by 6 obtaining 12, or 3. We add the digits in the product 3522756, obtaining 30, or 3, the same value that was obtained previously. This gives the required check. We have here used the principle that the product of the remainders is the remainder of the product. We should note that the test is not absolute; if we had interchanged two digits in the answer, the check would still have worked.

38. Congruences of the first degree

As we have already observed, there is a close analogy between equations and congruences. Analogous to equations of the first degree with one unknown, $ax = b$, are congruences of the first degree $ax \equiv b \pmod{m}$. However, there are differences. While the equation $ax = b$ always has a solution when $a \neq 0$, this is not the case with congruences. In fact, even though $a \neq 0$, there is no solution of $6x \equiv 4 \pmod 3$; for if x is an integer, it follows from the meaning of congruence that $(6x - 4)/3$ must be an integer; this is impossible because $6x$ is divisible by 3 while 4 is not. In solving the equation $ax = b$, on the other hand, we admit values of x that are not necessarily integral. Again, every congruence of the first degree that has a solution has more than one solution. For example, the congruence $2x \equiv 5 \pmod 7$ is true when $x = 6$, but is also true when $x = 13$ or 20. In fact it is true for $x = 6$ increased by any multiple of 7. All such solutions are considered the same, and, indeed, they all leave the same remainder 6 when they are divided by 7.

The most important results of congruences of the first degree

$$ax \equiv b \pmod m$$

are the following:

A. Let d be the greatest common factor of a and m. Then the congruence has no solution unless d is a factor of b.

B. If d is a factor of b, then the congruence has d solutions.

C. In particular, if $d = 1$, that is, if a and m are relatively prime, the congruence has exactly one solution.

The following examples illustrate these results:

In $6x \equiv 4 \pmod 3$, $d = 3$ and d is not a factor of 4; hence there is no solution.

In $3x \equiv 12 \pmod 6$, $d = 3$ and since d is a factor of 12 there are three solutions: 0, 2, and 4. All other solutions will be congruent to these modulo 6.

In $2x \equiv 5 \pmod 9$, $d = 1$, and the only solution modulo 9 is $x = 7$.

There are various methods for solving congruences. If the modulus m is small, the answer may be found by trial, since we need to test only those values of x among $0, 1, 2, \cdots$ up to $m - 1$. Other methods for solving such congruences are illustrated by the problem in Section 39.

39. A problem leading to a congruence of the first degree

A farmer has seven baskets of eggs, with the same number of eggs in each basket. After selling all his eggs by the dozen, he finds that he has five eggs left over. What is the smallest possible number of eggs in each basket?

If x is the number of eggs in each basket, $7x$ when divided by 12 must give a remainder of 5. Stated in the language of congruences, this says that

$$7x \equiv 5 \pmod{12}.$$

Here 7 and 12 have only the factor 1 in common so that there is a solution. In a simple problem such as this we could substitute 0, 1, 2 all the way to 11 for x, and find that the only value of x which would make the congruence true is $x = 11$. Thus the smallest number of eggs that the farmer could have had in each basket is 11. We could also have worked this problem by substituting -7 for 5 in the right member of the congruence, since $5 \equiv -7 \pmod{12}$. The original congruence then becomes $7x \equiv -7 \pmod{12}$, and, since 7 and 12 have no common factor, we may divide by 7 and obtain $x \equiv -1 \pmod{12}$, or $x \equiv 11 \pmod{12}$, the same answer as before.

In general, we may reduce a congruence of the form $7x \equiv 5 \pmod{12}$ to a congruence in which the coefficient of x is 1. This is accomplished by the use of Euler's generalization of Fermat's theorem. (Here we cannot use Fermat's theorem itself since the modulus 12 is not a prime.) We know from Euler's generalization that $7^{\phi(12)} \equiv 1 \pmod{12}$, and since $\phi(12) = 4$, this becomes $7^4 \equiv 1 \pmod{12}$. Hence, if we multiply the congruence $7x \equiv 5 \pmod{12}$ by $7^3 \equiv 7^3 \pmod{12}$, we obtain $7^4x \equiv (5)(7^3) \pmod{12}$. But $7^4 \equiv 1 \pmod{12}$, so that the congruence of the first degree becomes $x \equiv (5)(7^3) \pmod{12}$. Since $7^2 \equiv 1 \pmod{12}$, $7^3 \equiv 7 \pmod{12}$ and $(5)(7^3) \equiv 35 \equiv 11 \pmod{12}$, so that $x \equiv 11 \pmod{12}$, and again the same result is obtained.

40. Chinese remainder theorem

As a final illustration, we consider a system of congruences each of the first degree. This is analogous to a system of n simultaneous equations, each of the first degree, in n unknowns. Such congruences arise frequently in puzzle problems. The method of solving these congruences was known to the ancient Chinese, and for this reason it is called the Chinese remainder theorem.

We shall introduce a problem leading to simultaneous congruences. To solve this problem we first give the method for solving a general system of simultaneous congruences, and then we apply it to the special case.

A centenarian was asked how many great-grandchildren he had. He replied that he could not remember exactly, but he recalled that if you put them in groups of three, there would be one left over; in groups of five, there would be three left over. However, when you put them in groups of eight, there would be none left over. What is the least number of great-grandchildren he could have had?

Here it would be a simple matter to try multiples of 8 until we find that one that fits the data. However if we put the problem in the form of congruences, we solve this problem and many more complicated ones of a similar nature.

If we let x equal the nun.ber of great-grandchildren, the stipulations are

$$x \equiv 1 \pmod{3}$$

$$x \equiv 3 \pmod{5}$$

$$x \equiv 0 \pmod{8}.$$

We must remember that in all these congruences the x stands for the same natural number. That is why we call them *simultaneous congruences.*

The method for solving four simultaneous congruences will be given, although it applies equally well to three or to any number of congruences:

$$x \equiv a \pmod{k}$$

$$x \equiv b \pmod{l}$$

$$x \equiv c \pmod{m}$$

$$x \equiv d \pmod{n}.$$

We assume here that the moduli k, l, m, and n, when taken two at a time, have no common factor.

We must keep in mind that the numbers $a, b, c, d; k, l, m, n$ are given, and we are seeking the number x for which all congruences will be true at the same time.

Let R be the product of all the moduli; that is, $R = klmn$. Let $K = lmn$, $L = kmn$, $M = kln$, $N = klm$; that is, each of these numbers K, L, M, N is found by dividing R by k, l, m, n in this order. Let us now find solutions of the individual congruences in the unknowns y, z, u, w, respectively:

$$Ky \equiv 1 \pmod{k}$$

$$Lz \equiv 1 \pmod{l}$$

$$Mu \equiv 1 \pmod{m}$$

$$Nw \equiv 1 \pmod{n}.$$

The congruence $Ky \equiv 1 \pmod{k}$ is a linear congruence; the modulus k has no factor in common with $K = lmn$; hence, there is a unique solution y. Similarly, there are unique solutions for the other three congruences.

The general solution of the system of the congruences

$$x \equiv a \pmod{k}, \quad x \equiv b \pmod{l}, \quad x \equiv c \pmod{m}, \quad x \equiv d \pmod{n}$$

is

$$x = Kya + Lzb + Muc + Nwd$$

where

$$K = lmn, \qquad L = kmn, \qquad M = kln, \qquad N = klm,$$

and y, z, u, w are the solutions of the congruences

$$Ky \equiv 1 \pmod{k}, \qquad Lz \equiv 1 \pmod{l},$$
$$Mu \equiv 1 \pmod{m}, \qquad Nw \equiv 1 \pmod{n}.$$

Any multiple of $R = klmn$ added to or subtracted from x will also be a solution.

Observe that the same value of x has to satisfy all four of the original congruences, while the value of each of the unknowns y, z, u, w is found by solving separately the four new congruences.

The derivation of the general solution is not obvious, but it may be obtained from the simple properties of congruences (12: 189–91).

From the start the stipulation was made that the k, l, m, n, when taken two at a time, have no common factor. If even two of the moduli have a common factor, the result must be modified (12: 184–89).

In the problem of the great-grandchildren $x \equiv 1 \pmod{3}$, $x \equiv 3 \pmod{5}$, $x \equiv 0 \pmod{8}$; $k = 3$, $l = 5$, $m = 8$; and the pairs 3, 5; 3, 8; and 5, 8 have no factors in common. Here $a = 1$, $b = 3$, $c = 0$; $R = (3)(5)(8) = 120$, $K = {}^{120}\!/_3 = 40$, $L = {}^{120}\!/_5 = 24$, $M = {}^{120}\!/_8 = 15$. We now seek solutions of the separate congruences $40y \equiv 1 \pmod{3}$, $24z \equiv 1 \pmod{5}$, $15u \equiv 1 \pmod{8}$. We readily find by trial that $y = 1$, $z = 4$, and $u = 7$. Hence, from the formula for the general solution,

$$x = (40)(1)(1) + (24)(4)(3) + (15)(7)(0) = 328.$$

Since we may subtract from 328 any multiple of $R = 120$, we see that the least positive value of x is $328 - (2)(120)$ or 88. This then is the number of great-grandchildren.

The following is another example of a puzzle problem solved by means of the Chinese Remainder Theorem. Determine the age of a person if the remainders obtained by dividing his age by 3, 4, and 5 are given. It may be shown that the person's age x is given by

$$x = 40r_1 + 45r_2 + 36r_3 \pmod{60}$$

Here, r_1, r_2, and r_3 are the remainders he gives you after dividing his age by 3, 4, and 5, respectively. For example, if r_1, r_2, r_3 are 2, 1, 2, the above formula gives $x = 197 \pmod{60}$, so that the person's age is 77 or 17, a choice that should present no difficulty.

41. Congruences of degree higher than the first

So far the only congruences considered are of the first degree in the unknown. We shall now state briefly some results for congruences of higher degree.

Let the congruence be

$$ax^n + bx^{n-1} + \cdots + l \equiv 0 \pmod{m},$$

where $a, b, \cdots, l$ are integers or zero. If the coefficient a of the highest power of x is not divisible by m, the congruence is said to be of the nth degree. (If a is a multiple of m, then the term ax^n may be suppressed, from the very definition of congruences. The congruence would then be of degree lower than n.) *The integer $x = r$ is said to be a root of the congruence if the value of $ar^n + br^{n-1} + \cdots + l$ is exactly divisible by m.*

If the modulus m is a composite, it may be shown that the properties of this congruence may, with few exceptions, be obtained by considering the given congruence with each of the distinct prime factors of m as modulus (12: 192–96). We may, therefore, confine our study to congruences whose moduli are primes. The properties of these congruences bear a great similarity to those of ordinary algebraic equations.

One such property is known as Lagrange's theorem on congruences (12: 197–98). *The congruence*

$$ax^n + bx^{n-1} + \cdots + l \equiv 0 \ (mod\ p)$$

has at most n roots. For example:

A. $x^2 - 1 \equiv 0 \pmod{5}$ can have at most two roots, and, in fact, it has exactly two roots, $x = \pm 1$.

B. $x^3 - 1 \equiv 0 \pmod{5}$ can have at most three roots, but actually has only one, $x = 1$.

C. $x^2 - 1 \equiv 0 \pmod{4}$ has the four solutions, $x = \pm 1, x = \pm 3$; but Lagrange's theorem does not apply, since the modulus is not a prime.

While every algebraic equation has at least one root, this is not necessarily true for congruences. For example, there is no solution of $x^2 - 2 \equiv 0 \pmod{5}$, as we can see by substituting for x the values 0, 1, 2, 3, 4.

CHAPTER

VIII

Diophantine Equations

Diophantine equations are so named because they were first considered by the Greek mathematician Diophantus, who lived in the third century A.D. They are also called *indeterminate equations* for reasons that will be apparent presently. Just as in equations in algebra, Diophantine equations may be of the first degree or of higher degree; they may be in one or more unknowns; or there may be simultaneous systems of such equations.

42. Diophantine equations of the first degree in two unknowns

We begin with a simple example. Suppose a person spent 40 cents in a hardware store. If he gave the clerk a one-dollar bill, how could he get his change in dimes and quarters? The solution here is obvious, but it will be instructive to see how such a problem leads to a Diophantine equation. The equation in terms of the number of dimes d and the number of quarters q is

$$10d + 25q = 60,$$

when the transaction is expressed in cents. Simplified, this equation becomes

$$2d + 5q = 12.$$

We should bear in mind that d and q must be positive integers or zero. If we solve the equation for d, we obtain

$$d = 6 - \frac{5q}{2}.$$

Since d and 6 are integers or zero, $\frac{5q}{2}$ is also an integer, which we shall call r. Then $\frac{5q}{2} = r$ or $5q = 2r$, an even integer. Since 5 is odd, q must be an even integer, say, $q = 2s$. Hence $d = 6 - \frac{5q}{2}$ becomes $d = 6 - 5s$, where s is an arbitrary integer. Furthermore $5s$ must be less than 6 since d must be positive or zero. Thus we arrive at the two possibilities shown in the following table:

s	d	q
0	6	0
1	1	2

Our problem has two solutions: no quarters and six dimes, or two quarters and one dime.

In algebra every equation in two unknowns such as $x + y = 7$ has an infinite number of solutions, since we may assign any arbitrary value to one of the unknowns and determine the value of the other unknown by means of the equation. In Diophantine equations the problem is similar, except for the additional restriction that *both* unknowns must be integers or zero. A Diophantine equation may have no solution, a finite number, or even an infinite number of solutions. A Diophantine equation is sometimes referred to as an indeterminate equation, since it does not determine x and y completely.

It may be shown more generally that if $x = r$ and $y = s$ is a particular (integral) solution of the Diophantine equation

$$ax + by = c,$$

then all integral solutions may be found from the formulas

$$x = r + bt,$$

$$y = s - at,$$

where t may have any positive or negative integral value. The value $t = 0$ gives the particular solution $x = r$, $y = s$ (12: 56).

A Diophantine equation of the first degree is a congruence of the first degree. For example, the equation already considered, $2d + 5q = 12$, may be written either as $2d \equiv 12 \pmod{5}$ or $5q \equiv 12 \pmod{2}$. For $2d \equiv 12 \pmod{5}$ says that the expression $12 - 2d$ when divided by 5 must be an integer, say q. It is not surprising, therefore, to observe that a Diophantine equation may have no solution just as a congruence may have no solution. In fact, in order for the Diophantine equation $ax + by = c$ to have a solution, any common factor of a and b must also be a factor of c. This condition becomes the criterion for the solvability of congruences. For if $ax + by = c$, then $ax \equiv c \pmod{b}$; and to say that the Diophantine equation $ax + by = c$ is solvable implies that any factor of a and b is also a factor of c. This is precisely the condition for the solvability of the congruence $ax \equiv c \pmod{b}$.

One final point is worth noting. We sometimes require, as in the problem of dimes and quarters, that the solution be positive or zero. This means that the t in the formula above must be chosen in such a way as to make both $r + bt$ and $s - at$ positive or zero (12: 58).

43. The problem of determining a particular solution

So far nothing has been said about finding a particular solution $x = r, y = s$ of the Diophantine equation $ax + by = c$. We have already done this for congruences in disguised form by the use of Fermat's theorem or by Euler's generalization. The following method is a variant of what is known as the Euclidean algorithm (1: 26–27). Although the method to be given here is longer than the classical form of the Euclidean algorithm, it is more easily remembered and may be understood by pupils in an elementary algebra class.

We illustrate the method in the following problem. Suppose a man cashes a check in a bank, and the teller, in giving him his money, interchanges the dollars and cents. Suppose also that the amount he received is more than the amount of the original check. After accepting the money, the man spends 67 cents and finds that what he has left is twice the amount of the original check. What was this amount?

Let x equal the number of cents and y the number of dollars in the original check. From the nature of the problem x and y are both natural numbers, and each must be less than 100 to be interchangeable. The problem, written symbolically, is to find the solution of the Diophantine equation

$$100x + y - 67 = 2(100y + x),$$

which, when simplified, becomes

$$98x - 199y = 67.$$

If we solve this equation for x, we obtain

$$x = 2y + \frac{3y + 67}{98} \text{ or } x - 2y = \frac{3y + 67}{98}.$$

Since x and y are natural numbers, so are x and $2y$. It follows that $x - 2y$ or $\frac{3y + 67}{98}$ is a positive or negative integer or zero. Denoting the difference $x - 2y$ by the letter z, we have

$$z = \frac{3y + 67}{98},$$

or

$$3y = 98z - 67,$$

so that

$$y = 32z + \frac{2z - 67}{3}.$$

Again, since y and z are integers, so is $y - 32z$, as well as $\frac{2z - 67}{3}$. Then let

$$u = \frac{2z - 67}{3},$$

or

$$2z = 3u + 67,$$

and

$$z = u + \frac{u + 67}{2}.$$

Since $\frac{u + 67}{2}$ is to be an integer v, that is, $v = \frac{u + 67}{2}$, we finally obtain

$$u = 2v - 67.$$

The expression for u contains no denominator, so that u is an integer for every integral value of v.

We have thus arrived at the following conclusion: If there are two integers x and y satisfying the original Diophantine equation, there must be an arbitrary integer v in terms of which u, z, x, and y are expressible. By simple algebraic manipulations, we find

$$x = 199v - 4489,$$
$$y = 98v - 2211.$$

Since x and y are to be natural numbers, it follows that the arbitrary integer v must be chosen so that $98v$ is greater than 2211 and at the same time $199v$ is greater than 4489. The least value of v which fulfills both these requirements is $v = 23$. Inserting this value of v in the above expressions for x and y, we find $x = 88$ and $y = 43$, so that the amount of the original check is $43.88. Any integral value of v greater than 23 will make both x and y greater than 100 and hence will not comply with the conditions of the problem. There is only one solution to this problem.

The procedure given is perfectly general. We should observe that in solving the above equations we always solved for that unknown with the smaller coefficient. Thus, in the equation $98x - 199y = 67$, we solved for x.

44. Diophantine equations of the second degree

We have already referred to the equation $x^2 + y^2 = z^2$, which arises from the Pythagorean theorem for right triangles. This is a Diophantine equation of the second degree in the three unknowns, x, y, and z, each of which is to be a natural number. By a solution we mean a triple of numbers such as 3, 4, 5; 5, 12, 13; or 8, 15, 17; which, when substituted in the equation for x, y, and z, respectively, will satisfy the equation. Thus $8^2 + 15^2 = 17^2$, since $64 + 225 = 289$.

Particular solutions were known to the ancient Babylonians, Chinese, and Hindus. The Plimpton Library at Columbia University contains a Babylonian tablet which dates back at least 1000 years before the Pythagoreans. It gives many solutions, some involving large numbers (10: 175–78). The first general solution of the Pythagorean equation is found in Euclid's *Elements*, where it is couched in geometric terms.

In algebraic form, the general solution of $x^2 + y^2 = z^2$ is $x = 2mn$, $y = m^2 - n^2$, $z = m^2 + n^2$, where m and n are arbitrary natural numbers (13:317). If we further require that x, y, and z have no factor in common, we must stipulate that m and n have no factor in common, and also that m and n be of opposite parity; that is, one of them is even while the other one is odd. In this case, x, y, and z will have no factor in common, and the solution is said to be *primitive.*

We shall now indicate how the formulas for the general solution are derived. If the equation $x^2 + y^2 = z^2$ is to have a primitive solution, x, y, z, it is not hard to see that x and y cannot both be odd, so that one of them must be even and the other odd, and hence z must be odd. For if x is odd, it is of the form $2a + 1$ and x^2 is of the form $4b + 1$. Similarly, if y is odd, y^2 is of the form $4c + 1$, so that $x^2 + y^2$ must have the form $4d + 2$. But, since z is even, z^2 is of the form $4f$, and not of the form $4d + 2$.

Let x be even; then y is odd and so is z. Write the original equation in the form

$$x^2 = z^2 - y^2 = (z + y)(z - y).$$

Since z and y are both odd, $z + y$ and $z - y$ are both even, so that we may put $z + y = 2k$ and $z - y = 2l$, and

$$x^2 = (2k)(2l) = 4kl.$$

Suppose that x, y, z is a primitive solution so that x, y, and z do not have any factors in common; then k and l cannot have a factor in common. By solving $z + y = 2k$ and $z - y = 2l$, we obtain $z = k + l$ and $y = k - l$. If k and l had a common factor, so would y and z.

From $x^2 = 4kl$, it follows that the product kl is itself a square, and since k and l have no factor in common, k and l must themselves be squares. We may therefore say that there are integers m and n, such that $k = m^2$ and $l = n^2$. Again there is no common factor in m and n, since k and l have no factor in common. By substituting these expressions for k and l in $x^2 = 4kl$, $y = k - l$, and $z = k + l$, we obtain

$$x = 2mn, \qquad y = m^2 - n^2, \qquad z = m^2 + n^2.$$

We have shown then that all primitive solutions of $x^2 + y^2 = z^2$ must be of this form.

Finally, by substituting these values of x, y, z in $x^2 + y^2 = z^2$, we obtain

$$(2mn)^2 + (m^2 - n^2)^2 = (m^2 + n^2)^2,$$

which when simplified gives an identity in m and n; that is, it is true for all values of m and n. This shows that the above expressions for x, y, z in terms of m and n are indeed solutions.

A list of all primitive solutions (sometimes called primitive Pythagorean numbers), where the hypotenuse z does not exceed 3000, was computed in 1912. The 15 primitive solutions in the following table were obtained from the formulas given:

m	n	x	y	z	m	n	x	y	z
2	1	4	3	5	7	6	84	13	85
3	2	12	5	13	7	4	56	33	65
4	3	24	7	25	7	2	28	45	53
4	1	8	15	17	8	7	112	15	113
5	4	40	9	41	8	5	80	39	89
5	2	20	21	29	8	3	48	55	73
6	5	60	11	61	8	1	16	63	65
6	1	12	35	37					

The following fact concerning Pythagorean numbers may be verified for the special triples given in the table. It has been proved in general (13: 319).

In every triple of Pythagorean numbers, one is divisible by 3, *one is divisible by* 4, *and one is divisible by* 5. In the triple 5, 12, 13, the 12 is divisible by both 3 and 4; in the triple 8, 15, 17, the 15 is divisible by 3 and by 5. We may state this result in a more compact form: *The product of any triple of Pythagorean numbers is always divisible by* 60.

45. Some Diophantine equations of higher degree

In this section we give some examples which lead to Diophantine equations of degree higher than the second, involving three unknowns and having infinitely many solutions.

The first problem, one proposed and solved by Fermat, is to find a right triangle whose hypotenuse is the square of a natural number, while the sum of the other two sides is also a perfect square (12: 414–19).

If we denote the two legs by x and y, and the hypotenuse by z, the problem may be stated in the form of three equations

$$x + y = v^2, \qquad x^2 + y^2 = z^2, \qquad z = u^2.$$

Now let $w = x - y$ and eliminate the x and y, obtaining

$$2u^4 - v^4 = w^2,$$

which is a Diophantine equation of the fourth degree in the three unknowns u, v, w.

It is known that this equation has infinitely many solutions. In terms of x, y, z, the smallest solution is

$$x = 4{,}565{,}486{,}027{,}761$$

$$y = 1{,}061{,}652{,}293{,}520$$

$$z = 4{,}687{,}298{,}610{,}289.$$

The magnitude of the least solution is very large in contrast to the least solution, $x = 3$, $y = 4$, $z = 5$, of the Pythagorean equation $x^2 + y^2 = z^2$.

There are other Diophantine equations for which the least solution contains numbers with even more digits. The celebrated cattle problem (10: 140) attributed to Archimedes (287-212 B.C.) leads to the Diophantine equation of the second degree in two unknowns

$$x^2 - (4{,}729{,}494)(y^2) = 1.$$

In the least solution of this equation the number y has 41 digits (5: 121–24).

The next Diophantine equation is of special interest because its solution is closely related to the well-known law of cosines in trigonometry (see *The American Mathematical Monthly* 62: 251–52; April 1955).

Consider a Diophantine equation of the third degree in the three unknowns x, y, z:

$$x^2 + y^2 + z^2 + 2xyz = 1.$$

We write this equation in the form of a determinant of the third order

$$\begin{vmatrix} y & x & -1 \\ -1 & z & y \\ z & -1 & x \end{vmatrix} = 0.$$

In order that there be rational solutions x, y, z of the Diophantine equation we know that there must exist three rational numbers a, b, c for which

$$bx + ay = c$$
$$cy + bz = a$$
$$cx + az = b.$$

From these linear equations we readily find the solutions

$$x = \frac{b^2 + c^2 - a^2}{2bc}, \quad y = \frac{a^2 + c^2 - b^2}{2ac}, \quad z = \frac{a^2 + b^2 - c^2}{2ab}.$$

We may verify that these expressions for x, y, and z satisfy the Diophantine equation. They give the well-known formulas for the law of cosines in the triangle whose sides are the rational numbers a, b, c. It is in this sense that the Diophantine equation characterizes the law of cosines.

We should note that there are infinitely many solutions of this Diophantine equation. However, these solutions for x, y, z are rational numbers and not integers as they have been in all previous illustrations. But even for this Diophantine equation the general solution in integers has been found (see *The American Mathematical Monthly* 64: 101–103; February 1957).

46. A Diophantine equation having no solution

By using the method of infinite descent, Fermat proved the remarkable geometric fact that *the area of a right triangle with natural numbers for sides can never be the square of a natural number.*

The area of a right triangle is one-half the product of its two legs x and y, or

$$\text{Area} = \tfrac{1}{2}xy.$$

Using the explicit formulas for the sides of a primitive right triangle $x = 2mn$, $y = m^2 - n^2$, we may write, Area $= mn(m^2 - n^2)$. From the table of primitive solutions in Section 44, we see that the areas of the fifteen right triangles are 6, 30, 84, 60, 180, 210, 330, 210, 546, 924, 630, 840, 1560, 1320, and 504; not a single one of these is a perfect square. The proof of the general result depends on the method of infinite descent (see Section 29), which was first given by Fermat (10: 200–202).

As a consequence of this result we are able to prove that *the difference of two fourth powers of natural numbers is never a perfect square.*

If we again consider the right triangle with integral sides, we know that its legs are given by $2mn$ and $m^2 - n^2$. Since m and n may be chosen arbitrarily, we may set $m = u^2$ and $n = v^2$; the area of the triangle is then $\frac{1}{2}(2mn)(m^2 - n^2) = mn(m^2 - n^2) = u^2v^2(u^4 - v^4)$. If it were possible for $u^4 - v^4$ to be a square, say w^2, the expression for the area $u^2v^2(u^4 - v^4)$ would become $u^2v^2w^2$, which is the perfect square of the natural number uvw. However Fermat proved that the area cannot be a perfect square; it follows, therefore, that the difference between two fourth powers cannot be a perfect square.

47. Pell's equation—a quadratic Diophantine equation in two unknowns

We have already referred in Section 45 to a type of Diophantine equation of the second degree in two unknowns,

$$x^2 - Dy^2 = 1,$$

where D is a positive integer, not a perfect square.

This is known as *Pell's equation*. The name of the English mathematician John Pell (1610–1685) was given to this equation by Euler, who was under the impression that Pell was the first to solve it. It was Fermat, however, who was the first to state that there is an infinite number of solutions of this equation, while Lagrange was the first to publish a proof. However, a method for solving this equation was already known to the Hindus about A.D. 600.

Without giving the method for deriving the solution of $x^2 - Dy^2 = 1$, we shall give a formula for obtaining all the integral solutions of this equation. First, it may be shown that the equation $x^2 - Dy^2 = 1$, where D is an integer and not a perfect square, always has a least positive solution, $x = r$, $y = s$, where r and s are both natural numbers. By this we mean that any other positive values of x and y which satisfy this equation will have x greater than r, and y greater than s (1: 105).

From the least positive solution *all* other solutions are obtainable (1: 111) by the use of the following formula:

$$x + y\sqrt{D} = \pm(r + s\sqrt{D})^k, \qquad (k = 0, \pm 1, \pm 2, \cdots).$$

To illustrate how the formula gives solutions, consider a particular equation

$$x^2 - 3y^2 = 1.$$

Obviously the least positive solution is $r = 2$, $s = 1$. (A table in *The Higher Arithmetic* by Harold Davenport (1: 105) gives the least positive solutions for all values of D from 1 to 50.) The formula for the general solution now becomes

$$x + y\sqrt{3} = \pm(2 + \sqrt{3})^k, \qquad (k = 0, \pm 1, \pm 2, \cdots).$$

For $k = 0$, $x + y\sqrt{3} = \pm 1$ or $\pm(1 + 0\sqrt{3})$, so that $x = 1$ and $y = 0$, is a solution, but it is not the least positive solution, since 0 is not a natural number.

For $k = 1$, $x + y\sqrt{3} = \pm(2 + \sqrt{3})$. By equating the rational terms and the irrational terms of both members we obtain $x = \pm 2$, $y = \pm 1$. This value of k gave us the least positive solution as well as the solutions $x = 2$, $y = -1$; $x = -2$, $y = 1$; $x = -2$, $y = -1$. Each k, other than 0, yields four sets of solutions, but only one of them will be given here.

For $k = 2$, $x + y\sqrt{3} = (2 + \sqrt{3})^2 = (4 + 3 + 4\sqrt{3})$, so that $x = 7$, $y = 4$ is another solution.

For $k = -1$, $x + y\sqrt{3} = (2 + \sqrt{3})^{-1} = \dfrac{1}{2 + \sqrt{3}} = 2 - \sqrt{3}$, so that $x = 2$ and $y = -1$ is another solution.

For $k = 3$, $x + y\sqrt{3} = (2 + \sqrt{3})^3 = 8 + 12\sqrt{3} + 18 + 3\sqrt{3} = 26 + 15\sqrt{3}$, so that $x = 26$, and $y = 15$, is another solution.

In this way, we may obtain an infinite number of solutions.

We have seen that the general solution depends on finding the least positive solution. This task may be extremely difficult, since even for a small value of D, say $D = 19$, $r = 170$ and $s = 39$. The method for finding such a solution requires a special technique of mathematics known as continued fractions (1: 79–114).

48. A general result in Diophantine equations

We close our account of Diophantine equations with a statement of a remarkable result discovered by the Norwegian mathematician Axel Thue in 1908. It has no analogue in the theory of algebraic equations. Thue's theorem deals with Diophantine equations in two unknowns which have a limited number of solutions. This is in contrast to Diophantine equations of the second degree in two unknowns which have an infinite number of solutions, an example of which is the Pell equation just considered.

Before stating Thue's theorem, let us recall that a polynomial in a single variable x, with integral coefficients has the form

$$ax^n + bx^{n-1} + cx^{n-2} + \cdots + l,$$

where $a, b, c, \cdots, l$ are integers or zero. If a is not zero, then the degree is n. Analogously, a polynomial in two variables x and y, with integral coefficients, has the form

$$ax^n + bx^{n-1}y + cx^{n-2}y^2 + \cdots + ly^n,$$

where $a, b, c, \cdots, l$ are integers. This is a special type of polynomial, called *homogeneous*, because the sum of the exponents of x and y in each term adds up to the degree n of the polynomial. If this polynomial cannot be factored into two other polynomials with integral coefficients and of lower degree, we shall say that the polynomial is *irreducible*.

Thus it is not difficult to show that $x^3 + 3xy^2 + y^3$ is an irreducible homogeneous polynomial of the third degree while $x^3 + 3x^2y + 3xy^2 + y^3$ is obviously reducible since it may be written as $(x + y)(x + y)(x + y)$.

We are now ready to state Thue's theorem.

The Diophantine equation

$$ax^n + bx^{n-1}y + cx^{n-2}y^2 + \cdots + ly^n = K,$$

where the left-hand side is an irreducible polynomial of the third degree or higher with integral coefficients, and where K is an integer, has either no solution or at most a finite number of solutions (9: 263).

Clearly, the result would not apply to $(x + y)^3 = 1$ since the left side is reducible and, we readily see, there are infinitely many solutions. For if y is assigned any arbitrary integral value, $x = 1 - y$ will satisfy the equation, and thus infinitely many solutions are obtained.

Nor would Thue's theorem apply to $x^2 - Dy^2 = 1$, since the degree is *less* than the third.

Although the Diophantine equations discussed in Section 45 were of degree higher than the second, the number of unknowns was three, and we can draw no conclusions from Thue's theorem.

Following are two illustrations of Thue's theorem:

> The equation $x^4 - 2y^4 = 1$ is homogeneous in x and y, and of the fourth degree. It is not hard to show that the left member is irreducible, and it has been proved (12: 406) that the only solutions are $x = \pm 1$, $y = 0$.
>
> The equation $x^3 + 3x^2y + y^3 = 1$ is of the third degree, and its left member is homogeneous and irreducible. By Thue's theorem there can be only a finite number of solutions. It has been shown that this equation has exactly three solutions, viz.: $x = 1, y = 0$; $x = 0, y = 1$; $x = -3, y = 1$. (See Bibliography, 14: chap. vi.)

The following result is related to Thue's theorem.

The equation

$$ay^2 + by + c = dx^n, \qquad n \geqq 3$$

has only a finite number of integral solutions if a and d are natural numbers, if b and c are integers or zero, and if the discriminant $b^2 - 4ac$ *is not zero*. In $y^2 - 17 = x^3$, the conditions of the theorem are satisfied since $n = 3$, $b^2 - 4ac = 68$, so that the equation has a finite number of solutions. In fact there are exactly eight solutions, and these are $x = -2$, $y = 3$; $x = -1$, $y = 4$; $x = 2$, $y = 5$; $x = 4$, $y = 9$; $x = 8$, $y = 23$; $x = 43$, $y = 282$; $x = 52$, $y = 375$; $x = 5234$, $y = 378661$ (9: 265).

CHAPTER IX

Generalizations of Number Theory

Many topics of elementary number theory that are important for a deeper understanding of the subject will not be included in our treatment because of their technical nature. However we cannot omit a generalization that is entirely a creation of the nineteenth century. Since it is a generalization, it is bound to be abstract, but the beauty of the results should repay the reader for the extra effort required to understand them.

49. Complex integers

Mathematicians of the nineteenth century, Gauss and Kummer (1810–1893) in particular, extended some of the concepts of number theory. Gauss saw that many of the results of number theory would still remain valid if, for natural numbers, he substituted numbers of the form

$$a + bi,$$

where a and b are ordinary integers or zero and i is the well-known imaginary unit. We recall that $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, and so on. It must seem surprising that these so-called complex integers, $a + bi$, should possess many of the characteristics of the rational integers (the positive and negative integers). The sum, difference, or product of two rational integers is a rational integer; the sum, difference, or product of two complex integers is a complex integer. If $a + bi$ and $c + di$, for example, are two complex integers, their sum,

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

is also a complex integer. Similarly, their difference,

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

is a complex integer, and their product,

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

is also a complex integer.

Since complex integers were first systematically considered by Gauss, we often refer to them as *Gaussian integers,* to distinguish them from the rational integers. The rational integers are obtained from the Gaussian integers when the coefficient b of the imaginary unit is zero. We shall now extend some arithmetical concepts to complex integers.

50. Divisibility of complex integers

Although we usually denote a complex integer by $a + bi$, we shall sometimes find it convenient to use a single letter such as x, y, or z. For rational integers we say that 6 is divisible by 2 if there is a rational integer r for which $6 = (2)(r)$. Analogously, we also define divisibility of complex integers in terms of product. We say that *a complex integer x is divisible by a complex integer y if there is a complex integer z, for which $x = (y)(z)$; and y or z is said to be a divisor, or factor, of x.*

Not every complex integer is divisible by another. For example, we shall show that $3 + 2i$ is not divisible by $1 + 3i$. Otherwise, there would exist rational integers a and b such that

$$3 + 2i = (1 + 3i)(a + bi) = (a - 3b) + i(3a + b).$$

By equating real and imaginary parts of the two members, we find that

$$3 = a - 3b, \qquad 2 = 3a + b.$$

By the usual method for solving two simultaneous equations of the first degree, we find $a = 9/10$ and $b = -7/10$. These values are not integral; hence $3 + 2i$ is not divisible by $1 + 3i$. On the other hand, $3 + i$ is divisible by $1 + 2i$ and by $1 - i$, since $(3 + i) = (1 + 2i)(1 - i)$.

Every complex integer x always has the eight "trivial" divisors:

$$1, x, -1, -x, i, ix, -i, -ix,$$

since $1, -1, i, -i$ are complex integers. For example, $x = (ix)(-i)$; therefore ix and $-i$ are two divisors of x.

In rational integers (including both the positive and the negative), 1 and -1 are the only divisors of 1. In complex integers, 1 has four divisors: $1, -1, i, -i$, which are called the *unities* of the complex integers. (We call them unities rather than units to distinguish between the complex divisors of 1 and the rational integral divisors of 1.) We also say that a complex integer x multiplied by a unity is an *associate of x,* so that every complex integer has four associates: $x, -x, ix, -ix$.

Although we did not introduce the concept of associates among the rational integers, it is true that every rational integer has two associates, itself and its negative.

51. Gaussian primes

We are now ready to describe certain Gaussian integers which we shall call Gaussian primes. These primes, as we shall see, have many of the properties of rational primes.

A Gaussian prime is a Gaussian integer, not 0 or a unity, which is divisible only by the complex integers associated with itself or with 1.

We shall designate a Gaussian prime by the Greek letter π, not to be confused with the π of the circle.

From the definition of a prime, it is clear that the prime π has no divisors except the eight trivial ones,

$$1, \pi, -1, -\pi, i, i\pi, -i, -i\pi,$$

provided π is neither zero nor a unity. This definition of a Gaussian prime reduces to that of a rational prime. Since there are only two units, 1 and -1, among the rational integers, and since the rational prime is neither zero nor a unit, it follows that the only divisors of a rational prime p are the trivial ones, p, $-p$, 1, -1.

52. The norm of a complex number

In order to see how Gaussian primes are related to rational primes, we introduce the concept of *norm*, familiar to anyone who is acquainted with complex numbers. We recall, first of all, that if $a + bi$ is a complex number (not necessarily a complex integer), then $a - bi$ is called its *conjugate*, and the product of the complex number by its conjugate is the real number $a^2 + b^2$. *The number $a^2 + b^2$ is called the norm of the complex number $a + bi$.* If, now, $a + bi$ is a Gaussian integer, then the norm $a^2 + b^2$ will be a positive rational integer or zero.

For example, the norm of the complex number $2 - \frac{1}{2}i$ is $2^2 + (\frac{1}{2})^2 = 4\frac{1}{4}$, a real number; while the norm of $\sqrt{3} - i$ is $(\sqrt{3})^2 + (-1)^2 = 4$, a rational integer.

It is easily seen that the norm of a unity of a Gaussian integer is 1; that is, the norms of 1, -1, i, $-i$, are 1. The norm of $-i$ is $(0)^2 + (-1)^2 = 1$.

Conversely, *any Gaussian integer whose norm is* 1 *is a unity*; *that is, it is one of the four numbers* 1, -1, i, $-i$. To see this, we suppose the norm of $a + bi$ is 1, so that $a^2 + b^2 = 1$. Since a and b are positive or negative integers or zero, the only possibilities are: $a = 1, b = 0; a = -1,$

$b = 0$; $a = 0, b = 1$; $a = 0, b = -1$. When these values of a and b are substituted in $a + bi$, we obtain the four unities $1, -1, i, -i$.

We shall now state a general result about the norms of complex numbers, not necessarily complex integers.

If $a + bi$ and $c + di$ are two complex numbers, the norm of the product $(a + bi)(c + di)$ is the norm of $(a + bi)$ multiplied by the norm of $(c + di)$.

Since the product $(a + bi)(c + di)$ is the complex number $(ac - bd) + (ad + bc)i$, the norm of this product is $(ac - bd)^2\ (ad + bc)^2$, which may also be written in the form $(a^2 + b^2)(c^2 + d^2)$. But this last product is precisely the norm of $(a + bi)$ multiplied by the norm of $(c + di)$. This completes the verification.

53. The relation between rational primes and Gaussian primes

We shall now state the theorem which relates rational primes to the Gaussian primes.

A complex integer whose norm is a rational prime is a Gaussian prime.

Suppose that x is a Gaussian integer and that the norm of x is a rational prime p. If x is not a Gaussian prime, it may be factored into a product of two Gaussian integers, y and z, neither of which is a unity. But p is the norm of the complex integer x, that is, p = norm of x = norm of (yz); therefore p = (norm y)(norm z) by the previous result on the norm of a product. Since the norm of y and the norm of z are rational integers and p is a rational prime, the norm of y must be 1 or the norm of z must be 1. Hence either y or z is a unity, and x is a Gaussian prime. The following examples illustrate the relationship between rational and Gaussian primes:

The norm of $(2 + 3i)$ is $(2 + 3i)(2 - 3i) = 13$, which is a rational prime, so that both $2 + 3i$ and $2 - 3i$ are Gaussian primes.

The norm of $(1 + i)$ is $(1 + i)(1 - i) = 2$, which is a rational prime, so that $1 + i$ is a Gaussian prime.

The converse of this result is not true; that is, *the norm of a complex integer may be composite and yet the complex integer may be a Gaussian prime*. For example, the norm of 3 is the composite $(3)^2 + (0)^2 = 9$; yet we shall show that 3 is a Gaussian prime. For if 3 is not a Gaussian prime, it could be factored into two complex integers, neither of which is a unity. Thus,

$$3 = (a + bi)(c + di).$$

We again make use of the previous result that the norm of a product is the product of the norms, so that the norm of 3 equals the norm of $(a + bi)$ multiplied by the norm of $(c + di)$. This results in the equation

$$9 = (a^2 + b^2)(c^2 + d^2).$$

Since neither $a^2 + b^2$ nor $c^2 + d^2$ can be 3 by any choice of a, b, c, and d, it would follow that either $a^2 + b^2$ or $c^2 + d^2$ must be 1. If $a^2 + b^2 = 1$, the only possibilities are $a = \pm 1, b = 0; a = 0, b = \pm 1$. In each one of these four cases $a + bi$ would be a unity so that we have proved that the only factorization of 3 is a complex integer multiplied by a unity, and hence 3 is a Gaussian prime by definition.

54. Determination of Gaussian primes

All the Gaussian primes are found among the following (4: 218–19):

A. The complex integer $1 + i$ is a Gaussian prime.

B. The positive rational primes of the form $4k + 3$ are Gaussian primes.

C. The factors $a + bi$ of the rational primes of the form $4k + 1$ are Gaussian primes.

The above list does not include the associates of the primes, which are, of course, also primes. Since the norm of the complex integer $1 + i$ is 2, we know that $1 + i$ is a Gaussian prime. The complex integers $1 + i, -1 + i, -1 - i, 1 - i$, which are associates of $1 + i$, are all Gaussian primes.

The method of showing that all rational primes of the form $4k + 3$, such as $7, 11, \cdots$, are Gaussian primes is analogous to the method used in proving that 3 is a Gaussian prime.

By Fermat's two-square theorem we know that every rational prime of the form $4k + 1$ may be expressed in essentially one way as the sum of two squares $a^2 + b^2$. Now, $a^2 + b^2 = (a + bi)(a - bi)$, and hence $a^2 + b^2$ is the norm of each of the complex integers $a + bi$ and $a - bi$. But this norm is a rational prime by hypothesis, so that its factors are Gaussian primes.

Let us find the Gaussian prime factors of the rational integer 195. While the factorization of 195 in rational primes is (3)(5)(13), in Gaussian primes $195 = (3)(2 + i)(2 - i)(2 + 3i)(2 - 3i)$. The fact that $2 + i$ and $2 + 3i$ are Gaussian primes follows from the result above, since 5 and 13 are rational primes.

Incidentally, we observe that *a rational prime is not necessarily a Gaussian prime*. For example, $5 = (2 + i)(2 - i)$.

55. The fundamental theorem of arithmetic for Gaussian integers

The fundamental theorem of arithmetic for Gaussian integers is the analogue of the fundamental theorem of arithmetic for rational integers.

Just as with rational integers, any Gaussian integer, not zero or a unity, is divisible by a prime. By repeated use of this result, we may show that *any Gaussian integer, not zero or a unity, is a product of Gaussian primes.*

It is more difficult to prove the uniqueness of the factorization in the case of Gaussian integers.

These two results are combined in the following statement. *Every Gaussian integer may be expressed as a product of Gaussian primes, and there is only one way of doing this* (4: 184–87).

The statement tacitly assumes that the order of writing the primes, the introduction of unities, and the substitution of the associates of the primes for the primes themselves are considered the same factorization.

Let us recall how we find the rational prime factors of a rational integer. We test each prime in turn as a possible divisor, and we obtain all the prime factors. This procedure is not feasible when the prime divisors are very large; other means have been devised for determining them in such cases (10: 54–64).

For Gaussian integers the method is somewhat similar. We have already illustrated the factorization of the rational integer 195 into Gaussian integers. Now we shall give an example of the factorization of a Gaussian integer into its prime factors.

Consider the integer $24 + 54i$. We see at once that 2 and 3 are rational integral factors of the Gaussian integer. The 2 may again be factored into the Gaussian primes $(1 + i)(1 - i)$, while the 3, as we know, is itself a Gaussian prime. The remaining factor of $24 + 54i$ is $4 + 9i$, whose norm is $4^2 + 9^2 = 97$, which is a rational prime of the form $4k + 1$. It follows then from our general result that $4 + 9i$ is a Gaussian prime. The Gaussian prime factors of $24 + 54i$ are 3, $1 + i$, $1 - i$, and $4 + 9i$.

Of course this factorization may take on various forms by the introduction of unities and associates. For example,

$$24 + 54i = 3i(1 + i)^2(-9 + 4i).$$

To find the Gaussian primes of a Gaussian integer, then, we must know the prime factors of the norm, which is a rational integer.

The divisibility properties of Gaussian integers follow from the fundamental theorem for Gaussian integers, just as the divisibility properties of the rational integers follow from the fundamental theorem for rational integers. For example, it is true that *if the product of two Gaussian integers is divisible by a Gaussian prime, then at least one of the integers is divisible by this prime* (4: 187).

56. Fermat's little theorem for Gaussian integers

We shall not pursue the subject of Gaussian integers any further except to state the analogue of Fermat's little theorem (4: 219).

Let $p + qi$ be any Gaussian prime different from $1 + i$, and let $a + bi$ be any Gaussian integer which does not contain $p + qi$ as a factor; then the expression

$$(a + bi)^{p^2+q^2-1} - 1$$

is exactly divisible by $p + qi$. This result may also be written: $(a + bi)^{\text{Norm }(p+qi)-1} - 1$ is divisible by $p + qi$.

The following examples illustrate this result:

The Gaussian integer $2 + i$ is obviously not divisible by the Gaussian prime 3. Hence, by the preceding result, $(2 + i)^{(\text{Norm of }3)-1} - 1$, or $(2 + i)^{9-1} - 1$, or $(2 + i)^8 - 1$ is divisible by 3. We find by computation that $(2 + i)^8 - 1 = -528 - 336i$, which is obviously divisible by 3.

The Gaussian integer $4 + 5i$ is readily shown not to contain the Gaussian prime $2 + i$ as a factor. To verify Fermat's theorem, we see that the norm of $2 + i$ is 5, so that $(4 + 5i)^{5-1} - 1$ must be divisible by $2 + i$, as we see from the computation

$$(4 + 5i)^4 - 1 = (-80)(19 + 9i)$$

$$= (2 + i)(-32 + 16i)(19 + 9i).$$

57. The integers associated with the cube root of unity

There is another class of integers which have many features in common with the Gaussian integers, in particular, the property of unique factorization.

We recall first that a Gaussian integer is of the form $a + bx$, where a and b are positive or negative integers or zero, and x satisfies the equation $x^2 + 1 = 0$. Now let us consider an integer of the form $a + b\rho$, where a and b are again rational integers or zero, and ρ satisfies the equation $\rho^2 + \rho + 1 = 0$. It is readily seen that ρ is an imaginary cube root of unity, and must satisfy the equation $\rho^3 = 1$ since $\rho^3 - 1 = (\rho - 1)(\rho^2 + \rho + 1)$.

The totality of integers of the form

$$a + b\rho$$

will be called the *integers associated with the cube root of unity*, just as the Gaussian integers are connected with the square root of -1.

Of the two imaginary cube roots of unity, we shall, for the sake of concreteness, take $\rho = \frac{1}{2}(-1 + i\sqrt{3})$. We might equally well have defined these integers as the numbers $a + b\rho^2$. In view of the relation $\rho^2 + \rho + 1 = 0$, or $\rho^2 = -1 - \rho$, we see that

$$a + b\rho^2 = a + b(-1 - \rho) = (a - b) - b\rho;$$

and since a and b range over all the rational integers and zero, so do

$a - b$ and $-b$. Actually, $a + b\rho^2$ is the conjugate of $a + b\rho$ as is shown by the following computation:

$$a + b\rho = a + b\{\tfrac{1}{2}(-1 + i\sqrt{3})\} = \left(a - \frac{b}{2}\right) + \frac{bi}{2}\sqrt{3};$$

$$a + b\rho^2 = a - b - b\rho = (a - b) - \frac{b}{2}\{-1 + i\sqrt{3}\} = \left(a - \frac{b}{2}\right) - \frac{bi}{2}\sqrt{3}.$$

The sum of two integers $a + b\rho$ and $c + d\rho$ is again an integer of this form. Similarly, if we multiply $a + b\rho$ by $c + d\rho$, the product is $ac + (ad + bc)\rho + bd\rho^2$, which, in view of the relation $\rho^2 = -1 - \rho$, becomes $(ac - bd) + (ad + bc - bd)\rho$.

We define the norm of the integer $a + b\rho$ to be the product of $a + b\rho$ and its conjugate $a + b\rho^2$. By a simple computation we see that

$$\text{Norm}(a + b\rho) = (a + b\rho)(a + b\rho^2) = a^2 + b^2\rho^3 + ab\rho + ab\rho^2$$
$$= a^2 - ab + b^2 = \left(a - \frac{b}{2}\right)^2 + \frac{3b^2}{4}.$$

From the above expression we see that the norm of $a + b\rho$ is the sum of the squares of two real numbers. It follows then that the norm of $a + b\rho$ is positive except when $a - \frac{b}{2} = 0$ and $b = 0$, in which case $a = b = 0$, so that $a + b\rho = 0$. We have shown that when the norm of $a + b\rho$ vanishes, $a + b\rho$ is the zero associated with the cube root of unity.

As in the case of Gaussian numbers, the norm of a product is the product of the norms, that is,

$$\text{Norm}(a + b\rho)(c + d\rho) = \text{Norm}(a + b\rho)\text{Norm}\,(c + d\rho).$$

This statement is true for all real values of a, b, c, d, not necessarily integral.

An integer $a + b\rho$ is said to be divisible by $c + d\rho$, if there exists an integer $e + f\rho$ for which $a + b\rho = (c + d\rho)(e + f\rho)$; and $c + d\rho$ and $e + f\rho$ are said to be divisors of $a + b\rho$.

For example, $1 + \rho^2$ and $2 - \rho$ are divisors of $-1 - 3\rho$, since the product $(1 + \rho^2)(2 - \rho)$, when simplified, reduces to $-1 - 3\rho$. In simplifying the product we have replaced ρ^3 by 1.

A unity associated with the integers $a + b\rho$ is any integer of this form which is a divisor of 1. This is equivalent to saying (compare Section 52) that the norm of a unity is 1, and any integer $a + b\rho$ whose norm is 1 is a unity. To find the unities in $a + b\rho$, we need only solve $\text{Norm}(a + b\rho) = 1$ or $a^2 - ab + b^2 = 1$; which is the same equation as $(2a - b)^2 + 3b^2 = 4$. All the solutions of this equation are given by $a = \pm 1, b = 0; a = 0, b = \pm 1; a = 1, b = 1; a = -1, b = -1$. There are six unities in $a + b\rho$, namely, ± 1, $\pm \rho$, $\pm(1 + \rho)$. These may also be put in the form ± 1, $\pm \rho$, $\pm \rho^2$.

We define an associate of $a + b\rho$ as the integer obtained when $a + b\rho$ is multiplied by a unity. The associates of $a + b\rho$ are $\pm(a + b\rho)$, $\pm\rho(a + b\rho)$, $\pm\rho^2(a + b\rho)$.

A prime in $a + b\rho$ is an integer, not zero or a unity, which is divisible only by its associates and by 1.

An integer $a + b\rho$ is a prime if its norm is a rational prime. The proof follows the same lines as that for Gaussian primes (Section 53).

For example, $\text{Norm}(1 - \rho) = (1 - \rho)(1 - \rho^2) = 1 + \rho^3 - \rho - \rho^2 = 1 + 1 + 1 = 3$, so that $(1 - \rho)$ is a prime, since $\text{Norm}(1 - \rho)$ is the rational prime 3.

The converse is false; the $\text{Norm}(a + b\rho)$ may be composite and yet $a + b\rho$ may be prime. The Norm(2) is the composite number 4, and yet 2 is a prime among the integers $a + b\rho$. To prove that 2 is a prime, suppose

$$2 = (a + b\rho)(c + d\rho),$$

so that $\text{Norm}(2) = \text{Norm}(a + b\rho)\text{Norm}(c + d\rho)$, or $4 = (a^2 - ab + b^2) \cdot (c^2 - cd + d^2)$. Now this equation is satisfied in three and only three ways:

A. $a^2 - ab + b^2 = \pm 1, c^2 - cd + d^2 = \pm 4$, so that $a + b\rho$ is a unity since its norm is 1.

B. $a^2 - ab + b^2 = \pm 4, c^2 - cd + d^2 = \pm 1$, so that $c + d\rho$ is a unity.

C. $a^2 - ab + b^2 = \pm 2, c^2 - cd + d^2 = \pm 2$. But $a^2 - ab + b^2 = \pm 2$ is the same as $(2a - b)^2 + 3b^2 = \pm 8$, which is easily shown to be impossible for rational integral values of a and b.

Thus only the possibilities A and B remain: either $a + b\rho$ or $c + d\rho$ is a unity; hence, by definition, 2 is a prime.

It can be proved (4: 220–21) that all the primes among the integers $a + b\rho$ are given by

A. $1 - \rho$ and its associates;

B. the rational primes of the form $3n + 2$ and their associates;

C. the factors $a + b\rho$ of the rational primes of the form $3n + 1$.

The analogue of the fundamental theorem of arithmetic for integers of the form $a + b\rho$ will now be stated. *The expression of an integer $a + b\rho$ as a product of primes is unique, apart from the order of the primes and the ambiguities arising from associated primes and from unities.*

For example, the rational integer 105 is also an integer of the form $a + b\rho$, where $a = 105$ and $b = 0$, and $105 = (3)(5)(7)$. Starting with 3, the smallest rational prime factor of 105, we note that $3 = (1 - \rho)(1 - \rho^2)$, which may be factored further; $3 = (1 + \rho)(1 - \rho)^2$. Since $-\rho^2$ is a unity, and $1 - \rho$ is a prime by A above, the prime factors of 3 are $1 - \rho$ taken twice.

Since 5 is a rational prime of the form $3n + 2$, it is also a prime among the integers $a + b\rho$ by B above. Finally, since 7 is a rational prime of the form $3n + 1$, its factors $3 + 2\rho$ and $3 + 2\rho^2$ are primes by C. Therefore, apart from unities and associates, the unique prime factors of 105 among the integers $a + b\rho$ are $(1 - \rho)^2$, 5, $(3 + 2\rho)$, and $(3 + 2\rho^2)$.

There is a Fermat little theorem among the integers $a + b\rho$ analogous to that for rational integers and for Gaussian integers. This theorem says that *if $p + q\rho$ is a prime and if $a + b\rho$ is not divisible by $p + q\rho$, then*

$$(a + b\rho)^{\text{Norm}(p+q\rho)-1} - 1$$

is divisible by $p + q\rho$. Written more explicitly, this says that $(a + b\rho)^{p^2-pq+q^2-1} - 1$ is divisible by $p + q\rho$.

Let $p + q\rho = 2 + 3\rho$, which is easily verified to be a prime. If now we take $a + b\rho = 1 - \rho$, we may show that $1 - \rho$ is not divisible by $2 + 3\rho$. Thus, the hypotheses of the Fermat little theorem are satisfied, and we obtain the result $(1 - \rho)^6 - 1$ is divisible by $2 + 3\rho$. In fact, $(1 - \rho)^6 - 1 = -28$, which has the factors -4, $2 + 3\rho$, and $2 + 3\rho^2$.

58. Algebraic integers

The expressions $a + bi$ and $a + b\rho$ are two examples of algebraic integers. *They are the integers x which are roots of the algebraic equation*

$$x^n + a_1x^{n-1} + \cdots + a_n = 0,$$

where the coefficient of the highest power of x is 1, and the other coefficients $a_1, a_2, \cdots, a_n$ are all rational integers or zero.

For example, if $x = a + bi$, then $x^2 = a^2 - b^2 + 2abi$, and since $2ax = 2a^2 + 2abi$, we find that $x^2 - 2ax = a^2 - b^2 + 2abi - (2a^2 + 2abi) = -a^2 - b^2$, so that

$$x^2 - 2ax + a^2 + b^2 = 0.$$

We have now verified that a Gaussian integer is a root of a quadratic equation with the coefficient of x^2 equal to 1, the coefficient of x equal to the integer $-2a$, and the constant term equal to $a^2 + b^2$. Therefore, we have shown that a Gaussian integer is an algebraic integer.

If now $x = a + b\rho$, then $x^2 = a^2 + 2ab\rho + b^2\rho^2 = a^2 + 2ab\rho - b^2 - b^2\rho$. Further, since $x = a + b\rho$,

$$(2a - b)x = a(2a - b) + b(2a - b)\rho.$$

Hence

$$x^2 - (2a - b)x = -a^2 - b^2 + ab,$$

so that

$$x^2 - (2a - b)x + a^2 + b^2 - ab = 0.$$

And again we have shown that $a + b\rho$ is an algebraic integer.

The two examples just given are instances of quadratic integers, since each satisfies an algebraic equation of the second degree with rational, integral coefficients, and with 1 as the coefficient of the x^2 term.

From this point of view the rational integers are also algebraic integers. For if x is a rational integer a, x satisfies the linear algebraic equation $x - a = 0$.

Since a quadratic integer x satisfies the equation $x^2 + rx + s = 0$, where r and s are rational integers, then

$$x = -\frac{r}{2} \pm \frac{\sqrt{r^2 - 4s}}{2}.$$

Each of these roots has the form

$$a + b\sqrt{m}$$

where a and b are rational integers or halves of rational integers; m is a rational integer different from zero, not a perfect square, containing no square factor. (We are excluding rational integers from our discussion of the quadratic integers.)

If m is of the form $4k - 1$, as in the case of the Gaussian integers where $m = -1$, then all the Gaussian integers will be obtained from the quadratic integers $a + b\sqrt{-1}$ by allowing a and b to range over all the rational integers (4: 207).

If, however, m is of the form $4k + 1$, as in the case of integers associated with the cube root of unity where $m = -3$, these integers will be identical with the quadratic integers $a + b\sqrt{-3}$, where a and b need not be integers but may be halves of integers. Actually, if c and d range over the rational integers, then it suffices to take $a = c - \frac{1}{2}d$ and $b = \frac{1}{2}d$ (4: 207).

We were able to state a theorem for unique factorization of the integers $a + bi$ and $a + b\rho$. We naturally wonder whether there are other types of quadratic integers for which there is such a fundamental theorem of arithmetic.

It has been proved that the negative values of m for which there is a unique factorization among the quadratic integers $a + b\sqrt{m}$ are the nine values

$$m = -1, -2, -3, -7, -11, -19, -43, -67, -163.$$

The only additional value may be an m less than $(-5)(10^9)$, but its existence is highly improbable (4: 212–13).

While we know as many as 16 positive values of m for which we have unique factorization, we cannot say that there is a finite number of such m's (4: 213).

Once we know that unique factorization is possible for a set of integers, quadratic or of higher degree, the results of number theory for the rational integers may be extended to these new integers.

59. Integers for which unique factorization fails

We now give a class of quadratic integers $a + b\sqrt{m}$ for which unique factorization fails. The integers to be considered are of the form $a + b\sqrt{-5}$, where a and b range over all the rational integers or zero. (Here m is of the form $4k - 1$.) Of course the $a + b\sqrt{-5}$ are complex integers $a + b\sqrt{5}i$; they reduce to rational integers when $b = 0$. It is clear that the sum, difference, or product of two such integers is an integer of the same type. We may define *prime*, *unity*, and *associate* in a fashion analogous to that of Gaussian integers or the integers of the form $a + b\rho$. Thus, to define *prime* for the integers $a + b\sqrt{-5}$, we first introduce the unities, those integers whose norm is 1. Since the conjugate of $a + b\sqrt{-5}$ is $a - b\sqrt{-5}$, the norm of $(a + b\sqrt{-5})$ is $a^2 + 5b^2$. If the last expression is to equal 1, it is obvious that $a = 1$, $b = 0$ or $a = -1$, $b = 0$, so that the only unities are 1, and -1. A prime, then, is an integer $a + b\sqrt{-5}$, not zero or a unity, which is divisible only by itself and its negative.

We shall now verify that 7 and $1 + 2\sqrt{-5}$ are primes. Let $7 = (a + b\sqrt{-5})(c + d\sqrt{-5})$, where neither factor is a unity. But since the norm of a product is the product of the norms, we obtain $49 = (a^2 + 5b^2)(c^2 + 5d^2)$. It is impossible for $a^2 + 5b^2$ to equal 7 when a and b equal any rational integers, or zero. Hence, 7 is a prime among the integers of the form $a + b\sqrt{-5}$. Similarly, if $1 + 2\sqrt{-5}$ were not a prime, we would have $1 + 2\sqrt{-5} = (a + b\sqrt{-5})$ $(c + d\sqrt{-5})$, so that $\text{Norm}(1 + 2\sqrt{-5}) = 21 = (a^2 + 5b^2)(c^2 + 5d^2)$. If we examine the factors of 21, we see that neither 3 nor 7 is of the form $a^2 + 5b^2$, and our assertion is proved.

Now, consider the factorization of the integer 21 which is a generalized integer of the form $a + b\sqrt{-5}$, when $a = 21$ and $b = 0$. We see at once that 21 has two factorizations among the integers of the type $a + b\sqrt{-5}$:

$$21 = (3)(7) = (1 + 2\sqrt{-5})(1 - 2\sqrt{-5}).$$

We have already verified that 7 and $1 + 2\sqrt{-5}$ are primes among these integers, and we could show similarly that 3 and $1 - 2\sqrt{-5}$ are also primes. Thus we see that the integer 21 has been factored in two different ways into a product of primes, and our fundamental theorem is false for such integers.

In addition, we observe that the prime factor 3 is a divisor of the product $(1 + 2\sqrt{-5})(1 - 2\sqrt{-5})$, but 3 does not divide either factor, since each is a prime. Here again the results of number theory of rational integers do not hold.

60. Conclusion

It seems natural to ask several questions at this point. First, for what types of "integers" will we have unique factorization? This was partially answered in Section 58. Secondly, if we do not have unique factorization, is it still possible to repair this defect by attaching a new meaning to primes of such integers? The answers to these questions and to others form the subject matter of what is known today as *algebraic number theory*.

It is of interest to add that in his attempt to find a complete proof of Fermat's last theorem, Kummer presented a proof in which he tacitly assumed unique factorization for the integers he used. The error was pointed out to him, and, in trying to rectify it, he was led to the discovery of a sort of substitute for unique factorization which became useful not only in the theory of numbers but in many other branches of mathematics as well.

BIBLIOGRAPHY

1. Davenport, Harold. *The Higher Arithmetic*. London: Hutchinson's University Library, 1952.
   This is an extremely well-written book containing many original features, but no problems. It is highly recommended.

2. Dickson, Leonard Eugene. *Modern Elementary Theory of Numbers*. Chicago: University of Chicago Press, 1939.
   A rather technical treatment of the subject. There is a long appendix which contains a carefully worked out proof by Dirichlet on arithmetical progressions containing an infinite number of primes.

3. Gelfond, A. O. *The Solution of Equations in Integers*. Translated from the Russian by Leo F. Boron. Philadelphia: Universal Correspondence School of Mathematics (Post Office Box 5225, Oak Lane Station), 1960.

4. Hardy, G. H., and Wright, E. M. *An Introduction to the Theory of Numbers*. Third edition. New York: Oxford University Press, 1954.
   This book is primarily for mathematicians and gives a masterly exposition of certain special topics of an advanced nature. Most of the material of Chapter IX of this pamphlet was taken from this book.

5. Heath, Sir Thomas L. *Diophantus of Alexandria*. Cambridge, England: Cambridge University Press, 1910.

6. Jones, Burton W. *Theory of Numbers*. New York: Rinehart, 1955.

7. Kramer, Edna E. *The Main Stream of Mathematics*. New York: Oxford University Press, 1952.
   While this book is not primarily on number theory, it has a good account of the Fibonacci Series in plant growth; other topics of mathematics not usually found in an elementary textbook are treated here in an interesting fashion.

8. Lovitt, William V. *Elementary Theory of Equations*. Englewood Cliffs, N. J.: Prentice-Hall, 1939.

9. Nagell, Trigve. *Introduction to Number Theory*. New York: John Wiley and Sons, 1957.

10. Ore, Oystein. *Number Theory and its History*. New York: McGraw-Hill, 1948.
    This text contains an excellent account of the historical origin of many of the topics of number theory, as well as a clear exposition of the topics themselves. The book also contains a good treatment of constructions with ruler and compasses.

11. Stewart, B. M. *The Theory of Numbers*. New York: The Macmillan Co., 1952.

12. Uspensky, J. V., and Heaslet, M. A. *Elementary Number Theory.* New York: McGraw-Hill, 1939.

This text contains many of the geometric proofs concerning certain progressions of polygonal numbers. It also has a very readable account of magic squares, calendar problems, and card shuffling.

13. Young, J. W. A., editor. *Monographs on Topics of Modern Mathematics.* New York: Dover Publications, 1955.

This set of monographs is a reprint of a book originally published in 1911 and gives an excellent account of the topics of algebra, geometry, number theory, and the calculus.

14. Delone and Faddeev. *The Theory of Irrationalities of the Third Degree.* Providence: American Mathematical Society, 1964.

Index